Journal of Emerging Technologies in Web Intelligence

ISSN 1798-0461

Volume 6, Number 1, February 2014

Contents

Special Issue: International Conference on Intelligent Information and Network Technology
Guest Editors: M. Bahaj and I. Cherti

Guest Editorial 1
M. Bahaj and I. Cherti

SPECIAL ISSUE PAPERS

Reasoning on Hybrid Ontology Using Extended SWRL Rules 3
Souad Bouaicha and Zizette Boufaida

Fuzzy Ontology Evolution: Classification of a New Individual 9
Djellal Asma and Boufaida Zizette

Semantic Technologies Applying to Data Warehouses Federation 15
Mouhni Naoual and Elkalay Abderrafiaa

Vocabulary Persona by using Ontologies 20
Gaou Salma and El Kadiri Kamal Eddine

Unified Approach for Building Heterogeneous Artifacts and Consistency Rules 26
Mounir Zekkaoui and Abdelhadi Fennan

Crypto-Security Contribution in WSNs 32
Nourreddine Mitta, Rachid Gouri, Hlou Lamari, and Brahim Moalige

MGI-LEACH: Multi Group LEACH Improved an Efficient Routing Algorithm for Wireless Sensor Networks 40
Alami Chaibrassou and Ahmed Mouhsen

A WIFI Antenna Radiation Effects on Human Head in the ISM Band 45
Radouane Karli and Hassan Ammor

Development of a New Slit-Slotted Shaped Microstrip Antenna Array for Rectenna Application 49
Mohamed Adel Sennouni, Jamal Zbitou, Abboud Benaissa, Abdelwahed Tribak, Otman Elmrabet,and Mohamed Latrach

Performance Evaluation of WideMac Compared to ALOHA in term of Energy Consumption for IR-UWB based WSN 54
Anouar Darif, Rachid Saadane, and Driss Aboutajdine

Euclidean 3D Reconstruction of Unknown Objects from Multiple Images 59
Soulaiman El Hazzat, Abderrahim Saaidi, and Khalid Satori

Comparative Study of PCA, ICA, LDA using SVM Classifier 64
Anissa Bouzalmat, Jamal Kharroubi, and Arsalane Zarghili

SVM Classification of High Resolution Urban Satellites Images using Composite Kernels and Haralick Features 69
Aissam Bekkari, Soufiane Idbraim, Driss Mammass, Mostafa El Yassa, and Danielle Ducrot

3D Normalization Based on the Barycentric Coordinates 75
El Mostafa Rajaallah, Lahcen Moumoun, and Taoufiq Gadi

Optimization of the Organized KOHONEN Map by a New Model of Preprocessing Phase and Application in Clustering 80
Es-safi Abdelatif, Harchli Fidae, and Ettaouil Mohamed

Adaptive Learning 85
Soumaya El Mamoune, Marouane El Mabrouk, Loubna Cherrat, Ahmed Hadioui, Omar Kassara, and Mostafa Ezziyani

A Comparative Study between the Pseudo Zernike and Krawtchouk Invariants Moments for Printed Arabic Characters Recognition 89
Rachid Salouan, Said Safi, and Belaid Bouikhalene

People Tracking using Color Control Points and Skin Color 94
K.Laaroussi, A.Saaidi, K.Satori

Fuzzy Control System for Autonomous Navigation of Thymio II Mobile Robots 101
Fatma Boufera, Fatima Debbat, Francesco Mondada, and M. Faycal Khelfi

SURVEY PAPERS

A Survey on Prevention Approaches for Denial of Sleep Attacks in Wireless Networks 106
G.Mahalakshmi and P.Subathra

A Comprehensive Study of Finding Copy-and-Paste Clones from Program Source Codes 111
Kamran Khan, Saif Ur Rehman, Kamran Aziz, and Simon Fong

REGULAR PAPERS

An Optimized Parallel Computing Paradigm for Mobile Grids Based on DSPOM 119
Aghila Rajagopal and M.A. Maluk Mohamed

DWDE-IR: An Efficient Deep Web Data Extraction for Information Retrieval on Web Mining 133
Aysha Banu and M. Chitra

An Adaptive Weight Calculation based Bandwidth Allocation Scheme for IEEE 802.16 Networks 142
Akash Deep and KS Kahlon

Tumor Detection by Color Segmentation of PET/CT Liver Images 148
Neha Bangar and Akash Deep

A Model for Implementing Security at Application Level in Service Oriented Architecture 157
Said Nabi, Saif Ur Rehman, Simon Fong, and Kamran Aziz

Special Issue on International Conference on Intelligent Information and Network Technology

Guest Editorial

The International Conference on Intelligent Information and Network Technology (IC2INT) 2013 was held in Faculty of Sciences and Technology Settat Morocco, 13-14 November 2013. IC2INT is an international forum for discussions on recent advances in the fields of intelligent information and network technology specially related to the design, development, modeling, and how to use the information systems in organizations with multidisciplinary perspective. It also serves to foster communication leading academic scientists, scholars and practitioners to exchange and share their experiences and research results about all aspects of Computer Science and Information Engineering.

In response to the call for papers, we received 130 submissions. After an initial screening, 11 submissions were declared "out of scope" (or with inappropriate technical content) and the remaining 119 were sent to reviewers. All manuscripts underwent a very rigorous peer review process. We finally selected 78 full papers, 11 Poster and 19 technical notes for this special issue substantiallty validated and evaluated.

The full papers in this issue can be broadly organized into three main categories

Ontologies Engineering, Artificial Intelligence

Mobile & Wireless Technology, Network Telecommunication

Information Systems & Natural Language Processing

The papers were selected on the basis of **Novelty and Originality of the contribution as well as their Technical Content.**

- **Reasoning on Hybrid Ontology Using Extended SWRL Rules**
  Souad Bouaicha, Zizette Boufaida
- **Fuzzy Ontology Evolution: Classification of a New Individual**
  Djellal Asma and Boufaida zizette
- **Semantic Technologies Applying to Data Warehouses Federation**
  MOUHNI Naoual, ELKALAY Abderrafiaa
- **Vocabulary Persona by using Ontologies**
  GAOU Salma, EL KADIRI Kamal Eddine
- **Unified Approach for Building Heterogeneous Artifacts and Consistency Rules**
  Mounir ZEKKAOUI, Abdelhadi FENNAN
- **Crypto-Security Contribution in WSNs**
  Nourreddine Mitta, Rachid Gouri, Hlou Lamari, Brahim Moalige
- **MGI-LEACH : Multi Group LEACH Improved an Efficient Routing Algorithm for Wireless Sensor Networks**
  Alami Chaibrassou, Ahmed Mouhsen
- **A WIFI Antenna Radiation Effects on Human Head in the ISM Band**
  Radouane KARLI, Hassan AMMOR
- **Development of a New Slit-Slotted Shaped Microstrip Antenna Array for Rectenna Application**
  Mohamed Adel SENNOUNI, Jamal ZBITOU, Abboud Benaissa, Abdelwahed TRIBAK, Mohamed Latrach
- **Performance Evaluation of WideMac Compared to ALOHA in term of Energy Consumption for IR-UWB based WSN**
  Anouar Darif, Rachid Saadane, Driss Aboutajdine
- **Euclidean 3D Reconstruction of Unknown Objects from Multiple Images**
  Soulaiman El hazzat, Abderrahim Saaidi, Khalid Satori
- **Comparative Study of PCA, ICA, LDA using SVM Classifier**
  Anissa Bouzalmat, Jamal Kharroubi, Arsalane Zarghili
- **SVM Classification of High Resolution Urban Satellites Images Using Composite Kernels and Haralick Features**
  Aissam Bekkari, Soufiane Idbraim, Driss Mammass and Mostafa El yassa, Danielle Ducrot
- **3D Normalization Based on the Barycentric Coordinates**
  El Mostafa RAJAALLAH, Lahcen MOUMOUN, Taoufiq GADI
- **Optimization of the Organized KOHONEN Map by a New Model of Preprocessing Phase and Application in Clustering**
  Es-safi Abdelatif, Harchli Fidae, Ettaouil Mohamed
- **Adaptive Learning**

doi:10.4304/jetwi.6.1.1-2

Soumaya EL MAMOUNE, Loubna CHERRAT, Omar KASSARA, Marouane EL MABROUK, Ahmed Hadioui, Mostafa EZZIYANI

- **A Comparative Study between the Pseudo Zernike and Krawtchouk Invariants Moments for Printed Arabic Characters Recognition**
  R. SALOUAN, S. SAFI, B. BOUIKHALENE
- **People Tracking using Color Control Points and Skin Color**
  K.LAAROUSSI, A.SAAIDI, K.SATORI
- **Fuzzy Control System for Autonomous Navigation of Thymio II Mobile Robots**
  Fatma BOUFERA, Fatima DEBBAT, Francesco MONDADA, M. Faycal KHELFI

We hope you enjoy these papers. We would like to thank the authors for their contributions.

Special thanks to the organizing committee members, program committee members as well as the reviewers contributing so richly to the scientific ambiance.

We are indebted to Editor-in-Chief, Simon Fong for offering us this opportunity and for patiently waiting the completion of the special issue.

Guest Editors

M. Bahaj
Department of Mathematics and Computer Sciences
University Hassan 1st Faculty of Sciences & Technology, Settat, Morocco

I. Cherti
Department of Mathematics and Computer Sciences
University Hassan 1st Faculty of Sciences & Technology, Settat, Morocco

M. Bahaj He is a Full Professor in Department of Mathematics and Computer Sciences from the University Hassan 1st Faculty of Sciences & Technology Settat Morocco. He is co chairs of International Conference on Software Engineering, Databases and Expert Systems (SEDEXS'12)" , NASCASE'11. He has published over 60 peer-reviewed papers.

His research interests are intelligents systems, Ontologies Engineering, Partial and differential equations, Numerical Analysis and scientific computing .

I. Cherti He is a Full Professor in Department of Mathematics and Computer Sciences from the University Hassan 1st Faculty of Sciences & Technology Settat Morocco. His research interests are intelligents systems, partial and differential equations. He has organized various national and international events.

Reasoning on Hybrid Ontology Using Extended SWRL Rules

Souad Bouaicha
LIRE Laboratory, University of Constantine 2
Constantine, Algeria
bouaicha.souad@gmail.com

Zizette Boufaida
LIRE Laboratory, University of Constantine 2
Constantine, Algeria
zboufaida@gmail.com

***Abstract*—In this paper we present the principal steps to create a semantic rule which integrates Object Oriented Techniques to the comportment of the concepts.**

The set of rules obtained from the extension of SWRL rules will be posted to Jess Engine using rewrite meta-rules. The reasoning on this combination allows inferring new knowledge and stores it in the knowledge base. We propose an implementation of the method of extending SWRL rules with the adaptation of the different Object Oriented Techniques. The engine executes the transformation of these techniques to the Jess model. Our demonstration clarifies the importance of this kind of reasoning. We use a case study inherent to interpret a check up in preventive medicine.

***Index Terms*—SWRL rules, Reasoning, Object Oriented Technique, Hybrid Ontology.**

I Introduction

One of SWRL's most useful features is its ability to incorporate uses refined built-in libraries. This extension mechanism provides very powerful means of expanding SWRL's expressiveness and increasing the types of the reasoning information using rules. In particular, this mechanism can be used to tackle the issue of data integration, which is one of the central challenges of the Semantic Web. The ability to meet this challenge requires the development of a variety of mapping technologies to allow interoperation between the various formats that will be encountered when developing Semantic Web applications [1] While we embrace this proposal, we will argue in this paper that it is not sufficiently expressive for our own needs, and we have therefore proposed an extension to SWRL [25]. We have proposed an approach for enriching the comportment (behavior) of the concepts of an ontology with rules [2]. This enrichment is done through the two manners: 1) Binding attributes and roles of concepts with conditions, 2) Creating rules to combine the individuals and the atoms of concepts. The latter is, in a large scale, in literature.

This internal and external enrichment of the concepts of an ontology give rise to hybrid antologies, which permit amelioration of the reasoning results. To clarify the interest of such reasoning, we apply a case study inherent to interpretation of check up in preventive medicine, using the standard and current tools of semantic Web.

In this paper we present a prototypical implementation of our approach which integrates the rules inside a concept using Object Oriented Techniques and the Jess[1] reasoning engine. Our integration allows posing a query to a knowledge base using an nRQL language[26]. We assume that the hybrid ontology which is handled can contain both OWL axioms and SWRL rules and reasoning with extended SWRL rules.

During the development and research process, we have proposed and implemented two main methods of transforming OWL ontologies and SWRL rules into rules expressed in Jess language. In this work, we are focused on the implementation and the evaluation of the latter methods.

The paper is organized as follows. In the next section, we detail the approaches and strategies to combine existing OWL DL with another language rules and gave birth to languages such as SWRL, DLP and some work to combine reasoning about OWL and that on SWRL rules. In Section 3 we present the classification of rules. Section 4 has a discussion about some related work. Proposed process and reasoning methods are provided in Section5. Section 6 provides an example evaluation and application to illustrate the reasoning process. Finally, Section 7 contains concluding remarks and future work plans.

II Approaches of Integration of Owl with Language Web of Rule

In this section, we briefly overview the approaches, both practical and theoretical, for the utilization of onto-

[1] JESS : http://www.jessrules.com/

doi:10.4304/jetwi.6.1.3-8

logical knowledge in rule engines. More details can be found in the surveys [3][4].

2.1. Interfacing External Ontology Reasoners with Rule Engines

In this scenario, the rule engine calls an external OWL reasoner, e.g. a DL reasoner, whenever is needed ("on the fly"). The hybrid approaches can be classified into bidirectional and unidirectional. In unidirectional frameworks, the information goes from the ontology reasoner to the rule engine and thus, the ontology knowledge remains unmodified [5][6][7]. In bidirectional frameworks, the ontology predicates can be used both in the body and in the head of rules and thus, the ontology knowledge may be modified [8][9].

2.2. Mapping Ontology Reasoners on the Data Model of Rule Engines

In this case, the results of the external OWL reasoner are mapped to the data model that the rule engine supports, e.g. on triple-based facts [10]. In that way, the rule engine can operate without calling further the ontology reasoned (one-time mapping), since all the ontological knowledge exists in its knowledge base.

2.3. Strong Coupling of Ontologies and Rules

In this approach, also known as homogeneous, there is no external OWL reasoning module. The ontology semantics are partially mapped on rule formalism [11], e.g. Datalog, that coexists in the rule base with rule predicates, enhancing the expressivity [12]. Therefore, a new reasoner is needed, able to handle the new homogeneous language that emerges [13][14][15][16].

III CLASSIFICATION OF RULES

The rules are necessary to represent knowledge, which is required for various tasks of reasoning or of meta-reasoning. Classification below is a non thorough classification of the rules according to their use on a conceptual level. They are classified according to the role which they play in the various tasks [21]: deductive rules, meta-reasoning rules, connecting rules, mapping rules, querying rules. [22, 23,24]

3.1. Deductive Rules

Are needed for inferences based on dependencies between some ontologies properties, such as the transfer of properties from parts to wholes (e.g. a disease located in an organ part, is located in the organ), For a long time, rule-based expert systems have shown the usefulness of deductive rules in health care e.g. for diagnosis, decision making etc.

3.2. Connecting Rules

Aare required for connecting ontology to allow reasoning across several domains such as Genomics, Proteonomics, Pathology, for example when searching for correlations between diseases and the abnormality of a function of a protein coded by a human gene.

3.3. Mapping Rules

For mapping ontology in information integration, to allow answering queries over heterogeneous sources e.g., patient data scattered in many Hospital Information Systems.

3.4. Querying Rules

Expressing complex queries upon the Web or querying heterogeneous sources.

3.5. Meta-Reasoning Rules

Are needed to facilitate meta-reasoning, either to support ontologies engineering e.g., acquisition, validation, maintenance of huge reference ontologies, or control of reasoning.

IV LIMITATIONS OF SWRL

The current SWRL specification omits some typical rule language constructs, sacrificing some expressiveness to ensure decidability and/or efficiency. We have previously identified some limitations and workarounds in writing teaching strategies in SWRL; in this work, we directly address these limitations.

- *Flat list*. A SWRL rule body (or head) is a flat list of atoms. No block structure is supported (as there are no constructs that would need them). Note that either of #E or #F would also entail adding block structure.
- *E. Conjunction only*. Disjunction (E1) and negation (E2) are excluded.
- *F. No quantifiers*. Explicit universal and existential quantifiers are excluded. All atoms are implicitly universally quantified.
- *No user-defined functions*. SWRL provides a library of built-in functions for primitive math, string, and date operations. It doesn't allow an external function to be defined or called (which fails 1.1#C).
- *Assertions only.* A SWRL rule head "makes" its atoms true; the standard implementation is to add these atoms as new facts to the knowledge base. Existing knowledge cannot be changed.[27]

V RELATED WORK

- DLEJena [17] is inspired from the hybrid and homogeneous approach for the integration of rules and ontologies. The architecture of DLEJena has four modules: the first charges the ontology and separates its terminological and assertional parts. The second reasons on the T.box part, the third reasons on the A.box part and the last one, reasons with the help of the applied rules to T.box and A.box parts.
- In [18], the plugin of SWRL Jess Tab is illustrated on the Family ontology rules to show how certain reasoning could be provided for interoperability between SWRL and OWL.
- In [19], the symbolic knowledge is represented in an ontology of the cortical structures. For the representation standard languages of the semantic Web (OWL, SWRL) are used. In order to enrich ontology, it is ex-

tended by Horn clauses. These rules permit to propagate relations and to infer new facts from those existing.

- Alloy [20] is a system for reasoning on an ontology written in OWL and SWRL or SWRL-Fol. It is based on the transformation of OWL ontology and SWRL rules in a program called Alloy's program and its analysis starts from a charged model to verify automatically its uniformity.

VI Discussion

The study of the previous work shows that there are two different, fundamental approaches. The first approach, called homogenous, where the rules and the individuals of the ontology are compiled without separation between them. This leads to a base of facts, compiled in the same inference engine [20]. The hybrid systems are the second approach. They are made of several subsets, each of them treats a different part of the knowledge base and use new formalisms of representation and specific procedures of reasoning [17, 18, 19]. In such systems, the problem of the complementary in answers is asked. Really, if the inference is executed separately with OWL reasoner and an inference engine, some inferences are evidently missed.

Finally the integration of description logic and logic programming is very difficult because these two paradigms are semantically different. Therefore the interoperability between the SWRL rules and the ontology requires a close integration.

The following table summarizes the characteristics of each use case.

For this reasons, this type of integration has limits such as:

- The inferences of rule are based only on the component of rule, since Tbox is not integrated in the knowledge base of rule (for example Jess). Because the knowledge base Jess is incomplete, some inferences are necessarily missed.
- Language LD and of LP are basically and semantically different.
- A loose interoperability between SWRL rules and ontology (OWL LD) is not satisfied; therefore interoperability between the two components requires a tight integration. And to ensure a decidable reasoning, valid and complete it is necessary to obtain a Web rule language which has a very clear semantics.

In conclusion of these studies, the following table summarizes the characteristics of each use case.

TABLE 1.
DIFFERENT TYPES OF RULES AND INTEGRATION APPROACHES USED FOR EACH CASE STUDY

Use Case	Rule types	Integration Approaches
DLEjena[17]	Meta-reasoning, Deductive	loose (limit of Mapping)
FamilyOntology[18]	Meta-reasoning, Deductive	loose (limit of Mapping)
Medical ontology[19]	Deductive	Extension of OWL LD with rules
Alloy [20]	Deductive, Mapping	Extension of OWL LD with rules

In [2] we are interested in complete answers of a hybrid system by proposing an approach that allows enriching the comportment of concepts. This enrichment is done via rules reasoning on the attributes. Concretely, we propose a complete process of creation a rules base to the terminological part of ontology.

VII Proposed Reasoning Methods

In this paper we will introduce the new extension of SWRL from another angle to illustrate what makes it different, and why we believe that it provides good solutions to many real-world requirements. Our main point is that this extension (it extension and its formal definition in [2]) is borrowing good practices from object-oriented programming and modeling languages and integrates object-oriented techniques with the flexible architecture of Semantic Web to produce a new way of working with linked data.

A key contribution of this new kind of rules is to introduce a mechanism that allows users to organize those SQWRL (Semantic Query Web Rule Language) queries in a natural, object-oriented way. These rules are not just plain lists of rules like in comparable rule languages (SWRL etc). That you can arrange the rules in the class hierarchy where they belong. This follows the Object-Oriented principles of abstraction and encapsulation. Since the rules (and constraints) are attached to classes, any human or agent who looks at the ontology can quickly understand the meaning of the classes and properties. Furthermore, the rules are "scoped" so that tools are better guided when they need to execute the rules and constraints.

Our main goal of this work is to research novel reasoning techniques which allow for efficient reasoning in hybrid ontology. A Hybrid ontology is an ontology enriched by SWRL rules. this enrichment is effected by two ways.

1. Connect the attributes and roles of the concept which rules that adapt the object-oriented programming techniques. At this level we propose to enrich the language SWRL by adding new keywords.
2. Combine individuals with simple SWRL rules.

We have led to the development of highly optimized systems that support efficient reasoning. This novel reasoning methods and techniques which improve the efficient reasoning. It will present in the following architecture.

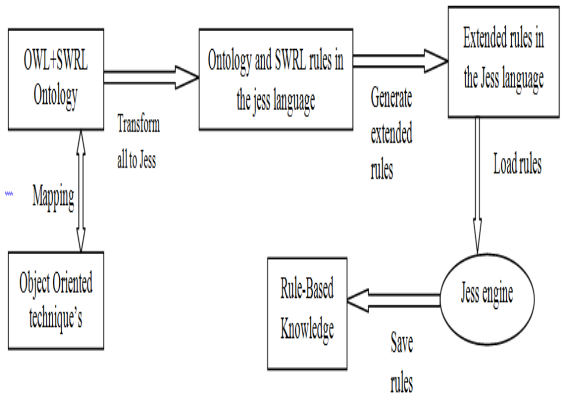


Fig. 1. The Reasoning Process

The Figure 1 concentrated to make clear the reasoning part in the system proposed in [2]; it presents the integration schema of an OWL ontology with SWRL rules, Object Oriented Techniques and the Jess engine. We assume that the ontology is in the Horn SHIQ language and contains SWRL rules (Horn-like clauses). Such OWL+SWRL ontology is transformed into a set of rules in the Jess language. The set of rules is stored as a Jess script file (*.clp). The script is then transformed into a set of extended rules (ExRScript.clp).

Our tool is implemented in Java language; its module provides the following functionalities:

- Reading OWL ontology (and viewing of concepts/roles hierarchies; These hierarchies are calculated by the RACER engine) and Jess scripts, Jess scripts generation (forward and backward chaining, extended rules, Horn-SHIQ transformation) from OWL ontology
- Mapping between ontology concepts/roles and Object Oriented techniques
- Executing a Jess query which consists of the concepts and roles from OWL ontology or templates defined in Jess language,
- Rule-based inquiry answering methods: hybrid and extended rules, Jess engine reasoning management (in forward and backward chaining).

7.1 SWRL extension to Jess Transformation Methods

The proposed tool supports two main methods of transforming OWL ontologies and SWRL rules into rules expressed in Jess language: Simple and Horn-SHIQ.

7.1.1 The Simple Method

This method transforms taxonomies of concepts and roles into Jess rules. These taxonomies are calculated by the RACER engine first. SWRL rules and SWRL extension predicates are also transformed into rules and Jess expressions. The simple transformation can be done in the following modes:

- Jess script assigned to forward chaining.
- Jess script assigned to backward chaining.
- Jess script assigned to forward chaining with extended rules.

7.2.2The Horn-SHIQ Transformation

The Horn-SHIQ transformation method is an extension of the simple one. In this case, additional rules are generated according to OWL axioms. Rather than transforming the semantics of the OWL language into rules we create rules according to this semantics and a given ontology.

VIII EVALUATION EXAMPLE

To validate the preceding methods, we use a preventive medicine domain ontology. In this demo we present an example where we start with a simple ontology (conceptualization steps [2]). It defines a class Person with the following characteristics:

Person is a class that has five properties. The values of Weight, size, sex, age and BMI (Body Mass Index), are specified by the user, and the new SWRL rule is used to compute and calculate the value of the BMI (The Body Mass Index is a measurement tool that compares your height to your weight and gives you an indication of whether you are overweight, underweight or at a healthy weight for your height) property by using the following formula: weight in kilograms divided by height in meters squared (weight (kg) / [height (m)]2). Let's have a look at the rule first. We need to add a new keyword as SWRL rule attached to a class using the CONSTRUCT rule for example to do mathematical calculations.

Whenever someone changes the of values size weight, then the value of BMI will update automatically, as shown with the following figure the editing of rule Construction on Protégé.

Fig. 2. Rule Construction

We use the built-in bridge to create the new concepts (keywords). It provides a mechanism for defining and dynamically loading built-in implementation written in Java [2].

Having edited the rules, we convert the rules to Jess by opening the "rules-only" .owl file. This automatically invokes XSLT, SweetRules, and rule transformations in Jess to achieve the conversion, detects which rules have changed, and updates only those rules in Jess's working memory.

A key contribution of this new kind of rules is not just plain lists of rules like in comparable rule languages (SWRL etc). That you can arrange the rules in the class hierarchy where they belong. This follows the Object-Oriented principles of abstraction and encapsulation. Since the rules (and constraints) are attached to classes, any human or agent who looks at the ontology can quickly understand the meaning of the classes and properties. Furthermore, the rules are "scoped" so that tools are bet-

ter guided when they need to execute the rules and constraints.

The use of the Rule Construction is presented in the figure 3.

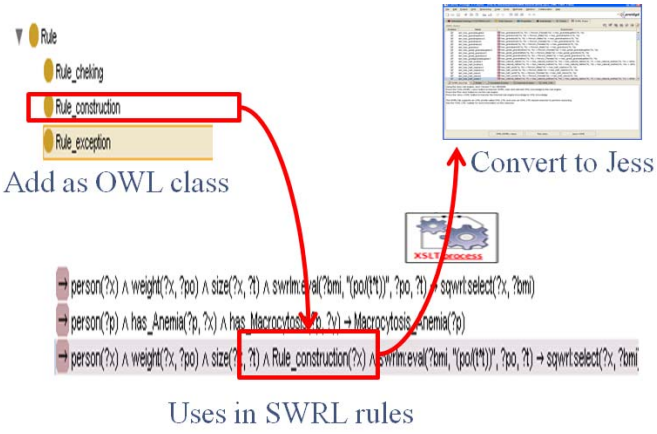


Fig. 3 Using RuleConstruction to Create Datatype Property

In this case we use Jess as our primary inferencing engine. Consequently we need to convert SWRLp rules into Jess syntax. The XSLT script to perform this operation is quite similar to the swrlp2prolog script.

The following is the code for the rule Construction as generated by swlp2jess.xsl:

```
(defrule RuleConstruction
(Patient(Name ?x))
(weight (?x ?w))
(size(?x ?s))
(bind ?s (* ?s ?s))
(bind ?BMI (/ ?W ?s))
 => assert (?BMI (?x ?BMI)))
```

When the system execute the rule construction, the BMI data property well be created in the Person concept, then the reasoning inside the concept we are permitted to create a new data property in the same concept Person.

The Person concept is enriched by the new data property (BMI), which is can be inherited. The following figure represents the BMI calculated for each person.

?x	?bmi
Nassima	40.0
Fateh	19.37716154272828
souad	20.0

Fig.4. Create and Calculate the BMI

Our system comes with a mode in which it does incremental inferencing. This means that whenever someone changes the values of width or size, then the value of BMI will update automatically, as shown with the example instance below:

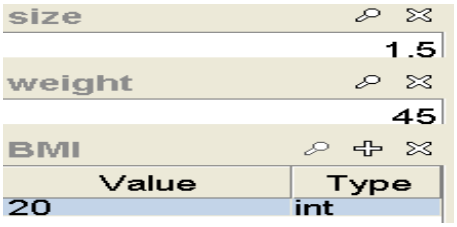


Fig.5. Update Automatically the BMI

To verify user input, and to provide warnings if values violate a constraint; we will add the constraint definitions to the editing tool. These tests can be performed very efficiently, only on the instance that the user is currently looking at.

IX Conclusion and Future Work

The reasoning on ontologies uses simple inference engine. But in view of insufficiencies of the supported languages, ontologies are extended by SWRL rules. We assess that SWRL isn't expressive enough to write rule's sets for inherent to interpretation of check up in preventive medicine. So we have estimated extending SWRL rules. We have also demonstrated a syntax-based extension to SWRL that supports more flexible and expressive rules.

In this paper we have presented a reasoning process, who adapted the proposed reasoning methods witch its goal is to convert SWRLp rules into Jess syntax. We have explained the motivation for this reasoning mechanism, provided an example of its use, offered some cases concerning the XSLT transformation scripts and described the rule construction and rule exception that have been developed thus far.

Reference

[1] Martin J. O Connor, Ravi Shankar, Mark A. Musen, Amar K. Das, Csongor Nyulas. The SWRLAPI: A Development Environment for Working with SWRL Rules. 7th International Semantic Web Conference (ISWC-2008), Karlsruhe, Germany, . V 432 October 26-27, 2008

[2] S. Bouaicha, Z.Bouaifaida Extending SWRL Rules to integrate Object-Oriented Techniques in Hybrid Ontologies in: International Journal of Computational Linguistics Research Vol.3, September 2012 pp 79-89.

[3] J. de Bruijn, P. Bonnard, H. Citeau, S. Dehors, S. Heymans, R. Korf, J. P¨uhrer, T. Eiter, State-of-the-art Survey of Issues, Tech. Rep. D3.1, EU-IST Integrated Project (IP) 2009-231875, ONTORULE (2009).

[4] G. Nalepa, W. Furmanska, Pellet-heart proposal of an architecture for ontology systems with rules, in: R. Dillmann, J. Beyerer, U. Hanebeck, T. Schultz (Eds.), KI 2010: Advances in Artificial Intelligence, Vol. 6359 of Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 2010, pp. 143–150.

[5] W. Drabent, J. Henriksson, J. Maluszynski, HD-rules: A Hybrid System Interfacing Prolog with DL-reasoners, in: Applications of Logic Programming to the Web, Semantic Web and Semantic Web Services, Vol. 287, CEUR-WS, 2007.

[6] T. Eiter, G. Ianni, T. Lukasiewicz, R. Schindlauer, H. Tompits, Combining answer set programming with de-

scription logics for the Semantic Web, Artif. Intell. 172 (12-13) (2008) 1495–1539.
[7] R. Rosati, Semantic and Computational Advantages of the Safe Integration of Ontologies and Rules, in: Principles and Practice of Semantic Web Reasoning, Springer 3703, 2005, pp. 50–64.
[8] H. Kattenstroth, W. May, F. Schenk, Combining OWL with F-Logic Rules and Defaults, in: 2nd International Workshop on Applications of Logic Programming to the Web, Semantic Web and Semantic Web Services, Vol. 287, CEUR-WS, 2007, pp. 60–75.
[9] S. Bragaglia, F. Chesani, A. Ciampolini, P. Mello, M. Montali, D. Sottara, An hybrid architecture integrating forward rules with fuzzy ontological reasoning, in: HAIS (1), 2010, pp. 438–445.
[10] M. Grobe, Rdf, jena, sparql and the 'semantic web'., in: G. FarallySemerad, K. J. McRitchie, E. Rugg (Eds.), SIGUCCS, ACM, 2009, pp. 131–138.
[11] P. F. Patel-Schneider, I. Horrocks, A Comparison of Two Modelling Paradigms in the Semantic Web, Web Semantics: Science, Services and Agents on the World Wide Web, Elsevier 5 (4) (2007) 240–250.
[12] G. Meditskos, N. Bassiliades, Combining a dl reasoner and a rule engine for improving entailment-based owl reasoning, in: International Semantic Web Conference, 2008, pp. 277–292.
[13] S. Heymans, J. D. bruijn, L. Predoiu, C. Feier, D. V. niewenborgh, Guarded hybrid knowledge bases, Theory Pract. Log. Program. 8 (3) (2008) 411–429.
[14] J. Mei, Z. Lin, H. Boley, ALCu p : An Integration of Description Logic and General Rules, in: Web Reasoning and Rule Systems (RR), Spinger 4524, 2007, pp. 163–177.
[15] G. Meditskos, N. Bassiliades, Dlejena: A practical forward-chaining owl2rl reasoner combining jena and pellet, Web Semantics: Science, Services and Agents on the World Wide Web 8 (1) (2010) 89 – 94.
[16] E. Kontopoulos, N. Bassiliades, G. Antoniou, Deploying defeasible logic rule bases for the semantic web, Data & Knowledge Engineering 66 (1) (2008) 116 – 146, including Special Section: Natural Language Processing and Information Systems (NLDB 2006) - Four selected and extended papers.
[17] G.Meditskos, N.Bassiliades A Practical Forward-Chaining OWL2RL Reasoner for OWL 2 DL Ontologies Combining Jena and Pellet (technical report) January 2009. This work was partially supported by a PENED program, jointly funded by EU and the General Secretariat of Research and Technology.
[18] C, Golbreich., Combining Rule and Ontology Reasoners for the Semantic Web, Invited talk, Rules and Rule Markup Languages for the Semantic Web, Hiroshima, Japan, G. Antoniou, H. Boley Editors,LNCS 3323, Springer, 2004.
[19] C, Golbreich., O, Bierlaire., O, Dameron., B. Gibaud. Use case: Ontology with rules for identifying brain anatomical structures. W3C Workshop on Rule Languages for Interoperability, 2005.
[20] H, Hai., J, Wanga., D, Song., S, Jing., S, Jun., Reasoning support for Semantic Web ontology family languages using Alloy. Multiagent and Grid Systems.V2.pp 455-471 (2007)
[21] A. Mechouche, C. Golbreich, B.Gibaud. Vers un système hybride pour l'annotation sémantique d'images IRM du cerveau. 7èmes journées francophones Extraction et Gestion des Connaissances (01/2007) 433-444
[22] C.Golbreich, O. Dameron, B.Gibaud, A.Burgun A, Web ontology language requirements w.r.t expressiveness of taxononomy and axioms in medicine, 2nd International Semantic Web Conference, ISWC 2003, Sanibel Island, Florida, October 2003, LNCS 2870, Springer. 2003.
[23] C. Golbreich,.O.Bierlaire,. O.Dameron,. and B. Gibaud. What reasoning support for Ontology and Rules? the brain anatomy case study, OWL Experiences and Directions Workshop, collocated with the International Conference on Rule Markup Languages for the Semantic Web, Galway, Ireland, 2005
[24] O. Dameron, B. Gibaud, M.Musen. Using semantic dependencies for consistency management of an ontology of brain-cortex anatomy, KR-MED 20
[25] I. Horrocks, PF. Patel-Schneider, H. Boley, S. Tabet and B. Grosof SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C, May 21, 2004.
[26] V. Haarslev, R. Moller and M. Wessel, RACER User's Guide and Reference Manual, version 1.6. Technical report, University of Hamburg, Computer Science Department, 2001.
[27] Wang, E., and Kim, Y. SUsing SWRL for ITS through Keyword Extensions and Rewrite Meta-Rules. 13th Int'l. Conf. Artificial Intelligence in Education (AIED), 101–105, Marina Del Rey. . (2007).

Fuzzy Ontology Evolution: Classification of a New Individual

Djellal Asma[1]
Boufaida Zizette[1]
[1] LIRE Laboratory, Constantine 2 University, Constantine, Algeria
{asmadjellal, zboufaida}@gmail.com

***Abstract*—Ontologies help to conceive the real world with its semantic constraints. But this world has several uncertainties and imperfections that we cannot conceive using traditional ontologies. In our work, we are mainly interested in the imprecise knowledge representation problem. We believe that the most appropriate way is to use fuzzy logic in order to build ontologies called "Fuzzy ontologies". Even if there are several researches revolve around handling imprecise knowledge, however, there are still some issues to be further studied; one of them is the evolution problem which is still a tedious field. In this paper, and as a first step in the problem of fuzzy ontology evolution, we are interested in the classification of new individuals; for this reason, we use the classificatory reasoning mechanism to enable the classification of new individuals in a fuzzy hierarchy.**

***Index Terms*—Fuzzy logic, Fuzzy ontology, Classification reasoning, Individual classification.**

I INTRODUCTION

Ontologies represent a key focus of research for many applications in knowledge engineering, especially the Semantic Web project. They help to conceive the real world with its semantic constraints [16]. But this world has several uncertainties and imperfections that we cannot conceive using traditional ontologies. The formal representation of ontologies is generally based on the classical descriptions logic which shows its limits for all facts that are not expressed with "true" or "false" values.

Several fuzzy approaches have been proposed in order to simplify the possibilities of imprecise knowledge representation by assigning weights to different links. For example, we say that "the patient has a moderately high fever" rather than "the patient has or does not have a fever". In our work, we are mainly interested in the imprecise knowledge representation problem. We believe that the most appropriate way is to build ontologies called "Fuzzy ontologies". Imprecise knowledge representation is not our unique problem, since the real word is very dynamic; we are also interested in how to evolve these representations. So it will be very interesting to represent these knowledge using fuzzy ontologies as conceptual model and guarantee there evolutions.

In the literature we can find several researches revolve around imprecise knowledge, however, there are still some issues to be further studied; one of them is the evolution problem which is still a tedious field. Authors in [17] assured that a coherent process of ontology evolution is still rarely discussed, a reconstruction process is preferred to evolution one since the creation of ontologies, especially from large text corpus, is a well understood problem [2, 13].

Ontology evolution deals with the problem of incorporating new information in an existing ontology such as new individuals; in this paper, and as a first step in the problem of fuzzy ontology evolution, we are interested in the classification of new individuals in a fuzzy ontology. Classification is the main reasoning mechanism associated with the class-instance representation model. It is a process that, from a structured knowledge base and a new object, finds the proper location of the new object in the knowledge base.

The remainder of this paper is organized as follows. Section 2 provides an overview of the ontology evolution field. In Section 3, we discuss the problem of imprecise knowledge representation and how to use fuzzy logic and fuzzy set theory to take into account the representation of imprecise knowledge and then we try to define the concept of fuzzy ontology. In section 4, we present our proposed classification algorithm in which an individual is connected to his more specific concepts in a fuzzy hierarchy. Finally Section 5 concludes the paper and suggests some future work

II ONTOLOGY EVOLUTION

Ontology is a specification of a shared conceptualization of a domain [9], a change may be appeared in the domain knowledge, or for some need, a change may be caused by the conceptualization or even by the specification of the ontology. Changes in the specification refer to changes in the representation language [7]; this type of change is dealt with in the field of ontology translation (which is out of scope of this paper, for more details the reader is referred at [8]). Since there is no static domain and no perfect conceptualization, changes in the domain or its conceptualization are very common. An ontology evolution is the process of modifying ontology in response to a certain change in the domain or its conceptualization [6].

doi:10.4304/jetwi.6.1.9-14

Introducing new information to one part of the ontology can induce inconsistencies in other parts of the same ontology, the identification and resolution of any inconsistencies that may arise as a result of a change is one of the most important tasks to be performed during ontology evolution process [14], for that reason we can say that ontology evolution is the process of modifying an ontology while maintaining its validity and consistence.

Ontology evolution is mainly of two types: Ontology Population and Ontology Enrichment [11]. Ontology Population refers to the fact when the new information is the adding of new individuals for already existing concepts in the ontology. Ontology Enrichment refers to the fact when the new information is the changing of the ontology structure, for example adding (deleting) new concepts or properties. Then ontology is enriched by the structure change and populated by the new individuals. Figure 1 is a graphical illustration of these two types of ontology evolution.

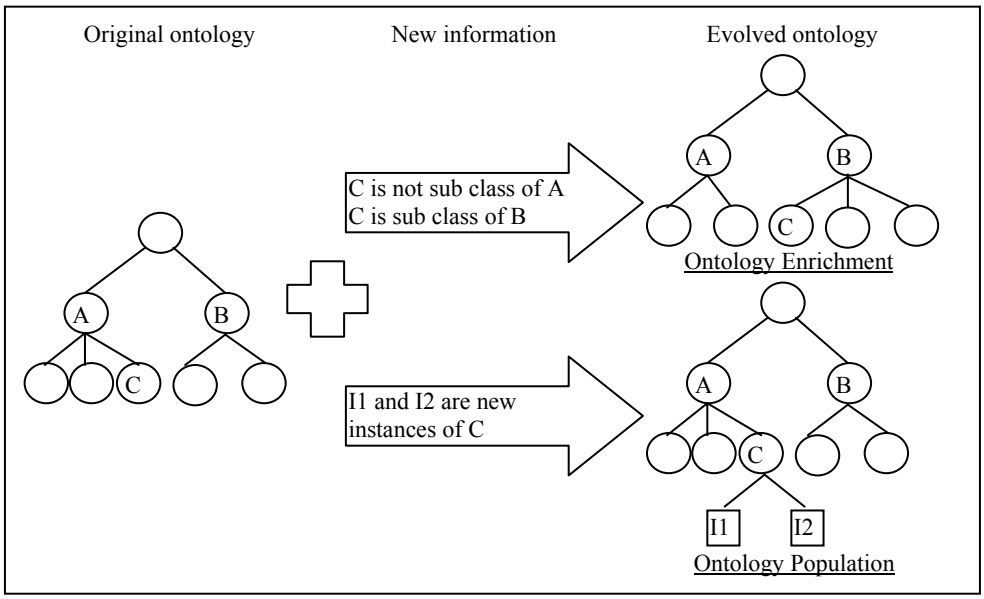


Fig 1: Ontology enrichment and population.

The process of evolution accepts as input a consistent[1] ontology and a set of new information, after a series of phases, the process generates as result a new consistent version of the same ontology. Maedche et al. identified six phases of the ontology evolution process [18]:

Change capturing: In this phase, changes to be applied are identified.

Change representation: the identified changes are formally represented by a finite sequence of elementary changes [2] for example Add (Delete) Concept/ Instance_Of/ Axiom… this decomposition is not always desirable as this might cause a set of unnecessary changes if each change is applied alone. To avoid these needless changes, it should be possible to represent changes in a highest level using the so called composite changes[3] that represent a group of elementary changes applied together for example Merge concepts, Move properties...

Semantics of change: The resulting effects of the required changes are identified in this phase and if there are problems caused by these changes, they will be also identified and resolved in order to guarantee the consistency of the ontology at the end of the process.

Change implementation: When the changes are approved by the user, they will be physically applied to the ontology.

Change propagation: after the modification of the ontology, it will be necessary to propagate the changes to all dependent applications.

Change validation: this phase allows reviewing the changes and possibly undoing them, if desired.

III HANDLING IMPRECISE KNOWLEDGE

The human being reasoning is often based on fuzzy knowledge. To solve everyday problems, he uses knowledge he doubts their validity (uncertain) or poorly expressed due to the complexity of the problem (imprecise). Despite this, it is often possible to solve these complex problems without needing to model them. According to [1], it is often useful to model the behavior of a human operator with the system rather than modeling the system itself. It is also preferable to describe this system with global quantifiers rather than using precise numerical values. Fuzzy logic was introduced as an extension of Boolean logic [19], this logic is not to be precise in the statements, but instead to respond to vague proposals, that requiring some degree of uncertainty.

[1] A consistent state of an ontology is defined in [18] as the state in which all constraints, which are defined on the structure and content of an ontology are satisfied. An example of the structural constraints is the need to define the domain and the range for each relation in the ontology. Content constraints are related to the axioms in the ontology

[2] In the same reference, authors have identified 17 elementary changes.

[3] Always in the same reference, authors have identified 12 composite changes.

III.1 Fuzzy Set Theory and Fuzzy Logic

Fuzzy logic is designed to solve the problem of the representation of uncertain and imprecise knowledge. It allows the characterization of elements in a "gradual" way. It was introduced by LA Zadeh in the late 60s as an extension of Boolean logic [19]. In classical set theory, two situations can be considered: Elements either belong to a set or not. The classical set theory does not take into account several situations frequently encountered in our daily life: It will be very difficult to say its hot today because heat is a progressive concept. If for a temperature of 25 °, we say it's hot, is it not hot with a temperature of 24.8 °?

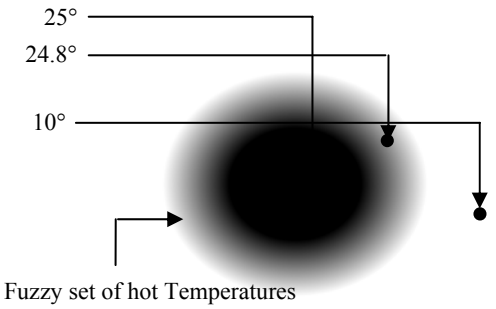


Fig 2: Fuzzy set and partial membership

This example shows that 25° belongs completely to the fuzzy set, so it's a hot temperature; 24.8° belongs partly to the fuzzy set, so this is a moderately hot temperature. When 10°, it does not belong to the fuzzy set, therefore it is not a hot temperature.

Definition: membership function: let X be a set of elements, a fuzzy subset A of X is defined by a function called *membership function* μ_A (x) or simply A(x) which take any value from the real interval [0, 1]. A membership function A(x) is characterized by the following mapping:

$$A(x): x \rightarrow [0, 1], \forall x \in A$$

An element belongs to a fuzzy set to some degree; this is in fact the value taken by the membership function of the fuzzy set at the considered point. As in the classical case, 0 means no-membership and 1 full membership, but now a value between 0 and 1 represents the extent to which x can be considered as an element of A. Membership degrees are calculated based on some specific functions (see Figure 3), we present here the most frequently used:

- *Crisp function:*

$$C(x; a, b) = \begin{cases} 1 & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

- *Trapezoidal function*:

$$T(x; a, b, c, d) = \begin{cases} (x - a)/(b - a) & \text{if } x \in [a, b] \\ 1 & \text{if } x \in [b, c] \\ (d - x)/(d - c) & \text{if } x \in [c, d] \\ 0 & \text{otherwise} \end{cases}$$

- *Right shoulder function:*

$$R(x; a, b) = \begin{cases} 0 & \text{if } x < a \\ (x - a)/(b - a) & \text{if } x \in [a, b] \\ 1 & \text{if } x > b \end{cases}$$

Fuzzy set theory is designed to take into account this kind of situations, where elements can belong to a defined fuzzy set with a certain degree. In [1] fuzzy set theory is defined as a theory based on the notion of partial membership. Each element is partially or gradually belongs to the defined fuzzy sets. The contours of each fuzzy set (see Figure 2) are not "net", but "fuzzy" or "gradual".

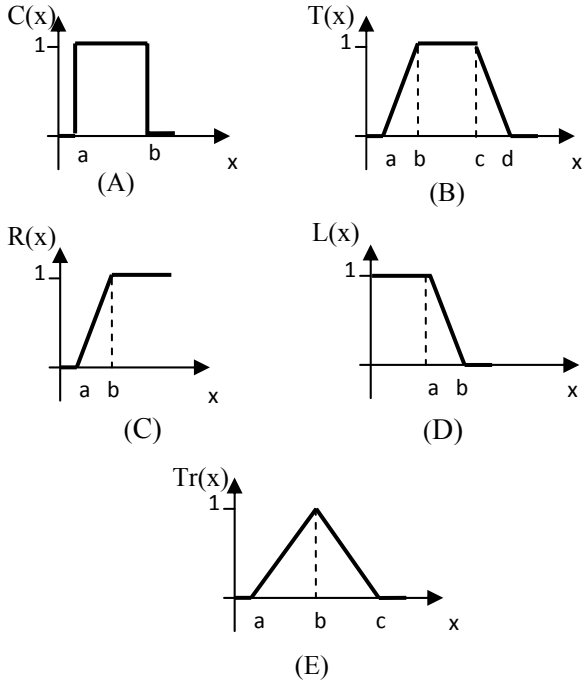


Fig 3: Crisp Function (A), Trapezoïdal Function (B), Right shoulder Function (C), Left shoulder Function (D), Triangular Function (E)

- *Left shoulder function:*

$$L(x; a, b) = \begin{cases} 1 & \text{if } x < a \\ (b - x)/(b - a) & \text{if } x \in [a, b] \\ 0 & \text{if } x > b \end{cases}$$

- *Triangular function:*

$$Tr(x; a, b, c) = \begin{cases} (x - a)/(b - a) & \text{if } x \in [a, b[\\ 1 & \text{if } x = b \\ (c - x)/(c - b) & \text{if } x \in]b, c] \\ 0 & \text{otherwise} \end{cases}$$

III.2 Fuzzy Ontology

Several definitions have been given to describe the term "ontology" [10, 20, and 15]. This diversity of definitions provides different viewpoints but especially complementary, they all revolve around the same goal: design the real world with its semantic constraints. When it comes to conceive imprecise or imperfect knowledge, this will be the role of *fuzzy ontologies* : building of these ontologies is based on fuzzy logic.

Fuzzy ontology consists of two types of components: crisp components (crisp concepts and roles, instances and axioms) and fuzzy components (fuzzy concepts and roles) this components are used to represent the vagueness and imperfection of the real world knowledge [3].

Crisp concepts: if a concept can have a clear and complete definition in which there is no fuzzy properties, this is a crisp concept like car, person, male…

Crisp roles: a crisp role represents the presence or absence of association between the instances of two crisp concepts such as "People *live* in houses", or tow fuzzy concepts like" Rich-People *Drive* Fast-Cars".

Fuzzy concepts: fuzzy concept is described as a concept defined on the basis of a particular value of a linguistic variable relative to the universe of discourse [12]. These linguistic variables represent the fuzzy properties of the concept (age, size, color degradation ...) so that we can represent the uncertainty of fuzzy concepts [3]. Each linguistic variable takes its values in a set of linguistic terms, consider the variable "Size" for example, we can define the following terms: "Small, Average, Tall…" which may become fuzzy concepts: "Small-person, Average-person, Tall-person …"

Fuzzy roles: fuzzy roles are the generalization of crisp roles in which we can allow various degrees of association between instances of crisp concepts such as "Hotels are *close to* the airport", or fuzzy concepts like "Small-People *appears to be* young".

Instances: The membership of an instance to a crisp concept is complete, in the case of fuzzy concept, the instance belongs partly to the concept, and its membership degree is determined by the value taken by the membership function of the instance to the fuzzy concept (modeled as a fuzzy set).

Taking the example of the fuzzy concept "Average-person", the membership degrees of its instances are determined by the values taken by its membership function, which is Trapezoidal type (see Figure 4).

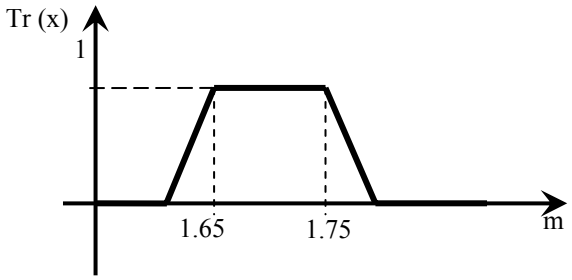


Fig 4: Membership function of the fuzzy concept "Average-person"

The membership degrees to this fuzzy concept are calculated using the following formulas:

$$T(x)=\begin{cases}(1.65 - x)/(1.65 - 1.60) & \text{if } x \in [1.60, 1.65[\\ 1 & \text{if } x \in [1.65, 1.75] \\ (x - 1.75)/(1.80 - 1.75) & \text{if } x \in]1.75, 1.80] \\ 0 & \text{otherwise}\end{cases}$$

Axioms: concept or role assertion axioms can be tainted with uncertainty as we have said; the membership of an instance to a fuzzy concept is partial. Same for the ontology hierarchy: sometimes, we cannot say that a concept is subsumed by another; this cannot be so sure. Axioms are also used to express the formulas of the membership functions [12] (given previously).

IV FUZZY ONTOLOGY EVOLUTION: CLASSIFICATION OF A NEW INDIVIDUAL

In this section we will present a classification algorithm of a new individual. This algorithm starts with an individual which we have a total or partial knowledge (some attributes may not be valued) and a hierarchy graph of the fuzzy ontology. The goal of this algorithm is to find the most specialized concepts to which the individual belongs by bringing him down as low as possible in the concept hierarchy.

Our algorithm is based on the multi-viewpoint classification algorithm of TROPS model presented in [4, 5], in our work, we are not interested in the multi-view point representation model, but how to find the appropriate place of the new individual in an imprecise knowledge hierarchy.

The classification algorithm is based on a loop; the new individual begins in an initial stable state and at each step of the loop, the ontology changes from one stable state to another stable state closer to the purpose which is the attachment of the new individual to his more specialized concepts.

The classification algorithm consists of five procedures: Obtaining-Information, Initial-State-Construction, Matching, Marks-Propagation and Choosing-Next-Concept.

Obtaining-Information
Initial-State-Construction
While not (finished) **do**
  Matching
  Marks-Propagation
  Choosing-Next-Concept
End while

IV.1 Obtaining Information

Before launching the classification, the user must provide some information of the new individual; this information is a set of pairs (attribute, value). The algorithm accepts the "unknown" value, which means that the attribute is not valued. Thus the algorithm can classify incomplete instances.

IV.2 Initial State Construction

This procedure consists in creating from the initial information about the individual a stable state by attaching this individual to its membership concepts (this information comes from the user), if the user has no such information, the individual is attached to the concept root of the hierarchy. Once the initial state constructed, the classification is to repeat the loop of the three procedures Matching, Marks-Propagation and Choosing-Next-Concept.

IV.3 Matching

The matching procedure checks the membership of the individual to the current concept. In what follows, we present two types of membership functions: membership function with two values and membership function with three values:

IV.3.1 Membership Function with Two Values

Membership function with two values (Boolean membership function) is used in Boolean logic (classical), it takes its values from the universe of discourse and returns the value *"True"* if the instance belongs to the class, and the value *"* false*"* otherwise. For example, taking the integer class "pair", its membership function can be described as follows:

$$Pair\ (x) = \begin{cases} True & if\ x\ mod\ 2 = 0 \\ False & if\ x\ mod\ 2 \neq 0 \end{cases}$$

The membership function with two values is well suited for the consideration of complete and precise knowledge for which we can say with certainty that they belong to a particular class or not, which is not the case in our knowledge base, for that reason, we present the membership function with three values

IV.3.2 Membership Function with Three Values

In a fuzzy ontology, we manage imprecise and incomplete knowledge. The scope of the membership function is increased here to accept the value "possible." The function returns the value "possible" for an incomplete instance if the knowledge we have of this instance does not allow affirming or denying its membership to the class. This is the principle of fuzzy logic using a multi valued membership (several membership degrees). A membership function with three values can be described as follows:

$$C(x) = \begin{cases} sure & if\ x \in C \\ impossible & if\ x \notin C \\ possible & otherwise \end{cases}$$

In our Matching procedure, we use the membership function with three values. The comparison between an individual and a concept can give three different results:

- *Sure* if the individual belongs to the concept.
- *Impossible* if the individual is in contradiction with the concept.
- *Possible* if it is not in contradiction with the concept, but missing information to be sure of its membership.

Based on this function, the purpose of the matching procedure is to mark the current concept by one of the three marks: "sure", "possible" or "impossible":

- A concept *C* is marked *"sure"* for the individual *A* (*A* belongs to *C)* if for each attribute of *C* the value of this attribute in *A* satisfies the constraints of *C* (interval, domain, etc...). So, this membership can be determined only if *A* is complete and satisfies the constraints of *C*.
- A concept *C* is marked *"impossible"* for the individual *A* (*A* do not belongs to *C*), when the value of an attribute of *A* does not satisfy the constraints set for this attribute in *C*, here we do not take into account the incompleteness of the individual *A*.
- If *A* has no value for some attributes defined in *C*, and if the valued attributes of *A* are not in contradiction with *C*, we say that the concept *C* is *possible* for the individual *A* (the membership of *A* to the concept *C* cannot be determined because it is missing information).

The marks allocation is based on the satisfaction of the attributes constraints in the concept, if it is a fuzzy attribute; the Matching procedure first begins to calculate the membership degree of the individual to the fuzzy concept based on the attribute value:

If membership degree >0 **then** constraint satisfied

If membership degree ≤0 **then** constraint not satisfied

The user can define another constraint satisfaction threshold; in case he needs to be more sure of the individual membership to the concept, for example, he can define the 0.3 threshold instead of 0:

If membership degree < 0.3 **then** constraint not satisfied

IV.4 Marks-Propagation

The purpose of this procedure is to minimizing the number of concepts to be tested by propagating marks to some concepts based on certain rules:

1- If a class is marked "impossible", all its sub-classes will be marked "impossible".
2- If a class is marked "sure", all its super-classes will be marked "sure".
3- If the current class C is marked "sure" (impossible), ∀D, D≡C (synonym), D will be marked "sure" (impossible).
4- If the current class C is marked "sure" (impossible), ∀D, D ≡ ¬ C (opposite), D will be marked "impossible" (sure).

IV.5 Choosing-Next-Concept

After the marks propagation, the classification algorithm chooses a new concept to the membership test. Unlike the classification algorithm of TROPS model, our algorithm is not based on the hypothesis of the exclusiveness of sisters classes.[4]

In our conceptualization, fuzzy concepts are modeled as fuzzy sets [3]. The strength of fuzzy logic in knowledge representation comes from the intersection between the fuzzy sets, thus an element can belong to several fuzzy sets with different membership degrees. Therefore, individual can belong to several concepts at the same level of the hierarchy.

Considering the following conceptualization, in which we define two fuzzy concepts as follows:

Medium-size-person: a person with a size between 1.60 and 1.80 m.

Tall–Person: a person with a height exceeding 1.75 m.

And an individual with a height of 1.77 m, he will belong to both concepts "Average–Person and Tall–Person" with two different membership degrees related to the concepts membership functions.

For this reason, our classification algorithm tests all the concepts of the current level before bringing the individual down in the hierarchy.

[4] In the TROPS model, sisters classes describe mutually exclusive sets. Thus, if an instance belongs to one of these classes, it cannot belong to any other class

IV.6 Stopping the Algorithm

The classification algorithm may terminate for one of the following reasons:

1- The matching is complete, and there are no more concepts to be tested:
   - Because the goal is reached and the individual is classified the lowest possible in the hierarchy.
   - Or because he does not belong to the hierarchy and he is not classified.
2- The user wants to stop the classification.

V CONCLUSION

In this paper, we have proposed an algorithm for reasoning with imprecise ontological knowledge. As a first step in the problem of fuzzy ontology evolution, we have proposed this algorithm in order to classify new individuals in such ontology. The underlying key of our algorithm is that it allows the classification of incomplete instances in a fuzzy ontology. The reasoning mechanism that we have used is the individual classification; this reasoning mechanism takes into account the characteristics of the conceptualization of fuzzy ontology: knowledge uncertainty and incompleteness.

As future work, we intend to validate and test the proposed algorithm in an application domain. We would like also to test it in a concept hierarch based on fuzzy subsumption.

REFERENCES

[1] Chevri, F, Guély, F, 1998: La logique floue, Cahier technique n°191, Schneider Electric www.schneider-electric.com, rubrique maîtrise de l'électricité.

[2] Cimiano, P., Vlker, J. 2005. Text2onto - a framework for ontology learning and data-driven change discovery. In Montoyo, A., Munoz, R., Metais, E., eds.: Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB). Volume 3513 of Lecture Notes in Computer Science. Springer. 227-238

[3] Djellal, A, Boufaida, Z. 2012. Conceptualisation d'une Ontologie Floue. 9eme Colloque sur l'Optimisation et les Systèmes d'Information (COSI). 62-73

[4] Djezzar, M, Hemam, M and Boufaida, Z. 2012. Ontological re-classification of individuals: a multi-viewpoints approach. Model and Data Engineering. Springer Berlin Heidelberg. 91-102

[5] Drews, O M. 1993. Raisonnement classificatoire dans une représentation à objets multi-points de vue. Doctoral Thesis. University of Joseph-Fourier-Grenoble I.

[6] Flouris, G and Plexousakis, D. 2005 On Applying the AGM Theory to DLs and OWL. Proceedings of the 4 th International Semantic Web Conference (ISWC-05). 216-231.

[7] Flouris, G. Plexousakis, D. and Antoniou, G. 2006. A Classification of Ontology Change. Poster Proceedings of the 3 rd Italian Semantic Web Workshop, Semantic Web Applications and Perspectives (SWAP-06).

[8] Flouris, G., Manakanatas, D, Kondylakis, H., Plexousakis, D and Antoniou, G. 2008. Ontology change: Classification and survey. The Knowledge Engineering Review, 23(2), 117-152.

[9] Gruber, T R. 1993. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition, 5(2)199-220,

[10] Gruber T R.2009. Ontology, the Encyclopedia of Database System, Ling Liu and M. Tamer Ozsu (Eds.),Springer-Verlag. Cloud Computing Economics: Democratization and Monetization of Services. Journal of Business & Economics Research, 7(6). 1-11.

[11] Khattak, A M. Latif, K. Lee, S Y. Lee, Y K. 2009. Ontology Evolution: A Survey and Future Challenges. The 2nd International Conference on u- and e- Service, Science and Technology (UNESST 09).

[12] Maâlej, S, Ghorbel, H, Bahri, A, Bouaziz, R. 2010. Construction des composants ontologiques flous à partir de corpus de données sémantiques floues. Actes du XXVIII° congrès INFORSID.

[13] Maedche, A. Staab, S. 2001. Ontology learning for the semantic web. IEEE Intelligent Systems and Their Applications 16(2) 72-79

[14] Plessers, P and De Troyer, O. 2006. Resolving inconsistencies in evolving ontologies. In The Semantic Web: Research and Applications. Springer Berlin Heidelberg. 200-214.

[15] Psyché, V., Mendes, O and Bourdeau, J. 2003. Apport de l'ingénierie ontologique aux environnements de formation à distance. Revue des Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation (STICEF), 10, 89-126.

[16] Roberts, R. 2008. Introduction aux technologies du Web Sémantique. AIDAinformazioni. http://www.aidainformazioni.it/pub/roberts122008.pdf. Italie

[17] Scharrenbach, T. Bernstein, A. 2009. On the evolution of ontologies using probabilistic description logics. In: First ESWC Workshop on Inductive Reasoning and Machine Learning on the Semantic Web.

[18] Stojanovic, L., Maedche, A., Motik, B., and Stojanovic, N. 2002. User-driven ontology evolution management. In Knowledge engineering and knowledge management: ontologies and the semantic web. Springer Berlin Heidelberg. 285-300.

[19] Straccia, U. 2001. Reasoning within Fuzzy Description Logics. Journal of Artificial Intelligence Research 14, Submitted 9/00; published 4/01. 137-166.

[20] Studer, R, Benjamins, VR, Fensel, D. 1998. Knowledge Engineering: Principles and Methods. IEEE Transactions on Data and Knowledge Engineering.

Semantic Technologies Applying to Data Warehouses Federation

Mouhni Naoual
University Cady Ayyad
Faculty of sciences and technology
Marrakesh-Morocco
nmouhni@gmail.com

Elkalay Abderrafiaa
University Cady Ayyad
Faculty of sciences and technology
Marrakesh-Morocco
elkalay@hotmail.fr

***Abstract*—Business intelligence is based on the existence of main components including : data warehouses .The data warehouse is a specialized database which main task is to provide quick access to data in analysis objective. But in some cases it is necessary to use a set of data warehouses to provide a complete information . This structure is what we call federation, and even if the components are physically separated, they are logically seen as a single component. Generally, these items are heterogeneous which make it difficult to create the logical federation schema ,and the execution of user queries a complicated mission. In this paper, we will fill this gap by proposing a model for logical federation schema creation based on ontology, in order to treat different schema types (star , snow flack) including the treatment of hierarchies dimension too.**

***General Terms*—Data warehousing, Database systems, Information systems, Algorithm**

***Index Terms*—Data warehouse Federation, Ontology, Hierarchical dimension, Schema Integration**

I INTRODUCTION

A Data Warehouse represents the enterprise-wide "single source of truth" and corporate memory of all business process data [3], it is "a subject oriented, non-volatile, integrated, time variant collection of data in support of management's decisions." as defined by Bill Inmon in 1990, the father of data warehouses.

In some cases, one data warehouse is not sufficient to provide a complete information about a fact, which makes grouping multiple data warehouses the only solution. e.g. in the context of a hotel chain that is geographically distributed in many countries, it may have several heterogeneous warehouses to store and analyse data about customers reservations.

this set of warehouses is what we call " a data warehouse federation".

Federated data warehouses are different than distributed Data warehouses, in order that distributed data warehouses can refer different subjects and there is a strict rule in data distribution (horizontal, vertical...) which make it easy to integrate the query results by using join or sum operations [7].

In federated system, the user send his query without having an idea about the location of data or its structure, the set of data warehouses is seen as a whole and the result is the combination of data warehouses components results.

The components in FDWS (Federated data warehouse system) can differ in aspects such as : data model, query language and data semantic [9].

So, a FDWS must contains the following elements [8] :

- An integration procedure of the schemas of the component warehouses giving the logical schema of the federation.
- A query language for user who does not need to know the schemas of the component warehouses.
- A procedure which enables decomposition of user queries to the federation into sub-queries which are sent to the component warehouses

The warehouse federation system management is first based on a logical schema called the federation schema , which integrates all the components schemas. to create this schema, we must integrate all the other local schemas, without loosing information. During this integration, it may be difficult to decide keeping or not an information by using the procedure shows in[8], which compare every measure to the one in the existing federation schema, if it exists only the location of this measure which is characterized by the couple $\left(D_0^i, b'\_name\right)$ is added, where :

D_0^i : represents the fact table in the data warehouse i.

doi:10.4304/jetwi.6.1.15-19

b'_name : represent the name of the measure
else, a new measure is added to the schema .

An algorithm is implemented to integrate dimensions attributes, respecting the same logic.

In fact, this algorithm presents its limits in case we have a measures or a dimension attributes that refer to the same subject, and represented by two different terms in data warehouses local schemas and it doesn't treat the relationship that could been between attributes and the case of hierarchical dimensions.

Our approach consist of using an application ontology defined in [10] as "a description of knowledge necessary to achieve a particular task and that allows to use the same programming language as the application programming language ", to fill this gap instead of using only Meta data that does not fully represent the semantic relationship between local schema measures and dimension attributes, and those of the federation schema.

In this article, we propose an ontology based data warehouses federation management system to solve the problem of semantic heterogeneity during federation schema creation, based on hotel chain data warehouse sources.

Moreover, in our knowledge, there is no studies that used ontology in a federation context to solve this problem, which justify our choice.

Then in section 2; we present and analysis in summary a set of related works.

II RELATED WORKS

In all domain research, It is always worth considering the others work , discuss it and check if we can refine and extend it for our particular purpose.

In computer sciences, reusing existing sources is one of the reasons that made the development of this domain possible.

Warehouses federation according to Sheth and Larson [9], and that appears in [7] and [4], is a set of data warehouses that are heterogeneous, autonomous and dispersed. Every component can continue its local operations and at the same time participate in federation.

It's for the better that all the integration operations be done without interrupting the process of component data warehouses.

There are no many studies on the data warehouse federation, however, R. Kern, K. Ryk, and Ngoc Thanh Nguyen, proposed a framework for building logical schema and query decomposition in data warehouse federations [7], they developed an algorithm to integrate component schemas into one global logical federation schema.

But this algorithm presents some limits in order to treat the case of warehouses with star schema only, and it doesn't consider the hierarchical dimensions and all the heterogeneity types, which are described in [9] as the difference in structure, where different data models provides two different structural primitives. then , differences in constraints ,differences in query languages and semantic heterogeneity.

Semantic heterogeneity, is one of the biggest problem that faces information integration nowadays, it occurs when two synonym terms from two different sources describe the same subject [1] (e.g: schedule and timetable are synonyms but we have to show it to the system) .

one of the solutions to fill this gap is using ontology, which is according to [6]" ontology is a formal explicit description of concepts in a domain of discourse (classes (sometimes called concepts)), properties of each concept describing various features and attributes of the concept (slots (sometimes called roles or properties)), and restrictions on slots (facets (sometimes called role restrictions)). An ontology together with a set of individual instances of classes constitutes a knowledge base. In reality, there is a fine line where the ontology ends and the knowledge base begins."

According to their use, we distinguish many types of ontologies, Generic Ontology, Domain ontology, Application ontology, Representation ontology, The ontology of methods, tasks and resolution of problems, Light ontology and rich ontology [2].

Even if using ontology may resolve the heterogeneity problem in federated data warehouses, it is not yet used in this context, and all the solutions proposed are based on Meta data repositories, which solve the problem of structure definition but not the semantic issues.

III OUR CONTRIBUTION

III.1 Presentation of the Solution

Our work is an extension to [7] algorithm to create the global logical federation schema .

R. Kern, K. Ryk, and N. Nguyen, proposed an algorithm of integration of component schemas into a federated logical schema. They assume that all warehouses are with star schema, so they do not deal with hierarchies in dimensions.

In fact, even with a star schema the hierarchy for dimension are stored are stored in the dimensional table itself.

Whereas, in a snow flack schema, a dimension table have more or more parent tables, and hierarchies are broken into separate tables in snow flake schema . These hierarchies helps to drill down the data from topmost hierarchies to the lowermost hierarchies [5].

Our objective is an improvement of this integration algorithm to cover heterogeneous schemas (snow flack or star schema). And use ontology as a tool to solve the semantic heterogeneity problem instead of using meta data only.

We propose a federation data warehouse management system (FDWS), which cover :

- Improved algorithm for schemas integration using application ontology
- A query analysis and decomposition tool .
- An ontology-based integration Algorithm for query results.

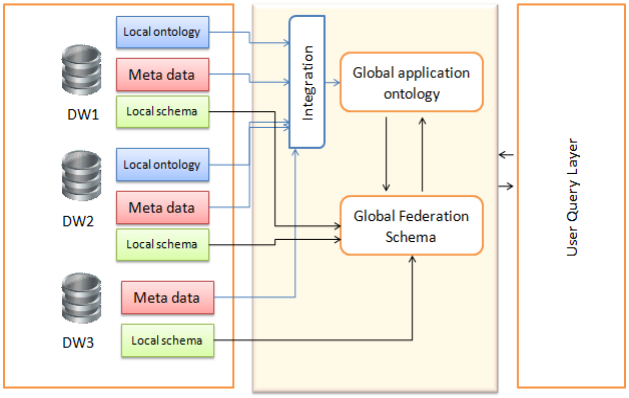


Fig. 1. Structure of the proposed Data warehouses federation management system

1. Every federation component may or not have its own local application ontology, which is written in OWL language describing the semantic of every attribute and measure, and describe the relationship between items and hierarchies of dimensions by using is_a and parentOf relations.
2. This local ontologies are exported to the logical layer ontology repository, besides that a meta data xml file is loaded into the federation system to describe data structure.
3. The user query is analyzed by the FDWS, decomposed, executed on the selected components
4. The query results are integrated using ontology to solve the heterogeneity problem .

III.2 Equations

In our case, the input can be with different schemas types (star , snow flack), so to treat the dimension hierarchies we propose the following algorithm:

Annotation:

We use the same notation as [7].

Input.

P_i^j as the set of parents of a dimension defined by $P_i^j = (D_1, ..., D_n)$

H_p a Data warehouse schema defined as $H_i = (D_0^i, D_1^i, ..., D_{\alpha i}^i)$

F an existing federation schema defined by $F = (D_0, D_1, ..., D_m)$

Output.

F the federation after integration with H_p.

Other notations are used:

a_name : name of attribute a

b_name : name of measure b

$D_x \sim D_y$: D_x is similar to D_y (based on ontology and meta data OR expert's decision)

$a_x \equiv a_y$: a_x is similar to a_y (based on ontology and meta data OR expert's decision)

$b_x \Leftrightarrow b_y$: similar measures (based on ontology and meta data OR expert's decision)

Recall of the Measure integration algorithm.

R. Kern, K. Ryk, and N. Nguyen in [7], defined a measure integration algorithm as follow:

For each measure from input data warehouse try to find corresponding measure in federation schema. If such a measure exists in federation schema add a mapping between them. If none of the federation measures corresponds to the current one add it to the federation and make a mapping between new measure and the current one.

Dimension Integration

In every iteration of the algorithm, the global schema is being updated by integrating parents of dimensions, then integrating the dimension it self .

1. For each dimension from a component schema, using ontology, we extract the set of this dimension parents , this set can be equal to $\varnothing$ or contains one or many items.

a. For each parents item, we look for similarity in F , if it contains a similar structure, we compare its attributes with the existing one, in case two attributes are similar, we add a new location to the attributes inventory represented by the couple (D_y^p, a_name) , else, we add the attribute as a new one to the dimension. In case that the attribute doesn't exist in the target dimension, we add a new attribute.

b. After integrating all the dimension parents, we integrate using the same operations the dimension it self.

foreach D_y^p in $H_p, y = 1,2,...,\alpha p$

if $P_y^p \neq \varnothing$

foreach dimension D_i in P_y^p

if $\exists D_t \in F : D_t \sim D_i$

foreach attribute $a' in D_i$

if $\exists a \in D_t : a \equiv a' \wedge a$ *is characterized by* $(a\_name, list)$

$list = list \bigcup \{(D_i, a'\_name)\}$

else

$D_t = D_t \bigcup \{(a'\_name, \{(D_i, a'\_name)\})\}$

endif

endforeach

else

$D_t = \varnothing$

foreach $a" in$ D_i

$$D_t = D_t \bigcup \left\{ \left(a"\_name, \left\{ \left(D_i, a"\_name \right) \right\} \right) \right\}$$

endforeach

$$F = F \bigcup \{ D_t \}$$

endif

if $\exists D_t \in F : D_t \sim D_y^p$

foreach attribute $a' in D_y^p$

if $\exists a \in D_t : a \equiv a' \wedge a$ *is characterized by* $(a\_name, list)$

$$list = list \bigcup \left\{ \left(D_y^p, a'\_name \right) \right\}$$

else

$$D_t = D_t \bigcup \left\{ \left(a'\_name, \left\{ \left(D_y^p, a'\_name \right) \right\} \right) \right\}$$

endif

endforeach

else

$$D_t = \varnothing$$

foreach a''' *in* D_y^p

$$D_t = D_t \bigcup \left\{ \left(a'''\_name, \left\{ \left(D_y^p, a'''\_name \right) \right\} \right) \right\}$$

endforeach

$$F = F \bigcup \{ D_t \}$$

endif

endforeach

III.3 Comparison and Discussion

The aim of our work, is to propose an improvement of an existing algorithm presented in [8]; in order to take by consideration multiple schema types and less human intervention during the integration process.

TABLE 1.
COMPARISON BETWEEN THE TWO SOLUTIONS

	Existing work	**Our solution using ontology**
Type of schemas covered	Star schemas only	Star, Snow flack, constellation schemas
Hierarchical dimensions integration	No	Yes
Human intervention in the integration process	Important: in every integration step	Only when preparing Component's ontologies

As we can see from the table above (table 1.), which situates our solution among the related works, we can conclude that the existing solution does not treats other schemas types such as snowflack or constellation, which makes it limited as a solution for federated heterogeneous data warehouses that may not always have a star schema type.

Otherwise, the related work used a simple stored taxonomy created manually over all the data warehouses, which could be a complicated mission if the number of components increases, also by using a manually created shared taxonomy that treat only synonyms, the problem of hierarchical dimension persists, that's why we proposed to exploit local ontologies which expresses all types of relations between attributes (parent/child, synonyms, ...) and improve the schema integration algorithm by comparing every two attributes before the integration in the global logical federation schema. the integration in the global logical federation schema is an automated process with less human intervention since the human control is already done during the global ontology creation. this ontology will be used also in other steps of the federation management process such as the query results integration.

IV EXAMPLE

We consider that we have two data warehouses which represent the sources of our federation system. The first component is with start schema, so hierarchies dimension are represented in dimension itself. e.g. the hierarchy $Country \rightarrow \text{Re}\, gion \rightarrow City$.

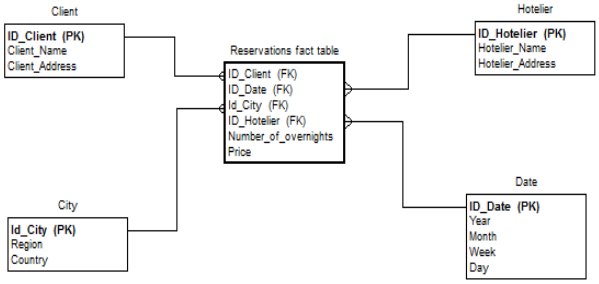


Fig. 2. A star schema of hotel reservations

The second component , is a snow flack schema representing Hotel reservations. this schema contains some hierarchies of dimensions.

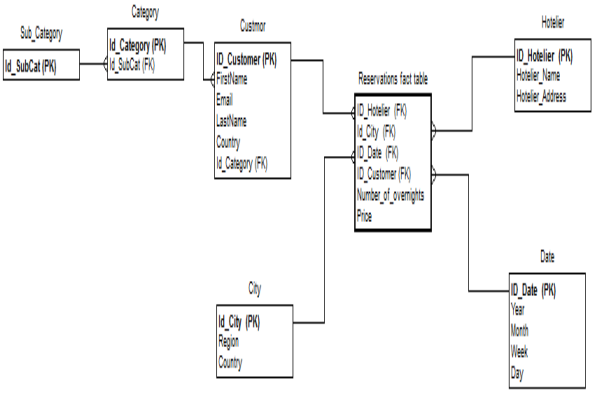


Fig. 3. A snow flack schema for hotel reservations

After applying the proposed integration algorithm we get the global schema as follow:

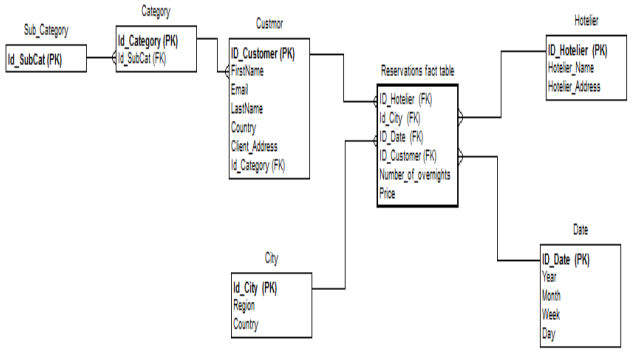


Fig. 4. The result of components schemas integration

Let consider two data ware houses, the first one (Fig2) with a star schema and the second one (Fig3) is a snow flack schema related to a reservation management in a hotel chain.

1. We first extract ontologies and metadata files from different nodes, in the integration layer of the FDWS , then include new entries into the global ontology repository .
2. Then we integrate fact tables by testing the existence of this table in the global federation schema, if it exists, we compare its measures to the existing ones referring to the ontology repository .
3. next step is to integrate dimensions and hierarchies dimension, e.g: we first integrate the client dimension from DW1 into the global schema, then when we try to include Customer dimension , witch is a synonym of client dimension, so referring to the ontology repository we don't add it as a new dimension, and we compare its attributes with clients attributes .

   Based on parrentof relationship mentioned in ontology files, between Customer/client and Category and sub_Category we integrate this hierarchy.

V IMPLEMENTATION

The integration schema algorithm was implemented using Java API Jena, to manipulate RDF language from java application. We are using two data warehouses; the first one with a star schema and has an ontology written in OWL/RD, the second data warehouse with a snow flack schema and has no local ontology.

Metadata files and OWL/RDF files are mapped into xml file and transferred into the network to the Federated data warehouses management system.

VI CONCLUSION

In this article, we have presented a part of our data warehouses federation management system. In particular the process of creating the federation schema based on the integration of local schemas using application ontology. Which makes possible to treat the hierarchies of dimensions by analyzing the parent of relationships, and make the it easy to automate the integration process in federation context.

REFERENCES

[1] D. AWAD, H. T., V. COURBOULAY, A. REVEL 2012. Ontology-based Solution for Data Warehousing in Genetic Neurological Disease *Proceeding of the world congress on Engineering*

[2] GAYO DIALLO, L. D. T. D. L. I., DE LA MODÉLISATION ET DE LA COGNITION 2006. Une Architecture à Base d'Ontologies pour la Gestion Unifiée des Données Structurées et non Structurées. *TIMC-IMAG — UMR,* 5525.

[3] INMON, W. H. 1999. *The Operational Data Store*.

[4] JINDAL, R., ACHARYA, A 2003. Federated Data Warehouse Architecture. *White paper, Wipro Technologies*.

[5] LLC, P. B. S. 2012. *Learn data modeling* [Online]. Available: http://www.learndatamodeling.com/star.php.

[6] MCGUINNESS, N. F. N. A. D. L. 2001. Ontology Development 101: A Guide to Creating Your First Ontology.

[7] R. KERN, K. R., N. NGUYEN 2011. A Framework for Building Logical Schema and Query Decomposition in Data Warehouse Federations. *Computational Collective Intelligence. Technologies and Applications,* Lecture Notes in Computer Science Volume 6922**,** pp 612-622.

[8] R. KERN, T. S., N. NGUYEN 2013. A formal framework for query decomposition and knowledge integration in data warehouse federations. *Expert Systems with Applications,* 40**,** 2592–2606.

[9] SHETH, A. P., LARSON, J.A. September 1990. Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases. *ACM Computing Surveys* Vol. 22, No. 3.

[10] VAN HEIJST, G., SCHREIBER, A. ET WIELINGA, B. 1997. Using explicit ontologies in kbs development. *Int. J. of Human-Computer Studies,* 46(2/3)**,** 183 292.

Vocabulary Persona by Using Ontologies

Gaou Salma 1
1. Faculté des sciences . Université Abdelmalek Essaadi.
Av Jbel Hayane Lot Oulianti N 7 ETG 3, BP 565. Tétouan,MAROC.
E-mail : salma_gaou@hotmail.fr

El Kadiri Kamal Eddine 2
2. Faculté des sciences de Tétouan. Université Abdelmalek Essaadi.
Mhannech II, B. P: 2121. Tétouan, MAROC.
E-mail : elkadiri@uae.ma

***Abstract*—Semantic Web add to all these resources that allow computer systems to "understand" the meaning by accessing structured collections of information and inference rules that can be used to drive reasoning automated to better satisfy user requirements. Standard description of Web resources proposed by the W3C, as the name implies, RDF (Resource Description Framework) is a meta-data used to guide the description of resources, to make it more "structured" information necessary for engines research and, more generally, to all necessary computer automated tool for analyzing web pages. The web is a new web sematique or all Web resources are described by metadata, which allows machines better use of these resources. Considering as a foundation specification FOAF (Friend Of A Friend), we use semantic structures (RDFa) to create an ontology and technologies in which it is implemented.Create a conceptual model (eg, an ontology) for personas and their uses in the context of human-computer interaction we will present some screenshots of execution of application.**

***Index Terms*—Semantic Web, FOAF, Persona, Vocabulary, Ontology, Persona.**

I Introduction

Semantic Web technologies and Semantic Web offers us a new approach to managing information and processes, the fundamental principle is the creation and use of semantic metadata. Using the semantics, we can improve the way information is presented. At its simplest, instead of providing a linear search results list, the results can be grouped by meaning. The use of semantic metadata is also crucial for the integration of information from heterogeneous sources, either within an organization or across organizations. [1]

For their implementation, management information, the integration of effective information and intergration of applications require that all information and underlying be semantically described and managed process, is that they are associated with a machine description of their meaning. This, the basic idea behind the Semantic Web has become very important to late 1990s [2] and in a more developed in the 2000s [3] form. The last half decade has seen intense activity in the development of these ideas, especially under the auspices of the World Wide Web Consortium (W3C) [4]. At the heart of all Semantic Web applications is the use of ontologies. A commonly accepted definition of an ontology is: "An ontology is an explicit and formal specification of a conceptualization of a domain of interest" [5] This definition focuses on two points, first, the conceptualization is formal and therefore allows reasoning by. the computer, and the ontology is designed to perform a particular area of interest. Ontology consists of concepts (also known as the class name), relationships (properties) and instances and axioms. Therefore, a brief definition was proposed as <C,R,I,A> 4-tuple, where C is a set of concepts, R a set of relations, I have a set of instances and A a set of axioms [6]. The first work in Europe and the United States on the definition of ontology languages has converged under the W3C to produce an OWL [7] Web Ontology Language. OWL provides mechanisms for the creation of all the elements of ontology concepts, instances, properties (or relations) and axioms [1]. OWL is based on the Resource Description Framework (RDF) [8], which is essentially a language for data modelling, also defined by the W3C. RDF is based on the graphs, but usually serialized as XML. Essentially, it's triplets: subject, predicate, object [1]. The Semantic Web is simply a web of data described and linked in order to establish the context or semantics that adhere to the defined grammar and language constructs [15]. In this article we try to create a conceptual model (eg, an ontology) for personas and their use in the context of human-computer interaction. These models are widely used in the design, development and science. They are powerful tools to represent the structures and relationships in order to better understand the complex, talk, or use. Without models, we have to make sense of raw unstructured data, without the benefit of an organizing principle. Good models emphasize the main characteristics of structures and relationships they represent and de-emphasize less important details [24].

II FOAF Vocabulary And Syntax

FOAF is not so much an application as an ontology used by many applications. The Friend of a Friend project (FOAF) was one of the first to recognize the simple

doi:10.4304/jetwi.6.1.20-25

power of social networks. The FOAF project provides tools to connect people so a model that contains typical social attributes such as name, email address, interests, etc. [15]. The FOAF project is based on the use of machine readable Web homepages for individuals, groups, businesses and other types of things. For this, we use the "FOAF vocabulary" to provide a set of basic conditions that can be used in these web pages. At the heart of the FOAF project is a set of definitions to use a dictionary of terms that can be used to express claims about the world. The initial objective of FOAF has been on the description of the people, because people are the things that connect most of the other kinds of things that we describe in the Web: they are documents, attend meetings, are shown photos, and so on. FOAF vocabulary definitions presented here are written using a computer language (RDF / OWL) which makes it easy for software to handle some basic facts about the terms of the FOAF vocabulary, and consequently the things described in the FOAF document. FOAF document, unlike a traditional web page, can be combined with other materials to create a FOAF unified database of information [17].

Example[20]

Here is a very basic document describing a person:

```
<foaf:Person>
  <foaf:name>Dan Brickley</foaf:name>

<foaf:mbox_sha1sum>241021fb0e6289f92815fc210f9e9
137262c252e</foaf:mbox_sha1sum>
  <foaf:homepage
rdf:resource="http://rdfweb.org/people/danbri/" />
  <foaf:img
rdf:resource="http://rdfweb.org/people/danbri/mugshot/d
anbri-small.jpeg" />
</foaf:Person>
```

FOAF is an application of the Resource Description Framework (RDF) because the area we describe - people - has so many competing needs a standalone size could not do them any justice. Using RDF, FOAF wins a powerful extensibility mechanism, allowing FOAF based descriptions can be mixed with claims made in any other RDF vocabulary [18].

FOAF, like the Internet itself, is an information system related. It is built using the Semantic Web technology decentralized, and was designed to allow the integration of data across a variety of applications, Web sites and services, and software systems. To achieve this, FOAF adopted a liberal approach to data exchange. It does not require you to say anything about yourself or others, or puts no limits on the things you say or the variety of the Semantic Web vocabularies that can be used to do this. The current specification provides a "dictionary" basic terms about people and the things they do and do [19].

III Persona

A persona is a typical user (the famous archetype), a fictional representation of target users, which can be used to set priorities and guide our design decisions interface [22].

The method is a technique personas Users centered design, initiated by Alan Cooper in 1999. This method can provide a common and shared vision of the users of a service or product, highlighting their goals, expectations and potential brakes, and offering a more engaging format. In the field of web persona is a fictional character who represents a targeted group. When designing a website, it may be necessary to define multiple personas that will represent each type of potential visitors. A good persona is not to stereotype users but to create users that seems real. That is why we have set goals and personality traits realistic. Based on the objectives of the persona and its specific characteristics (identity, age, familiarity with computers ...) you should check that the user interface to meet the needs of users represented by the personas [22].

Personas give us a precise way of thinking and communicating how users behave, how they think, what they want to accomplish and why. Personas are not real people, but they are based on the behaviors and motivations of real people that we have observed and represent them throughout the design process. They are composite archetypes based on behavioral data collected from many actual users encountered in the ethnographic interviews. Personas are based on patterns of behavior that we see in the research phase, so that we formalize in the modeling phase. Using personas, we can develop an understanding of the goals of our users in specific contexts - an essential tool for the use of user research to inform and justify our designs[23].

Personas are a model used to describe the objectives, skills, abilities, experience and technical context of the users. They are detailed descriptions of archetypal users built on understanding, very specific data models on real people. A character is not based on an individual - he is a construct developed through a detailed process, not the result of a search for the "right" (see the character creation for more details) . They are used by the design team (and largest project team) to describe and keep the foreground user (s) for which the system will be built [22].

Personas, like many powerful tools are a simple concept but must be applied with considerable sophistication. It is not enough to whip up a couple of profiles based on stereotypes and generalizations users, it is not particularly useful to include a photograph of a stock job title and call it a "persona." Personas to be effective tools for design, considerable rigor and finesse should be applied to the process of identification of significant and meaningful in user behavior trends and transform these into archetypes that represent a wide range of users [23].

IV Categories and Personas Construction

Persona is a technical approach to ensure the inclusion and optimizing the user experience in the design of an interactive medium. To be useful, the persona must come from real information about users without their creation which may be based on stereotypes. This approach allows one hand to filter and synthesize data users and secondly to unite all stakeholders around key profiles: the main

tasks that should be an answer, user's needs and priorities appear more easily identifiable.

Categories of persona:

- Primary Persona
  This persona is usually designated the primary persona. Indeed, each primary persona requires the presence of its own user interface in a particular application. Knowing that there will be more of a primary persona when their needs cannot be met by the same interface. The fewer the number of primary personas the better.
- Secondary Persona.
  By focusing on the primary persona, the secondary persona's goals and needs can mostly be met. Nevertheless, there are a few needs specific to them that are not a priority for the primary persona. To meet the needs of a secondary persona, there may be small additions to the interface necessary. However, these additions should not negatively affect the experience of the primary persona.
- Supplemental persona
  User persona which are neither primary nor secondary are called supplemental persona. The combination of primary and secondary personas represents completely supplemental persona's needs that are completely satisfied by the solution devised for one of our primaries.
- Customer persona
  Customer personas match customer needs, as discussed by ((auteur) et al., (année)) and their treatment is similar to that of secondary personas.
- Served persona
  Served persona some what differ from persona types discussed previously. Although they are directly affected by the use of the product, they are not users of the product at all.
- Negative persona
  Negative persona they aren't users of the product, like served personas. They mediate between stakeholders and product team members by informing them that there are specific types of users that the product is not being built to serve[23].
  The majority of studies that have been done on personas seems to focus on targets in the context of what distinguishes persona to another [24]. But with a narrative perspective objectives are part of what makes the persona act in a given situation. What difference personas are like in real life, personal characteristics possess persona (age, history, psyche, etc.).

Personas Construction

As a rounded character [25], the persona can be characterized by the following elements, namely:

- Body: body constitutes a human being. Sex, age, look helps the designer emphasise the Persona
- Psyche: to understand motivations for actions we need to understand what lies behind the motivation, the personality.
- Background : job position, family, education, social- and cultural positions explain motivations for actions.
- Emotional state : to know the emotional state furthers engagement in the Persona [26]. Inner needs and goals, ambitions and wishes create a foundation for the emotional state.
- Cacophony : two oppositional character traits [27]. The oppositional traits are what constitute the difference between a stereotype and a rounded character.

  Persona is static but is dynamic when inserted into the actions of the scenario. In the scenario, the persona is in a context, in a specific situation with a specific purpose [28].

Persona Elements :

- Goals ,
- Attitudes (related to your context) ,
- Behaviors & Tasks (in your context) ,
- Name ,
- Photo ,
- Tagline ,
- Demographic Info (brief just to help "humanize" them) ,
- Skill level ,
- Environment ,
- Scenarios (not all but perhaps the highest priority, most common or most telling about their needs) [29].

V The Persona Data Model and The Diagram

5.1 Use Case Diagram

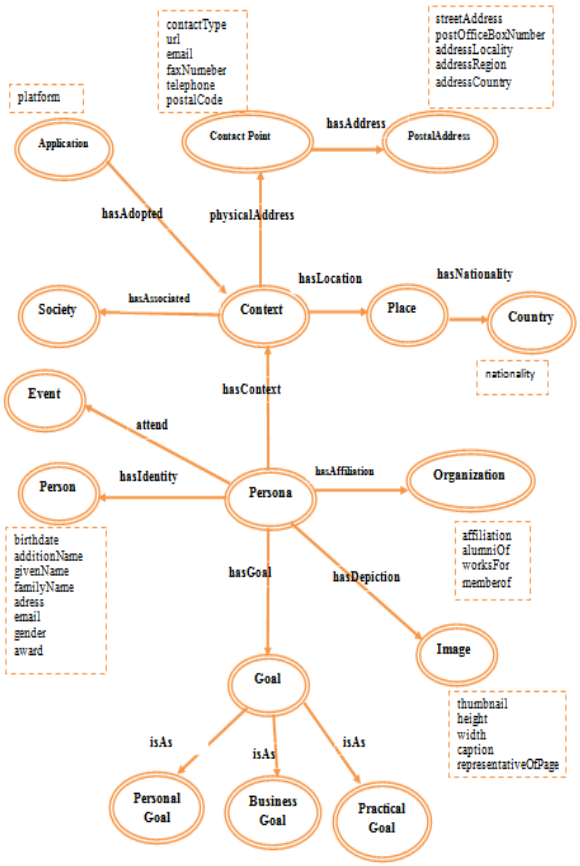


Fig. 1. Use case diagram of the Persona data model[37].

VI PROTEGE

Fig. 2. Gives the list of classes in a hierarchical view.

6.1 This is the List of Data Properties:

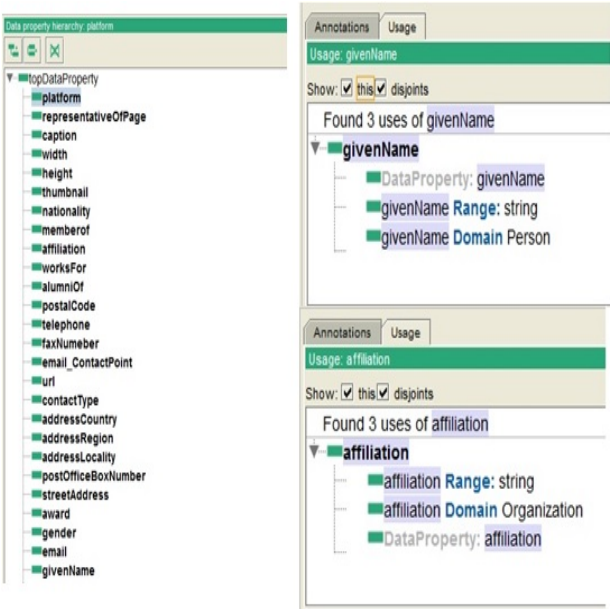


Fig. 3. Extract the Data property defined in the ontology Persona

6.2 This is the List of Object Properties:

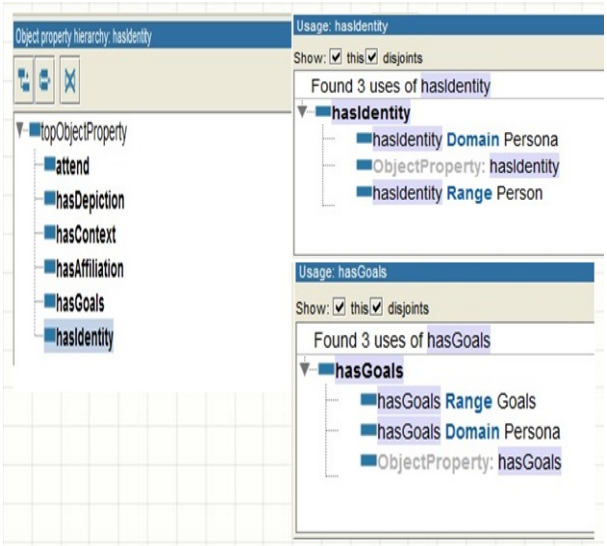


Fig. 4. Extract the Object property defined in the ontology Persona

VII DEPLOYING THE APPLICATION PERSONA

Persona web application developed using JSP and Servlet technologies. To be functional, it must be deployed in an HTTP application server with JSP / Servlet container. Apache Tomcat is both HTTP server (Apache) and Servlet / JSP container, which makes it an ideal candidate for the deployment of our application. The figure below shows the deployment scheme of Persona application.

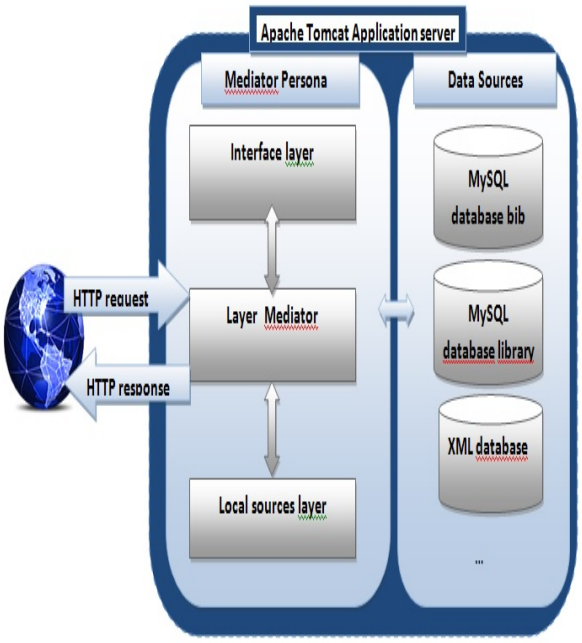


Fig. 5. Diagram of deployment Persona Application

The deployment of the application on Apache Tomcat persona is done by placing the directory containing the application files in the webapps directory of Tomcat.

VIII IMAGES OF APPLICATION EXECUTION

In what follows we will present some screenshots of execution of our application, the homepage is the first window website, where you can access the different

menus of the site persona. It contains five main menus (Identity, Organization, Context, Event, and Goals).

- Context menu contains three submenus (Place, Society and Contact point).
- Goals menu contains three submenus (Personal, Business and Practical).

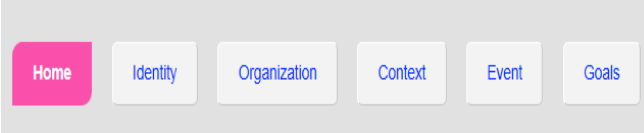


Fig. 6. Menus of the site persona.

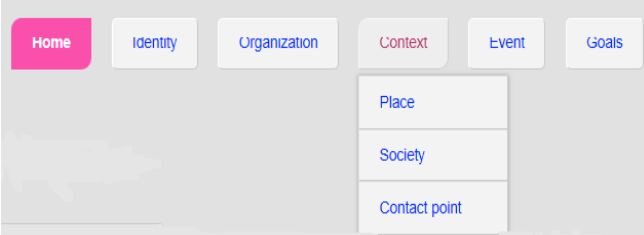


Fig. 7. Context menus active with submenus.

Fig. 8. The first page of website persona

To access this page below the identity, just click on the menu Identity This file contains all information (family name...) for each persona, also includes many features including:

The addition, modification and removal identity.

The same think for context (place, Society and Contact point), event, goals (Personal, Business and Practical), organization.

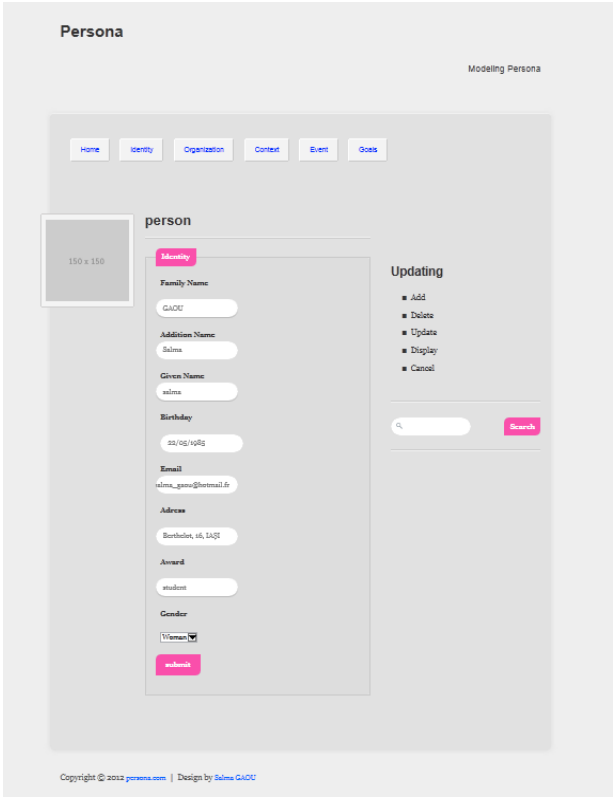


Fig. 9. Page Identity of website persona

CONCLUSIONS

In this article, we introduced the notion of ontology and several methods and tools for ontology engineering. In our study, taking into consideration the characteristics and benefits of ontologies, we focused on the ontology for building a persona vocabulary.

Like many powerful tools, personas are a simple concept but must be applied with considerable sophistication. It is not enough for some profiles based on stereotypes and generalizations users, it is not particularly useful to attach photos to a job title and call it a "persona." Personas to be effective tools for design, rigor and considerable finesse should be applied to the process of identification of significant and meaningful in user behavior trends and transform these into archetypes that represent a wide range of users.

In addition, ontologies have the potential to allow a true knowledge sharing and reuse of heterogeneous agents, both human and computer. A major challenge is still open alignment of different ontologies to provide interoperability between heterogeneous agents. Considering that we were able to achieve the objectives of the party in this document, and we made the right choices about the tools for implementation, so that our work will be a great way for other future projects.

Looking ahead, we plan to expand our study using another technique, namely: first, the development of other ontologies and combine with ours to enrich the vocabulary used for annotating and research. Then test the possibility of reasoning provided by OWL. Finally, our next task will be presented the implementation details of the application, and is going to use for the realization

of our application: Development Tool: JBuilder7, Virtual Machine, Java: J2SDK 1.4.2, Web Server JSP Tomcat 4.x, Libraries: Jena version 2.3 and Servlet / JSP Version 2.3/1.2 (included in Tomcat 4.x).

ACKNOWLEDGMENTS

Research reported in this paper has been partially financed by doctoral research within the Faculty of Computer Science University, Alexandru Ioan Cuza of Iasi in the program <<Eugen Ionescu >> 2011-2012, and thank you for Dr. Sabin-Corneliu Buraga Associate Professor Faculty of Computer Science in Iasi for helping.

REFERENCES

[1] Semantic Web Technologies ,trends and research in ontology- based systems,John Davies, BT, UK,Rudi Studer,University of Karlsruhe,Germany,Paul Warren,BT, UK

[2] Berners-Lee T.1999. Weaving the web. Orion Busines Books.

[3] Berners-LeeT , Hendler J, Lassila O.2001.The semantic web. In Scientific American, May 2001.

[4] W3C staff assigned to the W3C Semantic Web Activityare Sandro Hawke (see also private blog), Ivan Herman (see also private blog), and Eric Prud'hommeaux , Ivan Herman, Semantic Web Activity Lead, 2011-11-07. en ligne : http://www.w3.org/2001/sw

[5] Gruber T.1993.A translation approach to portable ontologie. Knowledge Acquisition 5(2) : 199-220, http://Ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html

[6] Staab S, Stuber R (Eds). 2004. Handbook on ontologies. International Handbooks on Information Systems. Springer : ISBN 3-540-40834-7.

[7] Ivan Herman 6 September 2007: W3C Launches the OWL Working Group, en ligne : http://www.w3.org/2004/OWL/

[8] 10 February 2004, RDF is a standard model for data interchange on the Web. In ligne : http://www.w3.org/RDF/

[9] Tim Berners-Lee, James Hendler, Ora Lassila The Semantic Web, Scientific American, May 2001

[10] Thomas R.Gruber. Ontolingua: A mechanism to mobile ontologies. Knowledge Systems Laboratory Technical Report KSL-91-66, Stanford University, version 3.0, CA, 1992.

[11] Xavier Lacot (2005) Introduction à OWL, un langage XML d'ontologies Web.

[12] Fensel D., Ontologies: A Silver Bullet for knowledge Management and Electronic Commerce, Springer, Berlin-Heidelberg-New York, 2004.

[13] Sticef.org (2004) Le rôle de l'ingénierie ontologique dans le domaine des EIAH.

[14] Sticef.org (2003) Apport de l'ingénierie ontologique aux environnements de formation à distance

[15] Semantic Web Programming,John Hebeler,Matthew Fisher,Ryan Blace,Andrew Perez-Lopez,septembrie 2009

[16] ATHENA WP4 SKOS Workshop Rome, ICCU, 16-17 July 2009 in ligne : http://www.w3c.it/talks/2009/athena/slides.html(52)

[17] D-FOAF - SECURITY ASPECTS IN DISTRIBUTED USER MANAGEMENT SYSTEM, Slawomir Grzonkowski, Adam Gzella, Henryk Krawczyk, Sebastian Ryszard Kruk, Francisco J. Martin-Recuerda Moyano, Tomasz Woroniecki , Gdansk University of Technology, ul. Narutowicza 11/12, 80-952 Gdansk, Poland,

[18] http://www.foaf-projet.org

[19] http://xmlns.com/foaf/spec/

[20] Semantic Web for the Working Ontologist Effective Modeling in RDFS and OWL, Second Edition, Dean Allemang,Jim Hendler,2006.

[21] Linked Data Evolving the Web into a Global Data Space,Tom Heath,Christian Bizer, 2011.

[22] PERSONAS: définition et démarche. J.C. GROSJEAN. Quality Street, Coaching Agile, Management Agile & Lean, Expérience Utilisateur, Tests. 29 décembre 2008

[23] About Face 3 The Essentials of Interaction Design Alan Cooper, Robert Reimann, and Dave Cronin, 2007.

[24] Pruitt, J. and J. Grudin (2003). Personas: Theory and Practice. 2003.

[25] Nielsen, L. (2002). From User to Character. DIS2002, London.

[26] Smith, M. (1995). Engaging Characters: fiction, emotion, and the cinema, Clarendon Press.

[27] Horton, A. (1999). Writing the Character-Centered Screenplay. L.A., University of California Press.

[28] Proceedings of the Third Danish Human-Computer Interaction Research Symposium,Morten Hertzum, Computer Science, Roskilde University ,Simon Heilesen, Communication Studies, Roskilde University ,Writings in Computer Science, No. 98 ,Roskilde University, Roskilde, Denmark, 2003.

[29] Persona Creation Added by Daphne Ogle, last edited by Allison Bloodworth on May 26, 2009.

[30] Personas Added by Gary Thompson, last edited by Daphne Ogle on May 25, 2010.

[31] Riichiro Mizoguchi. The role of ontological engineering in the field of ILE

[32] Natalya F.Noy and Deborah L.McGuinness. Ontology development 101: A guide to creating your first ontology.

[33] Michael k.Smith, Chris Welty, Deborah L.McGuinness. OWL Web Ontology Language-Reference. http://www.w3.org/TR/2004/REC-owl-ref-20040210/ (online June 16, 2005).

[34] Tr.Gruber:. A translation approach to mobile ontology specification. Knowledge Acquisition 5 (2): pp. 199-220, 1993.

[35] Gomez-Perez. Recent developments in matters of design, maintenance and use of ontologies. 3emes meetings TIA Terminology and Artificial Intelligence, Vol 19, pp. 9-20, 2000.

[36] Bachimont:. Commitment semantic and ontological commitment: design and construction of ontologies in knowledge engineering, in J. Charlet, M. Zacklad, G. Kassel, D. Bourigault, eds, Knowledge Engineering:. Recent evolutions and new challenges, Eyrolles, pp. 305-323, 2000.

[37] S. Gaou, K. E. El Kadiri, and *S.CORNELIU BURAGA*. Journal International Journal of Computer Science Information and Engineering Technologies l'article << USE OF ONTOLOGIES IN MODELING PERSONA>>.

Unified Approach for Building Heterogeneous Artifacts and Consistency Rules

Mounir Zekkaoui
Laboratory LIST
FST-Tangier
Tangier-Morocco
zekkaoui@gmail.com

Abdelhadi Fennan
Laboratory LIST
FST-Tangier
Tangier-Morocco
fennan@fstt.ac.ma

***Abstract*—It has become increasingly difficult to ensure consistency between all artifacts in complex software applications, and manage the impact of their development throughout the development process.**
Computer assistance in detecting and resolving inconsistency issues can help improve the quality of designs and development of software.
In this article, we propose a unified approach to representation of different heterogeneous artifacts and a uniform formalism to express methodological consistency rules based on traces of construction and we validated our approach by building a check engine in order to detect inconsistency.

***General Terms*—Design, Verification.**

***Index Terms*—Artifact, Meta-model Construction, Consistency, Software Engineering, Inconsistency Rules, Construction Operations, Check Engine.**

I. INTRODUCTION

We ask Software Engineering has been described as a discipline of description [2]. Software engineers make use of a large number of different artifacts, including source code, analysis models and design, unit tests, XML deployment descriptors, the user guides, among many others. Since these artifacts may evolve over the time through participation and collaboration of many engineers throughout the development process, [3] establishing and maintaining consistency among descriptions presents several problems:

- descriptions vary greatly in their formality and precision;
- individual descriptions may themselves be ill-formed or self-contradictory;
- descriptions evolve throughout the life cycle at different rates; and
- checking consistency of a large, arbitrary set of descriptions is computationally expensive.

We use the term inconsistency to denote any situation in which a set of descriptions does not obey some relationship that should hold between them [16]. The relationship between descriptions can be expressed as a consistency rule against which the descriptions can be checked. In current practice, some rules may be captured in descriptions of the development process; others may be embedded in development tools. However, the majority of such rules are not captured anywhere [3].

Spanoudakis and Zisman [1] define six inconsistency management activities that should be undertaken. The first activity, inconsistency detection, is of special interest as it defines the foundation of the whole process. Considering this activity, two families of approaches are identified: the logic-based approaches and the model checking approaches. The logic-based approaches are defined by the use of some formal inference techniques to detect any kind of model inconsistency. The model checking approaches deploy dedicated model verification algorithms that are well suited to detect specific behavioural inconsistencies but are not well adapted to other kinds of inconsistencies.

The approach called « consistency management » believes that it is impossible to ensure the global consistency of all software artifacts at all times. Any artifact can be temporarily inconsistent. The main problem of this approach lies in tracking inconsistencies. It is necessary to detect the introduction of new inconsistencies and removing existing inconsistencies in successive changes made by developers on artifacts, without impeding the progress of the development process.

It is important to note that current approach do not in general work on homogeneous artifacts (eg model objects [6, 7, 9, 11, 13, 14 et 15]), or using pivots formats (such as XML [4 et 10]) to hide the heterogeneity. Moreover,

doi:10.4304/jetwi.6.1.26-31

they cannot generally cope with the evolution of different heterogeneous artifacts.

In this article, we present CMAC, our approach to managing consistency which has the particularity of being based on construction operations of software artifacts. CMAC detects the presence or absence of inconsistency artifacts. Inconsistencies are specified by logical rules on construction operations. This representation has the advantage of supporting the implementation of incremental detection providing performance gains very interesting. Moreover, it allows the definition of methodological rules of inconsistency to specify temporal orders between construction operations.

The remainder of this paper is structured as follows. Section 2 describes, accurately, how we attacked the problem, the methods and tools we used and how we did (the meta-model construction, unified formalism for managing inconsistency). Section 3 presents the results of our approach (the prototype we built) and we conclude in the last section.

II METHODOLOGY

Understanding of the software and the acquisition of knowledge about the system are essential for all activities in software engineering. The term artifact means any entity falling within the scope of software development. However, it is both difficult and complex to identify the data inconsistency across all artifacts starting from the artifact changed. A change to any software artifact must be taken into account and it will treat by controlling the consistency rules defined in relation to other artifacts. It is then necessary to dispose an abstract and unified representation of software artifacts to facilitate the expression and management of consistency between these heterogeneous artifacts.

To check the consistency rules, we proposed CMAC approach that identifies the elements that do not comply with the consistency rules of artifacts. CMAC is composed a meta-model of structural and unified representation of the different software artifacts and a uniform mechanism for expressing consistency rules in these artifacts.

2.1 Artifact Construction

We considered that all artifacts, regardless of their type, are comparable to typed graphs [17]. An Artifact is then composed of elements. Each element is typed. It can reference other elements. All references are also typed. The artifacts are interrelated them at different levels of granularity [18]. All artifacts can be represented according to the hierarchisation in levels by the pair $< \sum_{Lv} , \sum_{art} >$ where $\sum_{Lv}$ is the set of all levels. The k^{th} artifact of the j^{th} level is represented by the pair $< Lv_j , art_k >$. For example, $<Lv_{Class}, Calculator>$ denotes the fact that Calculator is an artifact belonging to the class level.

It is important to note that the current approaches extract the data either by browsing all the artifacts with a listening on construction operations of developers at the end of to store in a database [5, 7, 8, 11 et 12], or through the conversion of all artifacts into XML before applying the consistency rules [4].

Extract all artifacts and listening different construction operations is a very heavy process to install it in a development environment. In our approach, we allow users (who are often responsible for the development) to extract artifacts by level who need and stored in the database at the end of applying different consistency rules (eg classes, methods, attributes, beans deployment descriptor).

For more complex artifacts such as compounds artifacts, texts files and documents of business rules. We adopt an approach that is defined manually by the expert of evolution. This means that artifacts are extracted manually or by specific algorithms implemented by artifact level such as java and xml files.

The figure (Fig 2.1.1) shows the meta-model we proposed, to present the structure of uniform representation of artifacts. Of specific algorithms defined by artifacts level and other added manually by experts following the approach adopted. The event listener adapted for each type of artifacts are listening to the construction operation (creation, modification, deletion) to represent all artifacts in meta-model, specifying temporal orders between construction operations, for better control of software evolution.

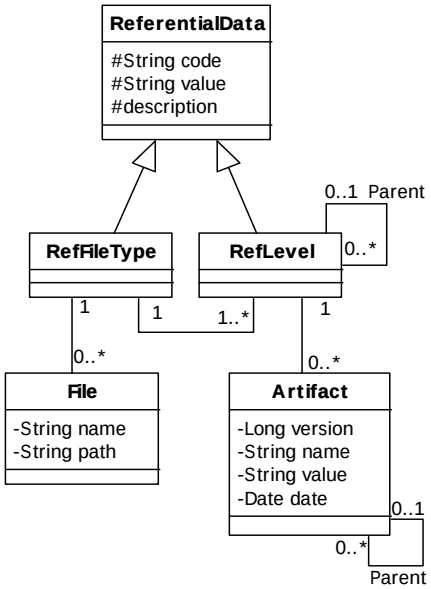


Fig 2.1.1:Meta-model of Artifact construction

We tried to provide a simple and scalable meta-model to represent all types of artifacts.

We considered that all artifacts are stored in files. Each file type must have hierarchical abstro-granular levels (eg class level which contains methods which can find the parameters). Artifacts can also be represented hierarchically, an artifact can have elements (artifacts), and each element can have other elements (artifacts), and so on. Each artifact belongs to a level (set manually by experts).

Taking the following example, a class java « *Calculator.java* » (*Fig 2.1.3*).

```
package ma.organization.calculation;
public class Calculator {
    private int operand1;
    private int operand2;
    public Calculator(int pOperand1, int pOperand2) {
        this.operand1 = pOperand1;
        this.operand2 = pOperand2;
    }
....
}
```

Fig 2.1.2: Sample java file "Calculator.java"

To meet these rules consistency, we propose a unified approach to extract only the artifacts which will need it, that means the developer responsible for the evolution specifies the different artifacts to extract by level, and it is through a form choosing the file type and artifacts levels to extract. For complex files (not supported by the application) such as text files and documents, specific algorithms can beings developed following the adopted approach end to enrich the form data.

Below is represented the different levels and artifacts they may be extracted from the previous example at the end of meet consistency rules.

***<Lv**$_{package}$,ma.organization.calculation>*
***<Lv**$_{Class}$, Calculator >*
***<Lv**$_{attribut}$, operand1>*
***<Lv**$_{attribut}$, operand1>*
***<Lv**$_{constructor}$, Calculator >*
***<Lv**$_{parameter}$, pOperand1>*
***<Lv**$_{parameter}$, pOperand2>*

2.2 Engine of Consistency Rules

Several classifications of consistency rules have been provided in [1] and others. Our goal is not to define a new classification of consistency rules, but rather to provide a uniform mechanism (engine of rules) for dealing with artifacts inconsistency regardless of their types.

For the consistency rules, we proposed to define them as of relationships between the elements defined in the meta-model construction (chapter 2.1). In the case of our model, we allow users (who are often responsible for the development) to define dependencies between software artifacts. These relationships are represented as of logical formulas defined by the development engineer through an intuitive interface (Eclipse plugin under construction). For more complex relationships we adopt a specific language that is to define them manually by the expert of evolution.

Modeling the inter-relationships artifacts is a complex and very important task. We consider three types of relationships (Fig 2.2.1).

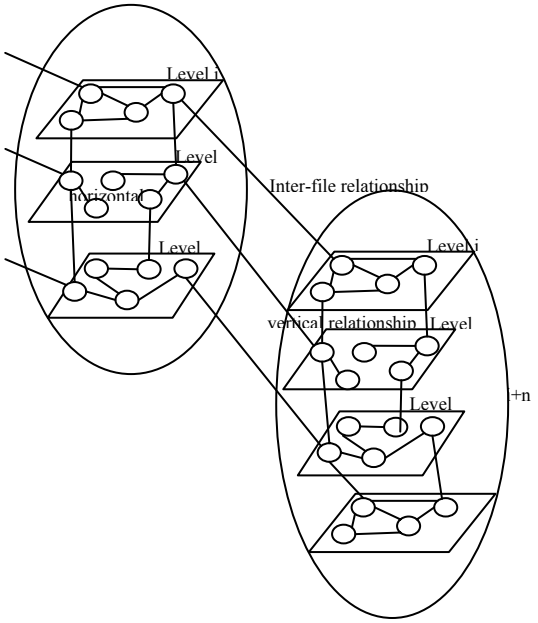


Fig 2.1.1: Classification of artifact relationships

They are:

1. **Inter-files relationships:** These relationships connect artifacts belonging to two different file systems. This is the case for example of the relationship between a UML class and a Java class that implements or the relationship between java class and the deployment descriptor (xml file).
2. **Horizontal relationships:** they represent different kinds of semantic links in the same file and linking artifacts of the same granular level. This is particularly the case of the call relationship between two methods or the inheritance relationship between two classes, ...
3. **Vertical relationships:** they connect two artifacts belonging to the same file at different granular levels. An example of this type of relationship is the one between a class of these attributes, a method body or block the instructions that compose it, ...

Figure 2.2.2 presents our proposal of a meta-model of construction and consistency rules between the different artifacts for monitoring the impact of their evolution (detect violations of methodological constraints)

We considered that a consistency rule « *Rule* » is composed of several conditions « *Condition* ».

A condition can be either:

1. A composition of several sub condition « *Condition* ».
2. Or in the form of two parameters « *ParameterA & B* » and one operation « *RefOperation* ».

« *ParameterA* » may have the exact value of artifact or after the application of the method « *RefMethode* ».

« *ParameterB* » same principle as the « *ParameterA* », except that it can also have a value entered by the user.

« *RefMethod* » specific methods that can be applied sue parameters (eg StringToInt, NumberOfChart and others).

« *RefOperation* » can have the value of the following symbols (>, <, = =, equals and others), usually all the signs used in the "if" statement of the Java language.

(e.g. the rule *"(StringToInt("23")> 3)*", the sign ">" represents the « *RefOperation* », "*23*" is the value of « *Artifact* », "*StringToInt("23")*" is the value of « *ParameterA* » and "3" is the value of « *ParameterB* »).
The user can specify an error message « *Message* » in case of inconsistency, as he may specify the severity level for each rule (in case of "*BLOCKING*" the system stops the application if it is started).
« *RuleState* » containing a state history of each rule (*successful* and *not succesful*) together with the current state (*historized: false*), this table is powered to each inconsistency check.

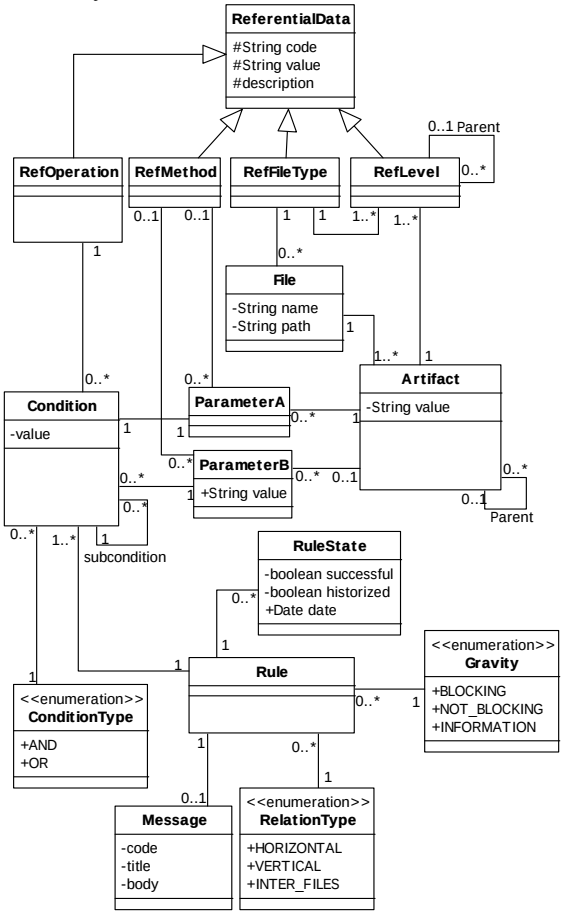


Fig 2.2.2: Meta-model of construction and consistency rules

2.3 Detection of Inconsistency

We have implemented a system for the automatic verification of inconsistency. The program is developed with Java language, it is based on the extracted artifacts and consistency rules defined by the developer before displaying the different specific messages to each inconsistency in the console. The key principle is to convert the rules defined in the meta-model to queries written in Java language.

A consistency rule is usually written in the following form :

IF ***conditions*** Then ***actions(A)*** *Else* ***actions(B)***

A condition can be either a set of other conditions

condition = $\sum_{\$condition}$

Or may have the following form :

Condition = *$ConditionType* ***(****$ParameterA $RefOperation $ParameterB****)***

- ***$ConditionType :*** can have two values {OR | AND}.
- ***$RefOperation :*** can have the following symbols *{<, >, =, ==, !=, equals, ...}* almost all the symbols used in the ***if*** condition of Java language.
- ***$ParameterA :*** *$RefMethode($Artefact)*, is the value of an artifact, be simple or well after application of a specific method (e.g. converting a string to a numeric value).
- ***$ParameterB :*** *{Value | $RefMethode($Artefact)}* the same principle as *$ParameterA*, except that *$ParameterB* can have a value entered by the developer.

actions(A) = *(update curent $RuleState) and (add new $RuleState)*

If the conditions of the rule are true then → update the current state *$RuleState(historized:false)* of the rule by changing the value of the attribute *historized* to *true*, then add a new object *$RuleState* with *successful:true*, *historized:false* and *date:new Date()*.

actions(B) = *(update curent $RuleState) & (add new $RuleState)*

If the conditions of the rules are wrong then → 1- Update the current state *$RuleState(historized:false)* of the rule by changing the value of the attribute *historized* to *true*. 2- Add a new object *$RuleState* with s*uccessful:false, historized :false* and *date:new Date()*. 3- if *$Gravity:blocking* → Stop processing and display specific message (or generic message if the developer has not specified an error message) in the console, else if *$Gravity:{not blocking|information}* display the message in the console without stop processing.

III RESULTS

As a proof of concept of our approach, we have built a prototype in the Java programming language. The key idea is that the artifacts and consistency rules are represented in a unified meta-model and a Java program is based on the meta-model for check the inconsistency. This java prototype will be integrated into the Eclipse development environment (CMAC eclipse plugin in progress) and development tool *Checkstyle* (in progress). Users can trigger the inconsistency check or set the start settings (at project start, activate the listeners, ...). Of listeners per file and level artifacts are listening to different construction operations at the end update the relevant data in the meta-model (e.g. if a file has been updated → update all artifacts of this file).

3.1 Architecture

Our prototype is composed of two main components: the *"Artifact Builder"* and the "*Rules Engine"* (see Figure 3.1.1).

"Artifact Builder" is responsible for the following two main tasks:

1. The extraction of artifacts from different file types and levels selected by developers, and then store them in a unified database of construction.
2. Listening to the various construction operations (add, update and delete) made by the developers on the files and monitored artifacts levels in order to update the database of construction.

"*Rules Engine"* allows users to specify different consistency rules through an intuitive graphical interface or through the specific language for complex rules. Also contains a program for detecting inconsistencies. It analyzes the rules stored in the database (conversion to the specific language before compilation) and produces an inconsistency detection report.

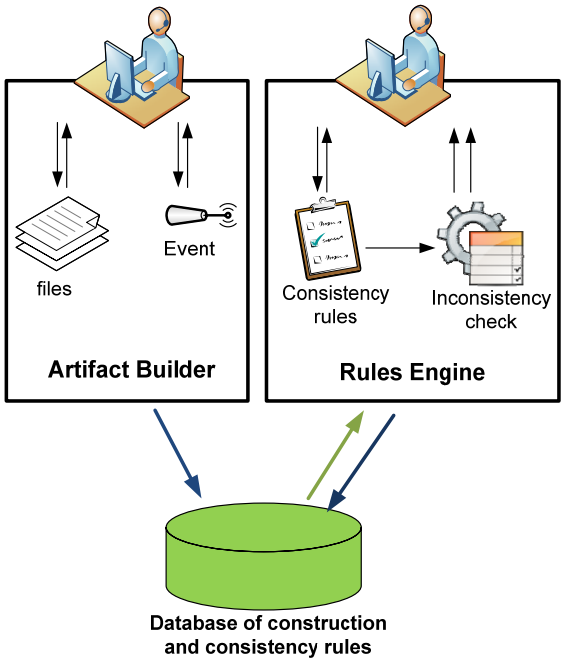


Fig 3.1.1: CMAC architecture

3.2 Artifact Builder

We have defined two kinds of Artifact Builder. One is a *file reader* and the other is an *event listener*. The file reader can scan all file type, and outputs all the artifacts that correspond to different levels selected by the user (existing levels or implemented by the developer). The event listener can receive events raised by the various modification made by developers in order to update artifacts in the database. This enables the incremental checking of inconsistencies.

The *file reader* artifact builder has been developed in Java on top of CMAC framework using Strategy pattern. Strategy is a software design pattern, whereby an algorithm's behavior can be selected at runtime based on the type of data [19], in our case the type of data is the level of artifact. We proposed to implement a specific class by artifact level, and for more complex levels the developer can add more classes by level following the adopted approach. In each class we defined the specific algorithm for extracting and recording the artifacts in the database with corresponding levels.

The *event listener* has been also developed in Java using the *WatchService* API of Java 7. The objective is to monitor the various changes to the files and Artifacts levels, then make the call to the file reader in order to update data in the database.

3.3 Rules Engine

The principle of the Rules Engine is very simple, it is a CMAC plugin interface allowing users to define consistency rules as logical operations between artifacts values, already extracted by the *Artifact Builder*.

For the simple rules (those defined between two artifacts for example), we have proposed to define them manually through an interface proposed by the plugin, and for more complex rules, the developers must go through an option of specific language that is available in the same plugin interface.

The two ways of declaring consistency rules deliver output recorded in a unified meta-model.

Rules Engine also contains a program for detecting inconsistencies. Our approach is to convert the different registered rules to queries in java language (section 2.3), and at the output of the execution of these conditions, the program update states history of each rule and display specific messages when inconsistency.

IV Conclusion and Perspectives

In this article, we presented an approach to managing consistency in the design and software development. This approach includes a unified meta-model representation of different heterogeneous artifacts and a uniform formalism to express methodological consistency rules to support information engineering in software development projects. And we validated our approach by building a system (check engine) to detect violations of methodological rules.

Due to its modular architecture, CMAC serves as a good basis to implement a whole chain of processing tools on top of it. Many tasks in the context of software development and reverse engineering require not only to retrieve information of interest, but also to process it in some meaningful way. Tasks like pre-processing, refactoring, consistency validation etc., all require a custom extraction tool. For example, block modification of any entity before updating the class in the UML model. This example shows that CMAC is a good basis for the implementation of tools and languages for the development of software engineering.

We are currently working to implement the plugin CMAC implementing our approach which can facilitate the management of inconsistency between heterogeneous artifacts. Finally, we wanted to integrate our approach in several object mapping and refactoring tools, citing for example "*Checkstyle"* and "*Dozer"* tools.

REFERENCES

[1] Spanoudakis, G. Zisman, A. 2001. Inconsistency Management in Software Engineering: Survey and Open Research Issues, Volume 1 (2001), pp. 329-380.

[2] Jackson, M. 1995. Software Requirements & Specifications: A Lexicon of Practice, Principles, and Prejudices, Addison-Wesley, Wokingham, England.

[3] Bashar, N. Easterbrook, S. and Russo, A. 2000. Leveraging Inconsistency in Software Development.

[4] Eichberg, M. Mezini, M. Ostermann, K. and Schäfer, T. 2004. XIRC: A Kernel for Cross-Artifact Information Engineering in Software Development Environments. WCRE 2004: 182-191.

[5] Blanc, X. Mounier, I. Mougenot, A. and Mens, T. 2008: Detecting model inconsistency through operation-based model construction. ICSE 2008: 511-520.

[6] Egyed, A. 2007. Fixing Inconsistencies in UML Design Models. ICSE 2007: 292-301.

[7] Blanc, X. Mougenot, A. Mounier, I. and Mens, T. 2009. Incremental Detection of Model Inconsistencies based on Model Operations, Conference on Advanced Information Systems Engineering, Springer-Verlag Berlin, Heideberg 2009.

[8] Mougenot, A. Blanc, X. and Gervais, M.-P. 2009. D-Praxis: A Peer-to-Peer Collaborative Model Editing Framework, Distributed Applications and Interoperable Systems, 9th IFIP WG 6.1 International Conference, DAIS 2009, 2009, p. 16-29.

[9] Caffiau, S. Girard, P. Guittet, L. and Blanc, X. 2011. Vérification de cohérence entre modèles de tâches et de dialogue en conception centrée-utilisateur.

[10] Nentwich, C. Capra, L. Emmerich, W. and Finkelstein, A. 2002. xlinkit: A Consistency Checking and Smart Link Generation Service.

[11] Egyed, A. 2009. Automatically Detecting and Tracking Inconsistencies in Software Design Models.

[12] Le Noir, J. Delande, O. Exertier, D. Silva, M. A. A. and Blanc, X. 2011. Operation Based Model Representation: Experiences on Inconsistency Detection.

[13] Liu, W. 2002. RULE-BASED DETECTION OF INCONSISTENCY IN SOFTWARE DESIGN.

[14] Liu, W. Easterbrook, S. and Mylopoulos, J. 2002. RULE-BASED DETECTION OF INCONSISTENCY IN UML MODELS.

[15] Silva, M. A. A. Mougenot, A. Blanc, X. and Bendraou, R. 2008. Towards Automated Inconsistency Handling in Design Models

[16] Nuseibeh, B. Kramer, J. and Finkelstein, A.C.W. 1994. A Framework for Expressing the Relationships between Multiple Views in Requirements Specification, IEEE Trans. Software Eng. pp. 760-773.

[17] Ehrig, H. Prange, U. and Taentzer, G. 2004. Fundamental Theory for Typed Attributed Graph Transformation, Graph Transformations, Second International Conference, ICGT 2004, Springer 2004.

[18] Basson, H. 1998. An integrated model for impact analysis of software change. In International Conference On Osftware Quality Management.

[19] Erich, G. Richard, H. Ralph, J. John, V. 1993. Design Patterns: Abstraction and Reuse of Object-Oriented Design

Crypto-Security Contribution in WSNs

Nourreddine Mitta
Laboratory of Electrical Genius & Energy Systems Ibntofail University-Faculty of Science –Kénitra-Morocco
Noureddine_mitta@yahoo.fr

Rachid Gouri, Hlou Lamari
Laboratory of Electrical Genius & Energy Systems Ibntofail University-Faculty of Science–Kénitra-Morocco
Elgouri.rachid@yahoo.fr

Brahim Moalige
Vice-President of Moroccan Cryptography Association
Ibntofail University-Faculty of Science-Kénitra-Morrocco
Bmoalige@hotmail.coml

***Abstract*—Securing a wireless communication has generally a vital importance, particularly when this communication is in a hostile environment like in wireless sensor networks (WSNs). The problem is how to create cryptographic keys between sensor nodes to ensure secure communications. Limited resources of sensor nodes make a public key cryptosystem such as RSA not feasible. So, most solutions rely on a symmetric cryptosystem. In this paper, we propose a new key mana-gement scheme based on symmetric cryptography which is well adapted to the specific properties of WSNs. The evalu-ation of our solution shows that it minimizes memory occupation, ensures scalability, and resists against the hardest attack: com-promised nodes.**

***General Terms*—Network Security.**

***Index Terms*—WSN, PDKR, Sink, TinyPK, Tiny ECC, SKNP, SPINS, LEAP.**

I Introduction

The convergence of technological advances in micro-electronics and wireless communications has enabled the emergence of a promising area: Wireless Sensors Networks (WSNs). WSNs come from the combination of embedded systems and distributed systems. WSNs have opened the way for a multitude of research areas and the huge interest generated by researchers activities calls for broad fields of applications in the near future.

Sensors appear as miniaturized systems, equipped with a processing unit and storage of data, a unit of wireless transmission and a battery. Organized as a network, the sensors (or nodes) of a WSN, despite resource constraints in computing capacity, storage, and energy, have to play an essential role in quasi all domains of human environment. They are primarily dedicated to collect data from physical phenomena such as monitoring global warming and send them to a base station (also called the sink) [1]. Figure 1 show an example of a WSN composed of ten sensor nodes deployed randomly around a base station. Depending on the size of the deployment area, the transmission range of the sensor nodes, and the base station, sensor nodes can communicate with the base station directly or indirectly by computing a hop-by-hop route to it. Many barriers to the common deployment of WSNs have to be overcome before they can reach their full maturity. Among these obstacles, the security problem is acute and must be addressed adequately and in accordance with the binding characteristics of WSNs. Because of their constraints and their deployment in unattended and hostile environments, the different nodes of a WSN are vulnerable to node compro-mising and also to physical damage [2]. In addition, the use of wireless transmission makes WSNs permeable to all sorts of malicious attacks. Consequently, security is a real challenge to rise.

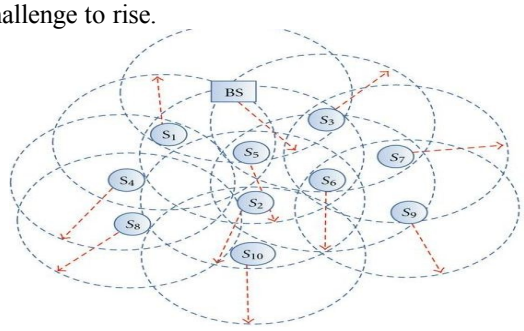


Figure 1 Example of WSN

Several key management protocols for WSNs [3-4-5-6-7-8] were proposed to respond to the security requirements of these environments. Unfortunately, node compromising is rarely or not enough investigated and most of these protocols have a weak resilience to this attack. In this paper, we present a symmetric-based key management solution for WSNs called PDKR (PROTOCOL DIAGRAM KEY RECOVERY SOLUTION for WSNs). PDKR is a simple and robust solution to secure node-to-node and node-to-base station communi-cations. PDKR assumes a random deployment

doi:10.4304/jetwi.6.1.32-39

of nodes. It builds a diagram that spans all the sensor nodes. This diagram allows key refresh with small costs. Simulation results show that PDKR is very resilient to node compromise while preserving energy consumption at the level of sensor nodes.

The rest of the paper is organized as follows. First, we discuss related work in Section 2. In Section 3, we present our solution and give its detailed algorithms. Section 4 is devoted to an analysis and a simulation of the proposed solution. Section 5 concludes our work.

II Background and Motivation

Key management is the process by which cryptographic keys are generated, stored, protected, transferred, loaded, used, and destroyed [3]. Figure 2 lists a selection of existing key mana-gement solutions proposed for WSNs in the literature, for a detailed state of the art see [3-4]. Most existing key mana-gement solutions are based on symmetric cryptography mainly because of its reasonable energy consumption. Asymmetric cryptography involves the use of a pair of keys (public key and private key) to encrypt and decrypt messages. Each node in the network has a public and a private key, the first is known throughout the network, the second is secret, that is, known only by the node. The source node encrypts messa-ges using the public key of the destination node, and this latter uses its private key to decrypt received messages. In symmetric cryptography, the source and the destination use the same key to encrypt and decrypt messages. Asymmetric cryptography offers better resistance against node compromise attack and allows scalability but requires an additional part on software and hardware of the nodes. Some researchers investigated asymmetric cryptographic tools and propose adapted solutions. Examples of such solutions are Tiny Public Key (TinyPK) [9] and Tiny Elliptic Curve Cryptosystem (Tiny ECC) [10].

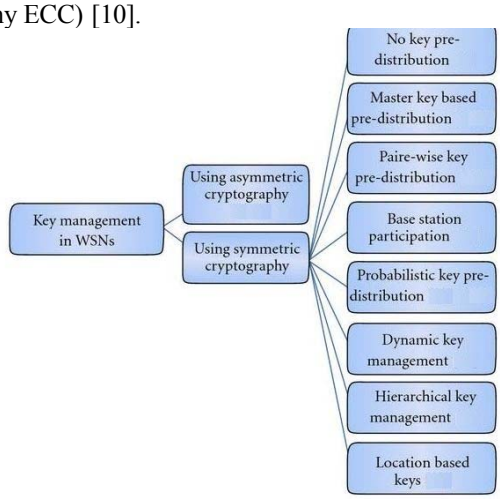


Figure 2 Some Existing key management solutions for WSNs

With symmetric cryptography, the simplest idea is to load secret information in the sensor nodes before their deployment in the network. This secret information deployed in the network may be the secret key itself or auxiliary information that helps nodes to derive the real secret key shared by the nodes. With this secret key, nodes can securely exchange messages [3]. The main disadvantage of this solution is that compromising one node (access to the pre-loaded key) might lead to compromise the entire network. To overcome this limitation, several researchers propose schemes that establish pairwise keys rather than a unique global key. For example in [11], the authors focus on developing cost-saving mechanisms while weakening the threat model. They propose Key Infection, a lightweight security protocol suitable for use in noncritical commodity sensor networks where an attacker can monitor only a fixed percentage a of communication channels. With Key Infection, a node wishing to communicate securely with other nodes simply generates a symmetric key [4] and sends it in the clear to its neighbors.

In [12], Blon describes an optimal class of symmetric key generation systems solution. In this solution, some of the possible link keys in a network of size N are represented as a $(\lambda+1) \times N$ key matrix. The scheme stores small amount of information in each sensor node, so that some pair of nodes can calculate corresponding field of the matrix, and uses it as the link key. This solution is λ-*secure*, meaning that keys are secure if no more than λ nodes are compromised. Another λ-*secure* solution is presented in [13] and called *Polynomial-based key pre-distribution scheme*. This scheme distributes a polynomial share (a partially evaluated polynomial of degreeλ) to each sensor. So, each sensor node stores a polynomial with $(\lambda+1)$ coefficients and every pair of sensor nodes can establish a key using the property of symmetry of polynomials. The solution is λ-*secure*, meaning that coalition of less than $\lambda+1$sensor nodes knows nothing about pairwise keys of others.

In [14], the authors propose *BROSK* (BROadcast Session Key negotiation protocol). With *BROSK* every node broadcasts a message containing its nonce. So, every two neighbouring nodes that hear each other can compute a common key which is function of their two nonces. Neighbouring nodes authenticate themselves with a pre-deployed key which is supposed to be unreachable in the case the node is captured. In [15], the authors propose a variation of this protocol where the pre-deployed key is used only for a restricted period of time during which nodes establish pairwise keys. Then, the pre-deployed key is erased. However, ***HELLO*** messages used to establish pairwise keys are sent in the clear. So, an attacker that captures a node and also eaves drops ***HELLO*** messages can use the IDs and nonces contained in these messages to derive established keys.

Perrig et al. propose in [16] *SPINS*, a key management protocol that relies on a trusted base station to distribute keys. SPINS contains two parts: *SNEP* (Secure Network Encryption Protocol) that protects communications between a node and the base station or between two nodes, and μ ***TESLA*** (micro time efficient streaming loss-tolerant authentication) that serves to authenticate packets coming from the base station. The first part is unsuitable to energy constraint of nodes because any

communication between two nodes must pass through the base station. The second part needs additional memory space to store authentication keys. In [17], the authors propose *LEAP* (Localized Encryption and Authentication Protocol); a key management protocol intended to support a several communication patterns. In this protocol, each node stores four types of keys: individual, pairwise, cluster, and group. An individual key is a key shared between a node and the base station. A pairwise key is shared between a node and each of its neighbours.

A cluster key is a key shared between a node and all neighbouring nodes. A group key is a key common to the entire network. The individual key is preloaded. After deployment, neighbouring nodes establish pairwise keys. They authenticate themselves using a pre-deployed key which is erased as soon as pairwise keys are established. To establish cluster keys and the group key, nodes use broadcasts and message relaying. The protocol uses ***μ Tesla*** [16] to authenticate broadcasts.

Liu et al. propose in [18] *LBKs* (location-based keys) that relies on location information to achieve key management. The keys are established according to the geographical location of sensor nodes. However, knowing the geographical location of nodes is not guaranteed with random deployment. Eschenauer and Gligor [19] propose a scheme based on a random key pre-distribution. In this scheme, each sensor randomly picks a set of keys and their identifiers from a key pool before deployment. Then, a shared-key discovery phase is launched where two neighbours exchange and compare list of identities of keys in their key chains. Basically, each sensor node broadcasts one message and receives one message from each node within its radio range where messages carry key ID lists. So, any pair of nodes has a certain probability to share at least one common key. The challenge of this scheme is to find a good trade off between the size of the key pool and the number of keys stored by nodes to achieve the best probability. The main drawback of this approach is that if the number of compromised nodes increases, the fraction of affected links also increases. Other solutions use the principle of probabilistic key pre-distribution [21, 22] introduced in [19], For example, the authors of [21] suppose that the deployment area is a grid-based structure of $t*n$ cells called groups. Groups contain the same number of sensor nodes. The protocol uses $t*n$ key pools such that neighbouring key pools have more keys in common. Sensor nodes are deployed with the key pool that corresponds to their group in the deployment area. After deployment, nodes sharing keys can communicate directly. Nodes that do not share keys must establish a path key using their neighbours. In [20] the authors propose to increase the amount of key overlap required in the shared-key discovery phase. Their scheme called ***Q-****composite* requires Q common keys to establish a link key. Link between a pair of sensor nodes is set as a hash of all common keys. The scheme improves resilience because the probability that a link is com-promised, when a sensor node is captured, decreases, but probability of key sharing also decreases because a pair of nodes has to share Q keys instead of one.

Eltoweissy et al. [22] propose *EBS* (Exclusion-based System), a dynamic key scheme that assigns each node k keys from a key pool of size $k + m$. If node capture is detected, rekeying occurs throughout the network. However, the authors [22] did not indicate a method for detecting a compromised node. More-over, even if a small number of nodes in the network are compromised, information in the entire network could be discovered as well as in [13].

Section 4 summarizes the properties of these different solutions together with the proposed one within a table.

In general, existing symmetric key management solutions for WSNs focus particularly on the efficiency of key establishment after the deployment of the network. However, they do not deal with key refresh which makes key management dynamic and adds a further difficulty to the task of attackers. Furthermore, existing solutions neglect the effect of captured node attacks.

We develop in this paper a key management framework well adapted to WSNs challenges especially scalability. We focus on establishing a key refresh scheme with minimum costs that allows dealing with the resistance against the hardest attack: node compromising.

III PDKR: A PROTOCOL DIAGRAM KEY RECOVERY SOLUTION FOR WSNS

In this section, we describe a new key management protocol for WSNs. Our main objective is to offer a robust and simple security framework that meets the resource constraints of sensor nodes. The main idea of *PDKR* is to build a diagram in a secure manner and while conserving energy after a random deployment of nodes. Thereafter, this diagram is used for rekeying to save communications. In fact, with a diagram only $log_2(n)$ messages are necessary to rekey a network of n nodes. We begin by presenting the assumptions and notations used in the design of the solution, and then we give the detailed algorithms.

3.1. Assumptions and Notation

Our solution relies on the following assumptions.

(i)The sensor network is static (nodes are not mobile).

(ii)The sensor nodes are homogeneous: the sensor nodes are similar in their processing capacity, communication, energy, and storage.

(iii)The deployment is random: the neighbours of any node are not known prior to deployment.

(iv)An attacker can listen to all traffic, reflect old messages, or inject its own messages.

(v)The compromise of a node implies that all information stored in its memory is known by the attacker.

(vi)The base station has no constraints on the capabilities of computing, storage and cannot be compromised.

(vii)The communication channels are bidirectional; if a node u can receive a message from node v, then u can send a message to v.

(viii)A base station which is generally the sink is responsible for initiating the key management process.

(ix)Each sensor node has a unique identifier.

Table 1 shows the notations that are used to write algorithms in the remaining of the paper. Each sensor node i, including the base station, maintains the following variables.

(i)$Father_i$: the father of the sensor in the final recovery diagram. The base station is the root of the diagram, so $Father_{BS}$ is set to ***null***.

(ii)$Level_i$: the level of the sensor in the diagram. The level of the root is zero, that is $Level_{BS}=0$.

(iii)$Sons_i$: a list containing the identifiers of the sons of node i within the diagram.

(iv)$Neighbours_i$: a list containing the identifiers of the neighbours of node i. This list is maintained to cope with node failure and node capture as described further in this section.

TABLE 1.
NOTATION.

Notation	Description
S_i	i^{th} sensor node in the network. S_i denotes the (unique) identifier of the sensor node.
$\{M\}_k$	The encryption of message M with key k.
$S \rightarrow *: M$	A node S broadcasts the message M, any node in the radius of perception of the BS receives the message M.
$MAC_k(M)$	The Message Authentication Code of the message M with the symmetric key k.
$A \| B$	The concatenation of the information A with information B.
N_i	A nonce generated by node S_i.

The protocol uses the following types of messages.

(1)$\{HELLO, Sender\_ID, Sender\_Level, MAC_{Kr}(Sender\_ID, Sender\_Level, Sender\_Father)_{Kr}\}$: this message is used to construct the recovery diagram. It is encrypted with the pre-deployed key K_r. ***Sender_ID*** is the identifier of the sensor node that sends the message. The ***Sender_Level*** is the position of the node in the diagram. So, the base station which is the root of diagram is at position ***0***.

(2)$\{REFRESH, BS, New\_K_r, Mal\_List, MAC_{KBS}, Si(BS, new\_K_r)\}_{KBS}$, ***Si*** :this message initiated by the base station is used to update key K_r which is shared by all the sensor nodes. ***Mal_List*** is a list of nodes that are suspected to be malicious (captured nodes).

(3)$\{REFRESH\text{-}REQ, S_i, <Mal_{ID}>_{Ki\text{-}BS}\}$: this message is sent by a sensor node to the base station to request a key refresh. A node S_i requests a key refresh when it suspects a neighbour ***Mal_ID*** to be captured.

(4)$\{JOIN, S_n, N_n, MAC_{Kr}(S_n, N_n)\}_{Kr}$: this message is used by a new node to join the network. S_n is the identifier of the new node and N_n is a nonce generated by him.

(5)$\{Join\text{-}Ack, S_i, Level_i, N_n, N_i, MAC_{Kr}(S_i, Level_i, N_i)\}_{Kr}$: this message is used by neighbouring nodes to acknowledge the receipt of a Join request.

(6)$\{Father, S_n, S_i, N_i\}_{Kr}$: this message is used by a new node S_n to inform surrounding nodes that its father in the recovery diagram is node S_i.

3.2. ALGORITHM

Each sensor node S_i is launched with three keys: $K_{i,BS}$, $K_{BS,i}$ and K_r. $K_{i,BS}$ and $K_{BS,i}$ are shared with the base station. They both serve to secure communications between the sensor node S_i and the base station. $K_{i,BS}$ serves to encrypt/decrypt messages sent by S_i to the base station while $K_{BS,i}$ serves to encrypt/decrypt messages sent by the base station to S_i. The aim of using two keys is to make more difficult the task of a cryptanalysis attacker. Key K_r is shared by all nodes of the network; this key is used to encrypt (decrypt) messages immediately after deployment.

After deployment, each node copies its key K_r in its RAM (volatile memory) and removes it from its non-volatile memory (EROM). If an attacker captures (physical access) a node after deployment, he will not have access to the key K_r rapidly. The base station initiates the construction of recovery diagram by broadcasting a ***HELLO*** message as follows:

$$BS \rightarrow *: \{HELLO, BS, 0, null, MAC_{Kr}(BS, 0, null)\}_{Kr} \quad (1)$$

The purpose of this message is to discover neighbouring nodes of the base station. Upon receiving the message for the first times each node S_i set, its father to ***BS*** and sets its level in the diagram to ***1***. Then, S_i broadcasts a similar message that is, $S_i \rightarrow *$: $BS \rightarrow *: \{HELLO, S_i, 1, BS, MAC_{Kr}(S_i, 1, BS)\}_{Kr}$, in its neighbouring hood to allow other nodes to join the diagram and so on until all the nodes join the diagram.

When receiving other ***HELLO*** messages, each node uses them to set its list of sons in the diagram and computes common keys with them or simply constructs its list of neighbours. ***Algorithm*1** describes the detailed actions achieved by a sensor node when receiving a ***HELLO*** message.

Algorithm 1

Recieve
$\{HELLO, Sender\_ID, Sender\_Level, MAC_{Kr}(Sender\_ID, Sender\_Level, Sender\_Father)\}_{Kr}$

If the message is received for the first time **Then**
    $FATHER_i := Sender\_ID$; /* Father receives the identifier of the sender. */
    $Level_i := Sender\_Level + 1$;
If ($Father_i \neq BS$) **Then**
/*Compute a common key with the father */
K_{Si}, $Sender\_node := H_{Kr}(S_i \| Sender\_node \| Level_i)$;

/* if the father is the *BS* a shared key already exists.
$S_i \rightarrow *$: $\{HELLO, S_i, Level_i, Father_i, MAC_{Kr}(S_i, Level_i, Father_i)\}_{Kr}$;

Else

If ($Level_i = Sender\_Level - 1$) **AND** ($Sender\_Father = S_i$) **Then**
Add the node *Sender_ID* to the list of sons;
If $S_i <> BS$ **then**
Compute a shared key with the son *Sender_ID*:
    $K_{Si\ Sender\_ID} := H_{Kr}(S_i \| Sender\_ID \| Level_i + 1)$;

EndIF
Else
Add the node S_i to the list of neighbours;
End IF

Figure 3 shows some steps of the construction of the recovery diagram of the WSN depicted on Figure 1. Figure 3(a) shows the first step where the neighbours of the base station join the diagram. In Figure 3(b), we only show the step during which neighbours of nodes S5 join the diagram. Finally, the complete diagram is depicted in Figure 3(c).

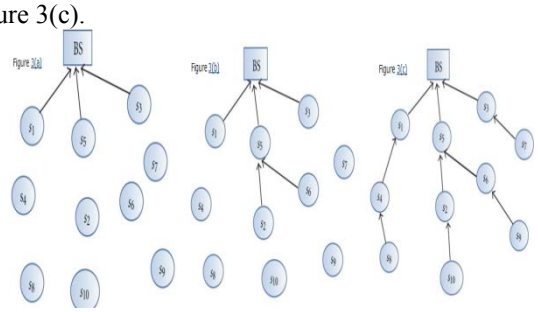


Figure 3. Diagram construction.

3.3. Diagram Maintenance and Rekeying

Once the diagram is constructed, each node shares a symmetric key with the base station, a symmetric key with its father in the Diagram, and the key K_r with the whole network. A rekeying process is launched by the base station periodically to refresh K_r as follows.

(1)The ***BS*** sends to each son S_i a ***Refresh*** message encrypted with their common key $K_{BS,Si}$ $\{REFRESH, BS, New\_K_r, null, MAC_{KBS,Si}(BS, new\_K_r)\}_{KBS,Si}$. Note that in a periodic refresh ***Mal_ID*** is set to ***null***. However, if the base station issues the ***REFRESH*** because of a captured node attack, it includes the ***ID*** of this node in the message.

(2)When a son node of the base station receives the Refresh message, it updates key K_r by new_K_r, *then* it forwards the ***Refresh*** message of the base station to its own sons. The message is encrypted with the symmetric key shared between father and son. So, the ***Refresh*** message goes downward within the Diagram till reaching all the sensor nodes. Thus, all sensor nodes get the new global key.

A rekeying process may also be trigged if malicious/ captured nodes are detected in the network. Because they are resource constrained, sensor nodes cannot implement mechanisms for detecting malicious nodes. For example, putting a sensor node in promiscuous mode is not feasible. Detecting captured nodes is conceivable only if the captured node issues attacks such as sending unnecessary messages to its neighbours to cause energy depletion. When a node S_i suspects one of its neighbours, say S_m, to be captured, two situations are possible.

(i)If the suspected node S_m is the son of S_i, node S_i ignores the messages of S_m and removes it from the list of sons. Then it generates a nonce N_i and sends $\{REFRESH\text{-}REQ, S_i, <S_m>_{Ki\text{-}BS}\}$to its father on the Diagram. This message goes upward until it reaches the ***BS***. This last broadcasts a refresh message $\{REFRESH, BS, New\_K_r, S_m, MAC_{KBS}, S_i(BS, new\_K_r)\}_{KBS,Si}$ in the Diagram. With this message, K_r is updated and all sensor nodes are aware that S_m is malicious.

(ii)If the suspected node S_m is the father of S_i, node S_i must first contact one of its neighbours in the list $neighbours_i$ to find another father and by the way another path to reach the base station. If it succeeds in this task, it issues a ***REFRESH-REQ*** as in the precedent case.

In both cases, it is the base station that decides if it issues a key refresh or not. We suppose that the base station has more means and resources to detect malicious/captured nodes. Captured nodes may also send ***REFRESH-REQ*** messages only to consume energy and to evict honest nodes from the network. If the base station detects this behaviour, it may issue a ***REFRESH*** message that evicts the node that sends the ***REFRESH-REQ*** from the network.

Adding a new node S_n is achieved as follows.

(1) S_n generates a nonce N_n and broadcasts a join message as follows: $S_n \rightarrow *:\ \{JOIN, S_n, N_n, MAC_{Kr}(S_n, N_n)\}_{Kr}$, The ***JOIN*** message is encrypted with the current value of K_r , the key shared by all the sensor nodes. So, the new sensor node must be launched with this key at deployment. Note here that deploying a new sensor must be synchronized with the periodic rekeying process to avoid deploying a new node with an obsolete key.

(2)When receiving the ***JOIN*** message, every node S_i in the transmission range of the new node generates a nonce N_i and responds with the following message:

$$\{JOIN\text{-}Ack, S_i, Level_i, N_n, N_i, MAC_{Kr}(S_i, Level_i, N_i)\}_{Kr} \quad (2)$$

(3)The new node declares the source of the first received message as a father and diffuses the following message:

$$\{Father, S_n, S_i, N_i\}_{Kr}. \quad (3)$$

(4)The father node adds the new son node in its sons list. The surrounding nodes that heard the ***Father*** message add the new node to their list of neighbours.

(5)The father and the son compute their shared keys.

A sensor node may fail or consume totally its energy. In this case, nodes that are attached to him, especially its sons, must search for another path to the base station. So, they ask the nodes in their list of neighbours to take the place of the failed father.

IV EVALUATION

4.1.Comparison with Some Existing Solutions

In this section, we evaluate the performance of our solution and compare it with some existing ones. We use the following metrics to achieve this evaluation.

(i)**Memory complexity**: memory needed to store keys.

(ii)**Communication Complexity**: number of messages exchanged for key management.
(iii)**Key connectivity**: the probability that two nodes (or more) share a key.
(iv)**Resilience against node capture**: this metric measures the impact of a node compromise on the security of the rest of the network. We quantify this metric with the three following values:
(a)Good resilience: the compromised node affects only its neighbours (local influence),
(b)Weak resilience: the compromised node affects its neighbours and also some non neighbouring nodes,
(c)Very weak resilience: if the compromise of one node leads to compromise the whole network.
(v)Scalability: this metric measures the flexibility of the protocol with the size of the network. In other words, the metric shows how the cost of the protocol, that is memory and message overhead, varies when the network becomes larger. Scalability is a very imp-ortant metric to consider when distributed algorithms are proposed, especially for dense WSNs. To quantify scalability, we use the following values:
(a)Very good: the protocol does not induce further costs when the number of nodes in the network increases,
(b)Good: the protocol induces reasonable costs when the number of nodes increases,
(c)Medium: the cost of the protocol depends on the number of nodes.
Table 2 presents the results of com-paring key management protocols described in Section 2 with the pro-posed solution. The different notations used in the table are described in Section 2. In the proposed scheme, a node needs to store initially three keys before deployment. After deployment, each node computes a number of keys that is the function of the number of its neighbours. If a node S_i has d neighbours, where d' are sons, the set of keys stored by the node S_i is two keys (with the base station) + one key (shared by the whole network) $+d'$ keys. The memory complexity of our solution is acceptable for nowadays sensors.

TABLE 2.
COMPARISON WITH SOME EXISTING SOLUTION.

Metrics / Schemes	memory Complexity	communication Complexity	Key connectivity	Resilience against node capture	Scalability
Key Infection [5]	Depends on the number of one hop neighbors (d)	For each node : $2 * d$	100 %	Weak	Good
BROSK [6]	1	$2 * d$	100 %	Very weak	Very good
Lightweight Key Management System [7]	$4+2g$, where g is : number of group in network	$2 * d$	100 %	Very weak	Very good
Blom Scheme [8]	$\lambda(\lambda + 1)$	$d + 1$	100 %	λ- secure	Medium
Polynomial scheme [9]	$\lambda+1$	$d + 1$	100 %	λ- secure	Very good
SPINS [10]	3 + the chain list of keys used by µTESLA	$3*n/2$	100 %	Weak	Medium
Random key pre-distribution [11]	Key pool size (m) + keys identifier s	$d + 1$	Probability that two nodes share a key, say p_1	Depends on m and p_1	Good
Q-composite [12]	$2 * m$	$d + 1$	Probability that two nodes share a key say $p_2 > p_1$	Depends on m and p_2	Medium
Key management using deployment knowledge [13]	d - 1	$d + 1$	Probability that two nodes share a key, say p_3	Depends on d and p_3	Good
Dynamic key management [14]	k keys + keys' identifiers	$d + 1$	Probability that two nodes share a key, say p_4	Depends on k and p_4	Good
LEAP [15]	$(3 * d) + 2$ + keys chain of µTESLA	$(2 * d) + 1$	100 %	Very good	Good
Location-based keys [16]	$2 * d + 1$	$2 * d$	Probability that two nodes share a key, say p_5	λ- secure	Good
Our solution	3 + number of sons	$d + 1$	100 %	Good	Good

The analysis of the communication complexity for the construction of Diagram is measured by the number of messages received and issued by each node. Each node sends a message and receives d messages from its neighbours, that is, $d + 1$ *message* by each node. Our solution has a low-complexity communication over other proposed solutions [11, 4-8], moreover, it is deterministic; key connectivity is equal to one (no concept of probability).

Preserving energy in WSNs is very important and having a low communication complexity means that the **solution minimizes energy**.

4.2. Simulation Results

To validate the results presented by table 2, we also measured the performance of our protocol by simulation. In this section, we provide an overview of our simulation model and some of the results we obtained. We implemented our scheme within the MATLAB framework (MATLAB for Matrix Laboratory is a matrix-based system for scientific and engineering calculation). We considered WSNs with 500 sensor nodes deployed randomly in an area of 250 * 250 meters square. Each sensor node has a signal range of 15 meters.

We first focused on evaluating the average number of messages received by a sensor node after a random deployment. This number of messages is equal to the number of neighbours of the sensor node as explained earlier. Multiple runs were conducted. Each run corresponds to a particular random deployment of the 500 identified sensor nodes. During each run, we measured the number of neighbours of each sensor node. We to-ok the average of the values registered at the level of each sensor node. Figure 4 illustrates the obtained results. It appears that our random deployments generate an average number of neighbours at most equal to 12. If the value of energy consumed by each type of message is known, we can compute the total energy consumed during Diagram construction.

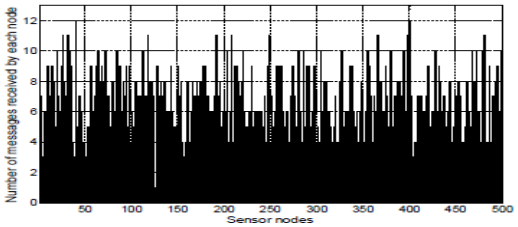


Figure 4 Average numbers of messages for each node in a network of 500 sensor nodes.

We then considered the resilience of our scheme to node capture. We compute the resilience to node capture by the fraction of total communication links that are compromised by the capture of x nodes. We assume that sensor nodes are randomly captured by an attacker. Figure 5 presents the obtained results. It shows that for a network of 500 sensor nodes, an attacker must capture at least 100 sensor nodes to reach all communications in the network. This corresponds to 20% of nodes in the network.

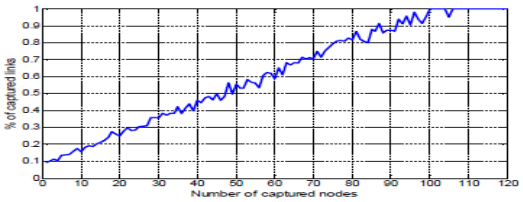


Figure 5. Percentage of corrupted links versus number of captured nodes.

We also carried out multiple experiments on networks of size ranging from 500 to 1000 nodes. In each experiment we computed the maximum number of messages that can be received by a node. We took the average number of this value for each network size. Figure 6 summarizes these results. It shows how the number of messages received by a sensor node evolves when the size of the network becomes larger. The curve is almost logarithmic which is very acceptable.

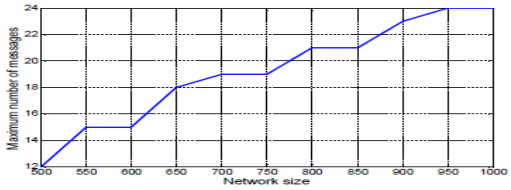


Figure 6. Maximum number of messages received by a node versus network size.

4.3. Security Analysis

In this section, we analyze the security of our solution. As mentioned in the assumptions (cf. Section 3.1), the base station will not be compromised. An outsider attacker, who does not know the key K_r, cannot discover the meaning of messages diffused by nodes after deployment. Nevertheless, an attacker can compromise one or more nodes, so he becomes an insider attacker. The keys of the compromised node can be used for forging wrong messages (reading message, for example,) and also consume nodes' energy by sending useless messages to his sons and father. If the base station can detect an abnormal behaviour of a node, it can lunch a rekeying to refresh K_r and revoke this node. In the following, we analyze the behaviour of our solution for three types of attack.

(1)*HELLO* flood attack: in our proposal, nodes discover their neighbours by sending a ***HELLO*** message encrypted with the key K_r. An attacker without knowing the key K_r could not launch a ***HELLO*** Flood attack.
(2)Sybil attack: in the algorithm, an MAC of the node's identifier, its level, and its father's identifier is calculated to authenticate the sender and the receiver. Therefore, a node cannot play a role of other nodes.
(3)Node capture attack: when a node is captured, this does not affect its neighbours. In fact, after a node is captured, what can an attacker do? Since he has the key shared with the base station, he may send wrong information (lectures) to that base station. The latter may have a mechanism to verify the behaviour of sender nodes. The attacker also has access to the keys of the sons of the victim enabling it thereafter to send unnecessary messages to these sons in order to consume their energy and cause battery depletion. Nodes that detect these useless messages can suspect the captured node and inform the base station with a ***REFRESH-REQ*** message.

V CONCLUSION

Security is a necessity for most applications using WSNs, especially if the sensor nodes are deployed in unsafe areas, such as battlefields, strategic places (airports, critical buildings ...). These sensor nodes operating in difficult access places, without protection and without possibility of recharging their batteries, may be subject to disruptive and malicious actions. Therefore, it is important to provide to them an acceptable security level. The primary objective of WSN nodes is to collect data and transmit them to a decision center. So, this must be done in a trustworthy and safe way. In this paper, we presented a key management solution to WSN that deals with one of the hardest attack: node capture. The main idea of the solution is to quickly and cheaply build a recovery diagram that serves to refresh the shared key with minimum costs. The solution is scalable and uses little memory.

As a perspective of our present work, we plan to use the NS3 simulator to compare the performance of our solution with other solutions from the literature and particularly those of [6-7-8]. We also work on **energy** and a mobile version of our scheme.

REFERENCES

[1] C. Buratti, A.Conti, D. Dardari and R. Verdone, An Overview on Wireless Sensor Networks Technology and Evolution. www.mdpi.com/ journal/sensors, 6869-6896 (2009).
[2] T Kavitha, D Sridharan, Security vulnerabilities in wireless sensor networks: a survey. Journal of Information Assurance and Security 5, 31–44 (2010)
[3] Y Xiao, VK Rayi, B Sun, X Du, F Hu, M Galloway, A survey of key management schemes in wireless sensor networks. Computer Communications 30(11-12), 2314–2341 (2007).
[4] Suman Bala, Gaurav Sharma and Anil K. Verma: Classification of Symmetric Key Management Schemes for Wireless Sensor Networks. International Journal of Security and Its Applications, Vol. 7, No. 2, March, 2013, 117-138.
[5] Tahira Laskar, Debasish Jena : A Survey on Key Management Issues in WSN. International Journal of Engineering and Innovative Technology (IJEIT), Volume 1, Issue 5, May 2012,74-77, ISSN: 2277-3754
[6] Dahai Du, Huagang Xiong, and HailiangWang, An Efficient Key Management Scheme for Wireless Sensor Networks. International Journal of Distributed Sensor Networks, Vol 2012, Article ID 406254, 14 pages
[7] Kamaljit Singh, Lalit Sharma, Hierarchical Group Key Manage-ment using Threshold, International Journal of Computer Applications (0975 – 8887), Volume 63–No.4, February 2013
[8] Saewoom Lee, Youngtae Noh, and Kiseon Kim, Key Schemes for Security Enhanced TEEN Routing Protocol in Wireless Sensor Networks. International Journal
[9] RJ Watro, D Kong, S-F Cuti, C Gardiner, C Lynn, P Kruus, TinyPK: securing sensor networks with public key technology. Proceedings of the 2004 ACM Workshop on

Security of Ad Hoc and Sensor Networks (SASN '04), October 2004, 59–64.

[10] P Ning, A Liu, TinyECC: elliptic curve cryptography for sensor networks (http://discovery), csc.ncsu.edu/ software/ TinyECC/ website

[11] R Anderson, H Chan, A Perrig, Key infection: Smart trust for smart dust. Proceedings of the 12thIEEE International IEEE International Conference on Network Protocols (ICNP '04), October 2004, 206–215.

[12] R Blom, An optimal class of symmetric key generation systems. Proceedings of the Eurocrypt 84 Workshop on Advances in Cryptology: Theory and Application of Cryptographic Techniques, 1985 (Springer), pp. 335–338.

[13] C Blundo, AD Santix, A Herzberg, S Kutten, U Vaccaro, M Yung, Perfectly-secure key distribution for dynamic conferences. Proceedings of the 12th Annual International Cryptology Conference on Advances in Crypto-logy, 1992, Berlin, Germany (Spring), pp. 471–486.

[14] B Lai, S Kim, I Verbauwhede, Scalable session key construction protocol for wireless sensor networks. Proceedings of the IEEE Workshop on Large Scale RealTime and Embedded Systems (LARTES '02), 2002.

[15] B Dutertre, S Cheung, J Levy, Lightweight key management in wireless sensor networks by leveraging initial trust. SDL Technical Report SRI-SDL-04-02 (2004).

[16] A Perrig, R Szewczyk, V Wen, D Culler, JD Tygar, SPINS: security protocols for sensor networks. Proceedings of the 7th Annual International Conference on Mobile Computing and Networking, July 2001 (ACM Press), pp. 189–199.

[17] S Zhu, S Setia, S Jajodia, LEAP: efficient security mechanisms for large-scale distributed sensor networks. Proceedings of the 10th ACM Conference on Computer and Communications Security (CCS '03), October 2003, 62–72.

[18] D Liu, P Ning, Location-based pairwise key establishments for static sensor networks. Proceedings of the 1st ACM Workshop on Security of Ad Hoc and Sensor Networks (CCS '03), October 2003, 72–82.

[19] L Eschenauer, VD Gligor, A key-management scheme for distributed sensor networks. Proceedings of the 9th ACM Conference on Computer and Communications Security, November 2002, 41–47.

[20] D Liu, P Ning, LI Rongfang, Establishing pairwise keys in distributed sensor networks. ACM Transactions on Information and System Security 8(1), 41–77 (2005).

[21] W Du, J Deng, YS Han, S Chen, PK Varshney, A key management scheme for wireless sensor networks using deployment knowledge. Proceedings of the IEEE International Conference on Computer Communications (IEEE INFOCOM '11), March 2004, 586–597.

[22] M Eltoweissy, M Moharrum, R Mukkamala, Dynamic key management in sensor networks. IEEE Communications Magazine 44(4), 122–130 (2006). of Distributed Sensor Networks, Volume 2013, Article ID 391986, 8 pages

[23] Xiaobing He, Michael Niedermeier and Hermann de Meer Dynamic Key Management in Wireless Sensor Networks: A Survey Journal of Network and Computer Applications, Volume 36, Number 2, pp. 611-622, 2013.

MGI-LEACH: Multi Group LEACH Improved an Efficient Routing Algorithm for Wireless Sensor Networks

Alami Chaibrassou
Research Laboratory in Mechanical, Industrial Management and Innovation Faculty of Science and Technology, Settat Morocco
BP.320 Driouch Nador
Alami70@yahoo.fr

Ahmed Mouhsen
Research Laboratory in Mechanical, Industrial Management and Innovation Faculty of Science and Technology, Settat Morocco
mouhsen.ahmed@gmail.com

***Abstract*—The wireless sensor networks (WSN) are composed by a large number of micro sensors that are randomly deployed in an area of interest to supervise or monitor various phenomena (temperature, pressure, humidity ..). The wireless sensor nodes are typically powered by non-rechargeable batteries which makes them constrained in terms of energy. As communication energy represents the largest part of the energy consumed in sensor nodes [1], so energy efficient routing is one of the main issues to prolong lifetime of wireless sensors networks.**

This paper proposes a hierarchical routing protocol called MGI-LEACH for homogeneous wireless sensor networks based upon the framework of LEACH such as the network is divided into multi groups to prevent the concentration of cluster head in a particular area of the network, and the cluster head selection is based on high residual energy and a short distance from the base station (BS) in each communication round, which minimizes the communication energy of wireless sensor nodes, and hence, increases lifetime of the whole network.

Simulations using MATLAB software show improved lifetime in different levels of Grouping compared to the classical algorithm LEACH.

***General Terms*—Algorithms .**

***Index Terms*—Wireless Sensor Networks, Network Lifetime, Sensor Node, grouping .**

I INTRODUCTION

A wireless sensor network is composed by a large number of micro sensors called nodes communicating with each other through radio links independently and randomly distributed over an area of interest. Sensors are powered by batteries, which is impossible to get recharged after deployment. As a large part of energy is consumed when communications are established, so it is imperative to develop an energy efficient routing protocol, taking into account the constraints by these sensors.

In flat routing, when the number of nodes increases, the traffic control dominates real communication. This increases the latency and overload routing tables. To overcome these limitations, hierarchical routing protocols have been introduced. These protocols allow the reduction of the number of messages transmitted, thus reducing energy consumption of sensor nodes. Furthermore, the transmission of data to BS via a single hop becomes impossible when the size of the network increases. To remedy this problem, the multi-hop routing is the mode of communication adopted to transmit data to the BS.

In this paper, we propose an energy efficient routing protocol based on the principle of clustering algorithm. We have modified the LEACH(Low-Energy Adaptive Clustering Hierarchy protocol as the CH selection is done by the predictive energy and the smallest path to BS.

The remainder of the paper is structured as follows: Section 2 describes The Sensor Node. In section 3 We propose an overview of LEACH protocol's. Section 4 describes problem statement, section 5 describes The Proposed Routing Protocol Algorithm. Simulation Setup and Scenarios and Simulation Results are interpreted in section 6 and 7. We draw the conclusion in Section 8.

II THE SENSOR NODE

2.1 Sensor Node Architecture

Principally a wireless sensor node consists of four basic units [2] (See Figure 1):

- Acquisition Unit: converts a physical signal into an electrical one.
- Processing Unit: responsible for all arithmetic operations and storage.
- Communication Unit: consists of a transmitter / receiver for communication between network nodes using a radio channel .

doi:10.4304/jetwi.6.1.40-44

- Power supply unit (battery): it powers the units that we have mentioned and it is generally not rechargeable or replaceable.

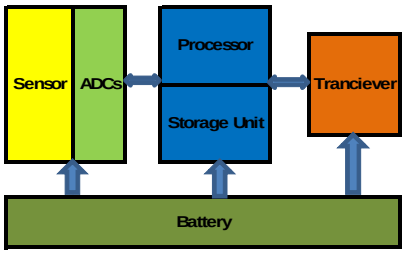


Fig 1: The components of sensor node

Based on the architecture above, the energy consumed by a sensor node is mainly due to the communication unit, it represents the greatest portion of the energy consumed by a sensor node .

III Overview of LEACH Protocol's

We start by giving an overview of LEACH protocol's which partitioned the network into groups (clusters). Nodes transmit their data to representatives of groups called cluster heads (CH), which in turn send the data after processing and aggregation to the desired destination or base station (See Figure 2).

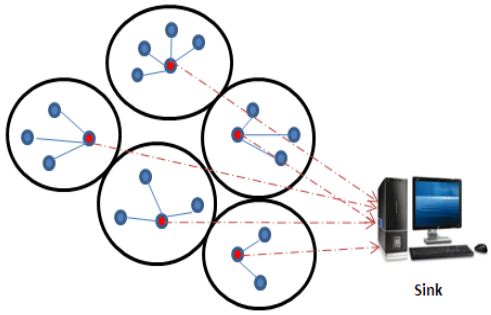


Fig 2: Clustering nodes in LEACH protocol's

In LEACH, the nodes self-elect to be CHs. They are based on the desired percentage of CHs and the number of rounds during which a node took the role of CH. Thus, a node n takes a random value between 0 and 1. If this value is less than the threshold T (n), the node becomes CH. The CHs inform their neighbors of their election. Each normal node decides to join the nearest CH[7].

$$T(n) = \begin{cases} \frac{P}{1-P*(r*mod\ 1/P)} & si\ (n \in G) \\ 0, \quad si\ non \end{cases} \qquad (01)$$

Where :

P : The percentage of the cluster head nodes in all nodes.

r : The number of rounds.

G : The collections of the nodes that have not yet been head nodes in the first 1/P round.

The optimal number of clusters is in general 5%.

3.1 Advantages

- Self-cluster configuration is independent of the BS (distributed algorithm).
- The data are aggregated to reduce the amount of information transmitted to the BS.
- Energy consumption is shared across all nodes, thus extending the lifetime of the network.

3.2 Disadvantages

- The further away from the BS, CHs die more quickly than those that are close to the BS.
- Without justifying their choice, the authors fix the optimal percentage of CHs for the network to 5% of the total number of nodes. However, the topology, the density and the number of nodes may be different in other networks.
- The use of probabilistic model to select CHs can generate CHs too close in an area of the network.
- The remaining energy and distance of nodes are not taken into account when electing CHs, it can give the CH role to node with discharged battery.

IV Problem Statement

The radio energy dissipation model used in our protocol (See Figure 3) is called first order radio model [3]. Each sensor node consists of a transmitter and receiver electronics. Energy is dissipated when nodes transmit or receive data.

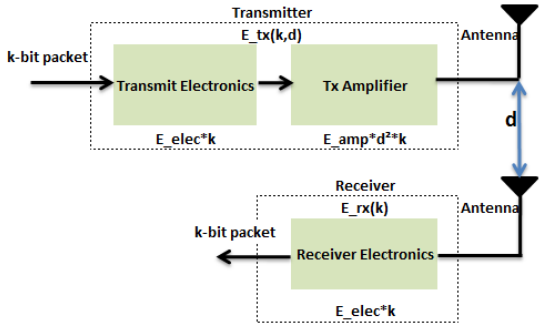


Fig 3: Radio energy dissipation model

The node radio will consume the following E_Tx amount of energy to transmit a k-bits packet over distance d [4]:

$$E_{\_Tx\,(k,d)} = \begin{cases} E_{elec} * k + E_{fs} * d^2 * k & \text{if } d < d_0 \\ E_{elec} * k + E_{amp} * d^2 * k & \text{if } d > d_0 \end{cases} \qquad (02)$$

And **E_rx** amount of energy to receive this k-bits packet:

$$E_{\_rx\,(k)} = E_{elec} * k \qquad (03)$$

Where ,

E_tx (k,d) : energy dissipated per bit at transmitter;

E_rx (k) : energy dissipated per bit at receiver;

Efs , Emp : amplifier parameters of transmission respectively in free space (with d2 power loss) and multi-path fading (with d4 power loss) channel models;

E_{elec} : Energy dissipation to run the radio ;
k : number of transmitted data bits;
d : distance from a sender node to a receiver node or BS.
d_0 : is the threshold distance between multi-path fading model and the free-space model.

$$\mathbf{d_0}=\sqrt{\frac{E_{fs}}{E_{amp}}} \quad (04)$$

Considering a wireless sensor network that is composed of N nodes randomly deployed over a square area M * M.
We will calculate the total energy consumed during one round of LEACH routing protocol :

$$E_{round} = \sum_{i=1}^{L}(E_{CHi}) + \sum_{j=1}^{N-L}(E_{NCHj}) \quad (05)$$

where $\mathbf{E_{CHi}}$ is the consumed energy when CH of the cluster labelled i, receives, aggregates, and transmits data to the base station. Whereas $\mathbf{E_{NCHj}}$ is the consumed energy by a non CH labelled j, and L is the total number of cluster heads.

the $\mathbf{E_{CHi}}$ is defined by:

$$E_{CHi}= E_{CHtoBS\,i} +E_{Recepti} +E_{Aggregi} \quad (06)$$

Where,

$$E_{CHtoBS\,i}=k.\ E_{elec} +k.\ E_{mp}.d^4_{toBSi} \quad (07)$$
$$E_{Recepti} =|Si|.k.\ E_{elec} \quad (08)$$
$$E_{Aggregi} =|S_i|.k.\ E_{DA} \quad (09)$$

With :

- $E_{CHtoBS\,i}$ is the energy required for the cluster head of the cluster labeled i to transmit its data to the base station.
- $E_{Recepti}$ is the total energy consumed when the cluster head of the cluster i receives data from its cluster member.
- $E_{Aggregi}$ is the total energy needed, by the CH of the cluster i, to process data.
- $|S_i|$is the cardinal of the set enclosing nodes of the cluster labeled k.

Hence,

$$\sum_{i=1}^{L}(E_{CHi})=\sum_{i=1}^{L}\left(E_{CHtoBS\,i} + E_{Recepti} + E_{Aggregi}\right)$$
$$=L.k.\ E_{elec} + L.k.\ E_{mp}.d^4_{toBS} +N.k.\ E_{elec} +N.k.\ E_{DA} \quad (10)$$

in the same way the $\mathbf{E_{NCHj}}$ is defined by:
there are N-L nodes non CHs, assume that l number of the latter work on multi-path fading mode while the others operate on the free space mode.

$$\sum_{j=1}^{N-L}(E_{NCHj}) =(N-L).k.\ E_{elec} +(\sum_{i=1}^{l} d^4_i).k.\ E_{mp} +(\sum_{j=1}^{N-L-l} d^2_j).k.\ E_{fs} \quad (11)$$

where $d_j < d_0$ and $d_i \geq d_0$

E_{round} is minimal if the following term :

$$L.\ E_{mp}.d^4_{toBS} +(\sum_{i=1}^{l} d^4_i).\ E_{mp} +(\sum_{j=1}^{N-L-l} d^2_j).\ E_{fs} \quad (13)$$

is minimal. However, it is known that the amplifier energy in a multipath fading channel model is greater than the amplifier one in a free space model (i.e $d_j < d_0 \leq d_i$ and $E_{fs} < E_{mp}$)
thus, to minimize the formula given in equation (13), all non CH nodes must operate in a free space model, it means direct communication between CHs and Cluster members .
And to minimize the term $L.k.\ E_{mp}.d^4_{toBS}$, the communication between cluster heads and the base station must be in multi-hop routing, thus multi hops-LEACH protocols is adapted.

V THE PROPOSED ROUTING PROTOCOL ALGORITHM

To overcome the constraints of LEACH protocols, we propose a new routing protocol called MGI-LEACH which consists, to partition the network into several equal areas A=A1+A2+A3….AN.

Deployed nodes are divided into sub groups (G1+G2+G3….GN) in which the modified LEACH protocol is applied. Number of groups mainly depends upon nodes density. These groups are created by the BS at the time of deployment.

The CHs selection in each group Gi is based on the selection of the cluster head with a high residual energy and a short distance from the base station in each communication round [5] .

In our protocol MGI-LEACH the threshold T(n) of LEACH protocol becomes :

$$T(n) = \begin{cases} f(cur).[\frac{p}{1-p.\left(r.mod\frac{1}{p}\right)} + p.h(D_{toBS})], & n \in Gi \\ 0, & n \notin Gi \end{cases} \quad (14)$$

With :

$$f(cur) = \frac{E_{cur}}{E_{ave}} \quad (15)$$

$$h(D_{toBs}) = \frac{D_{max}-D(n)}{D_{max}-D_{min}} \quad (16)$$

where ,

E_{cur} : the current energy of node.
E_{ave} : the average remaining energy of all nodes in the current round.
D_{max}, D_{min} :present the maximum and minimum of the distance between all nodes and the base station respectively.
f(cur) : function which measures the current residual energy of the node.
$h(D_{toBS})$: function which measures the distance between elected node and base station.
With this new threshold formula T(n) advantage is given to nodes that have a large amount of residual energy and close to the base station to be cluster head.

After the cluster head formation, normal nodes should optimize their choice to join optimal cluster head based on the cost function, this is done as follows :

- If the node is closest to the base station, no optimal cluster-head will be selected and it will directly send its data packages to the base station.
- Otherwise, the non-cluster head node chooses the optimal one among the candidate cluster-heads according to the cost function.
- the cluster head whose cost function is minimal, will be chosen as the optimal cluster-head.

The cost function of the normal node ni joins in the cluster with the cluster-head cj. This can be shown as:

$$cost(i,j) = \frac{d_{ij}}{d_{max}} \cdot \frac{E_{cur(i)}}{E_{cur(j)}} \cdot \frac{D_{(i)}}{D_{ave}} \qquad (17)$$

Where ,

d_{ij} : the distance from the node ni to cluster-head cj.

d_{max}:the maximum of the distances from ni to the candidate cluster-heads.

E_{cur} (i), E_{cur}(j) : the current residual energy of node ni and cluster-head cj .

$D_{(i)}$: the distance from cluster-head cj to the BS.

D_{ave} : the average distance between cluster-heads and the BS .

VI Simulation Setup and Scenarios

This paper uses Matlab7.14 as a simulation platform to emulate 200 immobile sensor nodes randomly deployed in 200m x 200m square area, the BS is fixed and located at (0,0) position. The coordinates of X and Y are measured in meters.

6.1 Simulation Parameters

Simulation scenarios in this paper are:

1. Sensor nodes are randomly distributed in a square region;
2. The energy of nodes which is limited, may be homogeneous. The nodes location is fixed after deployment;
3. Nodes communicate with base station via single-hop or multi-hops.
4. Sensor nodes are equipped with GPS to make them aware of their location.

TABLE 1.
SIMULATION ENVIRONMENT PARAMETERS

Parameter	Quantity
Nodes number	200
Rounds number	600
Node initial energy	2 joules
d_0	87.7 m
Packet size (k)	100 byte
Rounds number	200
BS location	(0,0)
E_{elec}	50 nJ/bit
E_{DA}	5nJ/bit
E_{fs}	10 nJ/bit/m²
E_{mp}	0.0013pJ/bit/m⁴

6.2 Simulation Scenarios

We propose to compare our proposed method MGI-LEACH to the classical algorithm LEACH with the same simulation parameters mentioned in Table 1. With which our method is tested for three levels of grouping (one, two and three group formation) (See Figure. 4 to 6).

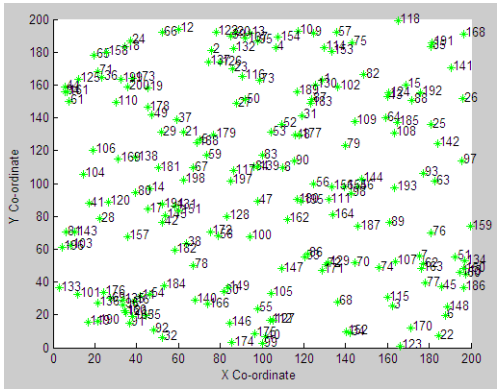


Fig 4 : One group formation

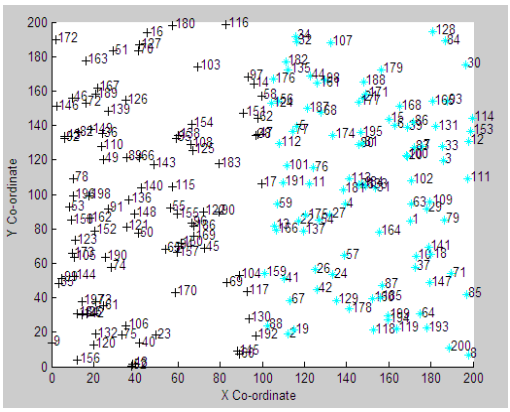


Fig 5 : Two groups formation

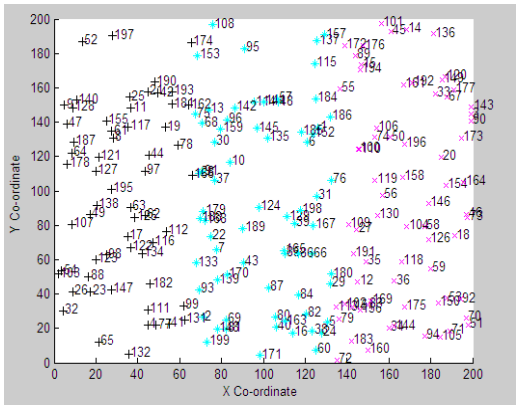


Fig 6 : Three groups formation

Using simulation parameters mentioned in Table 1, using the first order model equation of energy dissipation in sensor nodes, and applying the criteria to select the cluster heads and cluster formation to evaluate both MGI-LEACH and LEACH algorithms for one, two and three group formation. The cluster head(s) formed aggregates the data received from normal nodes and transmits them to the next hop cluster head closer to the BS or to the BS depending on the cluster formation and the shortest distance between the cluster head and the BS. At every

transmission or reception made, energy reduction occurs for every node, thereby cluster head rotation is utilized to extend the lifetime of the WSN.

VII Simulation Results

It can be proved that the proposed MGI-LEACH routing technique offers good results when compared to the classical LEACH. We investigated the advantage of the proposed technique by comparing the time in which the first node dies during the 600 rounds of simulation (network lifetime) to that of LEACH routing technique.

The network lifetime is shown in Figure 7. We observed that the first node dies faster in LEACH since all nodes tend to send captured data via their selected cluster head per round to the BS.

From Figure 7, we observe that the LEACH technique had an estimated lifetime of 20 rounds, MGI-LEACH for one group formation had an estimated lifetime of 80 rounds, MGI-LEACH for two group simulation had an estimated lifetime of 100 rounds and MGI-LEACH for three group simulation had an estimated lifetime of 150 rounds. The progressive increase of network lifetime employed by our proposed technique offers efficient energy usage for each node in the entire network.

The impressive increment in lifetime of the network from our proposed hierarchical technique is seen as a result of efficient routing decision and optimization of energy in cluster head selection of each cluster formed.

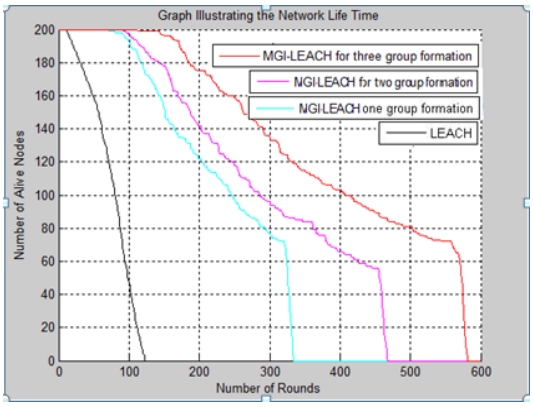


Fig 7: Network lifetime graph (number of alive nodes for a particular round of simulation) in one, two & three group formation

VIII Conclusion

In this paper we have proposed MGI-LEACH, an energy efficient routing technique for homogeneous wireless sensors based on LEACH technique. In which the network is partitioned into small equal areas to avoid the concentration of cluster heads in particular parts of the network. The cluster head selection is based on the residual energy prediction, for each sensor node and the shortest distance from sensor node to base station, this is the role of threshold T (n). Normal nodes choose their optimal cluster head in terms of energy using cost function. The simulation part indicates by comparing the time in which the first node dies, the MGI-LEACH performance is better than that of the classical LEACH.

References

[1] K. SHINGHAL, A. NOOR, N.SRIVASTAVA and R. SINGH, “Pouwer measurements of wireless sensor networks node” in International Journal of Computer Engineering & Science, 2011.

[2] P. Kaur and M. Katiyar "The Energy-Efficient Hierarchical Routing Protocols for WSN: A Review " in International Journal of Advanced Research in Computer Science and Software Engineering ,2012.

[3] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. J. 2000 Energy efficient communication protocol for wireless micro sensor networks in Proceedings of the Hawaii International Conference on System Sciences.

[4] A.Jorio, S. El fkihi, B. El Bhiri, D. Aboutajdine" A new clustering algorithm for wireless sensor networks",Journal of Theoretical and Applied Information Technology ,2013

[5] Qian Liao and Hao Zhu, ICSEM-2013.An Energy Balanced Clustering Algorithm Based on LEACH Protocol. In Proceedings of the 2nd International Conference On Systems Engineering and Modeling.

A WIFI Antenna Radiation Effects on Human Head in the ISM Band

Radouane Karli
Electronic and Communication Laboratory EMI, Mohammed V University-Agdal, Rabat, Morocco
radouane.karli@gmail.com

Hassan Ammor
Electronic and Communication Laboratory EMI, Mohammed V University-Agdal, Rabat, Morocco
ammor@emi.ac.ma

***Abstract*—This article describes in the first, the design of a new microstrip patch antenna for WIFI/WLAN applications (IEEE 802.11 b/g/n). Secondly, the article presents the effect of electromagnetic waves on a model of the human head exposed to the antenna designed in the ISM2450 band. By adopting a model tissue to seven layers of an adult. The objective is to evaluate the specific absorption rate (SAR) due to the propagation of electromagnetic waves along a human head for different antenna-head distances in simulation anatomic based model of the human head at a frequency of 2450MHz.**

All results, reflection coefficient, VSWR, radiation pattern, SAR and field distributions are presented. The simulation analysis was performed using the HFSS software.

***Index Terms*—WIFI/WLAN, microstrip patch, SAR, human head, HFSS.**

I Introduction

The recent advances in wireless communication systems, such as wireless local area networks (WLAN), wireless local loop (WLL), Broadband systems 3G etc, have induced a great interest in microstrip antennas. These have grown tremendously in recent years due to their ability to respond to the particular constraints of size, weight and especially cost imposed by the emerging mobile systems [1]. Future wireless systems will provide various services including the technology develops very quickly. A wireless communication technology has been an important topic for the telecommunication system because of its low cost and high opportunity. These systems specify the frequency range of 2.4 GHz to 2.5 GHz, which operate in the ISM band [2].The ISM (Industrial, Scientific and Medical) bands are frequencies that can be used for domestic or industrial scientists like medical applications, with the exception of radio applications, without requesting permission from the authorities. Frequency bands and possible limits emission levels are defined by the FCC Part 18 . The ISM frequency bands [3] assigned for multiple-user applications are suitable for, and expected for, the wireless LAN applications. The applications typically support a limited number of users in an indoor area. Thus the users are unconsciously exposed to EM waves which directly come from antennas or which are reflected and scattered from objects existing in the area [4]. The applications in these frequency bands also include portable telephones, microwave ovens, and so forth.

This development raises questions safety of new technologies. There has been an increasing public concern about possible health hazards due to exposure to EM waves. Accordingly, many international protection organizations and regulatory agencies have proposed the safety standards for exposure to EM waves [5–6]. These standards are based on the SAR, which is a measure of the EM power absorbed in the tissue. The advantages of analyzing specific absorption rate generated by cellular phones inside a human head are among others [7], as follows; verification of the compliance of phones with standards, electromagnetic solver, experimentally validated, gives not only reliable values of specific absorption rate but also locations inside a head [8], high resolution in field evaluation could be of interest, as input data for the analysis of athermal effects. Several methods have been used to study the effect of an antenna radiator on a human head such as homogenous or multilayered concentric spheres, multilayered planar model [9].

In this work, we propose to study and design a new microstrip patch antenna operating in the ISM band with a resonant frequency of 2450MHz. We will proceed by following an evaluation of specific absorption rate in the human head in a heterogeneous spherical model consists of seven layers for antenna designed. Thus, the effects of operating frequency (2450MHz) and gap distances between the antenna and the human head on distributions of SAR within the human head are systematically investigated.

II Antenna Geometry

Figure 2 illustrates the top, bottom and side views of the geometry of the antenna studied. The antenna is simulated on an FR4_epoxy substrate of 60 ×70 mm^2 with a dielectric constant ε_r=4.4. The thickness of the substrate is H=1.58 mm. A rectangular patch including technical slots with different dimensions shown in Table

doi:10.4304/jetwi.6.1.45-48

1. The antenna is fed by a microstrip line in order to increase the bandwidth and gain.

TABLE 1:
SPECIFICATIONS OF THE PROPOSED ANTENNA

ELEMENTS	DIMENSIONS
Patch	W_p=28.2mm ; L_p= 36.7mm
Microstrip line	W_{MI}=1mm ; L_{MI}=25mm
Ground plane	W_G=18mm ; L_G=60mm

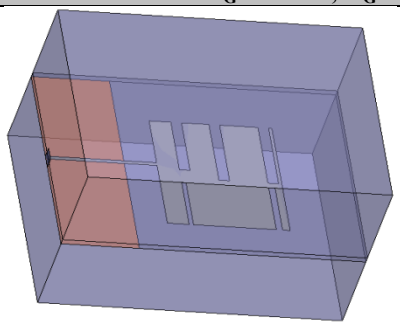


Fig 1: Ansoft HFSS generated antenna model

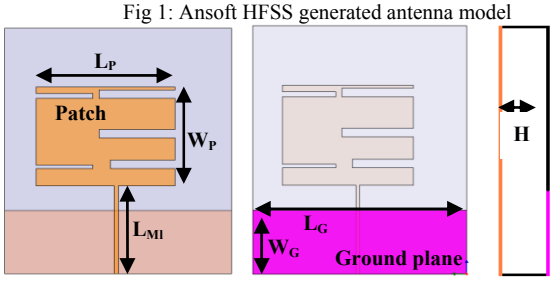


Fig 2: Geometry of the proposed antenna. (a) Top view, (b) Bottom view, (c) Side view

III RESULTS AND DISCUSSION

Figure 3 shows the different values of the reflection coefficient obtained.

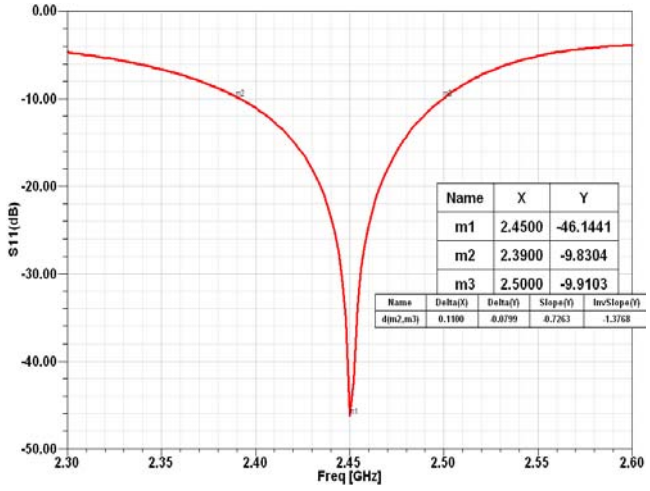


Fig 3: Reflection coefficient of the proposed antenna

It was found that the antenna resonates in the desired frequency band as shown in Figure 3. Indeed, for $|S_{11}|$ <-10 dB: band ranges from 2.39 to 2.5GHz with a resonant frequency 2.45GHz. The bandwidth is 110MHz (4.49%) which is used for WLAN applications namely the ISM band used by BLUETOOTH (2.4GHz-2.485GHz) and wireless systems (2.4GHz for 802.11b and 802.11g). Also, the band is sufficient for the intended width by the IEEE specifications which is 83.5MHz.

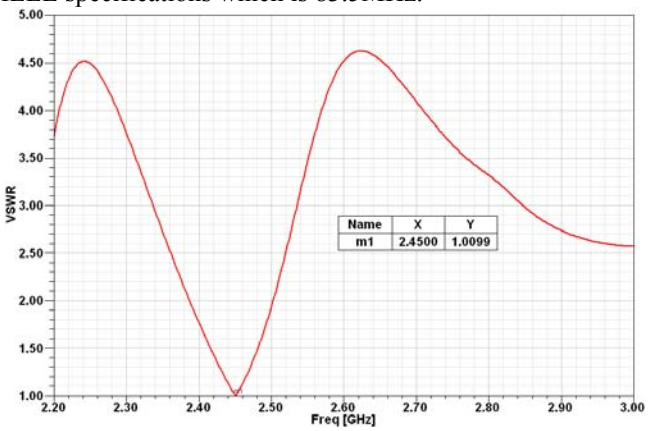
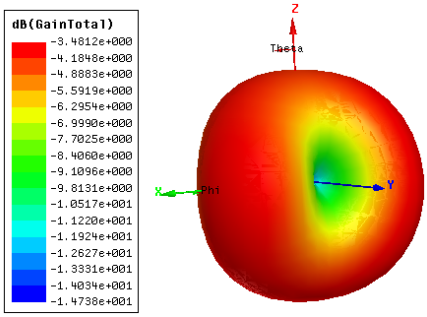


Fig 4: The VSWR of studied antenna

The variation of VSWR with frequency shows that it is less than 2dB for the resonance frequency which is 2.45GHz, as illustrated in Figure 4. A bandwidth equal to 110 MHz which is 4.49% from the center frequency. According to IEEE specifications, the width of the WLAN band represents 3.4% around 2.45 GHz. Our case is sufficient for this bandwidth and allows us to cover the ISM band.



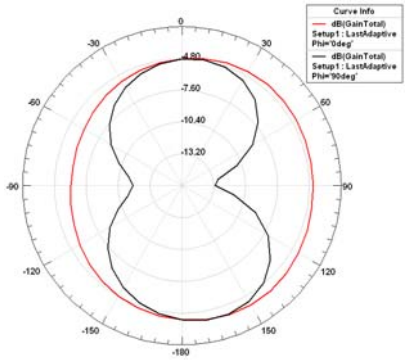


Fig 5: (a) 3D-Gain Total, (b) 2D Radiation Pattern

The radiation pattern of the antenna characterizes the variation of the radiated power over long distances in different directions in space. The radiation patterns of the antenna studied are shown in Figure 5. They are almost omnidirectional allowing use of this antenna for applications in the ISM2450 band and specifically for WIFI.

IV DOSIMETRY

4.1 Methods and Model

When electromagnetic waves propagate through the human tissues, the energy of electromagnetic waves is absorbed by the tissues. Interaction of electromagnetic fields with biological tissues can be defined in term of specific absorption rate (SAR). The specific absorption rate (SAR) is an index that measures the level of radio frequency electromagnetic field in the human head, as emitted by the mobile phone when operating at full power, in the worst conditions. Its unit is watts per kilogram (W/kg). Governments have put standards for the maximum SAR that should not be exceeded to avoid health hazards. This maximum is set to 1.6W/kg averaged over 1g of tissue, or 2W/kg averaged over 10g of tissue, in USA and Europe [10], respectively. The specific absorption rate is described by the following equation:

$$\mathbf{DAS} = \frac{\mathbf{1}}{\mathbf{2}} \cdot \frac{\boldsymbol{\sigma}}{\boldsymbol{\rho}} |\mathbf{E}|^{\mathbf{2}} \qquad \mathbf{(1)}$$

Where σ is electric conductivity (S/m) and ρ is the tissue density (kg/m^3). SAR is calculated as a function of position from the estimates of local fields and tissues properties.

In this study, we will evaluate the effects induced by the WIFI antenna located on human head side with different distances are considered as a source of radiating near field for the model of the human head

The figure 6 (a) shows the spherical model of the human head and the antenna position. This model consists of seven layers, as is shown in figure 6 (b), which are the skin, fat, muscle, skull, dura, CSF (cerebrospinal fluid) and brain. The dimensions we have chosen are as follows: for the sphere, we took a radius of 90 mm, skin thickness 2mm, fat 1 mm, 4 mm for muscle, skull 10mm, 1mm Dura, 2mm CSF and the last layer is the brain.

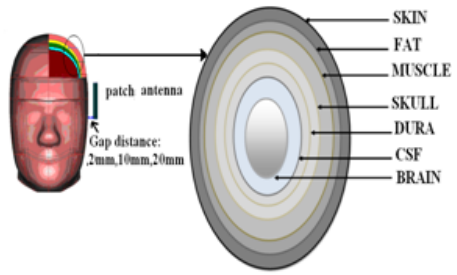


Fig 6: The Model of a human head. (a) Cross section human head model with patch antenna. (b) Model composed of seven layers.

The propagation depends on the dielectric properties of the layer, density and wavelength. Table 2 gives the values of these properties for the 2450MHz frequency [11].

TABLE 2:
THE PROPERTIES OF THE MATERIALS USED IN THE SIMULATIONS AT 2450MHZ

Type of tissue	2450MHz		
	ε_r	σ (S/m)	ρ (kg/m^3)
SKIN	42.85	1.59	1100
FAT	10.82	0.26	1100
MUSCLE	53.64	1.77	1040
SKULL	15.01	0.57	1850
DURA	42.03	1.66	1030
CSF	66.24	3.45	1030
BRAIN	42.61	1.48	1030

4.2 Specific Absorption Rate

The figure 7 shows the results obtained from the flow of local and average specific absorption on the spherical seven layer model with three different distances between the human head and antenna 2mm, 10 mm and 20 mm at 2450 MHz and a power of 100mW.

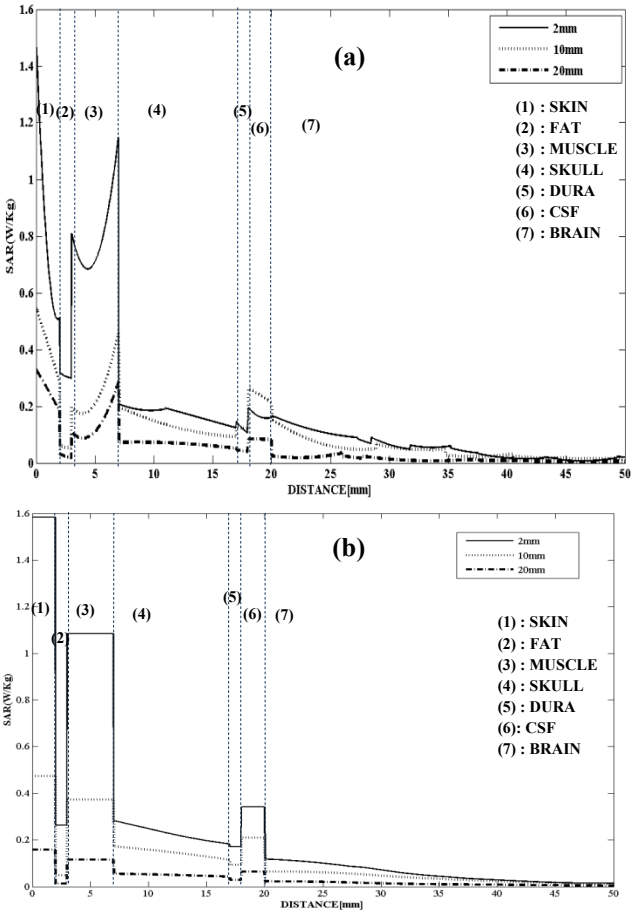


Fig 7: SAR variation at different distances WIFI antenna-head. (a)Local SAR, (b)Average SAR

The results allowed us to conclude that more than the antenna is closer to the human head, more than the SAR reached high values and vice versa. Thus, from Figure 7 the penetration of specific absorption rate in tissues and for different distances does not exceed the standards set by IEEE and FCC namely 1.6W/kg.

At layers, we note that the SAR has three peaks respectively located in the skin, muscle and CSF, which is mainly due to their relatively high dielectric properties. This allows us to say that these layers protect the brain because it absorbs a small radiation.

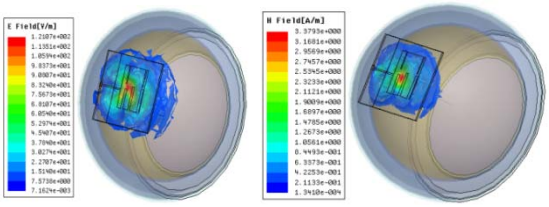


Fig 8-a: E_field and H_field at 2mm

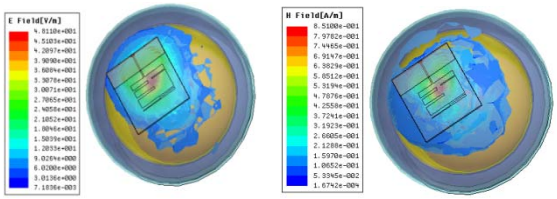


Fig 8-b: E_field and H_field at 10mm

Figure 8 shows the distribution of the fields E and H of the head model to seven layers with an antenna head distance of the order of 2 and 10mm. Note that the field is decreasing with distance and that these values are very important in the nearest antenna tissues.

V CONCLUSION

In this work we present in the first part a novel design of a miniature microstrip patch antenna. Light weight, low-cost, plain configuration are the advantages of this structure. Bandwidth enhancement has been improved by suitably cutting slots into rectangular patch which allowed us to cover ISM2450 band. The second area includes a study of the effects of electromagnetic waves on a model of the human head exposed to antenna for the wireless frequency at 2450MHz. This study shows that our patch antenna meets the standards when placed at various distances antenna-model human head at an incident power of 100mW. The numerical simulations shows several important features of the energy absorption in the human head and demonstrate that our patch antenna exhibits good electrical performance and thus can be considered as a suitable candidate for various applications in ISM-band (IEEE 802.11 b / g / n).

REFERENCES

[1] Best, S. R. Electrically Small Multiband Antennas, In: Multiband Integrated Antennas for 4G Terminals, D. A. Sanchez-Hernandez (ed.), 1-32, Artech House, ISBN 978-1-59693-331-6, Boston, USA. 2008.

[2] Juan P. Maícas., Recent developments in mobile communications– a multi-disciplinary approach, InTech, ISBN 978-953-307-910-3, pp.144-150, Croatia, 2011.

[3] Ahmed Al-Shaheen, New patch antenna for ISM band at 2.45GHz, ARPN Journal of Engineering and Applied Sciences, Vol.7, No.1, January 2012.

[4] Fung, L. C., S. W. Leung, and K. H. Chan, An investigation of the SAR reduction methods in mobile phone applications, IEEE International Symposium on EMC, Vol.2, pp.656-660, Aug. 2002.

[5] Lin, J. C., A new IEEE standard for safety levels with respect to human exposure to radio-frequency radiation," IEEE Antennas and Propagation Magazine, Vol. 48, No. 1, pp. 157-159, Feb. 2006.

[6] ICNIRP Guidelines, Guidelines for limiting exposure to time-varying electric, magnetic, and electromagnetic fields (up to 300 GHz), Health Phys., Vol.74, pp.494–522, 1998.

[7] A. Schiavoni, P. Bertotto, G. Richiardi and P. Bielli, SAR generated by commercial cellular phones-phone modeling, head modeling and measurements, IEEE transactions on microwave theory and techniques, Vol48, N°.11, pp 2064-2071, novembre 2000.

[8] M. R. Iqbal Faruque, M. T. Islam and N. Misran, SAR Analysis in Human Head Tissues for Different Types of Antennas, World Applied Sciences Journal, Vol. 11, N° 9, pp. 1089-1096, 2010.

[9] Akram, A. and T. Andrew, A multi layered model of human head irradiated by electromagnetic plane wave of 100MHz-300GHz, Int. J. Sci. Res., Vol. 15, 2005.

[10] T. Wessapan, S. Srisawatdhisukul, P. Rattanadecho, Numerical analysis of specific absorption rate and heat transfer in the human body exposed to leakage electromagnetic field at 915 MHz and 2450 MHz, ASME J. Heat Transfer 133, 051101, 2011.

[11] A. Drossos, V. Santomaa and N. Kuster, The dependence of electromagnetic energy absorption upon human head tissue composition in the frequency range of 300-3000MHz. IEEE transactions on microwave theory and techniques, Vol. 48, N° 11, pp. 1988-1995, November 2000.

Development of a New Slit-Slotted Shaped Microstrip Antenna Array for Rectenna Application

Mohamed Adel Sennouni[1]
Jamal Zbitou[1]
Abboud Benaissa[1]
1.Laboratoire LITEN FPK–
Khouribga/FSTS université Hassan
1er
Settat – Morocco
adelsennouni@gmail.com

Abdelwahed Tribak[2]
2.Institut national des postes
et télécommunication
Rabat – Morocco

Otman Elmrabet[2]
2.Electronics and Microwave
group,FS of Tétouan-Morocco

Mohamed Latrach[3]
3. RF & Hyper group ESEO,
Angers France

***Abstract*—These paper presents a new 3X3 array design using a microstrip patch array antenna to operate at 2.45GHz. The aim of this antenna array construction is to obtain high directivity. The element of the array is microstrip square patch antenna using V-shaped symmetric-slit along with rectangular slot in diagonal direction at the centre of square patch radiator to achieve the circularly polarized radiation and each element is fed by inset feed. The size and feed position of the single microstrip square patch is determined through the theoretical design and CST microwave studio software simulation. Based on which an array of six elements with equal sizes and equal spacing is designed on a planar substrate. The simulation results in this paper can be used as design reference for the practical design of the rectenna.**

***Index Terms*—Array antenna, microstrip antenna, rectenna, slit-slot, V shaped, square patch, feed, circular polarization.**

I Introduction

The rectifying antenna (rectenna) is an important component in wireless power transmission (WPT) system [1] Its antenna receives a wireless signal and then its rectifying circuit converts that AC energy into DC power. Most rectenna elements and rectenna arrays are developed for frequencies below 15GHz. Rectennas operating at millimetre-wave frequencies have the advantages of compact size and higher overall system efficiency for long distance transmission [2].

The antenna plays a significant role in the overall rectenna performance. It is one of the critical components needed to realize the desired high efficiency for different operating frequencies and for different input powers. For any WPT system high conversion efficiency is required.

There have been many rectennas reported,including different operating frequencies and different input powers. For example, a compact 5.8 GHz rectenna,which used a stepped-impedance dipole antenna, was presented in [3] with a high conversion efficiency of 82%.A 35 GHz rectenna array was reported in[4] that achieved a 35% efficiency when the input power density was 30 mW/cm2.Broad-band (2-18 GHz) rectenna arrays that recycle ambient microwave energy were presented in [5].A 2.45 GHz hybrid sensitive rectenna system that used zero-bias microwave Schottky diodes,was reported in [6].

Components for microwave-power transmission have traditionally been focused at 2.45 GHz considering that the frequencies below 3 GHz are not strongly attenuated by the atmosphere even under a severe weather condition [7] 2.45 GHz is thought of as a proper frequency for the application of power transmission between ground-to-ground, ground-to-space, and space-to-ground. However, for the space-to-space application the operating frequency can be increased to allow power transmission over much longer distances with the smaller antenna and rectenna.

Several antennas for CP radiation have been designed over the last few years using a circular ring, e.g., with stripline hybrid coupler feed at 2.45 GHz [8], with microstrip feed at 1.5GHz [9], and with coplanar waveguide and coaxial line feed at 2.45 GHz [10]. To the best of our knowledge, few rectenna for CP radiation have been proposed so far [11], [12]. Furthermore, these rectennas are intended for the operation at relatively high power densities. An annular ring-slot rectenna designed for low power densities and operating at 5.2 GHz was introduced in [13].

This paper presents a study of a 3X3 square patch antenna array with circular polarization and a V-shaped slot,designed as an RF receiver using CST microwave software. In this study, the design of the square antenna array was chosen because of its relevance in terms of power, size and directivity. The parameters discussed as follows include return loss, voltage standing wave ratio and radiation pattern.

doi:10.4304/jetwi.6.1.49-53

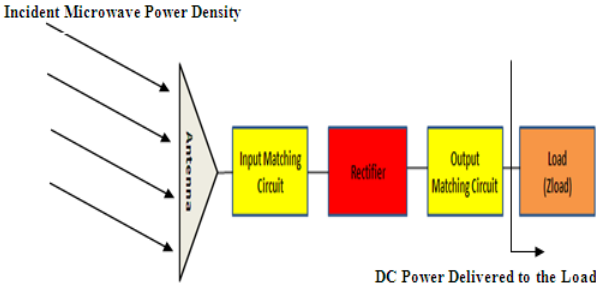


Figure 1. The Block diagram of the rectenna.

II Design Procedures

The antenna element is a square with an inclined slot in the center and V shaped slot at the corner. The slot technique is a way to obtain a circularly polarization. The antenna is feeding by a microstrip line having a characteristic impedance of 100Ω, this antenna was mounted on a FR4 substrate. The antenna dimensions are presented in Figure.2.

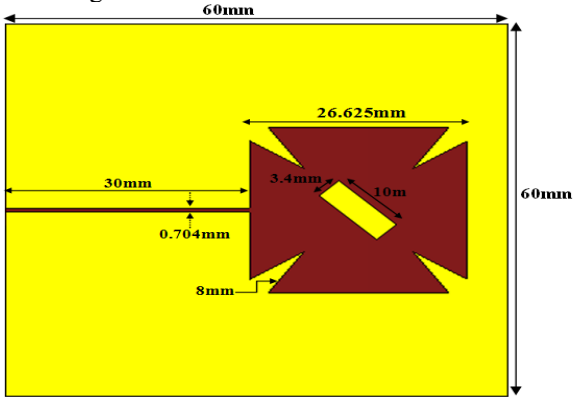


Figure.2: The square antenna structure

III Results and Discussion

3.1 Return Loss

The antenna is optimized and simulated by using CST Microwave Studio. The optimization was focused on the matching impedance and the slot dimensions. The substrate used in this conception is the FR4 substrate with:

- the relative permittivity $\varepsilon_r = 4.4$,
- the thickness $h = 1.58mm$
- and the dielectric loss $\tan(\delta) = 0.001$.

As shown in Fig.3, the S11 is equal to -21.52dB for the operating frequency 2.45 GHz which means good matching input impedance.

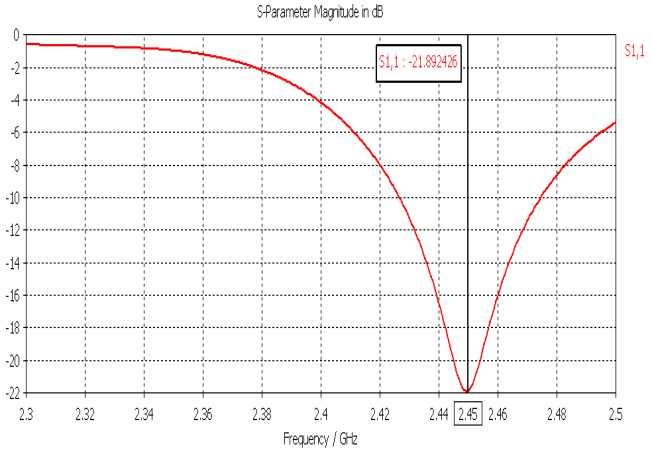


Figure .3: return loss vs frequency

3.2 Voltage Standing Wave Ratio

Fig. 4 shows the VSWR obtained, for 2.45 GHz it is 1.17, which is within the recommended range. The result obtained indicates that the transmitter and antenna is well matched. Maximum possible amount of energy is absorbed at the input terminal with a minimum reflected power.

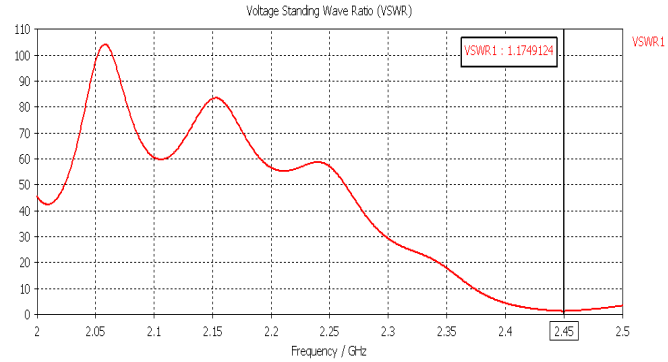


Figure.4: VSWR vs frequency

The Smith chart shown in Fig.5 indicates that the connection is well matched.

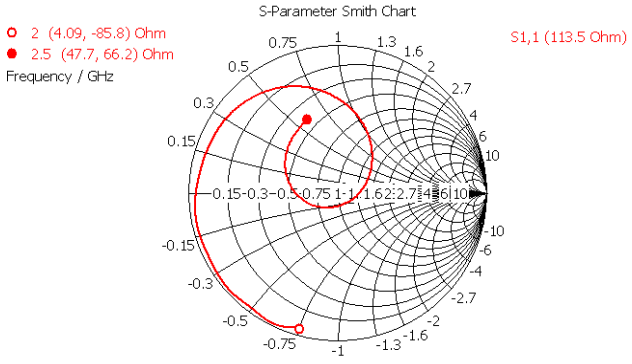


Figure.5: S (11) Smith chart

3.3 Radiation Pattern

Fig.6 presents the 3D radiation pattern with the structure below at 2.45 GHz. Fig.7 shows the polar view directivity of radiation pattern for simulation square patch antenna. The radiation pattern in the angular width (3dB) for theta component is 99.7˚ and main lobe direction is about 2.0 ˚.

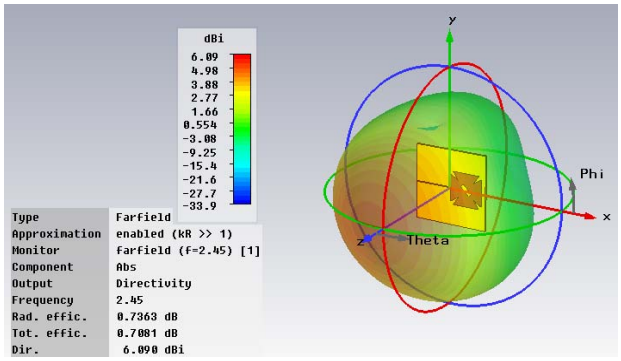


Figure.6: 3D radiation pattern @2.45 GHz.

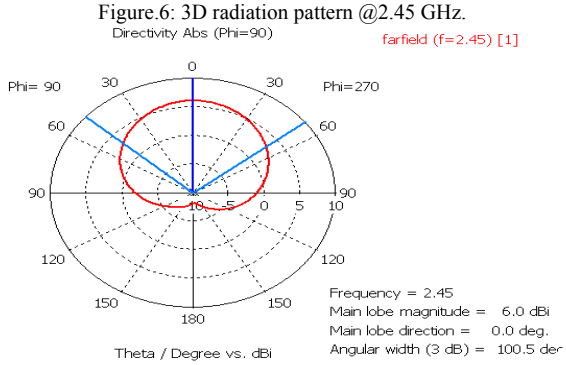


Figure.7: 2D polar radiation pattern @2.45 GHz.

IV ARRAY DESIGN

We have started from the elementary antenna optimized before; we arrived to model an array of six elements patches with the arrangement shown in fig.8, this new arrangement minimizes the effects of coupling and the generation of higher modes, also it can occupy a reduced area for such an antenna array of the same number of the elements.

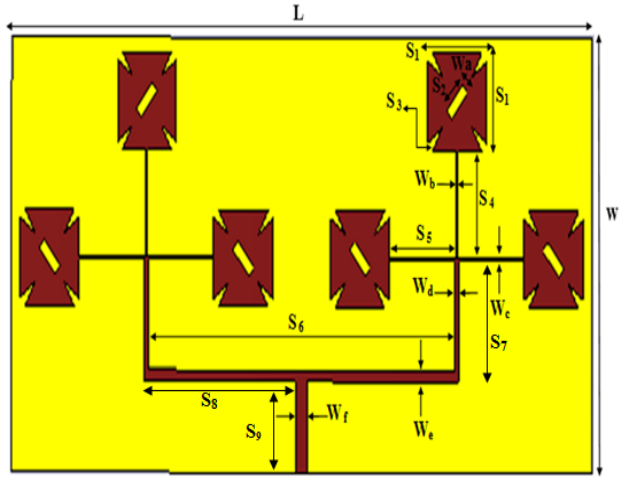


Figure.8: 3X3 slotted patches array .

The table.1 below shows the various parameters of the square patch antenna array developed. The characteristic impedances of the microstrip lines being used for feeding elements of the array is given in the following table.2.

TABEL.1:
ARRAY PHYSICAL DIMENSIONS

Parameters	W	W_a	W_b=W_c	W_d	We	W_f
Value(mm)	118.6	3.3	0.704	2.123	3.044	5.12
Parameters	L	S_1	S_2	S_3	S_4	S_5
Value(mm)	261.8	26.52	10	8	29.29	33.044
Parameters	S_6	S_7	S_8	S_9		
Value(mm)	137.87	33.04				

TABLE.2:
MICROSTRIP LINE IMPEDANCES

Impedance	Z_c	Z_d	Z_e	Z_f
Value(Ω)	100	61.237	50	35.35

4.1 Return Loss

Fig.9 shows the S11 result of the developed array for the operating frequency at 2.45 GHz with return loss of -29.22dB.

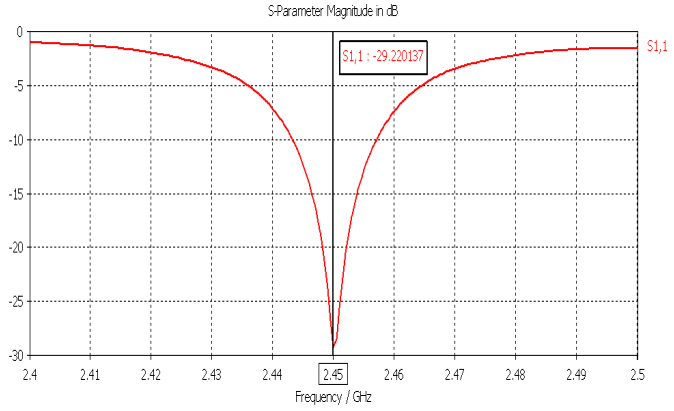


Figure .9: Simulation result for return loss for the 3X3 array

4.2 Voltage Standing Wave Ratio

As shown in Fig. 10 the VSWR obtained from the simulation at 2.45 GHz for the 3X3 array is 1.07, which is within the recommended range.

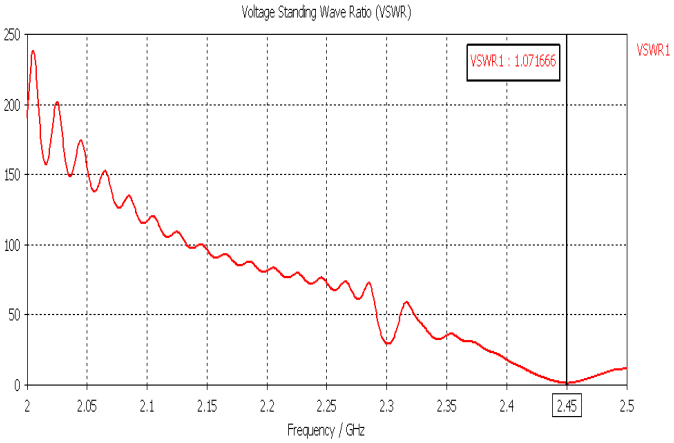


Figure.10. : 3X3 array Simulation result for VSWR.

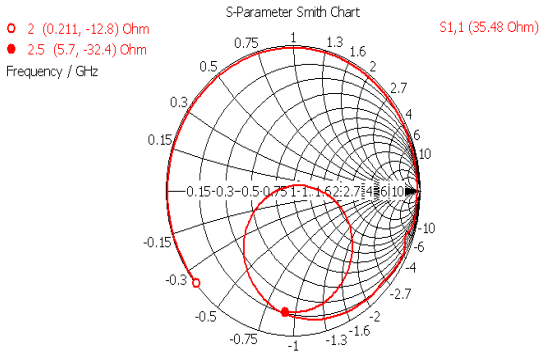


Figure.11: S(11) Smith chart

4.3 Radiation Pattern

The radiation pattern taken for the far-field at 2.45 GHz is indicated by the 3-D view in Fig. 12. The maximum normalized gain for simulation of 3x3 array square patch antenna that can be achieved is 9.153dBi.

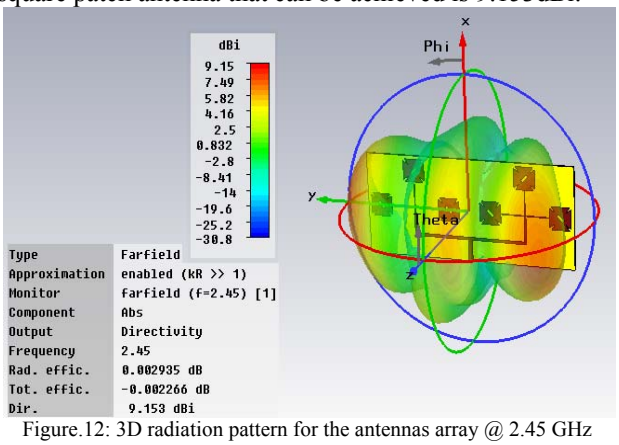


Figure.12: 3D radiation pattern for the antennas array @ 2.45 GHz

Fig.13 shows the polar view directivity of radiation pattern for simulation 3X3 array square patch antenna with 9.1 dBi of main lobe magnitude. The radiation pattern in the angular width (3dB) for theta component is 29.6˚ and main lobe direction is about 49.0 ˚.

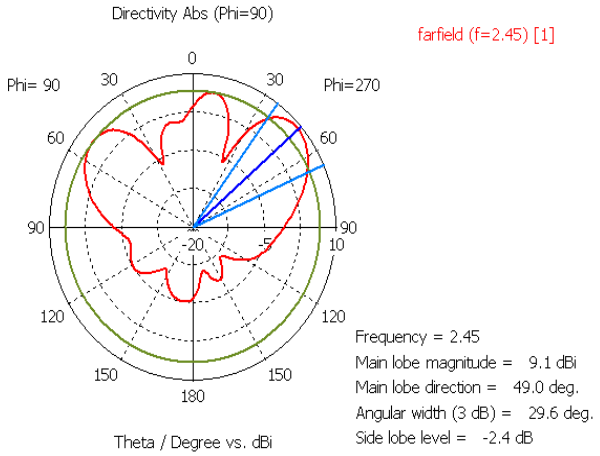


Figure.13: 2D polar radiation pattern @ 2.45GHz

The results indicate that the design has the ability to focus energy in a particular direction. However, which is perfectly suitable for wireless power transmission

CONCLUSION

This study permits to develop and validate array square antennas with circular polarization due to the use of the inclined slot in the center of the square patch. The array validated is then suitable for rectenna system which will permit to increase the efficiency of the RF-DC conversion system, by increasing the gain and directly increasing the received power at the input of the rectenna.

The array antenna optimized has been successfully designed and simulated using CST microwave software. The performance criteria extracted from the software includes return loss, VSWR and radiation pattern, provide clear indication that the array design is suitable for wireless power transmission. Hence, the research can be extended further for fabrication and testing of prototype as future work.

The procedure used in this study can be exploited for other frequency band suitable for wireless power transmission applications.

ACKNOWLEDGMENT

The authors would like to thank the staff members of the LITEN Laboratory, University HASSAN THE FIRST,FST OF SETTAT Morocco and the STRS Laboratory of INPT in RABAT Morocco for providing the facility and technical assistance in this research.

REFERENCES

[1] Strassner, B.H., and Chang, K.: 'Rectifying antennas (rectennas)' in'Encyclopaedia of RF and microwave engineering' (John Wiley & Sons,Inc., Hoboken, NJ, 2005), vol. 5, pp. 4418–4428

[2] Yoo, T.W., and Chang, K.: 'Theoretical and experimental development of 10 and 35 GHz rectennas', IEEE Trans. Microw. Theory Tech., 1992, 40,(6), pp. 1259–1266

[3] W.-H. Tu, S.-H. Hsu, and K. Chang, "Compact 5.8-GHz rectenna using stepped-impedance dipole antenna," IEEE Antenna and Wireless Propagation Letters, vol. 6, pp. 282-284, 2007.

[4] Y.-J. Ren, M.-Y. Li, and K. Chang, "35 GHz rectifying antenna for wireless power transmission," Electronics Letters 24th, vol. 43, no.11, pp., May 2007.

[5] J. A. Hagerty, F. B. Helmbrecht, W. H. McCalpin, R. Zane and Z.Popovic, "Recycling ambient microwave energy with broad-band rectenna arrays," IEEE Transactions on Microwave and Techniques, vol. 52, pp. 1014-1023, Mar. 2004.

[6] J. Zbitou, M. Latrach, and S. Toutain, "Hybrid rectenna and monolithic integrated zero-bias microwave rectifier," IEEE Trans.Microwave Theory Tech., vol. 54, pp. 147-152, Jan. 2006.

[7] W. C. Brown, "All electronic propulsion key to future space ship design, " presented at AIAA/ASME/SAE/ASEE 24th Joint Propulsion Conf., Boston, July 1988.

[8] W. C. Brown et al., U.S. Patent 3 434 678, Mar. 25, 1969.

[9] W. C. Brown, "Experiment in the transport of energy by microwave beam," in IEEE Inr. Conv. Rec., vol. 12, pt. 2, Mar. 1964, pp.

[10] W.C. Brown "The history of power transmission by radio waves," IEEE Trans. Microwave Theory Tech., vol. 32, no. 9, pp. 1230-1242, Sept. 1984.

[11] W.C. Brown "Experiments involving a microwave beam to power and position a helicopter," IEEE Trans. Aerosp.

Electron. Syst., vol. AES- 5, no. 5, pp. 692-702, Sept. 1969.
[12] W.C. Brown "Solar Power Satellite Program Rev. DOE/NASA Satellite Power System Concept Develop. Evaluation Program," Final Proc. Conf. 800491, July 1980. 8-18.
[13] W. C. Brown, "All electronic propulsion key to future space ship design, " presented at AIAA/ASME/SAE/ASEE 24th Joint Propulsion Conf., Boston, July 1988.

Performance Evaluation of WideMac Compared to ALOHA in term of Energy Consumption for IR-UWB based WSN

Anouar Darif
LRIT-CNRST (URAC29),
Faculté des Sciences,
Université Mohammed V-Agdal
Rabat, Maroc
anouar.darif@gmail.com

Rachid Saadane
SIRC2S/LASI
EHTP, Km 7, Oasis,
Route El Jadida
Casablanca, Maroc
saadane@ehtp.ac.ma

Driss Aboutajdine
LRIT-CNRST (URAC29),
Faculté des sciences,
Université Mohammed V-Agdal
Rabat, Maroc
aboutaj@fsr.ac.ma

***Abstract*—The introduction of the IR-UWB Technology in the field of WSN was promising for researchers especially for its low power consumption feature. To implement such a solution, we need a suitable MAC protocol to exploit the specific features of this technology. When introducing this Technology, ALOHA was the only candidate MAC protocol. Because of the high energy consumption of ALOHA, in this paper we present WideMac, a low power medium access control protocol designed specifically for Impulse Radio Ultra Wide Band transceivers. The IR-UWB channel offers ultra low power transmissions and unmatchable robustness to multiple access interference. WideMac takes advantage of these two key properties by using asynchronous periodic beacon transmissions from each network node.**
To test and evaluate the performance of WideMac protocol we used PhyLayerUWBIR class developed under MiXiM platform on OMNet++ as a physical layer. For the MAC layer we developed our own class WideMacLayer.

***Index Terms*—WSN, IR-UWB, ALOHA, WideMac, Power Consumption, OMNet++, MiXiM.**

I. Introduction

Wireless Sensor Networks (WSNs) is one of the most interesting networking technologies since its ability to use no infrastructure communications, it have been used for many applications, including military sensing, data broadcasting [1], environmental monitoring [2], Intelligent Vehicular Systems [3], multimedia [4], patient monitoring [5], agriculture [6], industrial automation [7] and audio [8] etc. This kind of networks has not yet achieved widespread deployments, though it has been proven able to meet the requirements of many classes of applications. Wireless sensor nodes have some limitations as lower computing capabilities, smaller memory devices, small bandwidth and very lower battery autonomy; these constraints represent the main challenges in the development or deployment of any solution using WSNs. Energy consumption is a very important design consideration in WSNs, New wireless technologies emerge in the recent few years, providing large opportunities in terms of low power consumption, high and low rate and are promising for environment monitoring applications. IR-UWB technology is one of these new technologies and is considered as a next generation of the IEEE802.15.4 standard; it is a promising solution for WSN due to its various advantages such as its robustness to severe multipath fading even in indoor environments, its potential to provide accurate localization, its low cost and complexity, and low energy consumption [9]. It is necessary to find a very adapt MAC layer protocol to this Technology for keeping his advantages.

The present paper is organized as follows. In Section 2 we introduced Wireless Sensor Networks. In section 3 we presented the IR-UWB technology. Section 4 presents the ALOHA MAC protocol. In section 5 we presented WideMac. The simulation and its results are presented in section 6; finally, Section 7 concludes the paper.

II Wireless Sensor Network

A wireless sensor network (WSN) in its simplest form can be defined as a network of (possibly low-size and low complex) devices denoted as nodes that can sense the environment and communicate the information gathered from the monitored field through wireless links; the data is forwarded, possibly via multiple hops relaying, to a sink (Base Station) that can use it locally, or is connected to other networks (e.g., the Internet) through a gateway (see Figure 1).

- The nodes can be stationary or moving.
- They can be aware of their location or not.
- They can be homogeneous or not.

doi:10.4304/jetwi.6.1.54-58

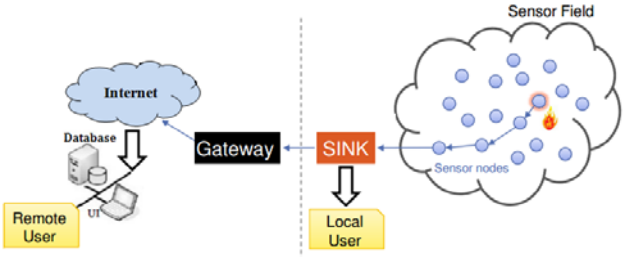


Figure 1: Sensor network architecture.

2.1 Sensor Node Architecture

A wireless sensor network (WSN) in its simplest form can be defined as a network of (possibly low-size and low complex) devices denoted as nodes that can sense the environment and communicate the information gathered from the monitored field through wireless links; the data is forwarded, possibly via multiple hops relaying, to a sink (Base Station) that can use it locally, or is connected to other networks (e.g., the Internet) through a gateway (see Figure 2). The nodes can be stationary or moving. They can be aware of their location or not. They can be homogeneous or not.

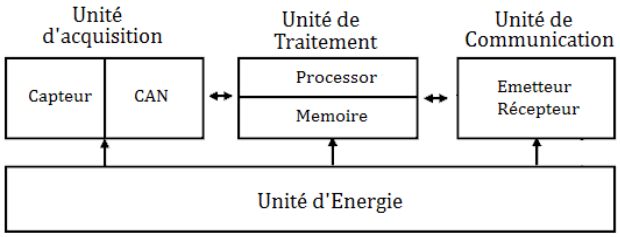


Figure 2: Sensor node architecture.

2.2 Wireless Technologies

Many wireless technologies are used in WSN; the most frequently used are ZigBee, the various forms of IEEE802.11 or Wi-Fi, Bluetooth, and UWB with the two standards IEEE802.15.3a/4a. The first standard for wireless local area networks (WLAN) was the IEEE802.11 specification in 1997. For applications requiring high data rate, the 802.11n standard is under development to achieve more than 100 Mbits/s. Recently Wi-Fi is widely adopted in various applications and due to its complexity and higher energy consumption compared to ZigBee and IR-UWB, this technology has been applied only to perform some particular functions in WSN [10].

III IR-UWB

Impulse Radio Ultra Wide Band (IR-UWB) is a promising technology to address Wireless Sensor Network constraints. IR-UWB signals are transmitted in form of very short pulses with low duty cycle (see Figure 3). The medium is divided into frames and each frame is shared in N_h chips. The frame and chip duration are T_f and T_c, respectively. The transmitted symbol can be repeated following a pseudo random sequence to avoid catastrophic collision under multiuser access conditions [11].

Using the Time Hopping Binary Pulse Amplitude Modulation (THBPAM) scheme for example, the k[th] user transmitted signal $S_{tx}^{(k)}(t)$ can be expressed as:

$$S_{tx}^{(k)}(t) = \sum_{j=-\infty}^{+\infty} \sqrt{\mathrm{E_{tx}}}\, X_{tx}\,(t - j\,.\,T_f\, - C_j^k.T_c\,)$$

Where E_{tx} is the transmitted pulse energy; $X_{tx}(t)$ denotes the basic pulse shape and C_j^k represents the j^{th} component of the pseudo random Time Hopping Sequence. The received signal r(t) when only one user is present can be expressed as:

$$r(t) = A.S_{tx}(t-\tau) + n(t)$$

$$r(t) = \sum_{j=-\infty}^{+\infty} A.\sqrt{\mathrm{E_{tx}}}\, X_{tx}\,(t - j\,.\,T_f\, - C_j^k.T_c\, -\, \tau) + n(t)$$

where represents the pulse propagation delay and n(t) is Additive White Gaussian Noise (AWGN) with $N_0/2$ power density and A represents the attenuation the signal experiences during propagation [11]. It depends on the considered channel model in terms of path loss, multipath, shadowing. In a multi user scenario where N_u users are active, the received signal is expressed as

$$r(t) = \sum_{k=1}^{k=N_u} A_K\,.\,S_{tx}\,(t - \tau_k\,) + n(t)$$

$$r(t) = A_1\,.\,S_{tx}\,(t - \tau_1\,) + \sum_{k=2}^{N_u} A_K\,.\,S_{tx}\,(t - \tau_k\,) + n(t)$$

where τ_k represents the delay associated to the propagation and a synchronism between clocks. A_k represents the attenuation of the k^{th} user's signal (k=1 represents the signal of the user interest). This formulation can be used to characterize the TH-IR-UWB PHY layer in a multi user scenario and directly reports to the network simulator [4]; however the used propagation delay does not represent the real propagation delay for the real deployment configuration. The used Bit Error Rate (BER) versus the Signal to Interference and Noise Ratio (SINR) is also based on a perfect power control assumption which is not always realistic.

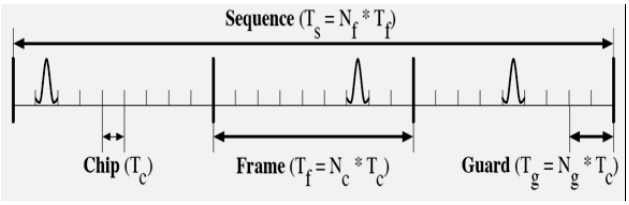


Figure 3: Classic IR-UWB signal and its parameters: Tc is the duration of a chip, Tf = Nc.Tc is the duration of a frame and Ts = Nf.Tf is the duration of a sequence. Tg = Ng.Tc is guard time used to prevent ISI.

IV ALOHA MAC PROTOCOL

ALOHA like Medium Access protocols for IR-UWB have shown their benefit [12]. A node immediately transmits once it has a packet to transmit without caring about the channel state (no need to Clear Channel Assessment: CCA). The MAC scheduler waits for the acknowledgement of the transmitted packet for a defined duration. When the acknowledgement is received before the expiration of this delay, it transmits the next packet. Otherwise, the transmitted packet must be retransmitted

until the number of retransmission exceeds the retransmission limit.

4.1 ALOHA Throughput

The probability that *n* packets arrive in two packets time is given by:

$$P(n) = \frac{(2G)^n e^{-2G}}{n!}$$

Where G is traffic load.

The probability P(0) that a packet is successfully received without collision is calculated by letting n=0 in the above equation. We get:

$$P(0) = e^{-2G}$$

We can calculate throughput S with a traffic load G as follows:

$$S = G.P(0) = G.e^{-2G}$$

The Maximum throughput of ALOHA, shown in Figure 4 (*G=1/2)* is:

$$S_{max} = \frac{1}{2e} \approx 0.184$$

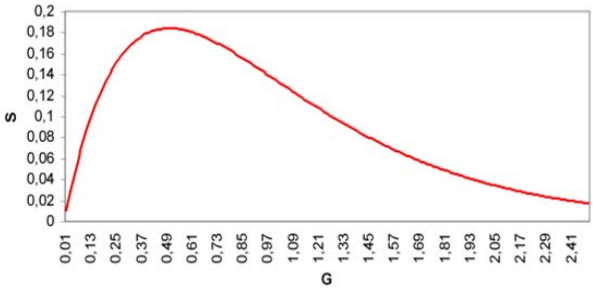


Figure 4: ALOHA throughput

4.2 Transition Diagram of ALOHA

With ALOHA the transmitter does not care about the channel state, once it has a packet to send, it transmits it on the medium, according to its own THS. As the received packets are not acknowledged here, no retransmission is needed. This protocol leads to low latency and gives a high priority to new events to be notified to the base station in a WSN application. It well suits applications where latency and new events notification is critical (see Figure 5).

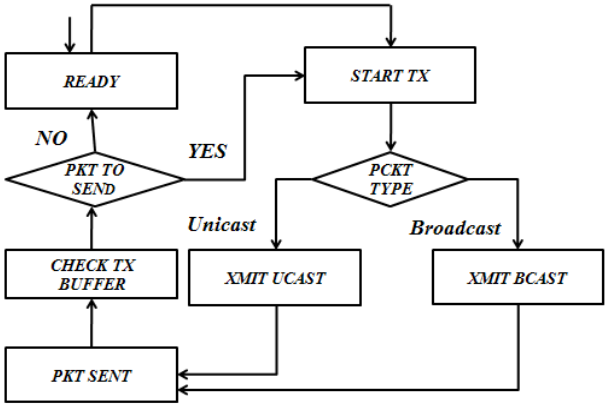


Figure 5: Transition diagram of ALOHA

V WIDEMAC

5.1 Presentation

WideMac was presented as a novel MAC protocol designed for wireless sensor networks using ultra wide band impulse radio transceivers. It makes all nodes periodically (period T_W, identical for all nodes) and asynchronously wake up, transmit a beacon message announcing their availability and listen for transmission attempts during a brief time T_{Listen}.

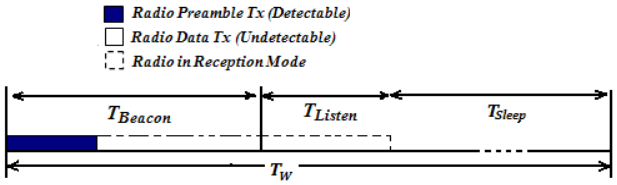


Figure 6: Detailed view of a WideMac period

Figure 6 illustrate a single period structure. It starts with a known and detectable synchronization preamble and is followed by a data sequence which announces the node address and potentially other information, such as a neighbor list or routing table information (for instance, cost of its known path to the sink). A small listening time follows T_{Listen}, during which the node stays in reception mode and that allows it to receive a message [13].

The whole period composed of T_{beacon} and T_{Listen} is called T_a (time of activity); and its very small compared to the time window T_W. This period is followed by a long sleeping period T_{Sleep} during which nodes save energy by keeping the radio in its sleeping mode.

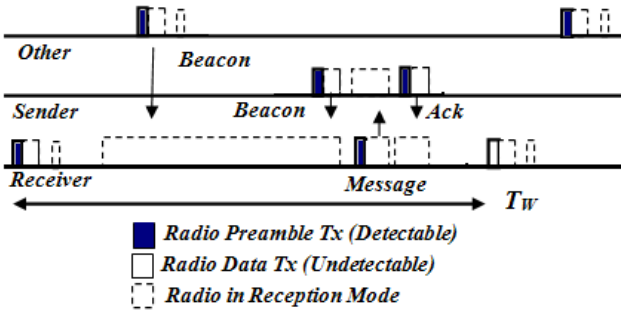


Figure 7: An initial WideMac data transmission.

When a node has a message to transmit, it first listens to the channel until it receives the beacon message of the destination node. This beacon message contains a backoff exponent value that must be used by all nodes when trying to access this destination. If this value is equal to zero, the source node can transmit immediately. Otherwise, it waits a random backoff time, waits for the destination beacon, and transmits its data packet. Because of the unreliability of the wireless channel, packets are acknowledged. If a packet is not acknowledged, or if the destination beacon was not received a retransmission procedure using the backoff algorithm is initiated, until the maximum number of retransmissions maxTxAttempts is reached.

The details of the backoff algorithm are described in subsection. Figure 7 depict a sender node listening to the channel, ignoring the beacon message of another node, and sending its message to the destination after receiving its beacon. The exchange ends with an acknowledgment

message transmitted by the receiver node and addressed to the sender node [14].

5.2 WideMac Backoff Algorithm

The backoff algorithm has a major effect on collision, latency and fairness. WideMac periodic beacons allow the sender nodes to get some information on the channel state at the destination. This can be used to reduce the hidden and exposed terminal problems. The WideMac transmission procedure works as follows: a candidate sender node first listens for the receiver node's beacon. Once it finds it, it can either immediately attempt transmission (default for lightly loaded networks) or it can start a backoff timer before sending (this is activated by a flag always Backoff in the beacon). In both cases, the sender node waits for an acknowledgment. If it does not arrive, a retransmission procedure begins. The sender node chooses a random time parameterized by the receiver node's Backoff Exponent (BE) which was broadcast in the beacon, using a binary exponential backoff:

$$T_{Backoff} = N_{Backoff}.T_W,$$
$$where \;\; N_{Backoff} \in [0, 2^{BE_{Receiver}} - 1].$$

The backoff time is thus a function of the wake-up interval T_W and of the channel state at the receiver node, as captured by $BE_{Receiver}$. Such a receiver-based backoff parameterization was also proposed in IR-MAC [15] . The use of a slotted backoff time based on T_W is natural since all candidate sender nodes are synchronized on the receiver node's wake up times: using a fraction of T_W would not change anything as the node would not transmit before receiving the destination beacon. Using an integer multiple of T_W for the unit backoff duration would increase latency and spread the traffic, but this can also be achieved by adapting the value of $BE_{Receiver}$ to the traffic conditions.

5.3 Power Consumption Models

Each normal T_W interval starts with a beacon frame transmission followed by a packet or a beacon reception attempt, during this start a node must enter transmission mode ($E_{SetupTx}$),transmit its beacon ($T_{Beacon}P_{Tx}$), switch to reception mode (E_{SwRxTx}) and attempt a packet reception ($T_{Listen}P_{Rx}$). These costs are regrouped in the beacon energy E_{Beacon}.

$$\mathrm{E_{Beacon}} = \mathrm{E_{SetupTx}} + \mathrm{T_{Beacon}P_{Tx}} + \mathrm{E_{SwTxRx}} + \mathrm{T_{Listen}P_{Rx}}$$

In addition, during a time L, a node must sometimes transmit a packet E_{Tx} or receive one E_{Rx}, and sleep the rest of the time E_{Sleep}, resulting to the following average power consumption:

$$\mathrm{P_{WideMac}} = \frac{1}{\mathrm{T_W}}(\mathrm{E_{Beacon}} + E_{Tx} + \mathrm{E_{Rx}} + \mathrm{E_{Sleep}})$$

Where:

$$E_{Tx} = K.C_{Tx}(P_{out}).V_B.T_{Tx}$$
$$E_{Rx} = K.C_{Rx}.V_B.T_{Rx}$$
$$E_{Sleep} = C_{Sleep}.V_B.T_{Sleep}$$

K represents the message length in bytes, P_{out} is the transmission power, C_{Tx}, C_{Rx} and C_{Sleep} represent the current intensities for the three modes, T_{Tx} and T_{Rx} are the time of transmission and reception.

VI Simulations and Results

6.1 OMNet++ and MiXiM Simulation Platform

OMNeT++ is an extensible, modular, component-based C++ simulation library and framework which also includes an integrated development and a graphical runtime environment; it is a discreet events based simulator and it provides a powerful and clear simulation framework.

MiXiM joins and extends several existing simulation frameworks developed for wireless and mobile simulations in OMNeT++. It provides detailed models of the wireless channel, wireless connectivity, mobility models, models for obstacles and many communication protocols especially at the Medium Access Control (MAC) level. Moreover, it provides a user-friendly graphical representation of wireless and mobile networks in OMNeT++, supporting debugging and defining even complex wireless scenarios [16].

6.2 Simulation Parameters

We performed the simulations in the MiXiM 2.1 release framework with the OMNeT++ 4.2 network simulator.

We used a grid network, where nodes transmit packets to a Sink node; also we ran several simulations with different nodes numbers and parameters values to evaluate our new protocol.

TABLE I:
ENERGY PARAMETERS

ıeter	
	0 mW
	mW
	mW
:x	0 mW
x	mW
.x	0 mW
'x	0 mW

TABLE II:
TIMING PARAMETERS

ıeter	
$T_{SetupRx}$	0.000103 s
$T_{SetupTx}$	0.000203 s
T_{SwTxRx}	0.000120 s
T_{SwRxTx}	0.000210 s
$T_{RxToSleep}$	0.000031 s
$T_{TxToSleep}$	0.000032 s
Bit rate	0.850000 Mbps

For the energy consumption we used the following radio power consumption parameters shown in TABLE I. For the radio timing we used the parameters shown bellow in Table II.

6.3 Results

In this section, we present the results obtained using the timing and energy parameters cited in section 6.2. The low power consumption of WideMac was concretized by the results shown in Figure 8. It shows that the power consumption of WideMac protocols is remarkably less than the ALOHA MAC protocol. This factor (power consumption) is considered as a key factor for WideMac protocol since it influences directly the Networks life time.

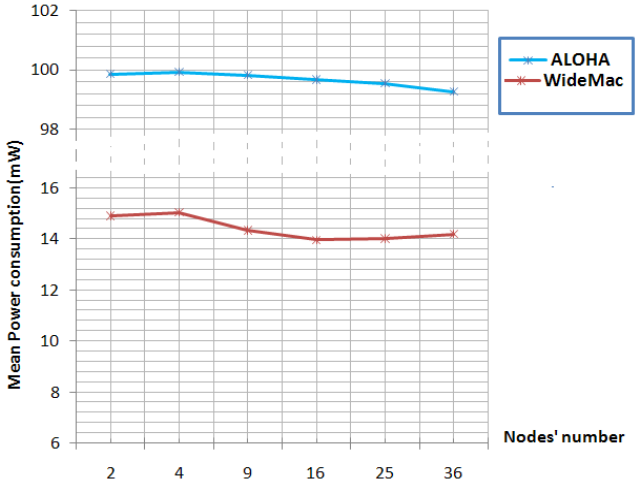


Figure 8: Nodes power consumption average

CONCLUSION

Power consumption was and is an interesting issue that is stills a factor in the development of WSN protocols especially in the physical and MAC layers; it is the primary metric to design a sensor node in wireless sensor network. The low power consumption is the main advantage of the WideMac protocol; it is also very close to an ideal energy consumption model for the IR-UWB based transceivers and gave a good result at this level. This result was achieved thanks to the fact that the network nodes are asleep in the T_{sleep} periods which occupy a wide range in the T_W periods.

We aim, as a future work, to develop a new adapted routing protocol that will be paired with WideMac to largely exploit the IR-UWB features.

REFERENCES

[1] Tien-Wen Sung, Ting-Ting Wu, Chu-Sing Yang, Yueh-Min Huang, ”Reliable Data Broadcast For Zigbee Wireless Sensor Networks”, International Journal on Smart Sensing and Intelligent Systems, VOL. 3, NO. 3, September 2010.

[2] Won Suk Jang, William M. Healyn, ”Assessment of Performance Metrics for Use of WSNs in Buildings”, International Symposium on Automation And Robotic in Construction (ISARC 2009), June 27-29, 2009, pp. 570575.

[3] Hussein T. Mouftah, M. Khanafer, M. Guennoun, ”Wireless Sensor Network Architectures for Intelligent Vehicular Systems”, In Symposium International for Telecommunication Techniques (2010)

[4] Changsu Suh, Zeeshan Hameed Mir, Young-Bae Ko, ”Design and implementation of enhanced IEEE802.15.4 for supporting multimedia service in Wireless Sensor Networks”, The International Journal of Computer and Telecommunications Networking, Volume 52 Issue 13, September, 2008 Pages 2568-2581

[5] N. Golmie, D. Cypher, O. Rebala, ”Performance analysis of low rate wireless technologies for medical applications”, Journal of Computer Communications 28(10), 12661275 (2005) ISSN:0140-3664

[6] H. Zhoul, X. Chen, X. Liu, J. Yang, ”Applications of Zigbee Wireless Technology Tomeasurement System in Grain Storage”, In: Computer and Computing Technologies in Agriculture II. IFIP International Federation for Information Processing, vol. 3, pp. 20212029 (2009), doi:10.1007/978-1-4419-0213-9 52

[7] A. Willig, ”Recent and emerging topics in wireless industrial communication”, IEEE Transactions on Industrial Informatics 4(2), 102124 (2008)

[8] Marcel Meli, Marc Sommerhalder, Martin Gysel, ”Using IEEE802.15.4 / ZigBee in audio applications”, Embedded World, Nuremberg (February 2006).

[9] Aubin Lecointre, Abdoulaye Berthe, Daniela Dragomirescu, Robert Plana, ”Performance Evaluation of Impluse Radio Ultra Wide Band Wireless Sensor Networks”, MILCOM’09 Proceedings of the 28th IEEE conference on Military communications, Pages 1191 1197.

[10] Jin-Shyan Lee, Yu-Wei Su, Chung-Chou Shen, ”A Comparative Study of Wireless Protocols: Bluetooth, UWB, ZigBee, and Wi-Fi”, The 33rd Annual Conference of the IEEE Industrial Electronics Society (IECON) Nov. 5-8, 2007, Taipei, Taiwan.

[11] G. Durisi and G. Romano, ”On the validity of Gaussian approximation to characterize the multiuser capacity of UWB-TH-PPM”, IEEE UWBST, Aug 2002. pp 157-161.

[12] N. Chirdchoo, W. S. Soh, and K. C. Chua, ”Aloha based MAC protocols with collision avoidance for underwater acoustic networks”, in Proceedings of the 26th IEEE International Conference on Computer Communications (INFOCOM ’07), pp. 22712275, June 2007.

[13] D. Piguet, J.-D. Decotignie, J. Rousselot, ”A MAC protocol for micro flying robots coordination”, European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement n 231855 (sFly).

[14] J. Rousselot, A. El-Hoiydi, J.-D. Decotignie, “WideMac: Simple and Efficient Medium Access for UWB Sensor Networks”, IEEE International Conference on Ultra-Wideband, 2008.

[15] Yanjun Sun, Omer Gurewitz, and David B. Johnson,”RI-MAC: a receiver- initiated asynchronous duty cycle MAC protocol for dynamic traffic loads in wireless sensor networks”, In SenSys ’08: Proceedings of the 6th ACM conference on Embedded network sensor systems, pages 1–14, New York, NY, USA, 2008. ACM.

[16] A.Kpke, M.Swigulski, K.Wessel, D. Willkomm, P.T.Klein Haneveld, T.E.V.Parker, O.W.Visser, H.S.Lichte, S.Valentin, ”Simulating Wireless and Mobile Networks in OMNeT++ The MiXiM Vision”, In Proc. Intl. Workshop on OMNeT++ (co-located with SIMUTools ’08), Mar. 2008.

Euclidean 3D Reconstruction of Unknown Objects from Multiple Images

Soulaiman El Hazzat
LIIAN, Dep. Of Mathematics and informatics
Faculty of Sciences Dhar-Mahraz
Fes, Morocco
soulaiman.elhazzat@yahoo.fr

Abderrahim Saaidi
LIIAN, Dep. Of Mathematics and informatics
Faculty of Sciences Dhar-Mahraz
Fes, Morocco
saaidi.abde@yahoo.fr

Khalid Satori
LIIAN, Dep. Of Mathematics and informatics
Faculty of Sciences Dhar-Mahraz
Fes, Morocco
khalidsatorim3i@yahoo.fr

***Abstract*—In this paper, we are interested in the problem of Euclidean 3D reconstruction of unknown objects by passive stereo vision method. Our method is based on the combination between Harris and Sift interest point detectors, to take advantage of the power of these two detectors, which will be useful when matching step, as a key step for 3D reconstruction, In order to have a sufficient number of matches distributed on the images. These matches will be used to estimate the 3D points (the projection matrices will be estimated after calibration using 3D Calibration Pattern). Finally, a 3D mesh is constructed by 3D Delaunay triangulation, applied to the 3D points reconstructed. Experimental results prove that this method is practical and gives satisfying results without going through the propagation step.**

***Index Terms*—Interest Points, Matching, Calibration, 3D reconstruction, 3D mesh, 3D Delaunay triangulation.**

I. Introduction

3D reconstruction from 2D images is an important problem in computer vision. Many approaches offer solutions to this problem : Stereo vision [1], Structure from motion [2 ,3], Shape from silhouettes[4], shape from shading [5, 6], and shape from texture [7].

In this paper, we are intersted in the problem of multi-view reconstruction from images by passive stereo vision methods.

In this method, the 3D information is estimated, only from images taken by cameras without any controlled light. However, the active stereo vision [8] uses a controlled light source such as a laser, or a structured light to find the 3D information.

The implementation of our method (Figure 1) involves four principal phases :

Camera calibration : Consists in estimating the cameras parameters.

Interest Points Detection : is a preliminary step in many computer vision processes, many methods have been proposed to extract points of interest. In this paper, we combined between Harris [9] and Sift [10] interest point detectors.

Matching : Finding in two images of the same scene, taken at different positions, pairs of pixels which are the projections of the same point of the scene. In this phase, the detected interest points are matched by ZNCC (Zero mean Normalised Cross Correlation) correlation measure [11].

To eliminate false matches, we used the global constraint given by the fundamental matrix F.

3D reconstruction of matched points: is to estimate the 3D coordinates from point matches and projection matrices estimated.

3D Delaunay triangulation : 3D mesh is constructed by 3D Delaunay triangulation, applied to the 3D points reconstructed. It is a triangulation that satisfies the Delaunay criterion (empty sphere)

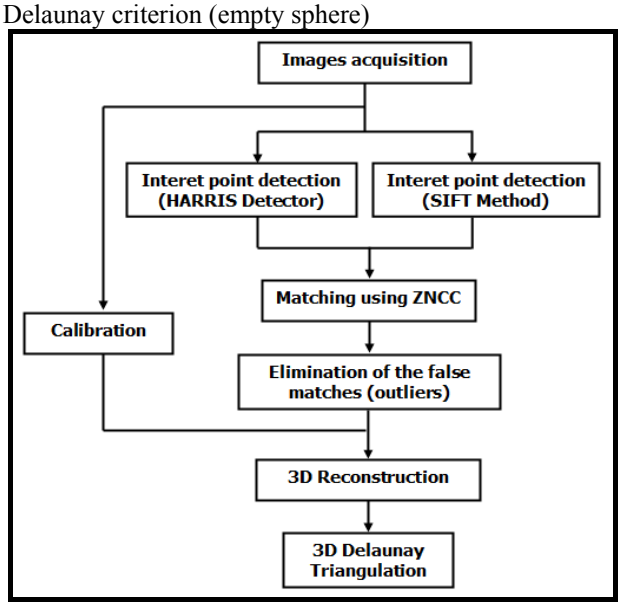


Figure 1 : Euclidien 3D Reconstruction Steps

II Camera Model and Calibration

2.1 Pinhole Camera Model

The pinhole model (Figure 2) consists of the image plane and the optical center O. A point $M_i = (X_i, Y_i, Z_i)^T$ of the 3D scene is projected onto the image plane at a $m_i = (u_i, v_i)^T$ point. This perspective projection is represented by the following formula :

doi:10.4304/jetwi.6.1.59-63

$$\begin{pmatrix} \lambda u_i \\ \lambda v_i \\ \lambda \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix} \quad (1)$$

λ : is a factor of homogeneity

$P = (p_{ij})_{i=1\ldots3}^{j=1\ldots4}$: is the perspective projection matrix

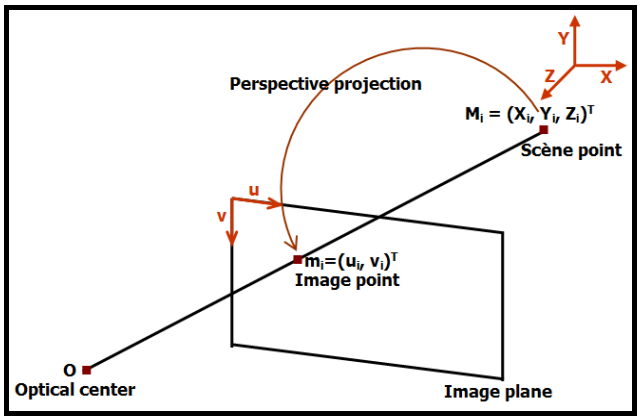


Figure 2 : Pinhole camera model

2.2 Calibration

The camera calibration consists in estimating the perspective projection matrix P that contains the parameters of the model.

3D Calibration Pattern (Figure 3) was used to estimate this matrix.

(a) (b)

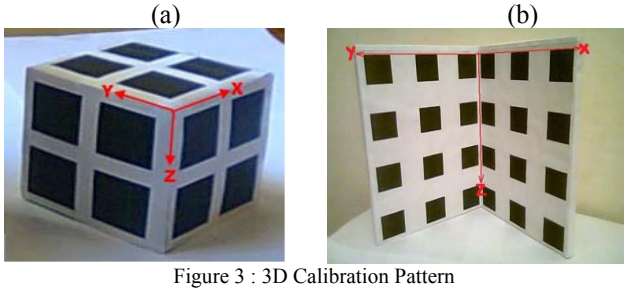


Figure 3 : 3D Calibration Pattern

From equation (1), we deduce :

$$\begin{aligned} X_i p_{11} + Y_i p_{12} + Z_i p_{13} + p_{14} - u_i X_i p_{31} - u_i Y_i p_{32} - u_i Z_i p_{33} = u_i p_{34} \\ X_i p_{21} + Y_i p_{22} + Z_i p_{23} + p_{24} - v_i X_i p_{31} - v_i Y_i p_{32} - v_i Z_i p_{33} = v_i p_{34} \end{aligned} \quad (2)$$

Each point M_i of our 3D Calibration Pattern with known coordinates (X_i, Y_i, Z_i) is projected onto the image plane at a coordinate point (u_i, v_i), provides the equations (2). These equations are linear over the coefficients of P. Therefore at least 6 non-coplanar points are needed to determine P.

III Detection and Matching

3.1 Interest Points Detection

There are many methods [13] of detection point, but they do not have the same performances. In this paper, we combined between Harris [9] and Sift [10] interest point detectors. Indeed, Sift is considered as one of the best performing detectors because of its robustness to scale, rotation, translation and lighting changes. Harris Detector can find points on objects, specifically near the corner.

3.1.1 Sift keypoints

Interest Points detection takes place in two steps : (1) Scale Space Extrema Detection, (2) Keypoint Localization.

- **Scale Space Extrema Detection**

The detection is done in a space called **Scale Space** that has three dimensions : x, y and σ.

The Scale Space of an image I(x, y) is defined by:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (3)$$

Where, ***G*** is the Gaussian function.

To find the extrema, we use the function **DoG** (Difference of Gaussians) defined by :

$$\begin{aligned} D(x, y, \sigma) &= L(x, y, k\sigma) - L(x, y, \sigma) \\ D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \end{aligned} \quad (4)$$

The extrema are pixels which present a maximum intensity, or minimum, compared with their immediate neighbours in the image as well as those in the space scale (26 neighbors).

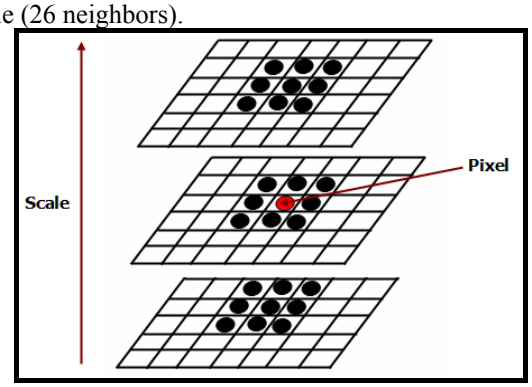


Figure 4 : Scale-space extrema detection

- **Keypoint Localization.**

In this step, key point candidates are localized and low contrast key points or those located on the image edge are removed.

3.1.2 Harris Corner Points

The detection is based on a three step process :

- Calculation of measurement of Harris R(x,y) (5) in each point of the image
- Thresholding to keep the larger values.
- Determination of local maxima in the values of R remaining.

$$R(x, y) = \det(M) - k \cdot [trace(M)]^2 \quad (5)$$

With $k = 0.04$

M : is the matrix of autocorrlation

$$M = G(\sigma) * \begin{bmatrix} (I_x)^2 & I_x I_y \\ I_x I_y & (I_y)^2 \end{bmatrix} \quad (6)$$

G : is the Gaussian filter

$I(x, y)$: is the image intensity

I_x : is the first derivative at x

I_y : is the first derivative at y

- $R < 0$: at the vicinity of an edge
- $R = 0$: in a homogeneous region
- $R >$ *seuil* : near a point of interest

3.2 Matching

The Matching of points is an important step in 3D reconstruction from images. It is to find for each point of the left image a corresponding in the right image.

In this paper, ZNCC [11] (Zero-mean Normalized Cross-Correlation) measure is used for matching interest points between different images.

The ZNCC value for a window size $(2N+1)\times(2P+1)$ is defined by :

$$ZNCC\ (m_1(u_1,v_1), m_2(u_2,v_2)) = \frac{A}{\sqrt{B\times C}} \qquad (7)$$

With :

$$A = \sum_{i=-N}^{N}\sum_{j=-P}^{P}(I_1(u_1+i,v_1+j)-\overline{I_1(u_1,v_1)})\times(I_2(u_2+i,v_2+j)-\overline{I_2(u_2,v_2)})$$

$$B = \sum_{i=-N}^{N}\sum_{j=-P}^{P}(I_1(u_1+i,v_1+j)-\overline{I_1(u_1,v_1)})^2$$

$$C = \sum_{i=-N}^{N}\sum_{j=-P}^{P}(I_2(u_2+i,v_2+j)-\overline{I_2(u_2,v_2)})^2$$

$$\overline{I_1(u_1,v_1)} = \frac{1}{(2N+1)(2P+1)}\sum_{i=-N}^{N}\sum_{j=-P}^{P}I_1(u_1+i,v_1+j)$$

$$\overline{I_2(u_2,v_2)} = \frac{1}{(2N+1)(2P+1)}\sum_{i=-N}^{N}\sum_{j=-P}^{P}I_2(u_2+i,v_2+j)$$

The value of the correlation ZNCC varies between -1 and 1. Consider a point in the image1 (left image), and its correlation is calculated with all points of the search box in the image2 (right image).

One retains only the maximum correlation point and above a threshold.

3.3 Removal of the False Matches

Not all extracted matches are accurate, can be of false matches (outliers). To avoid these matches, we used the global constraint given by the fundamental matrix F (matrice 3x3).

The RANSAC algorithm [12] was used to estimate this matrix and verify if the matches are correct.

This estimate is based on the equation :

$$m_2^T F m_1 = 0 \qquad (8)$$

With (m1, m2) is a pair of corresponding points.

IV Euclidean 3D Reconstruction

4.1 3D Reconstruction of Matched Points

3D reconstruction is to estimate, from the matches already made and projection matrices defined during the phase of calibration P and P', the coordinates (X, Y, Z) of a point M of the scene.

By applying the projection equation (1) for a couple of matches (m1, m2), we write:

$$m_1 = P\ M$$

$$m_2 = P'M$$

By developing these equations, we find:

$$\begin{pmatrix} p_{11}-p_{31}u_1 & p_{12}-p_{32}u_1 & p_{13}-p_{33}u_1 & p_{14}-p_{34}u_1 \\ p_{21}-p_{31}v_1 & p_{22}-p_{32}v_1 & p_{23}-p_{33}v_1 & p_{24}-p_{34}v_1 \\ p'_{11}-p'_{31}u_2 & p'_{12}-p'_{32}u_2 & p'_{13}-p'_{33}u_2 & p'_{14}-p'_{34}u_2 \\ p'_{21}-p'_{31}v_2 & p'_{22}-p'_{32}v_2 & p'_{23}-p'_{33}v_2 & p'_{24}-p'_{34}v_2 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = 0 \qquad (9)$$

The calculation of the coordinates (X, Y, Z) is obtained by solving the linear system (9) using the linear least squares sense. It is also possible to solve this system by a singular value Decomposition (SVD).

4.2 3D Delaunay Triangulation

3D mesh is constructed by 3D Delaunay triangulation (Figur 5), applied to the 3D points reconstructed. It is a triangulation that satisfies the Delaunay criterion (empty sphere)

- **The Delaunay criterion**

Delaunay criterion is satisfied for a couple of tetrahedra if no node of an element (tetrahedron) is contained in the Interior of the sphere circumscribing the other element.

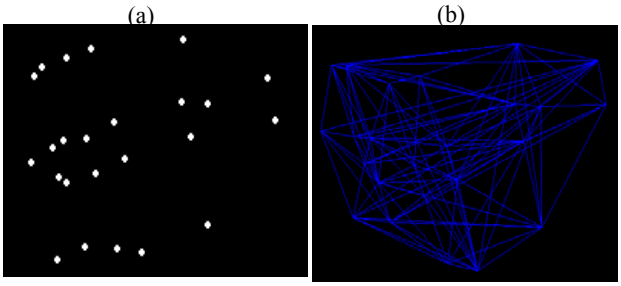


Figure 5 : (a) 3D point cloud (b) 3D Delaunay Triangulation

V Experiments and Results

The presented algorithm was implemented in Java, using both JAMA library and The Java 3D API.

5.1 Images Acquisition

A digital camera was used for these experiments. The resolution used for the images is 640 × 480.

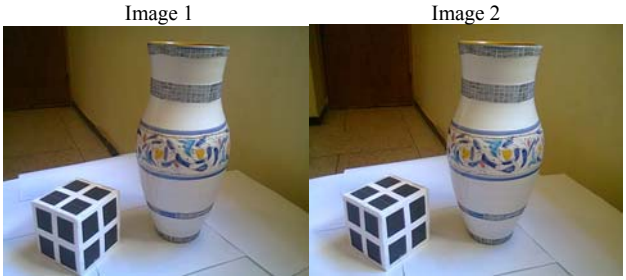


Figure 6. Images used for 3D reconstruction

5.2 Camera Calibration

In each teken view, our camera is calibrated (variable parameters). The coordinates of 10 well distributed corners in the 3D pattern are used.

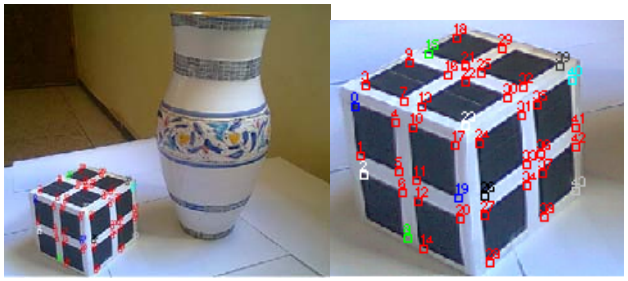


Figure 7. Detection of corners of our Calibration Pattern (Harris detector)

The first projection matrix estimated :

$$P1 = \begin{bmatrix} 14,8664 & -10,7459 & 1,8653 & 153,8133 \\ -0,5906 & -2,2124 & 16,2515 & 371,4603 \\ 0,0187 & 0,0061 & 0,0060 & 1,0000 \end{bmatrix}$$

The second projection matrix estimated:

$$P2 = \begin{bmatrix} 14,6487 & -8,6591 & 1,5661 & 146,6990 \\ -1,4943 & -2,4432 & 14,7060 & 362,3744 \\ 0,0136 & 0,0062 & 0,0038 & 1,0000 \end{bmatrix}$$

5.3 Matching

The ZNCC method was used to make the matching of points of interest detected by the Harris detector and the Sift method. The matches extracts are not all accurate, to remove false matches, we used the RANSAC algorithm [12].

Figure 8. Matching of points of interest (N = 7, P = 7 and thresholdZNCC = 0.85)

5.4 Euclidean 3D Reconstruction

We have the projection matrices and a set of matches, we can pass to the 3D reconstruction (we used a Singular Value Decomposition to solve the linear system (9)) :

5.4.1 3D Reconstruction of Matched Points

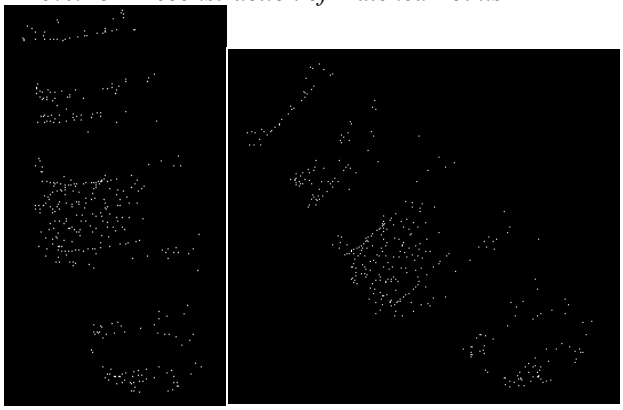


Figure 9. 3D Reconstruction of 389 Matched Points

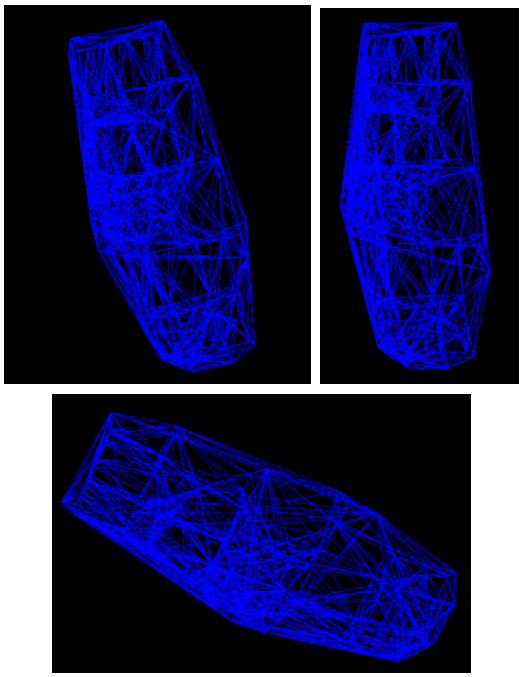


Figure 10. 3D mesh constructed by 3D Delaunay triangulation, applied to the 3D points reconstructed

5.4.2 *3D Reconstruction of Matched Points After propagation method* [14]

The propagation method is based on a set of reliable matches(germs). The germ with the best score ZNCC is removed from the current list of matches(germs), the new matches are searched in its neighborhood.

This method was applied to 167 germs for having 9622 matches.

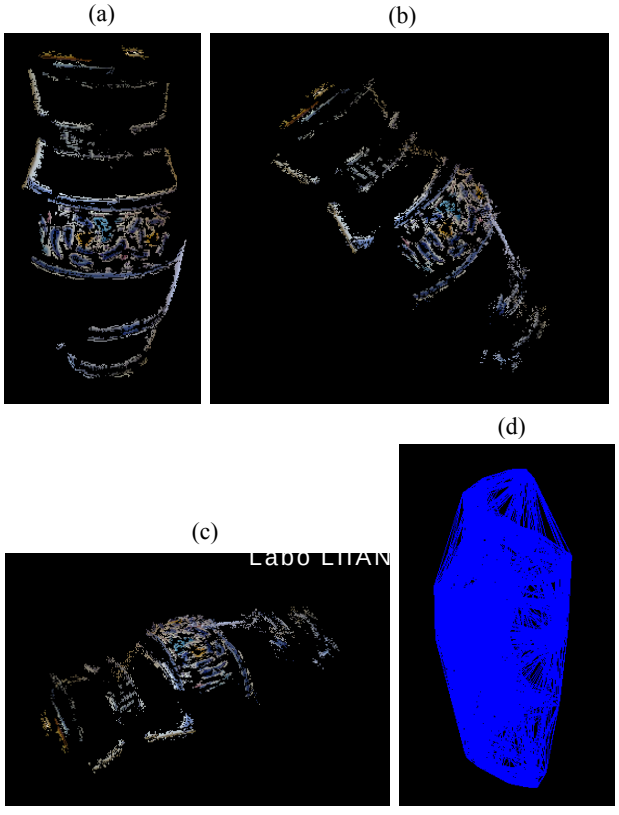


Figure 11. (a), (b) and (c) 3D Reconstruction of 9622 Matched Points (d) 3D Delaunay triangulation of 3D points reconstructed

5.4.3 COMPARAISON

The results are presented in the table below :

TABLE 1.
RESULTS OF 3D RECONSTRUCTION

	Proposed method	**Propagation method [14]**
Number of reconstructed points	389	9622
Time of calculations	15 s	3 min
Reprojection error (pixel) (10)	0.96	1.25

Based on the results of experiments, we can deduce :

TABLE 2.
COMPARAISON

	Proposed method	**Propagation method [14]**
Advantages	Rapid	The number of reconstructed points is large enough
	Quality of the resulting 3D model is accepted	Quality of the resulting 3D model is accepted
Disadvantages	Non-operational for non-textured objects	Non-operational for non-textured objects
		Slow due to the large number of reconstructed points

- **The reprojection error is defined by:**

$$Err = \frac{1}{2n}\sum_{i=1}^{n}\left(d(P_1M_i, m_{1i})^2 + d(P_2M_i, m_{2i})^2\right) \quad (10)$$

n : is the number of matches (Number of 3D points).
P_1 et P_2 : are the projection matrices.
(m_{1i}, m_{2i}) : is a pair of corresponding points.
M_i : is the 3D reconstructed point from (m_{1i}, m_{2i}).

CONCLUSION

The proposed 3D reconstruction method enables to achieve satisfactory results in a short time. The combination between Harris and Sift detectors allows to have a sufficient number of matches without passed to the propagation step, What has been very useful to have a 3D model of quality in short time.

REFERENCES

[1] N.Salman, M.Yvinec, Surface Reconstruction from Multi-View Stereo of Large-Scale Outdoor Scenes. The International Journal of Virtual Reality, 2010.

[2] M.Han, T.Kanade. Multiple Motion Scene Reconstruction with Uncalibrated Cameras. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 7, 2003.

[3] S. Christy and R. Horaud. Euclidean Shape and Motion from Multiple Perspective Views by Affine Iterations. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 18, NO. 11, 1996.

[4] K.Kolev, T.Brox, D.Cremers. Fast Joint Estimation of Silhouettes and Dense 3D Geometry from Multiple Images. IEEE Transactions On Pattern Analysis And Machine Intelligence, 2012.

[5] J.D.Durou, V.Charvillat, M.Daramy and P.Gurdjos. Résolution du shape from shading par apprentissage. ORASIS - Congrès des jeunes chercheurs en vision par ordinateur, Praz-sur-Arly : France (2011)

[6] M.B.Adm, A.M.Said. 3D Reconstruction using Interactive Shape from Shading. International Journal of Computer Applications Volume 28– No.1, August 2011

[7] T.Collins, J.D.DUROU, Pierre GURDJOS and Adrien BARTOLI. Single-view Perspective Shape-from-Texture with Focal Length Estimation: A Piecewise Affine Approach. 3DPVT'10, France, May 2010.

[8] A.Dipanda, S.Woo, F.Marzani, J.M. Bilbault, 3-D Shape Reconstruction In An Active Stereo Vision System Using Genetic Algorithms, Pattern Recognition, Vol.36, pp 2143-2159, 2003.

[9] C.Harris et M.Stephens. A combined Corner and Edge Detector. Alvey vision Conference. pp. 147-151, 1988.

[10] D. G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2):91–110, 2004.

[11] S.Elhazzat, A.Saaidi et K.Satori. Reconstruction tridimensionnelle semi-automatique. WoISA'13, 2013.

[12] M. A. Fischler and R. C. Bolles. Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM, 24(6) :381–395, 1981.

[13] N.El akkad, A.Baataoui, A.El abderrahmani, A.Saaidi et K.Satori. Etude comparative des détecteurs des points d'intérêt. WCCCS, 2011.

[14] M. Lhuillier et L.Quan. Appariement dense robuste à l'aide de contraintes géométrique locales et globales. RFIA, 2000.

Comparative Study of PCA, ICA, LDA using SVM Classifier

Anissa Bouzalmat
Sidi Mohamed Ben Abdellah University
Department of Computer Science
Faculty of Science and Technology
Route d'Imouzzer B.P.2202 Fez
30000 Morocco
anissabouzalmat@yahoo.fr

Jamal Kharroubi
Sidi Mohamed Ben Abdellah University
Department of Computer Science
Faculty of Science and Technology
Route d'Imouzzer B.P.2202 Fez
30000 Morocco
jamal.kharroubi@yahoo.fr

Arsalane Zarghili
Sidi Mohamed Ben Abdellah University
Department of Computer Science
Faculty of Science and Technology
Route d'Imouzzer B.P.2202 Fez
30000 Morocco
a.zarghili@ieee.ma

***Abstract*—Feature representation and classification are two key steps for face recognition. We compared three automated methods for face recognition using different method for feature extraction: PCA (Principle Component Analysis), LDA (Linear Discriminate Analysis), ICA (Independent Component Analysis) and SVM (Support Vector Machine) were used for classification. The experiments were implemented on two face databases, The ATT Face Database [1] and the Indian Face Database (IFD) [2] with the combination of methods (PCA+ SVM), (ICA+SVM) and (LDA+SVM) showed that (LDA+SVM) method had a higher recognition rate than the other two methods for face recognition.**

***Index Terms*—Face Recognition, SVM, LDA, PCA, ICA.**

I. INTRODUCTION

Face Recognition is a term that includes several sub-stages as a two step process: Feature extraction and classification.

Feature extraction for face representation is one of central issues to face recognition systems, it can be defined as the procedure of extracting relevant information from a face image.

There are many feature extraction algorithms, most of them are used in other areas than face recognition.

Researchers in face recognition have used, modified and adapted many algorithms and methods to their purpose . For example, Principle component analysis (PCA) was applied to face representation and recognition [3, 4, 5].

The PCA method [5] is obviously of advantage to feature extraction, but it is more suitable for image reconstruction because of no consideration for the separability of various classes. Aiming at optimal separability of feature subspace, LDA (Linear Discriminate Analysis) can just make up for the deficiency of PCA [6]. ICA (Independent Component Analysis) is a method that finds better basis by recognizing the high-order relationships between the pixels images [7], once the features are extracted, the next step is to classify the image .A large margin classifiers are proposed recently in machine learning such as Support Vector Machine (SVM) [8]. The method was used in this step is SVM (Support Vector Machines) which have been developed in the frame work of statistical learning theory, and have been successfully applied to a number of applications, ranging from time series prediction, to face recognition, to biological data processing for medical diagnosis [9,10]. VC (Vapnik-Chervonenkis) dimension theory and SRM (Structural Risk Minimization) principle based SVM can well resolve some practical problems such as small sample size, nonlinear, high dimensional problems, etc. [11,12] .

In this paper SVMs were used for classification using different method for feature extraction: PCA, LDA, ICA, the experiments were implemented on two face databases, The ATT Face Database [1] and the Indian Face Database (IFD) [2] .

The face recognition system is shown as Fig. 1.

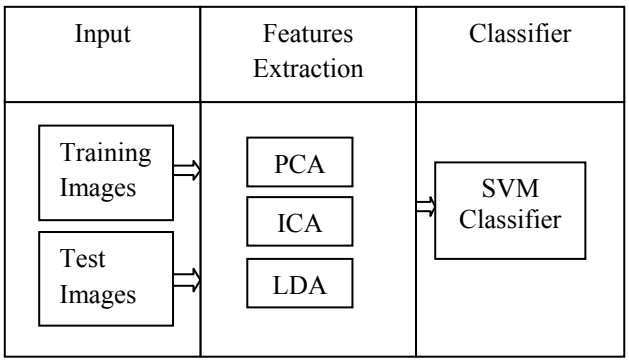


Fig 1: The face recognition system

The outline of the paper is as follows: Section 2 feature extraction and classification. In section 3 contains experimental results. Section 4 concludes the paper.

doi:10.4304/jetwi.6.1.64-68

II FEATURE EXTRACTION

Feature extraction involves several steps - dimensionality reduction, feature extraction and feature selection. We have a large features vector which considers the whole image that needs a reduction of dimension and selection the important features. Then these new features will be used for the training and testing of SVM classifier .In this paragraph we describe three techniques of extraction feature, Principal component analysis (PCA), independent component analysis (ICA) and linear discriminate analysis (LDA).

2.1 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a powerful tool for feature extraction as proposed by Turk and Pentland [13]. The main advantage of PCA is that it can reduce the dimension of the data without losing much information.
Suppose there are N images Ii(i=l,2,---,N), each image is denoted as a column vector xi , and the dimension is M. The mean of the images is given by:

$$\bar{x} = \sum_{i=1}^{N} \frac{x_i}{N} \quad (1)$$

the covariance matrix of images is given by

$$C = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^T = \frac{1}{N} XX^T \quad (2)$$

Where $X = [x_1 - \bar{x}, x_2 - \bar{x}, ..., x_N - \bar{x}]$ the projection space is made up of the eigenvectors which correspond to the significant eigenvalues when M>>N, the computational complexity is increased .we can use the singular value decomposition (SVD), theorem to simplify the computation .the matrix X, whose dimension is M*N and rank is N, can be decomposed as:

$$X = U\Lambda^{\frac{1}{2}}V^T \quad (3)$$

$$U = X \vee \Lambda^{\frac{1}{2}} \quad (4)$$

Where :
$\Lambda = diag[\lambda_1, \lambda_2, ..., \lambda_N], \lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N$, are the nonzero eigenvalues of XX^T and $X^T X$.
$U = [u_1, u_2, ..., u_M], V = [v_1, v_2, ..., v_N]$ are orthogonal matrices.
u_i is the eigenvector of XX^T , v_i is the eigenvector of $X^T X$ and the λ_i is the corresponding eigenvalue.
u_i is calculated by following :

$$U_i = \frac{1}{\sqrt{\lambda_i}} Xv_i \quad i = 1, 2, ..., N \quad (5)$$

The p eigenvectors $U = [u_1, u_2, ..., u_p] \quad p \leq N$ corresponding to the p significant eigenvalues are selected to form the projection space and the sample feature is obtained by calculating.

2.2 Analyse discriminate linear (LDA)

LDA also known as Fisher's Discriminate Analysis, is another dimensionality reduction technique, it determines a subspace in which the between-class scatter (extra personal variability) is as large as possible, while the within-class scatter (intrapersonal variability) is kept constant. In this sense, the subspace obtained by LDA optimally discriminates the classes-faces.

We have a set of C-class and D-dimensional samples $\{x^{(1}, x^{(2}, ... x^{(N}\}$
N_1 of which belong to class w_1, N_2 to class w_2 and N_c to class w_c, In order to find a good discrimination of these classes we need to define a measure of separation, We define a measure of the within-class scatter by Eq. (6):

$$S_i = \sum_{x \in w_i}^{c} (x - \mu_i)(x - \mu_i)^T \quad (6)$$

Where: $S_w = \sum_{i=1}^{c} S_i$ and $\mu_i = \frac{1}{N_i}\sum_{x \in w_i} x_i$

And the between-class scatter Eq. (7) becomes:

$$S_B = \sum_{i=1}^{C} N_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (7)$$

Where: $\mu = \frac{1}{N}\sum_{\forall x} x = \frac{1}{N}\sum_{i=1}^{C} N_i \mu_i$

Matrix $S_T = S_B + S_W$ is called the total scatter similarly, we define the mean vector and scatter matrices for the projected samples as:

$$\tilde{S}_W = \sum_{i=1}^{c}\sum_{y \in w_i}(y - \tilde{\mu}_i)(y - \tilde{\mu}_i)^T$$

$$\tilde{S}_B = \sum_{i=1}^{c} N_i (\tilde{\mu}_i - \tilde{\mu})(\tilde{\mu}_i - \tilde{\mu})^T$$

Where: $\tilde{\mu}_i = \frac{1}{N_i}\sum_{y \in w_i} y \quad , \quad \tilde{\mu} = \frac{1}{N}\sum_{\forall y} y$

From our derivation for the two-class problem, we can write: $\tilde{S}_B = W^T S_B W$ and $\tilde{S}_W = W^T S_W W$

Recall that we are looking for a projection that maximizes the ratio of between-class to within-class scatter. Since the projection is no longer a scalar (it has C−1 dimensions), we use the determinant of the scatter matrices to obtain a scalar objective function Eq. (8):

$$J(W) = \frac{|\tilde{S}_B|}{|\tilde{S}_W|} = \frac{W^T S_B W}{W^T S_W W} \quad (8)$$

And we will seek the projection matrix W* that maximizes this ratio it can be shown that the optimal projection matrix W* is the one whose columns are the eigenvectors corresponding to the largest eigenvalues of the following generalized eigenvalue problem Eq. (9):

$$w^* = [w_1^* | w_2^* | .. w_{c-1}^*] = \arg\max \frac{|W^T S_B W|}{|W^T S_W W|} \Rightarrow (S_B - \lambda_i S_W) W_i^* \quad (9)$$

S_B is the sum of C matrices of rank ≤1 and the mean vectors are constrained by : $\frac{1}{c}\sum_{i=1}^{C} \mu_i = \mu$

Therefore, S_B will be of rank (C−1) or less and this means that only (C−1) of the eigenvalues λ will be non-

zero. The projections with maximum class separability information are the eigenvectors corresponding to the largest eigenvalues of $S_W^{-1}S_B$.

We seek (C−1) projections $\left[y_{1,2},\ldots y_{c-1}\right]$ by means of $(c-1)$ projection vectors wi arranged by columns into a projection matrix $W=[w_1 \mid w_2 \mid \ldots \mid w_{c-1}]: y_i = w_i Tx \Rightarrow y = WTx$.

2.3 Independent Component Analysis (ICA)

The most common method for generating spatially localized features is to apply independent component analysis (ICA) to produce basis vectors that are statistically independent (not just linearly decorrelated , as with PCA) [14].it is an alternative to PCA which provides a more powerful data representation [15] and it's a discriminate analysis criterion, which can be used to enhance PCA.

ICA for face recognition has been proposed under two architecture by Barlett et. al. [16]. The architecture 1 aimed at finding a set of statistically independent basis images while the architecture 2 finds a factorial code. In this paper, the architecture 1 has been used. This process involves the following two initial steps :

1. The face images in the database are organized as a matrix X in which each row corresponds to an image.

2. The face database is processed to obtain a reduced dataset in order to reduce the computation efficiency of the ICA algorithm. The reduced dataset is obtained from the first m principal component (PC) eigenvectors of the image database. Hence the first step is applying PCA to determine the m PCs ,then the ICA algorithm is performed on the principal components using the mathematical procedure described in [17].

III CLASSIFICATION: SUPPORT VECTOR MACHINE (SVM)

SVMs (Support Vector Machines) are a useful technique for data classification and are still under intensive research [18],[19]. Although SVM is considered easier to use than Neural Networks, there are several kernels are being proposed by researchers, the four basic kernels as follow: linear, polynomial, sigmoid and radial basis function (RBF), We chose RBF kernel function for SVM classifier in our face recognition experiments Eq:10 which has fewer numerical difficulties [18] .

$$K\left(x_i, y_j\right) = \exp\left(-\gamma\left\|x_i - x_j\right\|^2\right) \qquad \gamma \succ 0 \qquad (10)$$

γ is kernel parameter and parameterized using $\gamma := \frac{1}{2\sigma^2}$

3.1 Maximal Margin Hyperplanes

After we change the representation of the training examples by mapping the data to a features space F where the optimal separating hyper plane (OSH) is constructed Fig 2, we limited our study to the case of two-class discrimination [8] and we consider the training data S a set of l vectors features each vector has n dimension, where each point xi belongs to one of two classes identified by the label -1 or 1 Eq 11.

$$S = \left\{\left(x_i, y_i\right) \middle| x_i \in R^n, y_i \in \{-1,1\}\right\}_{i=1}^{l} \qquad (11)$$

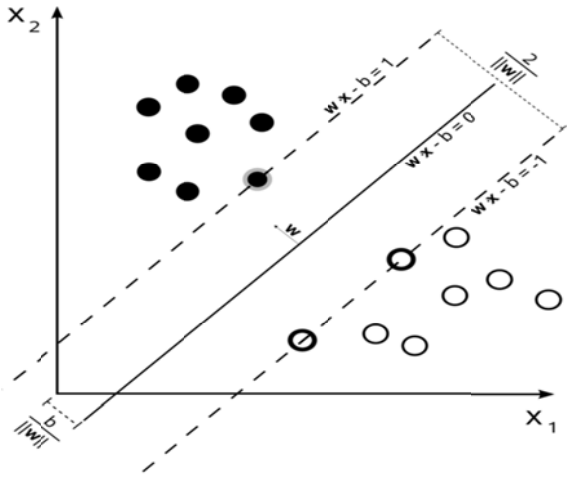


Fig 2: Maximum-margin hyperplane for SVM trained with samples from two classes.

We have solving a quadratic optimization problem with linear constraints that can be interpreted in terms of the Lagrange multipliers calculated by quadratic programming Eq: 12

$$\left\{\begin{array}{l}\max(\alpha_i): \tilde{L}(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j k\left(x_i, x_j\right) \text{(for any i=1,..,n)?} \quad \leq \alpha i \leq c \\ \sum_{i=1}^{n} \alpha_i y_i = 0\end{array}\right\} (12)$$

αi are the Lagrange multipliers parameters to be adjusted ,c is the penalty parameter of the classification error term it must be adjusted because the data are rarely completely separable, the xi are the training examples.

The solution of the optimization problem will be a vector $w \in F$, that can be written as a linear combination of the training inputs Eq: 13 $w = \sum \alpha_i y_i x_i \qquad (13)$

(w,b) define the hyperplane $OSH = \{x : w.x + b = 0\}$ b is the bias.

We use the separating (OSH), once we have trained it on the training set, The (OSH) divides the R^n into two regions: one where $w.x_i + b \geq 0$ and one where $w.x_i + b \leq 0$. To use the maximal margin classifier, we determine on which side the test vector lies and assign the corresponding class label. Hence, the predicted class of a test point x is the output of the decision function Eq 14.

$$d\left(x\right) = \operatorname{sgn}\left(\sum_{i=1}^{l} \alpha_i y_i k\left(x_i, x\right) + b\right) \qquad (14)$$

$$K\left(x_i, y_j\right) = \exp\left(-\gamma\left\|x_i - x_j\right\|^2\right) \quad \gamma \succ 0$$

3.2 Multiclass Classification

SVM was originally designed for binary classification. Face recognition is a multi-class classification problem. There are two basic methods for face recognition with SVMs: one against-one and one-against-all. The one-against-one method is

Classification between each pair classes . The one-against-all is classification between each class and all the

rest classes. In our experiments the one-against-all method was used for classification.

In real world problems we often have to deal with $n \geq 2$ classes. Our training set will consist of pairs (x_i, y_i), where $x_i \in R^n$ and $y_i \in \{1,...,n\}, i = 1..l$ for extending the two-class to the multiclass case this method will be described briefly below.

3.2.1 One vs. all approach

In the one-Vs-all approach n SVMs are trained. Each of the SVMs separates a single class from all remaining classes [20,21] ,A more recent comparison between several multi-class techniques [22] favors the one-vs-all approach because of its simplicity and excellent classification performance. Regarding the training effort, the one-vs-all approach is preferable over the one-vs-one approach since only n SVMs have to be trained compared to $n(n-1)/2$ SVMs in the pairwise approach (one-vs-one) [23], [24], [25] . The construction of a n-class classifier using two-class discrimination methods is usually done by the following procedure:

Construct n two-class decision functions $d_k(x), k = 1,..,n$ that separate examples of class k from the training points of all other classes,

$$d_k(x) = \begin{Bmatrix} +1 & \text{if x belongs to class k} \\ -1 & \text{otherwise} \end{Bmatrix}$$

In the face database of n individuals, 10 face images for everyone. 5 images among the 10 images of every one were taken to compose training samples and the rest 5 ones compose test samples.

Five images of first individual was taken and marked as positive samples, the all images of other training samples as negative samples. Both positive samples and negative samples were taken as input samples to train a SVM classifier to get corresponding support vectors and optimal hyperplane. The SVM was labeled as SVM1. In turn we can get the SVM for every individual and labeled as SVM1, … , SVMn respectively.

The n SVMs can divide the samples into n classes. When a test sample was in turn inputted to every SVM, there would be several cases:

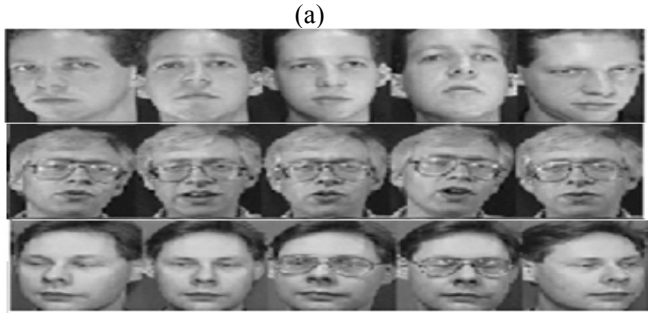
- If the sample was decided to be positive by SVMi and to be negative by others SVMs at the same time, then the sample was classified as class i.
- If the sample was decided to be negative by several SVMs synchronously and to be positive by other SVMs, then the classification was false.
- If the sample was decided to be negative by all SVMs synchronously, then the sample was decided not belonging to the face database.

IV. Experimentation and Results

Our experiments were performed on two face databases, The ATT Face Database [1] and the Indian Face Database (IFD) [2] the ATT database contains images with very small changes in orientation of images for each subject involved, while the IFD contains a set of 10 images for each subject where each image is oriented in a different angle compared to the other.

These two databases both contains 10 classes, each class have 5 images for training and 5 images for testing Fig 3 and Fig 4. We use these Databases for comparison of different face recognition algorithms such as PCA+SVM, LDA+SVM and ICA+SVM. We extract different features from a training set and testing set using PCA, LDA, ICA methods. Using these feature we trained the classifier SVM and find the accuracy of the three methods, the recognition rates of the three methods PCA+SVM, LDA+SVM, ICA+SVM were shown as Fig. 5.

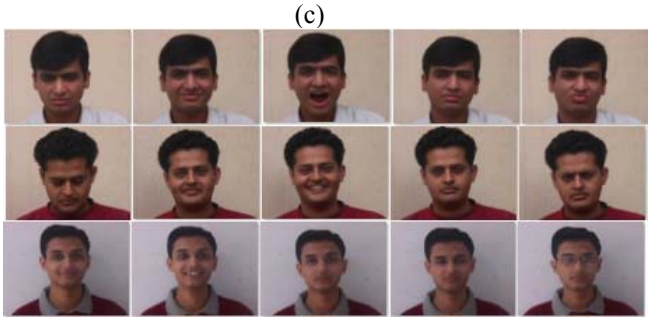
(a)



(b)

Fig 3: Examples of (a) training and (b) test images of (ATT) Face Database

(c)



(d)

Fig 4: Examples of (c) training and (d) test images of (IFD) Face Database

The comparison is done on the basis of rate of recognition accuracy. Comparative results obtained by testing the PCA+SVM, LDA+SVM, ICA+SVM algorithms on both the IFD and the ATT databases Fig.5.

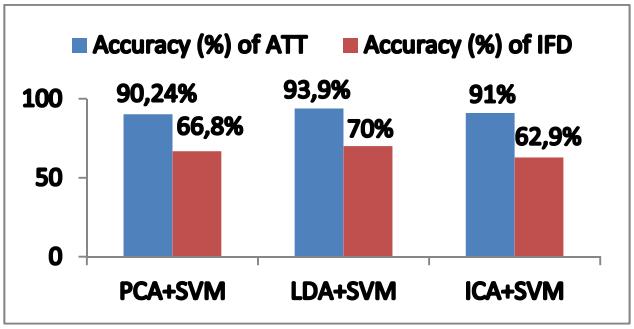


Fig 5: Comparative of the combination algorithms PCA+SVM,LDA+SVM,ICA+SVM On the basis Of recognition accuracy

It is observed that recognition rate of the method LDA+SVM is 93.9% obtained on ATT face database and 70% on IFD face database it is the higher as compare to PCA+SVM and ICA+SVM methods for both IFD and ATT databases.

CONCLUSION

We presented a face recognition method based on SVM classifier combined with LDA feature extraction. We implemented experiments on IFD and ATT face database. First, LDA for dimension reduction and SVM for classification. The experimental results showed that LDA+SVM method had a higher recognition rate than the other two methods for face recognition.

REFERENCES

[1] AT&T Laboratories Cambridge,http://www.cl.cam.ac.uk/Research/DTG/attarchive:pub/data/att_faces.tar.Z.
[2] Vidit Jain and Amitabha Mukherjee, The Indian Face Database.http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/, 2002
[3] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. Journal of the Optical Society of America A - Optics, Image Science and Vision, 4(3):519–524, March 1987.
[4] M. Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(1):103–108, 1990.
[5] M. Turk and A. P. Pentland, “Eigenfaces for recognition”, Journal of Cognitive Neuroscience, vol. 3, no. 1,pp. 71–86, 1991.
[6] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: recognition using class specific linear projection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7,pp. 711–720, 1997.
[7] M. Bartlett, J. Movellan, and T. Sejnowski. Face recognition by independent component analysis. IEEE Trans. on Neural Networks,13(6):1450–1464, November 2002.
[8] V. Vapnik, Statistical Learning Theory, John Wiley and Sons, New York, 1998.
[9] Vapnik V. “Statistical Learning Theory”, Wiley, New York, 1998.
[10] G. Paliouras, V. Karkaletsis, and C.D. Spyropoulos (Eds.), “Support Vector Machines: Theory and Applications,” ACAI ’ 99, LNAI 2049,pp. 249-257, 2001.
[11] Nello Cristianini, John Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, Cambridge, 2000.
[12] Vladimir N. Vapnik, The Nature of Statistical Learning Theory,Springer, New York, 2000.
[13] M. A. Turk and A. P. Pentland, "Face Recognition Using Eigenfaces," in IEEE CVPR), 1991,pp. 586-591.
[14] M. S. Bartlett, Face Image Analysis by Unsupervised Learning: Kluwer Academic, 2001.
[15] C. Liu and H. Wechsler. Comparative assessment of independent component analysis (ica) for face recognition. In Proc. of the Second In-ternational Conference on Audio- and Video-based Biometric Person Authentication, AVBPA’99, Washington D.C., USA, March 1999.
[16] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, Face Recognition by Independent Component Analysis, IEEE Trans. on Neural Networks, Vol. 13, No. 6, November 2002, pp. 1450-1464
[17] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, Face Recognition by Independent Component Analysis, IEEE Trans. on Neural Networks, Vol. 13, No. 6, November 2002, pp. 1450-1464
[18] Vapnik, V., 1995: The Nature of Statistical Learning Theory. Springer, N.Y.
[19] Vapnik, V., 1998: Statistical Learning Theory. Wiley, N.Y.
[20] C. Cortes, V. Vapnik, Support vector networks, Mach. Learning 20 (1995) 1–25.
[21] B. Sch€olkopf, C. Burges, V. Vapnik, Extracting support data for a given task, in: U. Fayyad, R.Uthurusamy (Eds.), Proc. First Int. Conf. on Knowledge Discovery and Data Mining, AAAI Press,Menlo Park, CA, 1995.
[22] R. Rifkin, Everything old is new again: a fresh look at historical approaches in machine learning,Ph.D. thesis, M.I.T., 2002.
[23] M. Pontil, A. Verri, Support vector machines for 3-d object recognition, IEEE Trans. Pattern Anal.Mach. Intell. (1998) 637–646.
[24] G. Guodong, S. Li, C. Kapluk, Face recognition by support vector machines, in: Proc. IEEE Int.Conf. on Automatic Face and Gesture Recognition, 2000, pp. 196–201.
[25] Platt J., N. Cristianini, and J. Shawe-Taylor. Large margin dags for multiclass classification. In Advances in Neural Information Processing Systems 12, pages 547-553. MIT Press, 2000.

SVM Classification of High Resolution Urban Satellites Images using Composite Kernels and Haralick Features

Aissam Bekkari, Soufiane Idbraim, Driss Mammass and Mostafa El Yassa
IRF – SIC laboratory,
Faculty of sciences
Agadir -Morocco

Danielle Ducrot
Cesbio, Toulouse- France

***Abstract*—The classification of remotely sensed images knows a large progress taking in consideration the availability of images with different resolutions as well as the abundance of classification's algorithms. A number of works have shown promising results by the fusion of spatial and spectral information using Support vector machines (SVM) which are a group of supervised classification algorithms that have been recently used in the remote sensing field.**

For this purpose we propose a methodology exploiting the properties of Mercer's kernels to construct a family of composite kernels that easily combine multi-spectral features and Haralick texture features as data source.

The proposed approach was tested on common scenes of urban imagery. The three different kernels tested allow a significant improvement of the classification performances and a flexibility to balance between the spatial and spectral information in the classifier. The experimental results indicate an accuracy value of 92.55% which is very promising.

***Index Terms*—SVM, Classification, Composite Kernels, Haralick features, Satellite image, Spectral and spatial information.**

I. INTRODUCTION

With the commercial emergence of the optical satellite images of sub-metric resolution (Ikonos, Quickbird) the realization as well as the regular update of numerical maps with large scales become accessible and increasingly frequent. The classification of such images is similar to that of other image types, it follows the same principle, and it is a method of analysis of data that aims to separate the image into several classes in order to gather the data in homogeneous subsets, which show common characteristics. It aims to assign to each pixel of the image a label which represents a theme in the real study area (e.g. vegetation, water, built, etc) [1].

Several classification algorithms have been developed since the first satellite image was acquired in 1972 [2-4]. Among the most popular and widely used is the maximum likelihood classifier [5]. It is a parametric approach that assumes the class signature in normal distribution. Although this assumption is generally valid, it is invalid for classes consisting of several subclasses or classes having different spectral features [6]. To overcome this problem, some non-parametric classification techniques such as artificial neural networks, decision trees and Support vector machines (SVM) have been recently introduced.

SVM is a group of advanced machine learning algorithms that have seen increased use in land cover studies [7, 8]. One of the theoretical advantages of the SVM over other algorithms (decision trees and neural networks) is that it is designed to search for an optimal solution to a classification problem whereas decision trees and neural networks are designed to find a solution, which may or may not be optimal. This theoretical advantage has been demonstrated in a number studies where SVM generally produced more accurate results than decision trees and neural networks [5, 9]. SVMs have been used recently to map urban areas at different scales with different remotely sensed data. High or medium spatial resolution images (e.g., IKONOS, Quickbird, Landsat (TM)/ (ETM+), SPOT) have been widely employed on urban land use classification for individual cities for ; building extraction, road extraction and other man-made objects extraction [10, 11].

On other hand, the consideration of the spatial aspect in the spectral classification remains very important, for this case, Haralick described methods for measuring texture in gray-scale images, and statistics for quantifying those textures. It is the hypothesis of this research that Haralick's Texture Features and statistics as defined for gray-scale images can be modified to incorporate spectral information, and that these Spectral Texture Features will provide useful information about the image. It is shown that texture features can be used to classify general classes of materials, and that Spectral Texture Features in particular provide a clearer classification of land cover types than purely spectral methods alone.

The proposed method consists of combining spatial and spectral information to obtain a better classification.

doi:10.4304/jetwi.6.1.69-74

We start with the extraction of spectral and spatial information (Haralick texture features) [12]. Then, we apply the SVM classification to the result file.

We exploit the properties of Mercer's kernels to construct a family of composite kernels that easily combine spatial and spectral information. The three different kernels tested demonstrate enhanced classification accuracy as compared to traditional approaches that take into account the spectral information only, and a flexibility to balance between the spatial and spectral information in the classifier

This paper is organized as follows. In the second section, we discuss the extraction of spatial and spectral information especially the Grey-Level Co-occurrence Matrix (GLCM) and Haralick texture features used in experimentations. In section3, we give outlines on the used classifier: Support Vector Machines (SVM). Section4 describe the three different composite kernels used in experimentations. In section5, the results are presented as well as the stating of numerical evaluation. Finally, conclusions are given in section6.

II Extraction of Information

2.1 Spectral Information

The most used classification methods for the multispectral data consider especially the spectral dimension. The set of spectral values of each pixel is treated as a vector of attributes which will be directly employed as entry of the classifier. According to Fauvel [13] this allows a good classification based on the spectral signature of each area. However, this does not take in account the spatial information represented by the various structures in the image.

2.2 Spatial Information

Information in a remote sensed image can be deduced based on their textures. A human analyst is able to distinguish man-made features from natural features in an image based on the 'regularity' of the data. Straight lines and regular repetitions of features hint at man-made objects. This spatial information is useful in distinguishing the different field in the remote sensed image.

Many approaches were developed for texture analysis. According to the processing algorithms, three major categories, namely, structural, spectral, and statistical methods are common ways for texture analysis. Grey-Level Co-occurrence Matrix (GLCM) [14] is one of the most widely used methods, which is a powerful technique for measuring texture features; it contains the relative frequencies of the two neighbouring pixels separated by a distance on the image.

Haralick assumed that the texture information is contained in the co-occurrence matrix, and texture features are calculated from it. A large number of textural features have been proposed starting with the original fourteen features described by Haralick et al [15], however only some of these features are in wide use. Wezska et al [16] used four of Haralick features. Conners and Harlow [17] use five features. Conners, Trivedi and Harlow [18] introduced two new features which address a deficiency in the Conners and Harlow set.

We found that the five features used by Conners and Harlow are commonly used because seen that the fourteen are much correlated with each other, and that the five sufficed to give good results in classification [19].

In this work, we have used these five features: homogeneity (E), contrast (C), correlation (Cor), entropy (H) and local homogeneity (LH), and co-occurrence matrices are calculated for four directions: 0°, 45°, 90°and 135° degrees.

Let us recall their definitions:

$$E = \sum_i \sum_j (M(i,j))^2 \tag{1}$$

$$C = \sum_{k=0}^{m-1} k^2 \sum_{|i-j|=k} M(i,j) \tag{2}$$

$$Cor = \frac{1}{\sigma_i \sigma_j} \sum_i \sum_j (i-\mu_i)(j-\mu_j) M(i,j) \tag{3}$$

Where μ_i and σ_i are the horizontal mean and the variance, and μ_j and σ_j are the vertical statistics.

$$H = \sum_i \sum_j M(i,j) \log(M(i,j)) \tag{4}$$

$$LH = \sum_i \sum_j \frac{M(i,j)}{1+(i-j)^2} \tag{5}$$

Each texture measure can create a new band that can be incorporated with spectral features for classification purposes.

III SVM Classification

In this section we briefly describe the general mathematical formulation of SVMs introduced by Vapnik [20, 21]. Starting from the linearly separable case, optimal hyperplanes are introduced. Then, the classification problem is modified to handle non-linearly separable data and a brief description of multiclass strategies is given.

3.1 Linear SVM

For a two-class problem in a *n*-dimensional space R^n, we assume that *l* training samples $x_i \in R^n$, are available with their corresponding labels yi = ±1, S = {(xi, yi) | i ∈ [1, l]}. The SVM method consists of finding the hyperplane that maximizes the margin, i.e., the distance to the closest training data points for both classes [22]. Noting $w \in R^n$ as the normal vector of the hyperplane and $b \in R$ as the bias, the hyperplane H_p is defined as:

$$\langle w, x \rangle + b = 0, \forall x \in H_p \tag{6}$$

Where $\langle w, x \rangle$ is the inner product between w and x. If $x \notin Hp$ then $f(x) = \langle w, x \rangle + b$ is the distance of x to H*p*. The sign of *f* corresponds to decision function *y* = sgn (f(x)).

Finally, the optimal hyperplane has to maximize the margin: $2/\|w\|$. This is equivalent to minimize $\|w\|/2$ and leads to the following quadratic optimization problem:

$$\min\left[\frac{\|w\|^2}{2}\right] \quad \text{subject to } y_i(\langle w, x_i\rangle + b) \geq 1 \ \ \forall i \in [1, l] \tag{7}$$

For non-linearly separable data, the optimal parameters (w, b) are found by solving:

$$\min\left[\frac{\|w\|^2}{2} + C\sum_{i=1}^{l}\xi_i\right] \quad \text{subject to } y_i(\langle w, x_i\rangle + b) \geq 1 - \xi_i, \xi_i \geq 0 \ \ \forall i \in [1, l] \tag{8}$$

Where the constant C control the amount of penalty and ξ_i are *slack* variables which are introduced to deal with misclassified samples. This optimization task can be solved through its Lagrangian dual problem:

$$\max_{\alpha} \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j \langle x_i, x_j\rangle \quad \text{subject to } 0 \leq \alpha_i \leq C \quad \forall i \in [1, l] \quad \sum_{i=1}^{l}\alpha_i y_i = 0 \tag{9}$$

Finally:

$$w = \sum_{i=1}^{l}\alpha_i y_i x_i \tag{10}$$

The solution vector is a linear combination of some samples of the training set, whose α_i is non-zero, called Support Vectors. The hyperplane decision function can thus be written as:

$$y_u = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i \alpha_i \langle x_u, x_i\rangle + b\right) \tag{11}$$

Where x_u is an unseen sample.

3.2 Non-Linear SVM

Using the Kernel Method, we can generalize SVMs to non-linear decision functions. With this way, the classification capability is improved. The idea is as follows. Via a non-linear mapping Φ, data are mapped onto a higher dimensional space F:

$$\Phi : R^n \to F \quad x \mapsto \Phi(x) \tag{12}$$

The SVM algorithm can now be simply considered with the following training samples: $\Phi(S) = \{(\Phi(x_i), y_i) \mid i \in [1, l]\}$. It leads to a new version of the hyperplane decision function where the scalar product is now: $\langle \Phi(x_i), \Phi(x_j)\rangle$. Hopefully, for some kernels function k, the extra computational cost is reduced to:

$$\langle \Phi(x_i), \Phi(x_j)\rangle = k(x_i, x_j) \tag{13}$$

The kernel function k should fulfill Mercers' conditions.

With the use of kernels, it is possible to work implicitly in F while all the computations are done in the input space. The classical kernels used in remote sensing are the polynomial kernel and the Gaussian radial basis function:

$$k_{poly}(x_i, x_j) = \left[(x_i \cdot x_j) + 1\right]^p \tag{14}$$

$$k_{gauss}(x_i, x_j) = \exp\left[-\gamma\|x_i - x_j\|^2\right] \tag{15}$$

3.3 Multiclass SVMs

SVMs are designed to solve binary problems where the class labels can only take two values: ±1. For a remote sensing application, several classes are usually of interest. Various approaches have been proposed to address this problem [23]. They usually combine a set of binary classifiers. Two main approaches were originally proposed for a k-classes problem.

- ***One versus the Rest***: k binary classifiers are applied on each class against the others. Each sample is assigned to the class with the maximum output.
- ***Pairwise Classification***: $k(k-1)/2$ binary classifiers are applied on each pair of classes. Each sample is assigned to the class getting the highest number of votes. A vote for a given class is defined as a classifier assigning the pattern to that class.

IV Composite Kernels

In the following section, we present three different kernel approaches for the joint consideration of spectral and textural information for multispectral image classification.

4.1 The Stacked Features Approach

The most commonly adopted approach in multispectral image classification is to exploit the spectral content of a pixel (x_i). However, performance can be improved by including both spectral and spatial information in the classifier. This is usually done by means of the 'stacked' approach, in which feature vectors are built from the concatenation of spectral and spatial features. Note that if the chosen mapping Φ is a transformation of the concatenation $x_i \equiv \{x_{i\text{-spat}}, x_{i\text{-spect}}\}$, then the corresponding 'stacked' kernel matrix is:

$$k_{\{Spect,Spa\}} \equiv k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j)\rangle \tag{16}$$

which does not include explicit cross relations between $x_{i\text{-spa}}$ and $x_{i\text{-spect}}$.

4.2 The Direct Summation Kernel

A simple composite kernel combining spectral and textural information naturally comes from the concatenation of nonlinear transformations of $x_{i\text{-spat}}$ and $x_{i\text{-spect}}$. Let us assume two nonlinear transformations $\varphi_1(.)$ and $\varphi_2(.)$ into Hilbert spaces H_1 and H_2, respectively. Then, the following transformation can be constructed:

$$\Phi(x_i) = \{\varphi_1(x_{i-spect}), \varphi_2(x_{i-spa})\} \tag{17}$$

and the corresponding dot product can be easily computed as follows:

$$\begin{aligned} k(x_i, x_j) &= \langle \Phi(x_i), \Phi(x_j)\rangle \\ &= \langle \{\varphi_1(x_{i-spect}), \varphi_2(x_{i-spa})\}, \{\varphi_1(x_{j-spect}), \varphi_2(x_{j-spa})\}\rangle \\ &= k_{spect}(x_{i-spect}, x_{j-spect}) + k_{spa}(x_{i-spa}, x_{j-spa}) \end{aligned} \tag{18}$$

4.3 The Weighted Summation Kernel

By exploiting properties of Mercer's kernels, a composite kernel that balances the spatial and spectral content in (19) can also be created, as follows:

$$k(x_i, x_j) = \mu k_{spect}(x_{i-spect}, x_{j-spect}) + (1-\mu) k_{spa}(x_{i-spa}, x_{j-spa}) \quad (19)$$

where μ is a positive real-valued free parameter ($0 < \mu < 1$), which is tuned in the training process and constitutes a tradeoff between the spatial and spectral information to classify a given pixel. This composite kernel allows us to introduce a priori knowledge in the classifier by designing specific μ profiles per class, and also allows us to extract some information from the best tuned μ parameter.

Note that solving the minimization problem in all kinds of composite kernels requires the same number of constraints as in the conventional SVM algorithm, and thus no additional computational efforts are induced in the presented approaches.

V Experiments

We developed a two stages classification process: the first one is the extraction of the spatial and spectral features, so we compute Grey Level Co-occurrence Matrix (GLCM) to extract Haralick texture features that we add to spectral information. The second one is the classification stage; with SVM, a supervised kernel learning algorithm widely used.We have selected SVMlight which is an implementation of Support Vector Machines (SVMs) in C language [24] with composite kernels.

To use jointly spatial and spectral information, we used three different kernel approaches presented in section 4; which are the stacked features approach (16), the direct summation kernel (18) and the weighted summation kernel (19).

In the case of the weighted summation kernel, μ was varied in steps of 0.1 in the range [0, 1]. For simplicity and for illustrative purposes, μ was the same for all classes in our experiments. The penalization factor in the SVM was tuned in the range $C = \{10^{-1} \dots 10^{7}\}$.

We used the Gaussian RBF kernel (15) (with $\sigma = \{10^{-1} \dots 10^{3}\}$) for the two kernels. k_{spect} uses a spectral information while k_{spa} uses Haralick features.

Concerning data we have used a multispectral satellite image (IKONOS) represented in Fig 1 (a), with size 800 by 600 pixels; at the last we will have, for this image 4 131 individuals (pixels) for learn, 4 952 for validation and 480 000 to classify, divided on six classes (Table 1).

The classification map presented on Fig 1(b), is obtained when the classification is performed using the stacked features approach. When the classification is performed using the direct summation kernel, we obtain the corresponding classification map which is presented on Fig 1(c). A visual analysis of classification maps shows those areas are more homogeneous for the maps obtained using the direct summation kernel.

TABLE 1.
DIFFERENT CLASSES

Class N°	Class name	Train samples	Validation samples
1	Asphalt	1 386	978
2	Green area	480	1 034
3	Tree	196	1 154
4	Soil	813	954
5	Building	920	688
6	Shadow	336	144

The fusion of the spectral and the spatial features using the weighted summation kernel give us the classification map presented on Fig 1(d). The classification map is less noisy and the classification performances are increased globally as well as almost all the classes. It matches well with an urban land cover map in terms of smoothness of the classes; and it also represents more connected classes. Table 2 lists the accuracy estimates for the study area, all models are compared numerically (overall accuracy and kappa coefficient), and table 3, table 4 and table 5 presents respectively the confusion matrix results for SVM classification using the stacked features approach (16), the direct summation kernel (18) and the weighted summation kernel (19).

In conclusion, composite kernels offer excellent performance for the classification of multispectral satellite images by simultaneously exploiting both the spatial and spectral information.

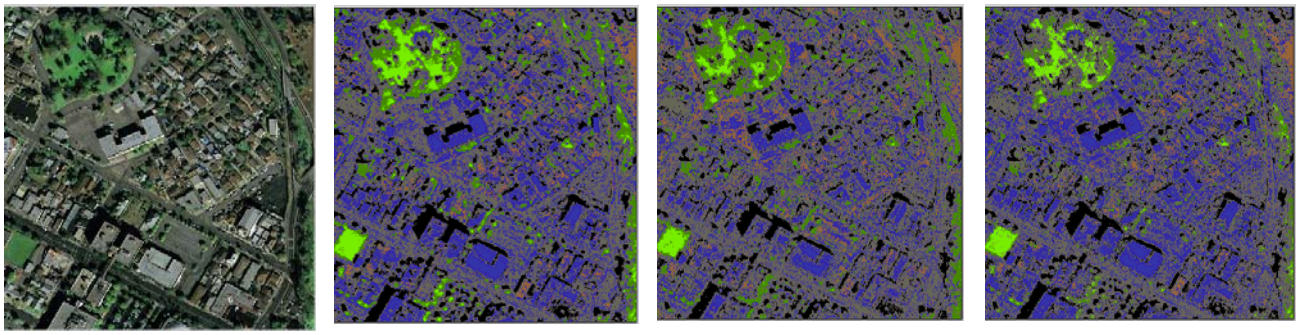
TABLE 2.
OVERALL ACCURACY (%) OF CLASSIFIED IMAGE

Method	OA	Kappa Coefficient
The stacked features approach	92.13%	0.91
The direct summation kernel	92.38 %	0.92
The weighted summation kernel	92.55%	0.92

VI Conclusion

Addressing the classification of high resolution satellite images from urban areas, we have presented three different kernel approaches taking simultaneously the spectral and the spatial information into account (the spectral values and the Haralick features). The weighted summation kernel allows a significant improvement of the classification performances when compared with the two other approaches.

As perspectives, the workflow of this study can be used in other remote sensing application, especially, in rural areas for thematic land cover, and more sophisticated texture techniques to describe the spatial structure of the classes.



(a) **(b)** **(c)** **(d)**

Fig1: (a) Original image, (b) Classification Map obtained using the stacked features approach, (c) Classification Map obtained using the direct summation kernel and (d) Classification Map obtained using the weighted summation kernel.(Asphalt, Green area, Tree, Soil, Building, Shadow)

TABLE 3.
CONFUSION MATRIX RESULTS (%) FOR SVM CLASSIFICATION USING THE STACKED FEATURES APPROACH.
GLOBAL ACCURACY = 92.13%

Class name	Asphalt	Green area	Tree	Soil	Building	Shadow
Asphalt	**90,12**	1,41	3,92	0	2,63	1,92
Green area	1,13	**94,99**	0	1,08	1,54	1,26
Tree	0,28	1,07	**90,82**	2,5	2,82	2,51
Soil	4,84	0,95	0	**91,87**	2,34	0
Building	3,01	1,16	2,69	2,47	**90,67**	0
Shadow	0,62	0,42	2,57	2,08	0	**94,31**

TABLE 4.
CONFUSION MATRIX RESULTS (%) FOR SVM CLASSIFICATION USING THE DIRECT SUMMATION KERNEL.
GLOBAL ACCURACY = 92.38%

Class name	Asphalt	Green area	Tree	Soil	Building	Shadow
Asphalt	**89,93**	2,34	0	3,62	2,12	1,99
Green area	1,13	**93,27**	4,71	0	0,53	0,36
Tree	1,18	2,62	**91,22**	1,08	0,6	3,3
Soil	0	0,98	0	**93,95**	5,07	0
Building	6,04	0,46	1,81	0,09	**91,58**	0,02
Shadow	1,72	0,33	2,26	1,26	0,1	**94,33**

TABLE 5
. CONFUSION MATRIX RESULTS (%) FOR SVM CLASSIFICATION USING THE WEIGHTED SUMMATION KERNEL.
GLOBAL ACCURACY = 92.55%

Class name	Asphalt	Green area	Tree	Soil	Building	Shadow
Asphalt	**89,36**	2,04	1,92	1,5	3,32	1,86
Green area	5,13	**92,21**	0	1,03	1,54	0,09
Tree	1,18	1,52	**93,15**	1,92	0,03	2,2
Soil	1,75	1,13	0,64	**93,04**	3,44	0
Building	1,96	2,78	2,72	0,87	**91,67**	0
Shadow	0,62	0,32	1,57	1,64	0	**95,85**

VII ACKNOWLEDGMENTS

This work was funded by CNRST Morocco and CNRS France Grant under “Convention CNRST CRNS” program SPI09/11.

REFERENCES

[1] Samson C. (2000) "Contribution à la classification des images satellitaires par approche variationnelle et équations aux dérivées partielles" : Thesis of doctorate, university of Nice-Sophia Antipolis.

[2] Townshend, J.R.G., (1992) "Land cover". International Journal of Remote Sensing 13:1319–1328.
[3] Hall, F.G., Townshend, J.R., Engman, E.T., (1995) "Status of remote sensing algorithms for estimation of land surface state parameters." Remote Sensing of Environment 51:138–156.
[4] Lu, D.,Weng, Q.,(2007) "A survey of image classification methods and techniques for improving classification performance. " International Journal of Remote Sensing 28:823–870.
[5] Huang, C., Davis, L.S., and Townshed, J.R.G.,(2002) "An assessment of support vector machines for land cover classification. "International Journal of Remote Sensing 23:725–749.
[6] Kavzoglu, T., Reis, S.,(2008) "Performance analysis of maximum likelihood and artificial neural network classifiers for training sets with mixed pixels." GIScience and Remote Sensing 45:330–342.
[7] Pal, M., and Mather, P. M. (2005). "Support vector machines for classification in remote sensing." International Journal of Remote Sensing, 26:1007−1011.
[8] Zhu, G., and Blumberg, D. G. (2002). "Classification using ASTER data and SVM algorithms: The case study of Beer Sheva, Israel." Remote Sensing of Environment, 80:233-240.
[9] Scholkopf, B., Sung, K., Burges, C., Girosi, F., Niyogi, P., Poggio, T., et al. (1997). "Comparing support vector machines with gaussian kernels to radial basis function classifiers." IEEE Transactions on Signal Processing, 45:2758−2765.
[10] Cao X., Chen J., Imura H., Higashi O., (2009) "A SVM-based method to extract urban areas from DMSP-OLS and SPOT VGT data", Remote Sensing of Environment 113:2205–2209.
[11] Inglada J., (2007) "Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features", ISPRS Journal of Photogrammetry & Remote Sensing 62:236–248.
[12] Bekkari A., Idbraim S., Mammass D. and El yassa M. (2011) "Exploiting spectral and space information in classification of high resolution urban satellites images using Haralick features and SVM" IEEE 2ed International Conference on Multimedia Computing and Systems ICMCS'11 , Ouarzazate, Morocco.
[13] Fauvel M., Benediktsson, J. A., Chanussot J. and Sveinsson, J. R., (2007) "Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles" IEEE International Geoscience and Remote Sensing Symposium, IGARSS 07, Barcelona Spain.
[14] Chiu, W. Y., and Couloigner I. (2004)"Evaluation of incorporating texture into wetland mapping from multispectral images" University of Calgary, Department of Geomatics Engineering, Calgary, Canada, EARSeL eProceedings.
[15] Haralick, R.M., Shanmugam K., and Dinstein I., (1973) "Textural Features for Image Classification." IEEE Transactions on Systems Man and Cybernetics.
[16] J.S. Weszka, C.R. Dyer, and A. Rosenfeld. (1976) "A Comparative Study of Texture measures for Terrain Classification." IEEE Transactions on Systems Man and Cybernetics.
[17] Conners, R.W. and Harlow, C.A., (1980) "A Theoretical Comaprison of Texture Algorithms." IEEE Transactions on Pattern Analysis and Machine Intelligence.
[18] Conners, R.W., Trivedi, M.M., and Harlow, C.A., (1984) "Segmentation of a High-Resolution Urban Scene using Texture Operators." Computer Vision, Graphics and Image Processing.
[19] Arvis V.; Debain C.; Berducat M.; Benassi A. (2004) "Generalization of the cooccurrence matrix for colour images: application to colour texture classification" journal Image Analysis and Stereology, 23:63-72.
[20] Chapel L. (2007) " Maintenir la viabilité ou la résilience d'un système : les machines à vecteurs de support pour rompre la malédiction de la dimensionnalité ? " : Thesis of doctorate, university of Blaise Pascal - Clermont II.
[21] Aseervatham S. (2007) " Apprentissage à base de Noyaux Sémantiques pour le traitement de données textuelles " : Thesis of doctorate, university of Paris 13 –Galilée Institut Laboratory of Data processing of Paris Nord.
[22] Bousquet O., (2001) "Introduction au Support Vector Machines (SVM)", Center mathematics applied, polytechnique school of Palaiseau. http://www.math.u-psud.fr/~blanchard/gtsvm/index.html.
[23] Fauvel M., Chanussot J. and Benediktsson J. A. (2006) "A Combined Support Vector Machines Classification Based on Decision Fusion" IEEE International Geoscience and Remote Sensing Symposium, IGARSS 06, Denver, USA.
[24] SVMlight Version: 6.02 (2008) Developed at University of Dortmund, Informatik, AI-Unit Collaborative Research Center on 'Complexity Reduction in Multivariate Data' (SFB475). http://svmlight.joachims.org/

3D Normalization Based on the Barycentric Coordinates

El Mostafa Rajaallah
Laboratory LAVETE,
Dept. Mathematics and Computer Science
Faculty of Science and Technology,
University Hassan 1st
Settat, Morocco
m.rajaallah@uh1.ac.ma

Lahcen Moumoun
Laboratory LAVETE,
Dept. Mathematics and Computer Science
Faculty of Science and Technology,
University Hassan 1st
Settat, Morocco
lahcenm@gmail.com

Taoufiq Gadi
Laboratory LAVETE,
Dept. Mathematics and Computer Science
Faculty of Science and Technology,
University Hassan 1st
Settat, Morocco
gtaoufiq@yahoo.fr

***Abstract*—The most common technique in the normalization of 3D objects is the Principal Component Analysis (PCA). However, it is well known that the principal axes generated by the PCA may be different for similar shapes. To overcome the limitations of the PCA we propose in this paper a normalization method to give robustness to remeshing, rotations and reflections of meshed 3D objects. This method is based on the integration of the mesh volume, the use of barycentric coordinates and the CPCA (Continuous Principal Component Analysis).**

***General Terms*—Image Processing.**

***Index Terms*—Normalization, PCA, barycentric coordinates.**

I. Introduction

Duplicating real-world object in a digital environment was an interesting task for many applications. The quality of the obtained models was often limited by the capacity of the existing hardware and software. However, recent advances in scanning technology and three-dimensional modeling helped to visualize and manipulate complex models with ease. 3D objects are usually given in an arbitrary orientation and scale. To improve the accuracy of the use of these objects, and more particularly the search by content systems, a near-normalization treatment is often necessary. Thus, the normalization of a 3D object is to orient it correctly in a canonical marker.

The normalization method proposed in this paper aims at addressing problems related to meshing resolution and the orientation of the principal axes of normalized objects. Our method is based on an integration of the mesh volume [15], using the barycentric coordinates of the vertex [14] to generate the matrix of moments (order 2). After generating the matrix of two-order moments, we apply the PCA [1] of the matrix to define a canonical marker for the object. At the end, we will consider a reflection coefficient inspired from the CPCA (Continuous Principal Component Analysis) [2].

We present in the beginning of this article an overview on 3D normalization. Next, we describe our proposal to standardize a 3D object. At the end of this article we present our experimental results with a qualitative and quantitative evaluation of the proposed method.

II A Overview On 3D Normalization

The most commonly used technique for normalization of a 3D object is based on Principal Component Analysis (PCA), in which the center of gravity is chosen as the origin, the size of the bounding box as a scale factor of the form and the determining canonical axes based on the calculation of the eigenvalues and eigenvectors of the covariance matrix of the set of points representing the object in question. The eigenvalues are sorted in a descending order, the eigenvectors are chosen on the basis of this ranking, the first vector is aligned with the first axis (x), the second with the second axis (y) and the third vector with the third axis (z) (Paquet et al). [1].

The PCA has been extended by Vranic et al. [2], leading to the PCA continues (Continuous Principal Component Analysis), the proposed approach is more accurate than traditional ACP (discrete), but it is a little more expensive in execution time.

Several Normalization approaches have been developed recently. Ricard et al. [12] proposed a method to integrate a 3D object without using a discrete representation but directly from its bounding. Their method is based on the contour's integration of the object for generating the moments' matrix; it is robust against remeshing of 3D objects. The proposed method uses the discrete PCA, on moments's matrix, with its limitations.

Tedjoknsumo et al. [3] proposed a normalization method based on bilateral symmetry planes (PSB). In their method, they calculate the three axes of the PCA and the three planes normal to its axis. Subsequently, they consider that the plan ACP generates the smallest error of symmetry (a feature introduced by the authors). Then, they pivot this plane around the three axes of rotation with predefined increments to generate the plane which minimizes the error of symmetry. After projecting all the points of the 3D object on this plan, and applying

doi:10.4304/jetwi.6.1.75-79

PCA 2D on the projected points for generate the first and second main axes, it turned out that their method is time consuming due to the procedure of research the planes of symmetry.

H. Fu et al. [4] presented a solution for detecting the vertical orientation of the object. Their method is based on the assumption that most objects in real life are symmetrical with respect to a plane, but this method is not appropriate when dealing with deformable shapes.

Chaouch et al. [5] proposed a method based on the symmetry properties of Minovic et al. [6] by considering the interesting properties of reflection symmetries of the PCA. The axes of the PCA are considered and the axes for initial's shape are studied. They introduced a measure for assessing local symmetry of translational invariance (CILT) whose main objective is to provide optimal directions (principal axes) for characterization compact and relevant of the 3D shape. The limitation of this method it is based on assumptions derived from human perception.

Yu-Shen et al. [7] proposed a method based on the LMS (Least Median of Squares) by considering the work of Fleishman et al. [8] to guide the calculation of the principal axes of the PCA. The proposed method gives good results for deformed objects. The main limitation of this method is that it is sensitive to the density of the samples.

Recently, as part of a project funded by the National Science Foundation of China, Chao Wang et al. [9] proposed a normalization method articulated volumetric 3D shapes. The main contribution of their work can be summarized in a proposed normalization algorithm to estimate the location and orientation of articulated 3D shapes, based on solving a problem using weighted least squares IRLS (Iteratively Reweighted Least Squares) and the implicit shape representation (IS-Rep: value introduced by the authors). A function of articulation insensitive natural weight is proposed to reduce the influence of the deformation articulated during the standardization process. The limitation of this method is that the orientation of the shape is not clear and is subjected to a large extent shape deformation.

It appears from this study, that there is currently no satisfactory method to both the constraints of normalization's good quality and low complexity. However, the PCA (and CPCA) remains the most adopted approach.

III The Proposed Method

In this paper we suggest a normalization method, based on barycentric coordinates, inspired from [2, 12, 14, 15, 16]. Using these coordinates to achieve robustness to rotations and reflections to remeshing of 3D objects.

3.1 Barycentric Coordinates

Definition: we consider a triangle T = <V_1; V_2; V_3> non degenerate representing one face of a 3D object, V a point of the tetrahedron D, composed from T and the center of gravity of the object. Consider a bi = bi(V) such as:

$$V = b_1V_1 + b_2V_2 + b_3V_3 \quad (1)$$

With b_1, b_2 and b_3 the barycentric coordinates of V relative to T.

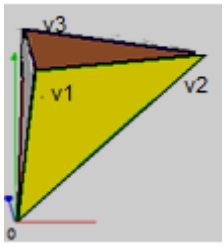


Fig. 1 Tetrahedron built from a surface of a 3D object

Note: The barycentric coordinates for a point V belonging to D with respect to T are unique. Consider a $V=(x, y, z)^T \in R^3$ and $V_i=(x_i, y_i, z_i)^T \in R^3$, with i=1,2,3. Then the system:

$$\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad (2)$$

admits a unique solution (since the triangle T is non-degenerate). Using Cramer's method we obtain:

$$b1 = \frac{\begin{vmatrix} x & x_2 & x_3 \\ y & y_2 & y_3 \\ z & z_2 & z_3 \end{vmatrix}}{\begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}}, b2 = \frac{\begin{vmatrix} x_1 & x & x_3 \\ y_1 & y & y_3 \\ z_1 & z & z_3 \end{vmatrix}}{\begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}}$$

$$et\ b3 = \frac{\begin{vmatrix} x_1 & x_2 & x \\ y_1 & y_2 & y \\ z_1 & z_2 & z \end{vmatrix}}{\begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}} \quad (3)$$

Theorem 1. Let T = <V_1; V_2; V_3> and TR = <RV_1;RV_2;RV_3> two non-degenerate triangles, with a diagonal matrix R. is a point of the tetrahedron formed by the center of gravity of the 3D object and the triangle T. Let b_i, b_{iR}, with i=1, 2, 3 ; the barycentric coordinates of V on T and TR; satisfy the following equation: $b_{iR}(RV)=b_i(V)$. (4)

The previous theorem shows that the barycentric coordinates are invariant under rotation [14].

3.2 Principal of Our Approach

Normalization parameters of the 3D object are calculated by the integration of the tetrahedra containing the three points of the faces in addition to the center of gravity of the object. The geometrical moments of order 1 are coordinates of the center of gravity g of the object. The alignment is done by calculating the eigenvectors of the moment's matrix M (order 2), respecting the logic of the PCA.

$$g = (M_{100}\ M_{010}\ M_{001}) \quad (5)$$

$$M=\begin{bmatrix} M_{200} & M_{110} & M_{101} \\ M_{110} & M_{020} & M_{011} \\ M_{101} & M_{011} & M_{002} \end{bmatrix} \quad (6)$$

To ensure the invariance considerations, we compute the signed distances from the surface of the object for the three planes (YOZ) (ZOX) and (XOY) defined as follows:

$f_x = \iint_S \text{sign}(p_x)|p_x|ds$, , (7)

p =(p_x, p_y, p_z) point of meshed 3D objects.

Ditto for f_y, f_z

Where, p_x, p_y and p_z are respectively the projections of p on the plans (YOZ) (ZOX) and (XOY)

A diagonal matrix defines the reflection matrix:

$$F = \text{diag}(\text{sign}(f_x), \text{sign}(f_y), \text{sign}(f_z)) \quad (8)$$

.

3.3 The Calculation of the Geometrical Moments

D_i is a tetrahedron compound the points g(0, 0, 0), $p_1(x_{i1},y_{i1},z_{i1})$, $p_2(x_{i2},y_{i2},z_{i2})$, $p_3(x_{i3},y_{i3},z_{i3})$, the geometrical moments of order (p+q+r) can be calculated by integrating:

$$m_{pqr} = \int_{Di} x^p y^q z^r dxdydz = \int_{Di} f(x,y,z)dxdydz \quad (9)$$

With $\int_{Di} f(x,y,z)dxdydz = \int_{di} |Ti|f(X,Y,Z)dXdYdZ$

And di =<(0,0,0), (1,0,0), (0,1,0), (0,0,1)> is the orthogonal tetrahedron unit.

Ti is the triangle formed by the points p_1, p_2 et p_3 and X, Y et Z are the barycentric coordinates of V(x, y, z) with respect to Ti=<p_1 ; p_2 ; p_3>.

$$f(X,Y,Z) = (x_{i1}X + y_{i1}Y + z_{i1}Z)^p(y_{i2}X + y_{i2}Y + z_{i2}Z)^q(x_{i3}X + y_{i3}Y + z_{i3}Z)^r \quad (10)$$

The moment of a 3D object can be seen as the sum of the moments tetrahedra compounds and center of gravity of the object and the faces of the meshing of the object [15] [16].

$$m^i_{pqr} = |Ti| \int_{di} f(X,Y,Z)dXdYdZ \quad (11)$$

$$M_{pqr} = \sum_{i=1}^{N_T} m^i_{pqr} \quad (12)$$

With N_T is the number of faces of the mesh and :

$$|Ti| = \begin{vmatrix} x_{i1} & x_{i2} & x_{i3} \\ y_{i1} & y_{i2} & y_{i3} \\ z_{i1} & z_{i2} & z_{i3} \end{vmatrix} \quad (13)$$

therefore

$$\begin{bmatrix} M_{200} & M_{110} & M_{101} \\ M_{110} & M_{020} & M_{011} \\ M_{101} & M_{011} & M_{002} \end{bmatrix} = \sum_{i=1}^{N_T} \begin{bmatrix} m^i_{200} & m^i_{110} & m^i_{101} \\ m^i_{110} & m^i_{020} & m^i_{011} \\ m^i_{101} & m^i_{011} & m^i_{002} \end{bmatrix} \quad (14)$$

.

3.4 Evaluation Measure: Rectilinearity Normalized Objects

Definition 1: A 3D mesh is rectilinear if the angles between each two faces belong to the set {0, π / 2, π, 3π / 2} [19].

Definition: For a 3D object mesh (B) consisting of N triangle {T1, T2, …, TN}, its measurement of rectilinearity:

$$\text{Rect}(B) = \frac{S(B)}{P(B)} \quad (15)$$

and :

- $S(B) = \sum_{i=1}^{N} S(T_i)$ is the sum of surfaces of component triangles of the object..
- $P(B) = P_x + P_y + P_z$
- $P_x = \sum_{i=1}^{N} S(T_{ix})$ is the sum of the surfaces of the projections of the triangles Ti on the plane (ZOY).
- $P_y = \sum_{i=1}^{N} S(T_{iy})$ is the sum of the surfaces of the projections of the triangles Ti on the plane (XOZ).
- $P_z = \sum_{i=1}^{N} S(T_{iz})$ is the sum of the surfaces of the projections of the triangles Ti on the plane (XOY).

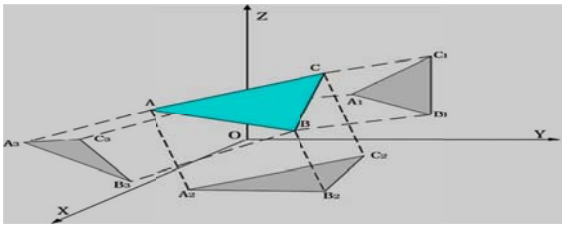


Fig. 2 Projections of a triangle with respect to the three plans [19]

Theorem 2: A 3D mesh (B) is rectilinear if and only if:

$\text{Rect}(B, \alpha, \beta, \gamma) = \frac{S(B)}{P(B,\alpha,\beta,\gamma)} = 1$, for $\alpha, \beta, \gamma \in [0, 2\pi]$ angles of rotation relative to the planes (ZOY), (XOZ), (XOY) respectively.

Proof. See [19] page 135.

Theorem 3: For a 3D object meshed (B):

$$\frac{1}{\sqrt{3}} \leq \text{Rect}(B, \alpha, \beta, \gamma) \leq 1$$

Proof. See [19] page 136.

IV EXPERIMENTAL RESULTS

The databases we have used for our tests are based on SHREC'07 [10] and 3D Segmentation Benchmark [17].

SHREC'07 which was created as part of the contest "3D Shape Retrieval Contests", was used to evaluate research methods 3D. The database contains 400 models SHREC'07 triangular mesh in format "OFF" divided into 20 categories (male, glasses, plane ...).

Benchmark "3D Segmentation Benchmark" was created within the framework of the project "3D Models And Dynamic Representation And segmentation models" [18]. The purpose of this benchmark is to provide an automated tool to evaluate, analyze and compare different algorithms for automatic segmentation of 3D meshes.

4.1 Qualitative Evaluation

The main advantages of the PCA are its simplicity and speed. It can be applied to most of the 3D models.

A first limitation of the PCA is that it is not robust to the deformation of objects. Principal axes generated by the PCA may be different for similar shapes. This limitation is illustrated in Figure 3. This figure shows the results of applying PCA to the object "12.off" [10] before and after deformation.

To highlight the invariance of the proposed method compared to deformations, we present in Figure 4, the results obtained by our method for the same object.

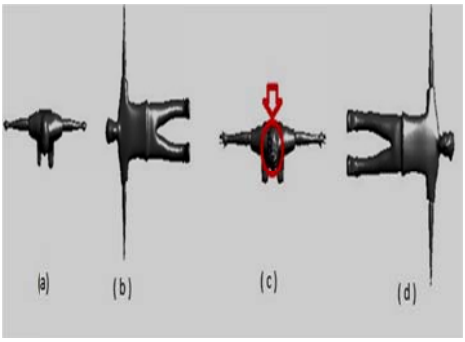


Fig. 3 PCA Normalization (b) for object (a), applying a deformation (c) and normalization after deformation (d)

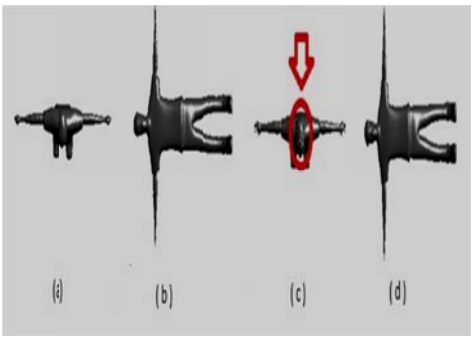


Fig. 4 Barycentric Normalization (b) for object (a), applying a deformation (c) and normalization after deformation (d)

The application of PCA on two clouds of points where the only difference between them is a rotation and / or translation can lead to the same axis directions PCA but not necessarily the same direction [13, 11] (Fig. 5).

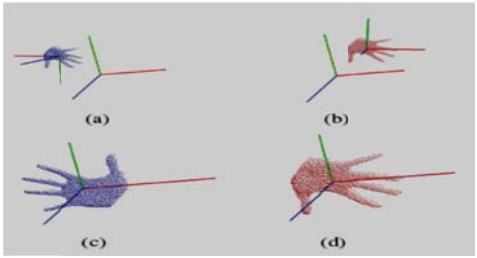


Fig. 5 Problems related to the axis direction of the ACP for the same object that has undergone to rotations [12]

Figures 6 and 7 show a comparison between the results obtained by our method with PCA standard. The object used is "octopus" [17], with and without the application of a 120 ° rotation.

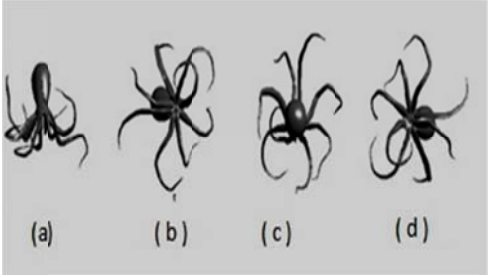


Fig. 6 PCA Normalization (b) for objet (a), applying a 120° rotation (c) and normalization after rotation (d)

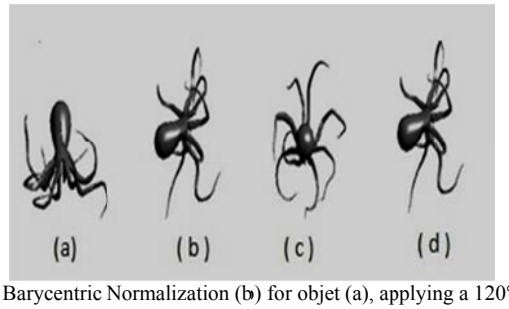


Fig. 7 Barycentric Normalization (b) for objet (a), applying a 120° rotation (c) and normalization after rotation (d)

The obtained results thus show the invariance of the proposed method to rotations. Indeed, whatever the angle of rotation of the initial object is, we obtain, unlike the PCA, the same direction and ditto for axes direction.

4.2 Evaluation of the Rectilinearity for Normalized Objects

To quantitatively evaluate our method, we are based on the criterion of rectilinearity "Rect" presented above. We used for the parameters, α, β and γ, the following values: $\alpha = \pi/3$, $\beta = \pi/3$ et $\gamma = \pi/3$.

Table 1 presents measures of rectilinearity obtained respectively for an initial object, the object after the implementation of the PCA and the object after applying our method. The objects used in this comparison are extracted from the database [17].

TABLE 1
MEASURING RECTILINEARITY RECT

Object	Initial RECT	After ACP Normalization	After Barycentric Normalization
alie	0,674	0,7047	0,7063
armadillo	0,6701	0,6687	0,6705
boy	0,6873	0,7094	0,7094
bunny	0,6698	0,6631	0,6802
homer	0,6678	0,6952	0,6954
robot	0,6818	0,7294	0,7333
vaselion	0,6655	0,6913	0,6925

The results obtained by our method, for objects: alien, armadillo, boy, homer, and robot vaselion are almost similar with respect to PCA. the rectilinearity measurements obtained for the object "bunny" are relatively distinct. For this purpose, our method guarantees more rectilinearity.

V CONCLUSION

Our method can be seen as a hybrid method that is based on the study made by the the integration of the mesh volume [15] and adopted by [16, 12], barycentric coordinates [14] and the CPCA [2].

Extractiing (Extracting) the normaliization's parameters on the surface of the object is independent of the choice of the resolution discretization and provides a normalization independent of the meshing and small

deformations. Applying the factor of reflection provides normalized objects independent of any parameter. The limitation of this method is the computational cost, for this reason we plan to improve it in order to overcome this limitation.

REFERENCES

[1] E. Paquet, M. Rioux, A. Murching, T. Xaveen, and A. Tabatabai, Description of shape information for 2-D and 3-D objects. In Signal Processing: Image. Communication, vol. 16, pp. 103-122, 2000.

[2] D. Vranic and D. Saupe. 3D shape descriptor based on 3d fourier transform. In Proceedings of the EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services (ECMCS'01), pp. 271-274, Budapest, Hungary. September 2001.

[3] T.J and wee khengheour. Normalization and alignement of 3D objects Based on Bilateral symmetry plans. MMM 2007, LNCS 4351 Part1 PP. 74-85. 2007.

[4] H.Fu, D. Cohen - or, G. Dror and A. Sheffer. Upright orientation of man-made objects. ACM Transactions on Graphics, 27(3), 2008.

[5] M. Chaouch and A. Verroust Blondet. A Novel method for alignement of 3D models. IEEE International conference on shape Modeling and applications (SMI'08) PAGES 187-195, Stony Brook, NY, USA, June 2008.

[6] Minovic, S. Ishikawa and K. Kalto. Symmetry identification of 3D object represented by octree. IEEE Transaction on Pattern analysis and machine Intelligente (TPAMI), 15(5) 507-514, 1993. 53.

[7] Y-S. Liu and K. Ramani Robust Principal axes determination for point based shapes using least median of squares Computer. Aided design, 41(4): 293-305, 2009.

[8] Fleishman S, Cohen - or, Silva C. T. Robust Moving Least Squares fitting with sharp features. In Processing of SIGGRAPH'05. 2005.p.545-52.

[9] Chao wang, Yu. Shen Liu, Min Liu, Jun-Hai Yong, Jean-Claude Paule. Robust shape normalization of 3D articulated Volumetric models. The research was Supported by the national Science Formation of China (61003095) Chinese 973 program (2010CB328001).

[10] Network of Excellence Project AIM@ShAPE, IMATI-Ge CNR Shape Modeling Group SHREC'07 Watertight Database. http:// watertight.imati.anr.it/26,209

[11] M. Chaouch. Recherche par le contenu d'objets 3D. Pastel-00005168, version1 - 16 Novembre 2010.

[12] Julien Ricard, David Cœurjolly, Atilla Baskurt. Expression analytique d'un maillage 3D.RR-LIRIS-2006-012.

[13] Ouissem Ben-Henia Saida Bouakaz. Utilisation de l'ACP pour la reconnaissance des gestes 3D de la main. inria-00595488, version 1 - 24 May 2011.

[14] O. Davydov, L. L. Schumaker, Scattered Data Fitting on Surfaces Using Projected Powell-Sabin Splines, Mathematics of Surfaces 2007, R. Martin, M. Sabin, and J. Winkler (eds.), Springer, 2007, 138-153.

[15] Lien, S.L.C. Combining Computation with Geometry. PhD thesis, Technical Report. California Institute of Technology, Pasadena, Californie (1984).

[16] Zhang, D., Chen, T. Efficient feature extraction for 2d/3d objects in mesh representation. In : IEEE International Conference on Image Processing (ICIP '01). Volume 1, Thessaloniki, Gréce (2001) 935–938.

[17] H. Benhabiles, J-P. Vandeborre, G. Lavoué and M. Daoudi. A comparative study of existing metrics for 3D-mesh segmentation evaluation. In the Visual Computer International Journal of Computer Graphics, Springer Editions, vol. 26(12), pp. 1451-1466, December 2010.

[18] 3D Models And Dynamic models Representation And Segmentation. http://www-rech.telecom-lille1.eu/madras/

[19] Zhouhui Lian · Paul L. Rosin · Xianfang Sun. Rectilinearity of 3D Meshes. Int J Comput Vis (2010) 89: 130–151.

Optimization of the Organized KOHONEN Map by a New Model of Preprocessing Phase and Application in Clustering

Es-safi Abdelatif
Scientific computing and Computer Sciences, Engineering Sciences
Modeling and Scientific Computing Laboratory
University of SIDI MOHAMED IBN ABDELAH
Faculty of sciences and techniques
FES, MOROCCO
essafiabdelatifi@hotmail.com

Harchli Fidae
Scientific Computing and Computer Sciences, Engineering Sciences
Modeling and Scientific Computing Laboratory
University of SIDI MOHAMED IBN ABDELAH
Faculty of sciences and techniques
FES, MOROCCO
Fidae.harchli@gmail.com

Ettaouil Mohamed
Scientific Computing and Computer Sciences, Engineering Sciences
Modeling and Scientific Computing Laboratory
University of SIDI MOHAMED IBN ABDELAH
Faculty of sciences and techniques
FES, MOROCCO
mohamedettaouil @yahoo.com

***Abstract*—The needless of clustering is continually growing. This fact is due to the huge amount of information daily stored in the sites web. Those informations must be classified in order to facilitate their treatment. There are several methods to classify a set of data; those based on matching learning are more efficient. Among those systems, Kohonen algorithm is a useful system because of its famous proprieties. Some of them will be presented in this paper. Unfortunately, as other clustering algorithm, it suffers from the following problems: dependency of the result on the initialization phase which is performed randomly and the number of classes is unknown in the beginning. The overcome of these problems represents a great challenge in the clustering domain. In the present paper we expose an approach which allows performing suitably the initialization phase. This approach consists of conducting a pre-processing phase. In this latter we use a parameter r, we obtain an idea on distribution of examples. Then the initial weight vectors are chosen from the area which has a high density. This allows us to avoid an initialization with the isolated examples which decrease the performance of the system. Also we can determine approximately the number of classes. After measuring the quality of clustering obtained by Kohonen algorithm, we update the parameter r and we repeat the same process. This latter is arrested when we obtain a suitable quality of clustering. To show the performance of this approach, some experiments are conducted.**

***General Terms*—Clustering, Standard Kohonen algorithm, Machine learning; Evaluation**

***Index Terms*—Preprocessing phase, initial code-vectors, number of clusters, optimization**

I. INTRODUCTION

The clustering is an important tool in data mining. It consists of grouping a set of examples in some cluster. This plays a great importance in retrieval research (RI). It allows reducing the research space focusing on specific clusters. So the consuming times to access to the desired information is also reduced. This is more important than the case while size of the data set is great. Because of the enormous quantity of information available in tools communication, this is the case in almost actual situation. Adopting the matching learning techniques the performance of classifier are improved. According to the type of the learning used by the classifier, this latter can be supervised or unsupervised. In the first case, the class of the elements of the training set is determined. While in unsupervised case any information about the class is known. Among those algorithms the neural network Kohonen is an efficient algorithm: It has an important proprieties, among them we present the following:

- It is considered as a tool of reduction of dimensionality space, because each data will be associated to a dot in space with a low dimensionality. So it can be used as a technique of nonlinear projection and visualization of inputs as the PCA (Principal component analysis).
- It resist to noise.
- It is rapid in clustering of the new input.

As the other clustering system, the Kohonen algorithm suffers of the following drawbacks:

- Difficulty of determination of the suitable measure distance: In the numerical case (i.e. when the attributes tack numerical values), the standard measure is the Euclidian Manhattan and maximum distance measure. But in the case of categorical attributes the problem is more difficult.

doi:10.4304/jetwi.6.1.80-84

- Structure of database: Real life data may not always contains clearly identifiable cluster and contain several outliers.
- The determination of the number of clusters and the choice of the initial code-vectors are unknown: Opposite of the supervised case, the number of class in clustering system is unknown. Therefore the initial code-vectors are chosen randomly.

Outlier detection: The objects which are enormously dissimilar to other data are called outliers inputs. It is difficult to avoid such elements.

For more detail about the drawbacks of clustering system we refer the reader to [6].

In our work we focus on issue 2, 3 and 4 because identifying the number of class is a difficult task, a correct identification yield to good result. Otherwise a catastrophe can be expected. So it attracts a great interest in scientific research. Also the choice of initial clusters has a great effect in the result of a clustering system. Those systems choose the k initial cluster randomly. So it is likely that these classes are not suitable and can contain several outliers. This can produce empty classes or classes with few elements.

In this paper aiming to estimate a number of classes and suitable initial clusters, we conduct a study of data set. The precision increases during the process. The remaining of this paper is organized as follows: In section 2 we define the classical Kohonen algorithm. Our proposed method is introduced in section 3. During each iteration of the proposed algorithm, we evaluate the quality of the obtained clustering, so we reserve the section 4 to present the formula of evaluation. In section 5 we present the result obtained in experimental phase accompanied with discussion. In the last section we give a conclusion and some perspectives.

II KOHONEN ALGORITHM

In the literature, Kohonen's SOM is a well-known and widely used algorithm in clustering techniques.

The learning process of Self organizing map is based on a competitive and unsupervised artificial neural network. It is a clustering algorithm that is used to map high-dimensional data into a low-dimensional representation space.

The underlying idea of this algorithm is to project a set of data called input space on an output space which has a low dimension (generally 3, 2 or 1). This is performed by establishing a correspondence between the two spaces. This latter must preserve the topology of data set i.e. if x is close to y; f(x) is also close to f(y). This idea is invented by scientist Tuevo Kohonen in 1975. The goal of this neural network is to model the brain activity. In order to realize this task he associates to each vector x another vector g(x) called the weight of the neuron f(x) and noted w (f(x) is noted by n, the set of neurons forms a gird in the output space). Each neuron has a neighborhood. The weight vector is initialized randomly and updated using a competitive and unsupervised learning. This adaptation process can be described as: For each presentation of an input x, the index i of the neuron nearest from x is determined, using the following formula:

$$d(x, w_i) = \min\left(d(x, w_j)\right), 1 \le j \prec N$$

Then the weight of the neuron ni and those of its neighborhood are updating as follows:

$$w_j(t) = w_j(t-1) + \eta(t)\alpha_{i,g}(t)\left(x - w_j(t-1)\right)$$

$w(i)$ represents a neuron in the neighborhood of the neuron n_i, η and α represents respectively the learning factor and neighborhood function which must decrease during the process. Indeed, the learning is performed on two phases; the first is called organization phase. This latter, because of a high rate learning and extended neighborhoods, allows deploying map were data are concentrated. After the learning rate, the size of the neighborhoods begins to decrease progressively, in order to conduct a folding of the map. In This phase the weight converge to some vectors which are approximation of the inputs. In the literature there exist different definitions of the parameters $\eta(t)$ and $\alpha_{i,g}(t)$. Among this latter we find the following:

$$\eta(t) = \begin{cases} \eta_0 - \left(\frac{\eta_0 - \eta_r}{r}\right)t & if\ t \prec \tau \\ \eta_r & otherwise \end{cases}$$

$\alpha_{i,g}(t) = \frac{1}{t}$ with t is the number of current iteration

To update the neighborhood of neurons some authors use the following formula:

$$v_{i,g}(t) = \exp\left(-\frac{d_{i,g}^2}{2\sigma^2(t)}\right)$$

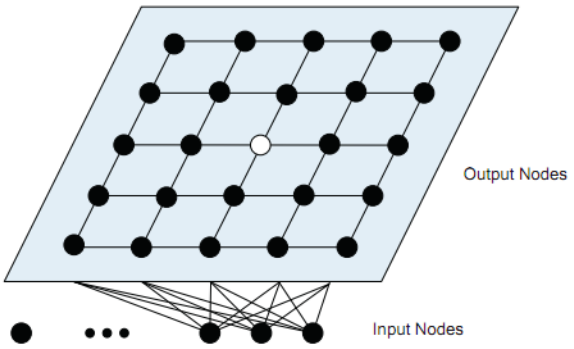


Fig 1: A self-organizing map with 25 neurons

III PROPOSED METHOD

In this section we present with more detail our approach. As is presented in previous algorithm the initialization of weight vectors is performed randomly. The number of initial cluster is also taken randomly. To overcome some of these difficulties, the searcher used different strategies. Some existing works hybrid supervised clustering and unsupervised one [2]. Some Latter works transform the problem of optimization of the

architecture of Kohonen map to a problem of linear programmatic to improve the efficiency of Kohonen algorithm [4]. Other works are post clustering i.e. they choose the set of codes which is the partition and which provides the best partition from a number of alternative ones [1].

As is reported above a bad initialization leads to bad result. In this approach in order to initialize suitably weight vectors, we study the structure of input set. So we locate the areas where the density is great. Thereby the initial vector weights are taken from those. Hence we can remove the outliers and estimate the number of class. In this approach we propose two phases: the pretreatment phase and the clustering phase.

Practically, we begin by calculating the distance between all inputs which will be stored in a matrix $M=\left(a_{ij}\right)_{i,j}$ where $a_{ij}=d\left(x_i,x_j\right)$ and we search the maximum and the minimum in this matrix. Then we fix a real value $\varepsilon_1=\frac{\max M-\min M}{n_{max}}$. After, according to the parameter k data set E is divided on many subsets. Each one of these latter is formed as follow:

$$E_x=\left\{x_j / d\left(x,x_j\right)\prec\varepsilon_1\right\}$$

where x is an element chosen randomly from E. if the cardinality of E_x is equal to m (number of elements in data set) the first iteration in the pretreatment phase is achieved. Otherwise, another subset E_y is constructed by choosing randomly an element $y\in E\setminus E_x$ where $E_y=\left\{x_j / d\left(y,x_j\right)\prec\varepsilon_1\right\}$.

The same process will be applied to $E\setminus E_x\setminus E_y$ if $\left|E_x\cup E_y\right|\neq m$. It is certain that after a number of steps the process will stop and a number n of subsets is generated $E_1,E_2,...,E_n$. We have conducted this phase in order to search the number of neurons in the map of Kohonen and the initial vector weights. So in the clustering phase, we turned the basic Kohonen algorithm with n neurons whose weights are the means of $E_i\ \forall i\in\left\{1,...,n\right\}$. According to a measure function, the quality of the clustering is estimated in order to know the performance of the algorithm. In this stage we have generated the first iteration of our algorithm. The next iterations are performed by changing the value of the parameter ε_1 by $\varepsilon_k=\frac{\max M-\varepsilon_{k-1}}{n_{max}}$ or $\varepsilon_k=\frac{\varepsilon_{k-1}-\min M}{n_{max}}$ where n_{max} is a fixed number.

This approach is summarized in figure 1.

IV Assessment of Clustering

The evaluation of a system clustering represents a major challenge for the searchers. The quality of classifier is measured by calculating the homogeneity of classes that are produced and the separation between them. There exist several indexes which can be used to measure the quality of clustering. For more information we refer the reader to [3] [5].

In our work, we have chosen the average silhouette which is defined by the following formula:

$$s(i)=\frac{b(i)-a(i)}{max\{a(i),b(i)\}}$$

This formula can be written as:

$$s(i)=\begin{cases}1-\frac{a(i)}{b(i)}, if\ a(i)<b(i)\\ 0, \quad if\ a(i)=b(i)\\ \frac{b(i)}{a(i)}-1, \quad if\ a(i)>b(i)\end{cases}$$

Where: $a(i)$ is the average dissimilarity of xi with all other data within the same cluster and $b(i)$ is the lowest average dissimilarity to xi of any such cluster.

We report that: $a(i)$. More that $s(i)$ is close to 1 more than xi is correctly classified. In contrast, if $s(i)$ is close to -1 xi is appropriately clustered. Finally, if this value is about zero, xi is on the border of two natural clusters.

The definition of Silhouette index doesn't care into count the number of classes (n). Since in this work we attempt to look for a suitable number of classes, we developed a new index Ni.

$$N_i=\frac{(\text{Silhouette})^2}{\sqrt{\left(\frac{n}{150}\right)}+\text{Silhouette}}$$

This proposed index establishes a trade-off clustering quality and the number of classes.

V Experiments and Commentaries

In order to show the performance of our proposed method, two experiments are performed. In this context we used the data set IRIS. This latter is widely used is the quality clustering area. The inputs of this data set are ranked in three groups and they are characterized by four components.

In order to find a suitable number of classes we run several tests according to the parameter ε_k or a fixed number of iteration. In this experiment the best number of classes (s) corresponds to the best clustering quality according to the proposed criterion.

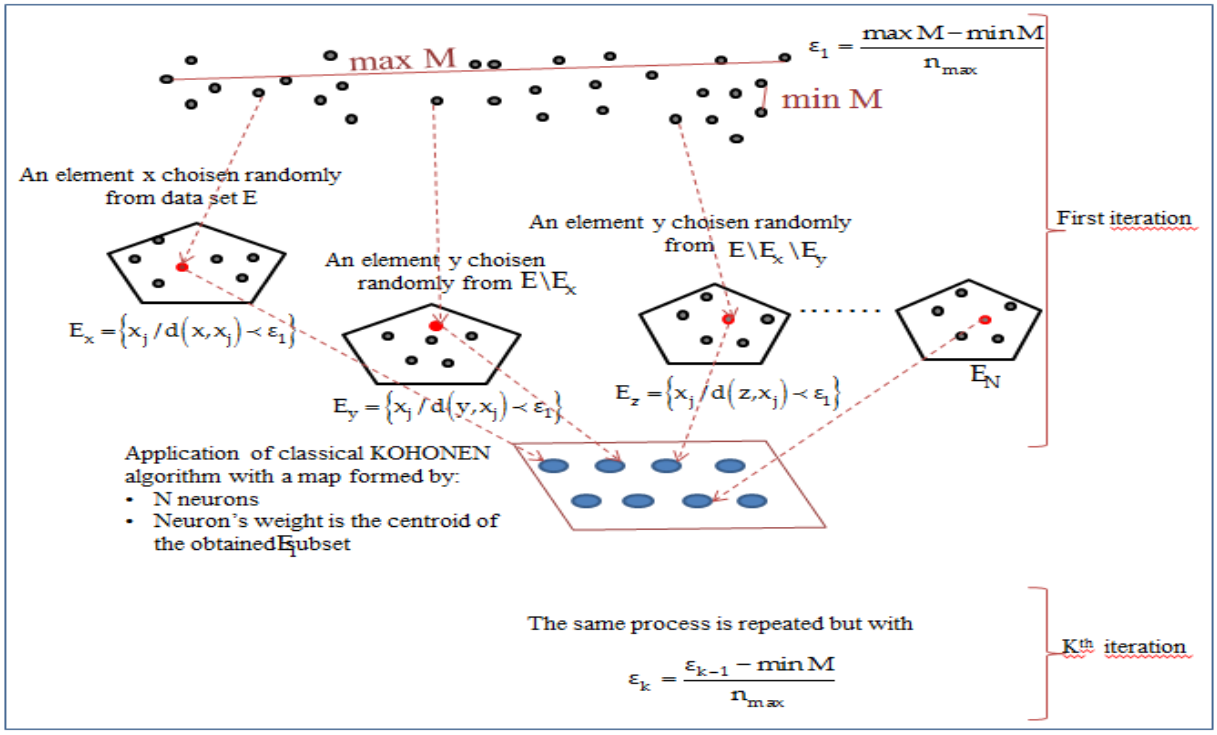


Fig 1: Summarization of the steps of the proposed algorithm

The following table reports the number of classes find by the system and clustering performance for different values of threshold ε_k.

TABLE 1.
OBTAINED NUMBER OF CLUSTERS IN THE PROCESS OF OPTIMIZATION

$\varepsilon_k = \frac{\max M - \varepsilon_{k-1}}{n_{max}}$		$\varepsilon_k = \frac{\varepsilon_{k-1} - \min M}{n_{max}}$	
Number of classes	**Proposed index**	**Number of classes**	**Proposed index**
2	0,25	2	0,25
3	0,28	3	0,77
3	0,77	5	0,77
2	0,25	15	0,69
3	0,28	14	0,7
3	0,77	16	0,69

In Table 1 we can see that whenever ε_k is close to maxM, the quality of the outputs become good but whenever ε_k is close to minM, inputs become dispersed and the system provides a large number of clusters. This experiment shows that with this method we achieved 0,77 of clustering performance with three clusters which is the desired number of classes. This satisfactory result proved the relevance of our approach.

To further investigate the relevance of the method we compare our results with those of classical Kohonen. In this latter we are forced to give a number of neurons and there weights. We built the map of classical Kohonen with s neurons (where s=3 is the result of the first experiment) and we affect a random initialization to these neurons. Results of this comparison are stored in Table II.

The proposed method is more powerful than conventional Kohonen algorithm and this is evident in Table II where we can see that the quality of result given by Kohonen classic is bad compared to what we found in the first experience. We can say that our approach gives a better result which is incomparable with the classical method.

TABLE 2.
COMPARISON BETWEEN THE RESULTS OF CLASSICAL KOHONEN AND THAT OF THE PROPOSED METHOD

Random initialization	**Silhouette index**	**Proposed index**
Choice 1	0,857	0,735
Choice 2	0,856	0, 734
Choice 3	0,853	0, 732
Choice 4	0,905	0, 782
Choice 5	0,906	0,784

Finally, those experiments show that the proposed method is able to identify suitable code vectors and give a number of classes equal to the real one. Therefore it gives a satisfactory clustering.

VI CONCLUSIVE REMARKS AND FUTURE WORKS

In this work a clustering method based on a clustering algorithm belonging to the Kohonen Self Organizing map family has been proposed. Despite its excellent properties whose some ones are represented above in this paper, the Kohonen algorithm is widely used in clustering area. Unfortunately, it suffers from some drawbacks whose

main stems from its dependency on initialization phase. In this latter a number of cluster and code vectors are chosen randomly. So an inappropriate choice leads to a bad result. In this work, we proposed a method which aims to overcome this problem by choosing automatically suitable initial code vectors.

The analysis was performed on dataset IRIS. The experiments carried show that this method gives good result and satisfactory rate of clustering.

This is an encouraging result to try to integrate other parameters and criteria attempting to more improve the performance of clustering of this method, extend the analysis to datasets with more samples, implement this method on other databases, and apply this approach in other domain as text clustering and opinion meaning.

REFERENCES

[1] G. W. Milligan, and M. C. Cooper, "An Examination Of Procedures For Determining The Number Of Clusters in a Data Set," Psychometrica, vol. 50, No. 2, pp. 159-179, June 1985.

[2] K. Nigam, A. K. Mccallum, S. Thrun and T. Mitchell, "Text Classification from Labeled and Unlabeled Documents Using EM", Machine Learning, vol. 39, pp.103-134, 2000.

[3] M. Ettaouil, A. Essafi, and F. Harchli, "Optimizing the architecture of Kohonen map and classification," JOURNAL OF COMPUTING, vol. 4,pp. 31-38, July 2012.

[4] M. Ettaouil, Y. Gannou, K. Elmoutaouakil and M. Lazaar, "A new Architecture Optimization Model for the Network and Clustering," Journal of Advanced Research in Computer Science, vol 3, no. 1, pp. 14-32, 2011.

[5] M. Qiu, S. Davis, and F. Ikem, "Evaluation of clustering Techniques in Data Mining Tools", Issues Information systems, vol. 5, No. 1, 2004.

[6] P. Agarwal, M. Afshar Alam, and R. Biswas, " Issues, Challenges and Tools of Clustering Algorithms," IJCSI International journal of Computer Sciences Issues, vol. 8, Issue 3, No. 2, May 2011.

Adaptive Learning

Soumaya El Mamoune
Phd Student, Information
Technologies System Laboratory
University Abdelmalek Essaadi
Soumaya.mgi@gmail.com

Loubna Cherrat
Phd Student, Information
Technologies System Laboratory
University Abdelmalek Essaadi
cherrat81@yahoo.fr

Omar Kassara
Phd Student, Information
Technologies System Laboratory
University Abdelmalek Essaadi
okassara@yahoo.fr

Marouane El Mabrouk
Phd Student, Information
Technologies System Laboratory
University Abdelmalek Essaadi
elmabrouk@gmail.com

Ahmed Hadioui
Phd Student, Information
Technologies System Laboratory
University Abdelmalek Essaadi
Hadioui@yahoo.fr

Mostafa Ezziyani
Professor, Faculty of Sciences
and Technologies of Tangier
University Abdelmalek Essaadi
ezziyyani@gmail.com

***Abstract*—The automatic programming is an essential phase in the adaptation of information systems because it allows the system to advance from its observations on the environment. The system is formed during the operation by creating gradually its own ontology of the domain; therefore it is necessary that the knowledge stored evolve in an automatic way and during operation without asking a large participation of the human user. In this sense, this article aims to analyze the different learning techniques and process the information relating to the context to which the objective of choosing the proper technique to the situation in order to facilitate the storage, access and exploitation of such information.**

***Index Terms*—Adaptive learning, automatic programming, scalable system, experimentation, adaptation.**

I. Introduction

The natural disasters shall designate the harmful effects of a sudden phenomenon, sustainable or intense, natural or human-caused. It distinguishes any particular situations including human losses and destruction on a large scale. In such situations, a reliable information and a good intervention can decrease the rate of damage it is for this reason we found several research for the development of systems of rescue and aid to decision making for the management of disasters but the particular problem in the field "decision Support" is that there is of artificial actors who will interact with a dynamic environment. In this environment, decision-making must be based on current state to adapt to the change from the outside. In this case, the system must be equipped with a large capacity of the automatic learning which refers to the analysis and to the implementation of the methods that allow a machine to evolve and fill the tasks difficult or impossible to fill by algorithmic means more conventional.

II Learning

2.1 What is?

In a general sense, learn corresponds to the adaptation of individuals to their environment. Learn, is to change in the course of the life to adapt to the environment. Childhood is the principal period of this adaptation but the individuals adapt to changes in their environment throughout their lives. For humans, the environment is essentially physical, living, social (or cultural), emotional and technological. What is prepared by the individual during this adaptation within either of knowledge, either of the memories.

Learning is a criterion for human adaptation to a changeable environment but this criterion remains difficult or impossible to achieve for the conventional systems of regular or just the need of automatic learning or learning artificial to give systems the ability to improve its performance through interactions with the environment.

2.2 Automatic Learning

The automatic programming (machine Learning in English), which is one of the sub-domains of artificial intelligence, has for objective to extract and automatically exploit the information present in a set of data. This is the scientific discipline concerned for the development, analysis and implementation of automation methods which allow you to a machine (in the broad sense) to evolve through a process of learning, and thus to fulfill the tasks that it is difficult or impossible to fill by algorithmic means more conventional.

III Adaptatif System

Our goal is create a system for the management of natural disasters which is a difficult mission view the amount of data non-structured and distributed that

doi:10.4304/jetwi.6.1.85-88

receives the system and the contents of the data provides a rich source of information that needs to be maintained, filter and organized to allow an efficient use of the objective to offer a solution or an appropriate action.

The establishment of such a system means the plunge in a dynamic environment where there are many dangerous situations that require a response adequate to perform it is for this reason our studies seek mechanisms to identify these situations and he proposed the actions to be taken to get by with less damage and all of this will be to base of learning for power to the system to adapt to the course of its exchanges with the environment in order to aim at any time to the appropriate function. The general operation of the system follows the following steps: (Fig1)

As it is stated before the operation of the system follows three essential steps that allow the system to be scalable. These steps are to adapt according to the situation encountered in passing of the virtual state to the state concrete which is made following an actual event gives: it creates an instance of the system by the definition of the disaster, its category according to comments (usually data) and the services offered. This instance adapted must be saved for it to be reusable in other similar cases subject to several learning techniques and do the update for self-improvement.

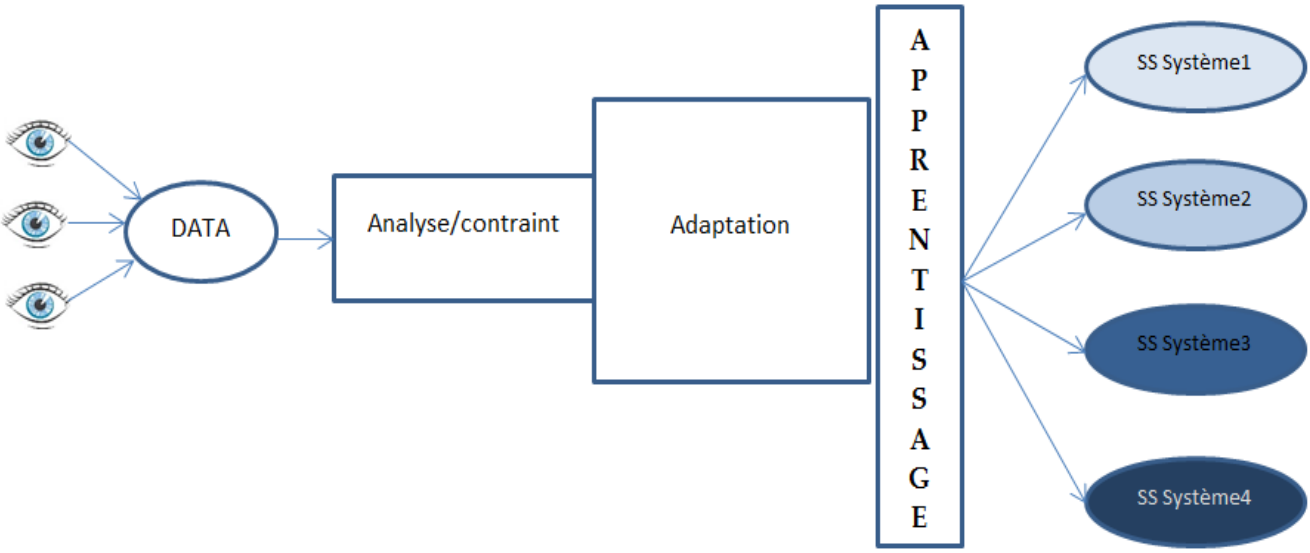


Fig 1: The steps of the operation of the system

IV Comparative Study of Techniques for Learning

The problem of decision in the case of relief is a complex problem and in this case, the programming of the system (to learn how to adapt with a situation face) must meet several requirements: interactive, narratives, online, manipulate not only the quantitative data but still qualitative parameters. Even, we must partition our decision problem into several sub problems and apply other learning algorithms to each, based on its particular characteristics.

There are several algorithms for programming different: network of neuron, decision tree, Bayesian network ... Each, with its model of decision, is granted with the types of problem and the different types of application . Our system intervenes in the different situations and to ensure an effective use, we propose to choose and adapt the learning algorithm according to the situation encountered at base of its constraints and the degree of its compatibility with the different techniques in order to make the right choice. To do this, we will do a comparative study following which allows us to deduct and choose the suitable technique.

Below, we examine the adaptability of a few algorithms of learning common to the requirements of our problem:

4.1 Decision Tree

Set of rules to classify based on tests associated with attributes, organized as a tree. A tool to help in the decision which represents the situation more or less complex to which we must face in the graphic form of a shaft in such a manner as to show at the end of each branch the different possible results depending on the decisions taken at each step.

4.1.1 Advantage:

The decision tree is a technique very popular and usable views its simplicity and its benefits:

- The results of the shaft is explicit and easily interpretable
- It is a model easily programmable to assign new individuals
- CART" Classification And Regression Trees " allows the use of variables of all types (Continuous, discrete, categorical)

- Treatment of a large number of explanatory variables for subdividing the individuals in heterogeneous groups (classes)
- The decision rules are simple
- Little disturbance of individuals extremes because the results are isolated in small sheets.

4.1.2 Disadvantage

Even if this technique has many advantages but it also has disadvantages which are:

- Use of variables not simultaneous but sequential.
- "Butterfly Effect" ➔ change a variable in the shaft, while the shaft changes.
- Use of heuristic rules : a method that provides rapidly (in temps polynomial) a workable solution, not necessarily optimal, for a problem ofoptimisation NP-difficile.
- Time extensive calculations (the search for criteria of division and pruning).

4.2 Neural Network

Network of neuronesest is a calculation model whose design is very schematically inspired of the functioning of biological neurons. The neural networks are usually optimized by methods of learning of probabilistic type, in particular Bayesian filtering. They are placed in a share in the family of statistical applications. The neural network works with real numbers, the response reflects a probability of certainty.

4.2.1 Advantage

- Ability to represent any function, linear or not, simple or complex.
- The learning from examples representative, by "retro propagation of Errors". The learning (or construction of the model) is automatic
- Resistance to noise or the lack of reliability of the data.
- Simple to handle, much less work staff to provide that in the statistical analysis classic. > Behavior is less bad in case of low amount of data.
- For the novice user, the idea of learning is more simple to understand that the complexities of statistics multi variables.

4.2.2 Disadvantage

- The neural network does not always provide rule of exploitable by a human.
- It does not fit the requirement "explanatory memorandum" because it does not provide justification easy to interpret.
- The absence of systematic method for defining the best network topology and the number of neurons to place in the (or the layer(s) hidden(s).
- The choice of the initial values of the weights of the network and the adjustment of the not of learning which play an important role in the speed of convergence.

4.3 Bayesian Network

In computer science and statistics, the Bayesian network is a graphical model probabilistic representative of random variables in the form of a directed graph acyclic. Intuitively, they are at once: models of knowledge representation and of machines to calculate the conditional probabilities. For a given domain, it describes the causal relationships between variables of interest by a graph.

4.3.1 Advantage

The Bayesian Networks are often used, since they have not bad advantages compared to other techniques.

- The Bayesian Networks can represent intuitively an area of knowledge, a lot of experiences show that it is often easier to formalize the knowledge in the form of a causal graph that in the form of a system based on the rules.
- The Bayesian Networks can manage the set of incomplete data. And more the Bayesian networks allow you to learn the causal relationship which can help us to make decisions.

4.3.2 Disadvantage:

The observation of a cause or of several causes does not lead systematically the effect or the effects that depend on them, but only modifies the probability of the observed so this technique is not interactive. This technique may not respond well to changes in the environment and in this case, the relief cannot be guaranteed on time.

V PROPOSED SOLUTION

From the point of view of the situations being addressed and view the advantages and disadvantages of each technique, we present the above tables or we have grouped our needs the acquisition, representation and use of knowledge. The representation adopted for this comparison is the following:

"Each line corresponds to a characteristic or need, which can be an advantage, or the taking into account of a given problem.

"If the technique considered allows to take into account this problem, or has the advantage, a + sign is placed in the corresponding box.

"A signed ⋆ is place in the box of the best technique from the point of view of the characteristic considered.

TABLE1.
ACQUISITION OF GIVEN

	Neural Network	Decision Tree	Bayesian Network
Data only			
Joint	*	+	+
Incremental	+		*
Incomplete Data	+		*

TABLE2.
REPRESENTATION

	Neural Network	Decision Tree	Bayesian Network
Uncertainty			*
Readable		+	*
Easy	+	*	
Homogeneity			*

TABLE3.
USE

	Neural Network	Decision Tree	Bayesian Network
Queries developed			*
Economic Utility	+		*
Performance	*		

The disasters are different situations to ensure a perfect functioning we propose to adapt the technique of learning according to the criteria presented previously. The adaptation will be through the automatic choice of the best technique which is represented by the star " * " and in the case where there are multiple criterion we will do a combination between the degree of satisfaction (+ , *).

CONCLUSIONS AND PROSPECTS

In general, the problem of the aid to decision is a complex problem, because of its characters. The system operates in the environment which changed a lot, it must confront with the altered situations or urgent, there are a lot of parameters and criteria to meet. The system of "simulation of relief" is able to learn the way to make the decision of the user to support the user in other situations. Usually, it must be split into several sub-problems and apply a particular algorithm for each.

The algorithm must satisfy several requirements: interactive, explanatory memorandum, online, "manipulate not only the quantitative data but still qualitative parameters". After the analysis of a few techniques we found that the use of adaptive learning is a proposal capable for having the best solution. But of course, it is good only with a part of the problem of general relief: that is the type of sub-problem as the ambulance rescueth them the victim or the fireman get rid fire ...

It has already used the character of the understanding to make the learning for a system which can approach to the capacity of human learning.

The search is still maintained in order to make a complete algorithm to give the possibility to a system to react as the human brain and putting as perspective the study of a method of classification of any type of a natural disaster and also the dangerous situations and criticism.

REFERENCES

[1] Avgoustinos Filippoupolitis, X. Q. (2007, March 14). An adaptive system for movement decision support in building evacuation. , 80(17), 3673–3677.
[2] GokceGorbil, AvgoustinosFilippoupolitis, &ErolGelenbe (2007, November15). Intelligent Navigation Systems for Building Evacuation. , 8(10), 3487–3492.
[3] AvgoustinosFilippoupolitis, X. Q. (2007, March 14). An adaptive systemfor movement decision support in building evacuation. , 80(17), 3673–3677.
[4] GokceGorbil, AvgoustinosFilippoupolitis, &ErolGelenbe (2007, November15). Intelligent Navigation Systems for Building Evacuation. , 8(10), 3487–3492.
[5] AvgoustinosFilippoupolitis, X. Q. (2007, March 14). An adaptive systemfor movement decision support in building evacuation. , 80(17), 3673–3677.
[6] GokceGorbil, AvgoustinosFilippoupolitis, &ErolGelenbe (2007, November15). Intelligent Navigation Systems for Building Evacuation. , 8(10), 3487–3492.
[7] Bang-Jensen, J.; Gutin, G. (2000). Digraphs: Theory, Algorithms and Applications. Springer.
[8] D. Maquin, 'Elements of graph theory and linear programmation', Éléments de théorie des graphes et programmation linéaire, Institut Nationale Polytechnique de Lorraine, École Nationale Supérieure d'Électricité et de Mécanique, 2008.
[9] P.E.Hart, A Formal Basis for the Heuristic Determination of Minimum Cost Paths, IEEE Transactions on Systems Science and Cybernetics SSC4, vol. 4, no 2, p. 100–107, 1968.
[10] Balakrishnan, V. K. (1997-02-01). Graph Theory (1st ed.). McGraw-Hill. ISBN 0-07-005489-4.
[11] Berge, Claude (1958) (in French).'Graph theory and its applications', Théorie des graphes et ses applications. Dunod, Paris: Collection Universitaire de Mathématiques, II. pp. viii+277. Translation: Dover, New York: Wiley. 2001 [1962].
[12] Biggs, Norman (1993). Algebraic Graph Theory (2nd ed.). Cambridge University Press.ISBN 0-521-45897-8.
[13] Bollobás, Béla (2002-08-12). Modern Graph Theory (1st ed.). Springer.ISBN 0-387-98488-7.
[14] Diestel, Reinhard (2005). Graph Theory (3rd ed.). Berlin, New York: Springer-Verlag. ISBN 978-3-540-26183-4.
[15] Peters, L. D, Saidin, H, 2000, IT and the Mass Customization of Services: the Challenge of Implementation, International Journal of Information Management, 20(2), 103-119

A Comparative Study between the Pseudo Zernike and Krawtchouk Invariants Moments for Printed Arabic Characters Recognition

Rachid Salouan
Department of Mathematic and Informatic, Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, MOROCCO
rachidsalouan@gmail.com

Said Safi
Department of Mathematic and Informatic, Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, MOROCCO
safi.said@gmail.com

Belaid Bouikhalene
Department of Mathematic and Informatic, Polydisciplinary Faculty,Sultan Moulay Slimane University,Beni Mellal, MOROCCO
b.bouikhalene@usms.ac.ma

***Abstract*—In this paper, we are focused on characters recognition, for this we present a comparison between the Krawtchouk Invariant Moment (KIM) and the Pseudo Zernike Invariant Moment (PZIM) for the recognition of printed Arabic characters (translated, rotated and contaminated by noise). In the preprocessing phase, we use the thresholding technique, and in the learning-classification phases, we use the supports vectors machines (SVM).The simulation results demonstrates that the KIM method gives more significant results that the PZIM for each Arabic character.**

***Index Terms*—Recognition, Printed Arabic characters, Krawtchouk invariant moments, Pseudo Zernike invariant moments, Support vectors machines.**

I. Introduction

In the pattern recognition many mathematical tools used to extract the primitives from pattern such the moments [1,2], and to train and classify these patterns such a support vector machines (SVM) [3,4].

In fact the moments are efficient methods to extract the primitives from the images that can be learned and classified by the SVM method. In this paper we are focused on the comparison between the performance of the KIM and the PZIM in the recognition of noisy printed Arabic characters. In the preprocessing of the images, representing each character, we use the thresholding technique and in the extraction phase of the primitives, we use the KIM [5] and the PZIM [6] which are used to transform the images to the vectors. The transformed images are used as classes of a SVM in the training phase and then for classify the images at the test phase. We describe our processes of recognitions as follow : In the learning phase we use the SVM method, whose the strategy is one against all for optimally separating each image (which modeled by a class characterized by a label which equal to 1) of the learning base to the rest of the other images that is modeled by another class characterized by a label which equal to 1.This separation (maximizing the margin between two classes) is therefore creating a decision function separating these 2 classes. We have 28 numerals each of them will be used as a class with a label which equal to 1 and the rest of the other characters will be fully accrued in another class with opposite label which equal to -1. So we built 28 decision functions each of them separating a pair of classes (1and-1) among the 28 pairs. In the classification phase, we calculate the image of the vector which models the character test preprocessed translated, rotated containing a noise by all 28 decisions functions, the recognition will be given to the character whose decision function separating its classes to another class containing the rest of the other character s which gives the largest value among all the calculated values of the 28 images of the test.

In this paper, we present firstly the technique of the Images preprocessing that we used, then we describe the process of extracting of primitives from images by the KIM and the PZIM then we will deduce the performances of those moments with using the SVM in the learning and classification phases.

II The Proposed System

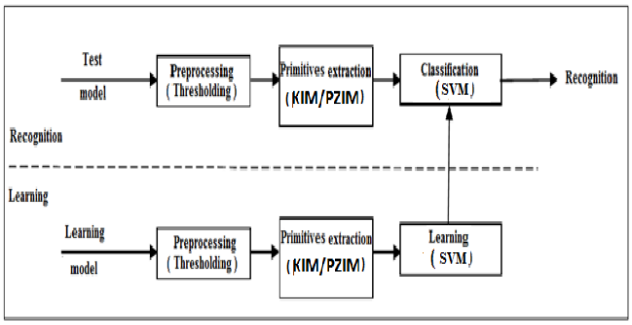


Fig1: The proposed system for Arabic characters recognition.

III The Pre-Processing

Before proceeding to the character recognition, it should be to preprocessing the images of these characters

doi:10.4304/jetwi.6.1.89-93

for to eliminate as much as possible the noise.existed in the images. In our approach, we preprocess the images by a thresholding technique which makes the images contain only black and white according a preset threshold.

IV THE PHASE OF PRIMITIVES EXTRACTION

4.1 The Krawtchouk Moments

4.1.1 The Krawtchouk polynomials

The definition of n-th order of classical krawtchouk polynomial is defined as:

$$K_n(x;p,N)=\sum_{k=0}^{N}a_{k,n,p}x^k={}_2F_1(-n,-x,-N;\frac{1}{p}) \quad (1)$$

Where : $x, n = 0,1,2 \ldots N, N \succ 0, p \in [0,1]$.

${}_2F_1$ is the hyper geometric function[7], defined as :

$${}_2F_1(a,b;c;x) = \sum_{k=0}^{\infty}\frac{(a)_k(b)_k}{(c)_k}\frac{x^k}{k!} \quad (2)$$

And $(a)_k$ is the pochhammer symbol (also rising factorial) defined by:

$$(a)_k = a(a+1)\ldots(a+k-1) = \frac{\Gamma(a+k)}{\Gamma(a)} \quad (3)$$

The Γ function is defined by:

$$\Gamma(x) = \int_0^{\infty} t^{x-1}e^{-t}dt \quad (4)$$

And : $\forall n \in N, \Gamma(n+1) = n!$

The set of $(N+1)$ Krawtchouk polynomials $\{k_n(x,p,N)\}$ forms a complete set of discrete basis functions with weight function:

$$w(x;p,N) = \binom{N}{x}p^x(1-p)^{N-x} \quad (5)$$

and satisfies the orthogonality condition:

$$\sum_{x=0}^{N} w(x;p,N)K_n(x;p,N)K_m(x;p,N) = \rho(n;p,N)\delta_{nm} \quad (6)$$

Where : $m, n = 0,1,2 \ldots N$, *and* $\rho(n;p,N)$

Is the squared norm, which is given by:

$$\rho(n;p,N) = (-1)^n\left(\frac{1-p}{p}\right)^n\frac{n!}{(-N)_n} \quad (7)$$

And δ_{nm} is the Kronecker symbol defined by

$$\delta_{nm} = \begin{cases} 1 & if\ n=m \\ 0 & others \end{cases} \quad (8)$$

4.1.2 The Krawtchouk moments

The Krawtchouk moments have the interesting property of being able to extract local features of an image. The krawtchouk moments of order $(n+m)$ in terms of weighted krawtchouk polynomials, for an image with intensity function, $f(x,y)$ is given by:

$$Q_{nm} = \sum_{x=0}^{N-1}\sum_{y=0}^{M-1}\bar{K}_n(x;p_1,N-1)\bar{K}_m(y;p_2,M-1)f(x,y) \quad (9)$$

The NxM is the number of pixels of an image $f(x,y)$ .The set of weighted Krawtchouk polynomials $\bar{K}_n(x,p,N)$ is defined by :

$$\bar{K}_n(x;p,N) = K_n(x;p,N)\sqrt{\frac{w(x;p,N)}{\rho(x;p,N)}} \quad (10)$$

4.1.3 The Krawtchouk invariants moments

The geometric moments [8,9]of an image $f(x,y)$ is defined by :

$$M_{nm} = \sum_{x=0}^{N-1}\sum_{y=0}^{M-1}x^n y^m f(x,y) \quad (11)$$

Then the standard set of geometric moment invariants which are independent to rotation, scaling and translation can be written as:

$$V_{nm} = M_{00}^{-\gamma}\sum_{x=0}^{N-1}\sum_{y=0}^{M-1}[(x-\bar{x})\cos\theta + (y-\bar{y})\sin\theta]^n$$

$$*[(y-\bar{y})\cos\theta(x-\bar{x})\sin\theta]^m f(x,y) \quad (12)$$

$$Where: \gamma = \frac{n+m}{2}+1, \quad \bar{x} = \frac{M_{10}}{M_{00}}, \quad \bar{y} = \frac{M_{01}}{M_{00}}$$

$$\theta = \frac{1}{2}arctg\ \frac{2\mu_{11}}{\mu_{20}-\mu_{02}} \quad (13)$$

And μ_{nm} are the central moments defined in as:

$$\mu_{nm} = \sum_{x=0}^{N-1}\sum_{y=0}^{M-1}(x-\bar{x})^n(y-\bar{y})^m f(x,y) \quad (14)$$

The krawtchouk moment invariant is :

$$\tilde{\Omega}_{nm} = \Omega_{nm}\sum_{i=0}^{n}\sum_{j=0}^{m}a_{i,n,p_1}a_{j,m,p_2}\tilde{V}_{ij} \quad (15)$$

$$\Omega_{nm} = [\rho(n;p_1,N-1).\rho(m;p_2,M-1)]^{-1/2} \quad (16)$$

$$With: \tilde{V}_{ij} = \sum_{p=0}^{i}\sum_{q=0}^{j}\binom{i}{p}\binom{j}{q}\left(\frac{N^2}{2}\right)^{\frac{p+q}{2}+1}\left(\frac{N}{2}\right)^{i+j-p-q}V_{pq} \quad (17)$$

$$And: \binom{x}{y} = \frac{x!}{y!(x-y)!} \quad (18)$$

For each character, the values calculated by the krawtchouk invariant moments are used as a vectors which will be considered as a class of a SVM which's making in the characters recognition.

4.2 The Pseudo Zernike Moments

For a image $f(x,y)$,the pseudo-Zernike moment of order n and repetition m is given by:

$$A_{nm} = \sum_{x=0}^{N-1}\sum_{y=0}^{M-1}f(x,y)V^*(x,y) \quad (19)$$

$$V^*(x,y) = R_{nm}(x,y)e^{jm\arctan(y/x)} \quad (20)$$

$$And: R_{nm}(x,y) = \sum_{s=0}^{n-|m|} \frac{(-1)^s (x^2+y^2)^{\frac{n-s}{2}} (2n+1-s)!}{s!(n-|m|-s)!(n+|m|+1+s)!} \quad (21)$$

where $x^2 + y^2 \leq 1$ and the symbol $*$ denotes the complex conjugate operator.

4.2.1 The pseudo Zernike invariants moments

The pseudo-Zernike moment is invariant under rotation but sensitive to translation and scale. Therefore a normalization must be done of this moments.

$$f(x,y) = f(\bar{x} + \frac{x}{a}, \bar{y} + \frac{y}{a}) \quad (22)$$

where $(\bar{x}, \bar{y})$ being centered of pattern function $f(x,y)$ and $a = (\beta / M_{00})^{1/2}$, β is a predetermined value for the number of object points in the pattern.

We note that The KIM formula is more complex than that of the PZIM.

V THE LEARNING PHASE

5.1 Principle of Functioning between Two Classes of the SVM

5.1.1 Linear case

Given a set of vectors $x_i \in \Re^n$ n is the dimension of the vector space and two classes. The first class containing a party of these vectors and bears the label 1,the second class contains the other party of vectors and bears the label -1.the goal of SVM [10] is to find a classifier that will separate these classes and maximize the distance between them. This classifier is called hyperplane (see figure 2)

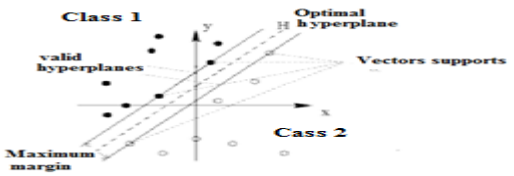


Fig 2: determination of optimal hyperplane, vectors supports, maximum marge and valid hyperplane.

The nearest points which alone are used for determination of hyperplane are called support vectors. The property of SVM is that this hyperplane must be optimal that is to say it must maximize the distance between the supports vectors of a class and those of the other class.The classifier is represented by :

$$f(x, w, b) : x \rightarrow y \quad (23)$$

Where w and b are the parameters of the classifier y is the label.

5.1.2 The primal/dual problems

5.1.3 The primal problem

For to maxim the distance between the supports vectors of a class and those of the other class, we must to solve a problem of minimization under the constraints called the primal problem:

To minimize $P(w,b) = \frac{1}{2}\|w\|^2$ (24)

Such that $y_i(wx_i + b) \geq 1$

5.1.4 The dual problem

To simplify the calculations, it is necessary to introduce a formulation called dual of the problem by using the Lagrangian operator L

$$L(w,b,\alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i [y_i(wx_i + b) - 1] \quad (25)$$

The dual variables α_i intervening in the Lagrangian are called Lagrange multipliers. The dual problem is:

To maximize $D(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j x_i x_j$

Such that $\sum_{i=1}^{n} \alpha_i y_i = 0 \quad \alpha_i \geq 0 \ , \ \forall i = 1,2,.....,n$ (26)

Only the α_i^* corresponding to points nearest to hyperplane is nonzero, we speak of support vectors. The decision function associated is:

$$f(x) = \sum_{i=1}^{n} \alpha_i^* y_i x_i . x + b \quad (27)$$

5.1.5 Non linear case

In the linear case (see figure 3), the classification of data is easy, but in the nonlinear case the representation space $\Re^n$ must be transformed to a other space of higher dimension is called space of re-description $\Re^p$ ($p \succ n$), this transformation is carried by virtue a special functions called the kernel functions:

$$K : \Re^n x \Re^n \rightarrow \Re^p \qquad p \succ n \quad (28)$$
$$(x_i, x_j) \rightarrow K(x_i, x_j)$$

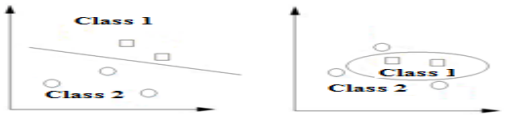


Fig 3: nonlinear separation between class 1 and class 2.

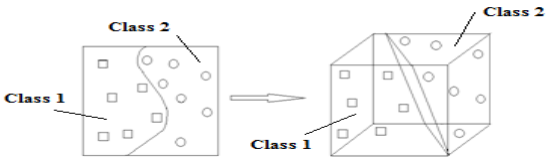


Fig 4: transformation of original data space to a re-description space.

We must solve therefore:

To maximize $D(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$

Such that $\sum_{i=1}^{n} \alpha_i y_i = 0$ (29)

$0 \leq \alpha_i \leq C \ , \qquad \forall i = 1,2,.....,n$

The parameter C which appears here is a positive constant fixed in advance, it's called the penalty constant. The decision function has the form:

$$f(x) = \sum_{i=1}^{n} \alpha_i^* y_i K(x, x_i.) + b \quad (30)$$

Some example of the kernel functions:

Kernel linear	xy
Kernel polynomial of degree n	$(axy + b)^n$
Gaussian Radial Basis Function(GRBF):	$e^{-\frac{\|x-y\|^2}{2\sigma^2}}$

5.2 Principle of Functioning between a Several Classes of the SVMs

The method described above is designed for two classes problem many studies treat a generalization of this method to multi-classification classes [11] among these studies we cite the two most frequently used strategies: The first approach is to use N decision functions (one against t all) allowing to make the discrimination of each class against all others. The decision rule used in this case is usually the maximum such that we assign an unknown vector x into the class associated with the output of SVM is the largest:

$$i = \arg \max_{i=1,2,\ldots,N} (f_i(x)) \quad (31)$$

The second method called one against one Instead of learning N decision functions, each class is discriminated against another So $\frac{N(N-1)}{2}$ decision functions are learned and each of them performs a voting for the assignment of a new unknown vector x. its class then becomes the majority class after the vote.

In our work, We use the kernel function GRBF with the standard deviation $\sigma = 0.2$ and the penalty constant: $C = 10^{30}$.

VI The Classification Phase

After having built the 28 decision functions between the 28 pairs of classes in the learning phase by the strategy of(one against all) we calculate the values of the images vector modeling the character test by all the 28 decision functions the recognition will be given to the character whose decision function separating its class to another class containing the rest of the other characters which gives the largest value among all calculated values of the 28 images of the test.

VII Experements and Results

We choose the size image 50x50. Each character was converted to vector of 7 components which are the KIM/PZIM that used as: a class having a label that is equal to 1, and the rest of all the other vectors modeling the other characters can be used. The separation of the 28 pairs of two classes by the construction of the 28 decision functions is fact by the strategy of one against all. First we present the test character translated, rotated…and not containing a noise, then we add increasingly an quantity of noise of type 'salt & pepper' for to know the effect of noise added on the rate recognition of each character, for each moment. We choose the parameters of the KIM equals to: p=q=0.96. For each test character noisy, translated, rotated...The noise used is of type 'salt & pepper', its values are: [0,0.01,0.02, 0.03, 0.04 ,0.05,0.06, 0.07,0.08, 0.09,0.1,0.11,0.12,0.13,0.14,0.15,0.16, 017,0.18, 0.19, 0.2]. We group the results obtained in the following tables:

TABLE 1:
THE RECOGNITION RATE OF EACH ARABIC CHARACTER.

Character	τ_c (PZIM)	τ_c (KIM)
أ	39%	100%
ب	48%	100%
ت	58%	58%
ث	81%	96%
ج	29%	34%
ح	34%	72%
خ	29%	72%
د	29%	100%
ذ	48%	100%
ر	100%	100%
ز	62%	100%
س	29%	34%
ش	100%	68%
ص	20%	39%
ض	34%	48%
ط	100%	29%
ظ	43%	39%
ع	39%	58%
غ	68%	34%
ف	53%	81%
ق	100%	100%
ك	100%	100%
ل	81%	100%
م	20%	100%
ن	100%	100%
ه	34%	39%
و	48%	100%
ي	29%	100%

TABLE 2 :
THE VARIATION OF THE GLOBAL RECOGNITION RATE IN FUNCTION OF NOISE ADDED. AND THE ASSOCIATED GRAPH

Noise	τ_g (PZIM)	τ_g (KIM)
²0	100%	100 %
0.01	100%	100%
0.02	100%	100%
0.03	100%	100%
0.04	93%	100%
0.05	93%	97%
0.06	75%	86%
0.07	65%	75%
0.08	58%	75%
0.09	54%	72%
0.10	43%	72%
0.11	40%	65%
0.12	36%	65%
0.13	36%	61%
0.14	33%	54%
0.15	33%	54%
0.16	33%	54%
0.17	25%	50%
0.18	25%	50%
0.19	25%	50%
0.20	25%	47%

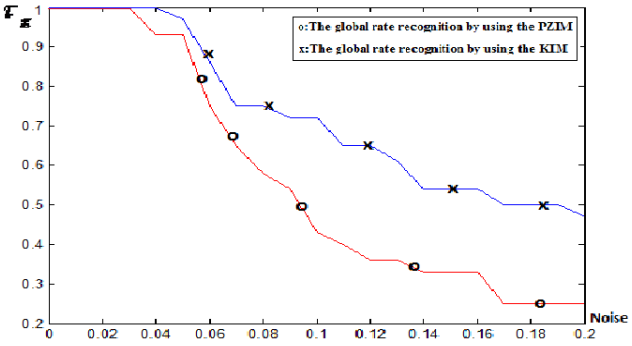


Fig 5: Evolution of global rate recognition τ_g of the PZIM and the KIM in function of noise added.

For two moments, the global rate recognition is a decreasing function according to noise added, but the important remark is that the falling of this rate of the PZIM is greater than the rate of the KIM, this that we check that the KIM is more robust than the PZIM against noise.

VIII CONCLUSION

In this paper we have concentrated on character recognition of noisy Arabic characters. The results, shows that reliable recognition is possible using a thresholding technique in the preprocessing phase and the invariants moments of Krawtchouk and those of pseudo Zernike in the primitives' extraction phase and the SVMs in the learning and the classification phase. The simulation result demonstrates that the KIM is more robust and more performing than the PZIM, despite the fact that the recognition time of the KIM is greater than that of the PZIM.

REFERENCES:

[1] Anass El affar Khalid Ferdous,Abdeljabbar Cherkaoui,Hakim El fadil and Hassan Qjidaa, "Krawtchouk Moment Feature Extraction for Neural Arabic Handwritten Words Recognition", IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.1, January 2009.

[2] Ibrahim A.El rube',Mohamed T. El Sonni,and Soha S.Saleh, "Printed Arabic Sub-Word Recognition Using Moments", World Academy of Science, Engineering and Technology 42 2010.

[3] Sami Gazzah and Najoua Ben Amara, "Neural networks and supports vectors machines classifiers for writer identification using Arabic scripts" ,The International Arab Journal of Information Technology, Vol. 5, No. 1, January 2008

[4] Mehmmood A. Abd , Sarab Al Rubeaai and George Paschos, "Hybrid features for an Arabic word recognition system", Computer Technology and Application 3 (2012) 685-691

[5] Pew-Thian Yap, Raveendran Paramesran, Senior Member, IEEE,and Seng-Huat Ong.:Image analysis by Krawtchouk moments, IEEE transactions on image processing , vol. 12, NO. 11, November 2003

[6] M. Teague, "Image analysis via the general theory of moments". Journal Optical Society of America,vol.70, pp: 920–930,1980.

[7] Jan Flusser Tomáš Suk Barbara Zitová "Moments and Moment Invariants in Pattern Recognition. © 2009 John Wiley& Son.

[8] R.Mukundan and K.R.Ramakrishnan, "Moment Functions in Image Analysis: Theory and Applications", Worl Scientific Publishing, 1998.

[9] V.N. Vapnik, "An overview of statistical learning theory", IEEE Trans. Neural Networks., vol. 10, pp. 988– 999, Sep. 1999.

[10] C. Singh ; E. Walia and N. Mittal "Rotation invariant complex Zernike moments features and their applications to human face and character recognition", Computer Vision, Volume 5, Issue 5, September 2011, p. 255 – 266,

[11] C. Cortes and V. Vapnik. "Support vector networks. Machine Learning", 20:pp 1–25, 1995.

People Tracking using Color Control Points and Skin Color

K.Laaroussi
[1]LIIAN
[1]Faculty of sciences Dhar-Mehraz, Sidi Mohamed Ben Abdellah University, B.P 1796, Atlas, Fez, Morocco
Laaroussi.khadija@yahoo.fr

A.Saaidi
[1]LIIAN, [2]LIMAO
[1]Faculty of sciences Dhar-Mehraz, Sidi Mohamed Ben Abdellah University, B.P 1796, Atlas, Fez, Morocco.
[2]Faculty polydisciplinary of Taza, Sidi Mohamed Ben Abdellah University, B.P 1223, Taza, Morocco.
saaidi.abde@yahoo.fr

K.Satori[1]
[1]LIIAN
[1]Faculty of sciences Dhar-Mehraz, Sidi Mohamed Ben Abdellah University, B.P 1796, Atlas, Fez, Morocco
khalidsatorim3i@yahoo.fr

***Abstract*—This paper presents a new approach to be used for people detection and tracking in image sequences based on color control points and skin color modeling. This method aims to track these people in complex situations such as players on a soccer field. Each person in the image is represented by several control points which are obtained using a color version of the Harris algorithm detector. Each control point is characterized by the local appearance which is a vector of local characteristics. Then we determine the rules that define the skin regions in three different color spaces such as RGB, HSV and YCbCr, and we apply these rules to our images to segment skin regions. Using a set of control points and skin regions allows us to track a person by matching control points based on the measure of ZNCC correlation «Zero mean Normalized Cross Correlation». The simulations and experimental results show the robustness of our algorithms in terms of stability and convergence. Performance is illustrated by some examples. Thus, our method fits well with the noise conditions.**

***Index Terms*—Color Control Points, Skin Color, Appearance Model, Matching, People Detection, People Tracking.**

I. Introduction

Detecting and tracking people in image sequences is a difficult task for several reasons namely the complex movements of people followed, the human body is highly articulated, the illumination of the scene changes can result in non-consistency pixel values representing a person, as well as image noise and phenomenon of occlusion (the person itself, with the other moving objects or objects in the background). However, the use of people tracking is pertinent in many applications of computer vision such as human-computer interaction, civil and military surveillance, video conferencing or analysis of human movement.

In its simplest form, tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. A large number of tracking methods have been proposed; in fact most of them are based on region segmentation [16, 17] or people tracking based on the subtraction of the background [1] which simplifies subsequent processing by locating regions of interest in the image from a model of the environment and an observation. We also find algorithms that are based on the extraction of contours with a particle filtering using color information or a filter Calman [2, 3, 18, 21] or tracking based on mean shift [13, 19]. Other algorithms are based on active contours [20] or normalized cross correlation and color space [22, 23], optical flow [24].

Or algorithms which using a variable search window (VSW) algorithm based on color and feature points is proposed [28].

Recently, Gouet and Lameyre [14, 25] presented a tracker that uses control points and Snacks but with images in grayscale and for a single object in the scene. However, to the best of our knowledge, there papers [26, 31] presents a new approach for tracking objects in complex situations using color interest points characterized both by local descriptors and by a geometric model.

In this work, we are interested in tracking people in image sequences, or more specifically to track a certain number of points detected on these people; the main purpose is to reduce the maximum information of an image in some control points and some skin regions. Thus, a person is represented by several control points which are obtained with a color version of the Harris algorithm detector [4, 33]. Each control point is characterized by its local appearance which is a vector of local characteristics. Then we determine the rules that define the skin regions in three different color spaces such as RGB, HSV and YCbCr, and we apply these rules to our images to segment skin regions. Using a set of control points allows us to track a person by matching control points from image to image based on the measure of ZNCC correlation «Zero mean Normalized Cross Correlation» [9, 34, 35].

doi:10.4304/jetwi.6.1.94-100

The main contribution of our method is to track multiple people in a color video sequence containing complex scenes where people can have different size. Indeed, the combination of these two primitives will develop a tracking method implementing the time parameter and quality of tracking. However, skin color is a primitive often used as a first estimate of location and segmentation to reduce the search area. The color information is clearly relevant to the presence of people in video since they are very specific and allow fast algorithms invariant to changes in orientation and scale. The strong point of this method is that it is possible to achieve a more robust tracking by using only one of the two primitives considered i.e. if one of the two failed, it is possible to keep track of a person using the other and vice versa.

Figure 1 explains our treatment procedure for detection and tracking people in image sequences:

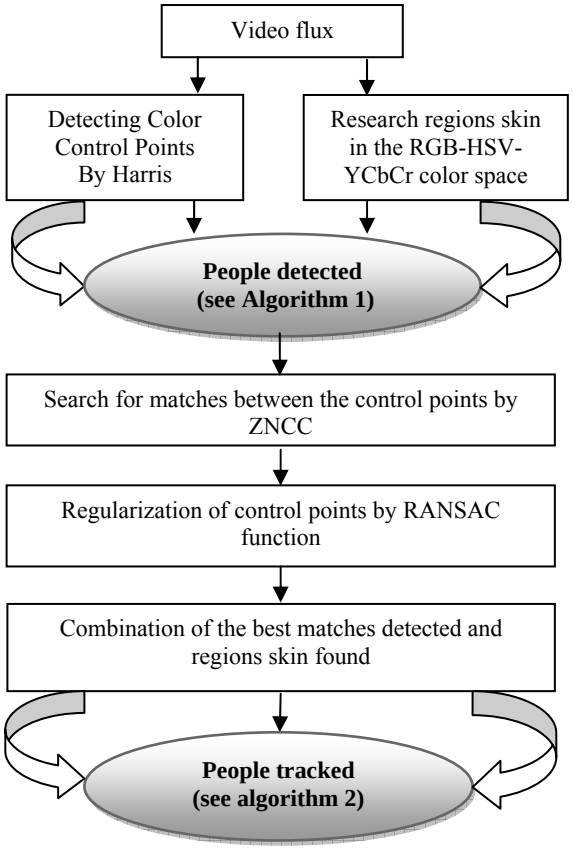


Fig1: Procedure for detection and tracking people.

This paper is organized as follows: in section 2 we describe the tracking tools, namely skin color detection, color version of Harris detector, and matching. In section 3, we present our proposed tracking system. In section 4, we show the results of our experiment. In section 5 we conclude with a summary.

II TRACKING TOOLS

2.1 Skin color Detection

The skin detection is to detect the pixels corresponding to human skin in a color image; including methods of detecting skin color are the explicit methods [5, 6, 12]. However, the skin color can be presented in different color spaces such as:

- **RGB Space:** This color space combines three color components (red, green and blue), it not need any model of skin and no color transformation [10, 15]. The threshold applied in this space is defined as: *(RGB*) is classified as a skin color pixel if:

(R>95) and (G>40) and (B>20) and (((max{R, G, B})-(min{R, G, B}))>15) and (ABS(R-G)>15) and (R>G) and (R>B) (1)

- **YCbCr space:** In this space there is Y represents the luminance and Cb Cr represent the two components of chrominance (Blue and Red). This space separates the luminance and chrominance making YCbCr attractive for modeling the skin color [15]. The threshold in this area is as follows:

$$77 <= Cb <= 127$$
$$133<=Cr<=173 \quad (2)$$

- **HSV space** *(Hue, Saturation, value*): Hue defines the dominant color in a region. The saturation of the color measurement of a region in proportion to its shine. The highlight of the use of this color space is the discrimination between the luminance and chrominance [15]. The Threshold in this region is as follows: [11]:

$$0 \le H \le 0.5$$
$$0.17 \le S \le 0.63 \quad (3)$$

However, we can make a hybridization of two or more spaces to have good results.

2.2 Color Harris Detector

Control points are the locations of an image which contain the maximum information, there are involved in many applications, like stereovision, image retrieval or scene monitoring. Several control points detectors have been developed over the last two decades. Schmid and Mohr [7] compare the performance of several of them. The most popular control point detector is the Harris detector [8] with its adaptations [27, 29, 30]. While this detector only applies to grayscale images, Montesinos et al. [4] generalized it to color images. The control points produced by their detector are defined as the positive local extrema of the intermediate grayscale image.

$$R(x,y) = \det(M(x,y)) - k \cdot trace(M(x,y))^2 \quad (4)$$

Where k is typically set to 0.04 and $M_{Color}(x,y)$ is the 2×2 matrix.

$$M_{Color} = \begin{bmatrix} R_x^2 + G_x^2 + B_x^2 & R_x R_y + G_x G_Y + B_x B_y \\ R_x R_y + G_x G_Y + B_x B_y & R_x^2 + G_x^2 + B_x^2 \end{bmatrix} \quad (5)$$

With R, G and B are three components such as Red, Green, and Blue.

After the matrix calculation M_{Color} for each pixel, we eliminate the values that are lower to a single modifiable

of response R and we make the extraction of the local maxima.

According to the comparisons made by Gouet and Boujmaa [32], the above detector appears to be the most stable among the popular color control points detectors with regard to illumination changes, noise with a Gaussian filter, rotation, and viewpoint changes.

2.3 Matching

Matching correlation ZNCC "Zero mean Normalized Cross Correlation" or centered normalized correlation [9, 34, 35] allows measuring the similarity of pixels between the two images by calculating the correlation between two windows. It provides a measure of the interval [-1, 1], the similarity between two points X and Y is high when ZNCC(X,Y) approaches to 1.

In the practice, we set a threshold S (in general S=0.5) such as the couples(X, Y) are considered the true correspondents if the following constraint is satisfied:

$$ZNCC\ (X, Y) > s \tag{6}$$

This function uses the following formula to measure the correlation between the points of images:

$$ZNCC\ (X, Y) = \frac{\sum_{i=1}^{n} (x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{X})^2 \sum_{i=1}^{n} (y_i - \overline{Y})^2}} \tag{7}$$

With: - $X = (x_i)_{i=1\ldots n}$ and $Y = (y_i)_{i=1\ldots n}$ the coordinates of points X and Y.

- $\overline{X}$ and $\overline{Y}$ are the averages of X and Y respectively such as:

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \overline{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

III TRACKING SYSTEM

The purpose of the proposed system is to track people in a color video sequence, for each image we apply the color Harris detector to extract the control points. Then we determine the rules that define the regions of the skin in three color spaces RGB, HSV and YCbCr and apply these rules to our images to segment skin regions. To keep track of people we are looking for connections between these points, calling the measure of correlation ZNCC.

The purpose of the proposed system is to track persons from frame to frame, for each image we use the Harris detector to extract color control points. Then we determine the rules that define the areas of the skin in three different color spaces RGB, HSV and YCbCr and apply these rules to our images to segment skin regions. To keep track of people we are looking for connections between these points, calling the measure of correlation ZNCC.

However, each control point is characterized by a vector containing some local attributes, it consists of three channels r, g and b plus Gaussian derivatives $r_x, g_x, b_x, r_y, g_y, b_y$, and we also add the local Cornerness R given by equation (4).

So, we characterize each control point j in frame n by the vector following characteristics:

$$V_j(n) = (r, g, b, r_x, g_x, b_x, r_y, g_y, b_y, R) \tag{8}$$

A person p in image n is then modeled by:

$$V^p(n) = \{V_j(n) / j = 1,...., P(n)\} \tag{9}$$

Where P(n) is the number of control points.

The basic problem is to find for each control point $i \in \{1,....,P(n-1)\}$ with $P(n-1)$ the number of control points detected in the image $n-1$, the best control point j which is corresponding with $j \in \{1,....,P(n)\}$ where $P(n)$ the number of control points for the image n. The search party is local and occurs in an area of research for the control point i is calculated all the correlation values on the corresponding epipolar line in the other image, the control point associated with the highest score is selected as appropriate (see figure 2).

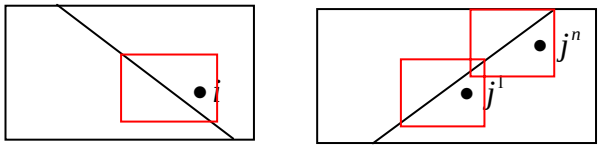


Fig2: Correlation windows

The degree of similarity between pairs $(i \quad j^1), \ldots, (i \quad j^n)$ is calculated using a window centered at i and a window that moves on the points $j^1, \ldots, j^n$, the robustness of these functions are classified according to their matching capabilities in the case of large displacement, change brightness, noise and low textured areas.

The performance of our tracking system is also closely linked with the results obtained in the detection of skin color. The segmentation procedure is done by scanning all the pixels of the image in search of those that follow the desired threshold in the space used. The pixel belonging to the range of desired threshold is 1 otherwise it is 0. This process produces a binary image highlighting segments with skin color (white) the other did not take the skin color black.

Our suggested method for tracking persons contains two algorithms, in the first; we propose a simple and effective method for detection of people in a sequence images based on the color control points detected by color Harris and skin color with the aim of eliminating the non- skin color regions. The second algorithm is to keep track of people detected by Algorithm 1 based on matching correlation ZNCC and the skin regions.

Algorithm 1: (People Detection)

Step1: *Image acquisition: Take a series of successive images by a digital camera or from a color video.*

Step2: *Apply color Harris detector for the first image* $n = 1$ *with the aim of identifying regions of interest detected on people.*

Step3: *Apply a skin detector in the YCbCr-RGB-HSV*

space to segment the image into skin regions and differentiate individuals with respect to other objects.

Step4: *Apply a skin detector in the YCbCr-RGB-HSV space to segment the image into skin regions and differentiate people with respect to other objects.*

Step5: *Combination of the detected control points and regions skin found.*

Step6: *Show the detection result: people detected are surrounded by rectangles containing the found color control points and sections of skin.*

The result of Algorithm 1 allows us to initialize the algorithm 2, so the detection step serves only to initialize people tracking.

Algorithm 2: (People Tracking)

Step1: *Initialization :*
Image acquisition: Take a successive series of images containing people to track.
Apply Algorithm 1 for the first image $n = 1$ *and* $n = 2$.
Determine the characteristic vector for each detected point control $V^P(1)$ *and* $V^P(2)$.

Step2: *Apply Algorithm 1 for each image n and fill vectors* $V(n)$ *with appearance features.*

Step3: *Match the color control points detected by Harris using correlation measure ZNCC.*

Step4*: Regularization of control points detected by RANSAC function.*

Step5: *Combining the results of the best matches detected and regions skin found.*

Step6: *Show results of tracking by combining color control points and skin color, such as persons tracking is surrounded by rectangles.*

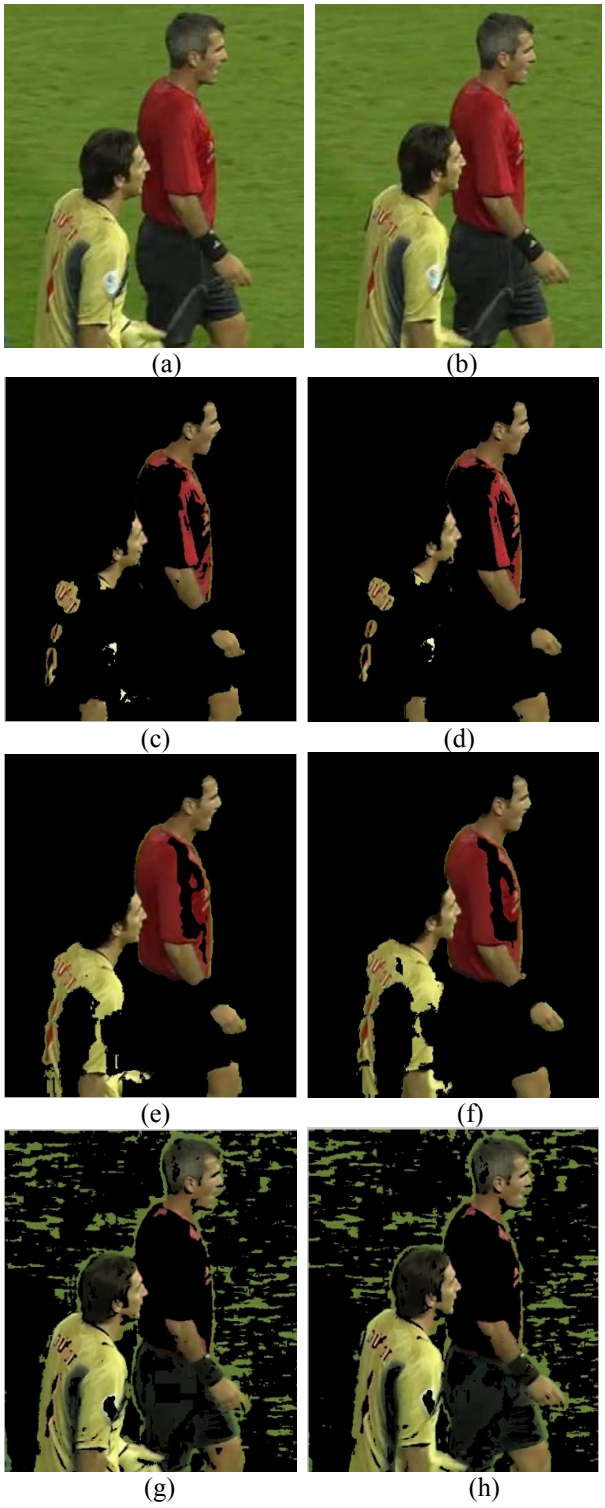
These two algorithms were tested on a large number of image sequences. The scenarios are varied, ranging from individual person walking alone to people in a crowd interacting (two or more people meet and walk side by side). The simulated results show that our algorithm is able to track more people, and therefore, detect and manage the interactions between them.

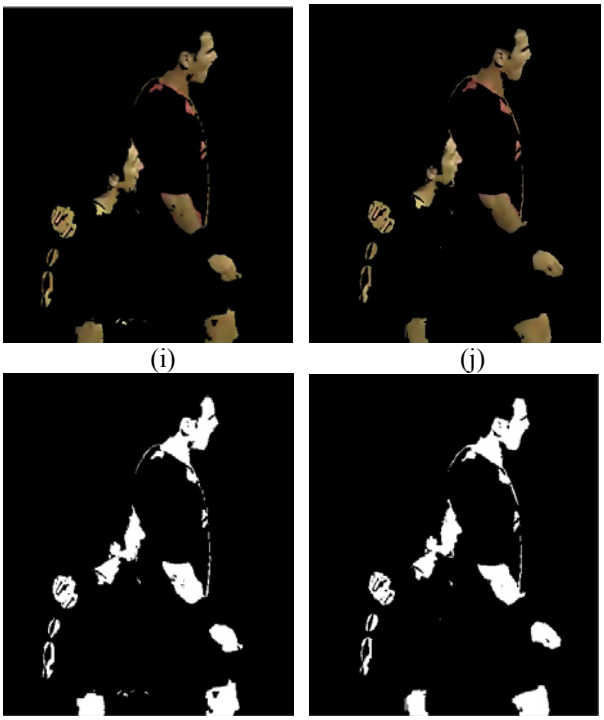
IV Experimentation

To work on a real case with our method, we have chosen two successive images with 374×421 size taken from a football match video time of 4:13 min (two people who meet and walk side by side). First, we detected the control points by Harris algorithm [4, 33] and matched to each pair of images by measuring correlation ZNCC [9, 34, 35], at this stage several matches are false, where a regularization step is very important to keep just the best matches found by calling the function (RANSAC: Random Sample Consensus) [36]. Then we determine the rules that define the areas of the skin in the hybridization of these three color spaces RGB, HSV and YCbCr and apply these rules to our images to segment skin regions.

All algorithms used in this article have been implemented in Matlab 7 Intel Core (TM) 2 CPU 2.00GHz, Microsoft Windows 7 with 3GB memories.

Fig.3 shows an example for detecting skin color for these two images in the three spaces RGB, YCbCr and HSV.



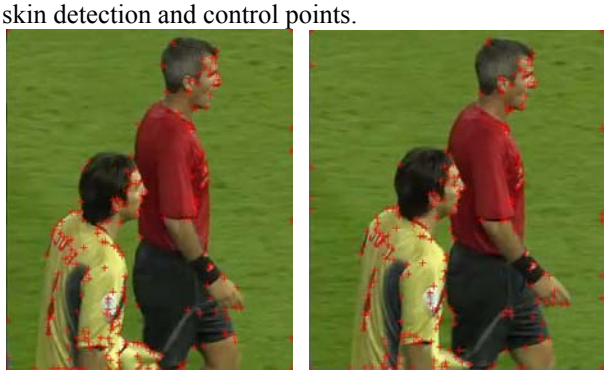


(i) (j) (k) (l)

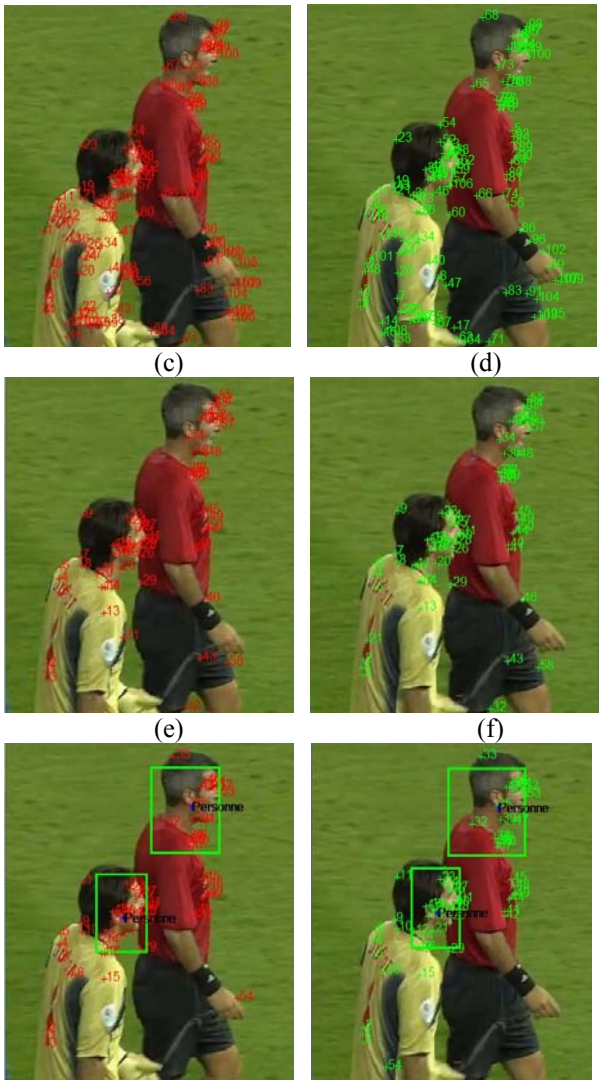
Fig3: Example of skin detection. (a) and (b) two successive original image, (c) and (d) skin pixels in the RGB space, (e) and (f) skin pixels in the YCbCr space, (g) and (h) in the HSV space, (i) and (j) in the RGB-YCbCr-HSV space, (k) and (l) the binary images in the YCbCr-RGB-HSV space.

Based on these results, the detection of skin is based on the choice of the color space and the phase thresholding, and the hybridization of these three spaces RGB, YCbCr and HSV gives us a better result compared to others, we can nevertheless see a higher error rate for certain spaces (YCbCr, HSV). The execution time of the program is 1.039306 seconds for each image.

In Figure 4, we illustrate the result of tracking by applying the two algorithms 1 and 2. First of all, we detect the control points then we determine the skin regions. Tracking consists of matching control points detected on these two images and skin color. We can notice that our system is able to track these people, by detecting several skin areas, despite the existence of false skin detection and control points.



(a) (b)



(c) (d) (e) (f) (g) (h)

Fig4: Results of tracking using both algorithms 1 and 2. (a) and (b) the red and green points indicate control points detected by Harris detector, (c) and (d) indicate control points detected by correlation ZNCC, (e) and (f) regularization of control points detected by RANSAC function, (g) and (h) results of tracking based on the color control points and skin color in the RGB-YCbCr-HSV space such as persons tracking is surrounded by rectangles.

According to the results in Figure 4, the two people are correctly detected and tracked; the result of matching is best as pictures show (e) and (f) after the elimination of false matches by RANSAC function. The skin regions are specified and surrounded by boxes as the pictures show (g) and (h). The total running time of our system is 72.256917 seconds for the two images.

Our proposed tracking system has been tested on a large number of images from several snuff videos of different length sequences. The scenarios are varied, ranging from simple person walking alone several people interacting (two or more people meet and walk side by side). The experiment results show that our algorithm is able to detect and track multiple people in image sequences, and that this system is robust in the case of the existence of noise.

In summary, according to the simulations, only the proposed method can correctly track people, other methods are disturbed by noise or the presence of control points belong to the background and use of skin color allows increase the performance of our approach. By combining these two primitives will develop a tracking method implementing the time parameter and quality tracking.

V CONCLUSION

In this paper we have presented a new approach for tracking people in image sequences using color control points and skin color. A person is defined by a set of control point detected with version color Harris detector. Each control point is characterized by local appearance i.e. a vector of local characteristics more cornerness. Using a set of control points and regions skin allows us to track a person by making connections between these points from one image to another based on matching correlation ZNCC. The combination of control points with local descriptors has been successfully used for matching points and the color of the skin allows us to increase the performance of our proposed method, reducing areas of research.

REFERENCES

[1] C. Stauffer and W.E.L.Grimson, Adaptive background mixture models for real-time tracking, Computer Vision and Pattern Recognition, pages 246–252, 1999.

[2] M.Isard and A. Blake. Condensation – conditional density propagation for visual tracking, International Journal of Computer Vision, pages 5–28, 1998.

[3] P.Li, T.Zhang, and A.E.C.Pece, Visual contour tracking based on particle filters, Image and Vision Computing, pages 111–123, 2003.

[4] P.Montesinos, V.Gouet, and R.Deriche, Differential invariants for color images, International Conference on Pattern Recognition, pages 838–840, 1998.

[5] M.HAMMAMI, Modèle de peau et application à la classification d'images et au filtrage des sites Web, L'ECOLE CENTRALE DE LYON, pages 69-80, 2005.

[6] F.Gianni, Suivi de parties du corps pour l'interprétation de gestes de communication à partir d'une séquence monoculaire, Université Paul Sabatier-Toulouse III, pages 7-10, 2010.

[7] C.Schmid, R. Mohr, and C.Bauckhage, Evaluation of interest point detectors, International Journal of Computer Vision, pages 151–172, 2000.

[8] J. Harris and M. Stephens, A combined corner and edge detector, 4th ALVEY Vision Conference, pages 147–151, 1988.

[9] A.Saaidi, H.Tairi et K.Satori, Fast Stereo Matching Using Rectification and Correlation Techniques, ISCCSP, Second International Symposium on Communications, Control and Signal Processing, Marrakech, Morrocco, March 2006.

[10] J.Kovac, P.Peer and F.Solina, 2D versus 3D color space face detection, 4th EURASIP Conference on Video/Image Processing and Multimedia Communications, Croatia, pages 449-454, 2003.

[11] S. Tsekeridou and I. Pitas, Facial feature extraction in frontal views using biometric analogies, Proc. of the IX European Signal Processing Conference, vol. I, pages 315-318, 1998.

[12] H.BOUIROUGA, S.ELFKIHI, A.JILBAB, D.ABOUTAJDINE, Skin detection in pornographic videos using threshold technique, Journal of Theoretical and Applied Information Technology, Vol.35 No.1, 15th January 2012.

[13] D. Comaniciu, V. Ramesh, and P. Meer, Real-time tracking of non-rigid objects using mean shift. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Hilton Head, SC, volume II, pages 142–149, June, 2000,.

[14] V. Gouet and B. Lameyre. SAP: A robust approach to track objects in video streams with snakes and points. British Machine Vision Conference, pages 737–746, 2004.

[15] V.VEZNEVETS, V.SAZONOV, A.ANDREEVA. A Survey on Pixel-Based Skin Color Detection Techniques. International Conference Graphicon, Moscow, Russia, 2003.

[16] M. Kim, J.G. Jeon, J.S. Kwak, M.H. Lee, and C. Ahn. Moving object segmentation in video sequences by user interaction and automatic object tracking. IVC, 19(5):245–260, April 2001

[17] S. Pateux. Tracking of video objects using a backward projection technique. In Visual Conference on Image Processing, volume 4067, pages 1107–1114, Australia, 2000.

[18] L.Peihua , W.Haijing , Object Tracking with Particle Filter Using Color Information, Third International Conference, MIRAGE, Rocquencourt, France, pages 534-541, 2007.

[19] F.Zheyi , L.Mo , L.Zhiwen. An Improved Video Target Tracking Algorithm Based on Particle Filter and Mean-Shift. International Conference on Information Technology and Software Engineering, pages 409-418, 2013.

[20] S.Lefèvre, N.Vincent. Real Time Multiple Object Tracking Based on Active Contours. International Conference, ICIAR, Proceedings, Part II, pages 606-613, 2004.

[21] Y.ZHAO, H.PEI. Object tracking based on particle filter with discriminative features. Journal of Control Theory and Applications, Volume 11, Issue 1, pages 42-53, February 2013.

[22] W.Jian; Y.Heng-jun; Yan-yan Cao ; Zhi-ming Cui. Video Object Tracking Method Based on Normalized Cross-correlation Matching. Distributed Computing and Applications to Business Engineering and Science (DCABES), pages 523 – 527, 2010.

[23] S.Patrick, V.Yap Vooi. Tracking Using Normalized Cross Correlation and Color Space. International Conference on Intelligent and Advanced Systems, pages 770 – 774, 2007.

[24] G. Castellano, J. Boyce, and M. Sandler. Regularized cdwt optical flow applied to movingtarget detection in IR imagery. Machine Vision and Applications, pages 277–288, 2000.

[25] V. Gouet and B. Lameyre. Object Tracking and Identification in Video Streams with Snakes and Points. Advances in Multimedia Information Processing - PCM 2004 Lecture Notes in Computer Science Volume 3333, , pages 61-68, 2005.

[26] P. Gabriel, J.-B. Hayet, J. Piater, and J. Verly. Object tracking using color interest points. Advanced Video and Signal Based Surveillance, AVSS. IEEE Conference on, pages 159–164, 2005.

[27] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 530–534, May 1997.

[28] H.Lim, D.Kang, Object tracking system using a VSW algorithm based on color and point features, EURASIP Journal on Advances in Signal Processing, September 2011.

[29] K.Mikolajczyk, C.Schmid. Indexing based on scale invariant interest points. In International Conference on Computer Vision, Vancouver, Canada, July 2001.
[30] David G. Lowe. Distinctive image features from scale-invariant keypoints. Accepted for publication int the International Journal of Computer Vision, 2004.
[31] P. Gabriel, J.-B. Hayet, J. Piater, and J. Verly. Utilisation des Points d'Intérêts Couleurs pour le Suivi d'Objets. Actes du Congrès ORASIS, pages 159–164, 2005.
[32] V. Gouet and N. Boujemaa. About optimal use of color points of interest for content-based image retrieval. Internal Report, INRIA Rocquencourt, 2002.
[33] J. Harris and M. Stephens. A combined corner and edge detector. 4th ALVEY Vision Conference, pages 147–151, 1988.
[34] M.Lhuillier and L.Quan. Quasi-dense reconstruction from image sequence. ECCV, 2002.
[35] S.Chambon, A.Crouzil, M.El miziani, G.Lemarie, P. Le callet. Complémentarité de mesures de corrélation pour la mise en correspondance: 4TReconnaissance des Formes et. Intelligence Artificielle4T"4T, RFIA, 2010.
[36] M.A.Fischler, R.C.Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Graphics and Image Processing, Juin 1981.

Fuzzy Control System for Autonomous Navigation of Thymio II Mobile Robots

Fatma Boufera
Mascara University
Faculty of Science and Technology, Algeria
fboufera@yahoo.fr

Fatima Debbat
Mascara University,
Faculty of Science and Technology, Algeria
Debbat_fati@yahoo.fr

Francesco Mondada
EPFL, Lausanne
Robotic Systems Laboratory
francesco.mondada@epfl.ch

M. Faycal Khelfi
RIIR Laboratory
Faculty of Exact and Applied Sciences,
University of Oran
Algeria
mf_khelfi@yahoo.fr

***Abstract*—This paper treats the autonomous navigation problem of robotic systems in a dynamic and uncertain environment. In particular, we are interested in determining the robot motion to reach the target while ensuring their own safety and that of different agents that surround it. To achieve these goals, we have adopted a fuzzy controller for navigation and avoidance obstacle, taking into account the changing nature of the environment. The approach has been tested and validated on a ThymioII robots set. As application field, we have chosen a parking problem..**

***Index Terms*—Robotics, Navigation Autonomous, Fuzzy Logic, ThymioII Robot.**

I. Introduction

In unstructured environments, robot controller must be able to operate under conditions of imprecision and uncertainty and efficient response to events [1]. Obstacle avoidance is an essential component to achieve successful navigation. Several research works have been reported in this area. Several trajectory tracking and path following algorithms have been proposed to steer the mobile robot along a path to a desired goal in order to prevent collisions with obstacles. Many researches turned their attention to the obstacle avoidance problem developing interesting real-time behavior for mobile robots in unknown environment. The most well know are, the potential field method, vector field histogram and the method of deformable virtual zone. The first one was introduced by [2] imagines the virtual forces acting on the robot. This method assumes that the robot is driven by virtual forces that attract it towards the goal, or reject it away from the obstacles. The actual path is determined by the resultant of these virtual forces, the second method is introduced in [3] which corresponds to local occupancy grid, constructed from the sensors of the robot; this method was improved in [4], landmark learning [5], edge detection, graph-based methods [6], Limit-cycles method [7] and many others. However, relatively few of them are suitable for real-time and embedded applications on very low-cost systems. Among them, the methods that use artificial intelligence approaches as:-Neural Network [8]: this approach is applied to determine the optimal neural networks structure for real-time obstacle avoidance task. In [9], the paper proposes a neural network that uses Qlearning reinforcement technique for solving the problem of obstacle avoidance. -Multi Agent System: In [10], a new local collision avoidance algorithm between multiple robots for real-time navigation is presented .This algorithm is based on multi agent system and quadratic optimization method for a collision free navigation and to compute the motion of each robot. -Hybrid Neural Network Genetic Algorithm: [11] used a hybrid neural net- work, genetic algorithm and local search method for solving the problem of finding the optimal collision free path in complex environments for mobile robot. -Particle Swarm Optimization: In [12], the author proposed two extensions of Particle Swarm Optimization (PSO) and Darwinian Particle Swarm Optimization (DPSO) to solve the problem of obstacle avoidance for multi robot.

All these methods require to be implemented on robots with a sufficiently complex functional structure. The robot must possess capabilities of perception, decision, evaluation of the action and a relatively high computational power and/or memory.

This paper mainly deals the navigation control and obstacle avoidance. Our objective is to develop a simple and reactive obstacle avoidance tool that can be implemented on an extremely compact such as the Thymio II [13]. In this context, we propose an intelligent and fast fuzzy controller system for navigating in real-time. The fuzzy logic is certainly one of the most adopted approaches in industry. It addresses such applications perfectly as it resembles human decision making with an ability to generate precise solutions from certain or approximate information. It fills an important gap in engineering design methods left vacant by purely mathematical approaches (e.g. linear control design), and purely logic-based approaches (e.g. expert systems) in system design. The advantage of using fuzzy logic for navigation is that it allows for the easy combination of various behaviors outputs through a command fusion process. The navigation system in this case consists of two behaviors an obstacle avoidance behavior and a goal seeking behavior[14][15].

A set of experimentations is realized to demonstrate the feasibility of this approach for navigation, avoiding

doi:10.4304/jetwi.6.1.101-105

static and dynamic obstacles (other robots) and aggregation (in our case, parking in the edges of environment) for ThymioII mobile robots.

Besides this introduction, the structure of the paper is as follows: Section 2 gives the specifications of the robot platform and its functioning under ASEBA tool. Section 3 presents the description of fuzzy logic controller. Section 4 shows different experimentations for testing the proposed approach. Conclusions and perspective are given in Section5.

II THYMIO II AND ASEBA PRESENTATION

The system consists of a set of ThymioII mobile robots (https://thymio.org) navigating in a dynamic and uncertain environment. The robots must explore the environment and avoid static and dynamic obstacles (other robots).

2.1 Thymio II Robot

Thymio II is an affordable educational robot (<100$). It provides three main features: a large amount of low-cost sensors and actuators,a specific interactivity based on light and touch, aimed at increasing the understanding of the robot functionalities and a very efficient programming environment based on Aseba (https://aseba.wikidot.com/en:start). Its interactivity is based on several functionalities: capacitive touch buttons,color of the body (full RGB spectrum) and LED associated with each robot functionality (https://aseba.wikidot.com/fr:thymioprogram). This robot is representative of low-cost mass-produced systems: A control method applied on the Thymio II robot can be deployed at very low cost: <2$ for the processor, <1$ for each proximity sensor, <1$ per motor.

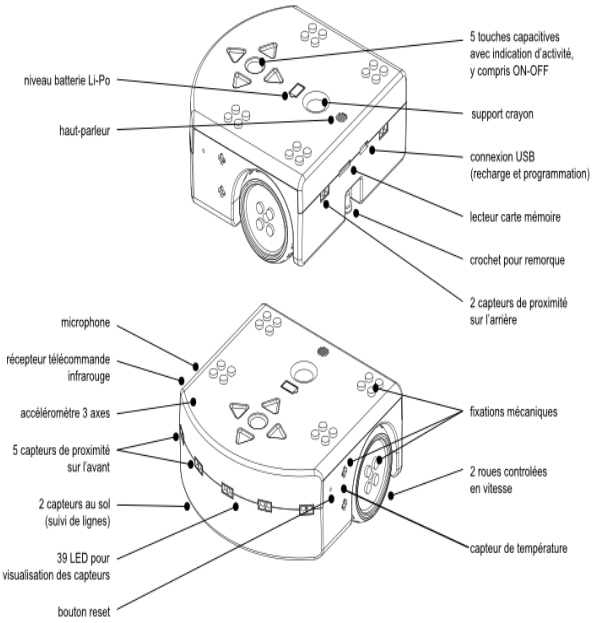


Figure 1. The ThymioII robot Platform and its structural component.

2.2 ASEBA Environment

Aseba is a set of tools which allow novices to program robots easily and efficiently. It is an event based architecture for real-time distributed control of mobile robots. It targets integrated multi-processor robots or groups of single-processor units, real or simulated. The core of Aseba is a lightweight virtual machine tiny enough to run even on microcontrollers. With Aseba, we program robots in a user-friendly scripting language using a cosy integrated development environment. (https://aseba.wikidot.com/en:description).

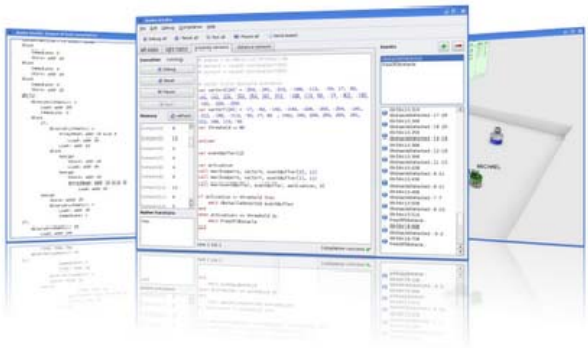


Figure 2. Aseba environnement (http://www.mobsya.org/robot/logiciels).

III NAVIGATION PROPOSED APPROACH

In order to avoid static and dynamic obstacles and permit parking in the edges of the environment by the robot Thymio II, we proposed and developed a reactive system control of navigation using fuzzy logic (Figure 3). The robot reacts according to the changes in environment. The fuzzy system determines the two wheels speed of

ThymioII robot so that it avoids obstacle and go to the reach.

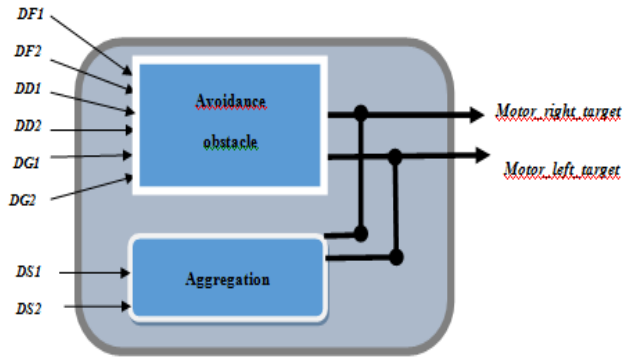


Figure 3. ThymioII robot navigation controller.

The fuzzy control system is composed of two modules based on inputs sensory of the ThymioII robot:

✓ Avoidance obstacle module
✓ Aggregation module.

A fuzzy system that determines the speed of right and left wheels: ***motor_right_target, motor_left_target*** based on the distance of the obstacles around the robot measured by the sensors: **C0, C7, C6, C5, C1, C2** and values given by the two sensors **DS1** and **DS2**.

TABLE 1.
THE MEASURED DISTANCES OBSTACLES-ROBOT

Obstacle direction	Sensors	Distances given
Frontal obstacle	C_0	DF1
	C_7	DF2
Left Obstacle	C_6	DD1
	C_5	DD2
Right Obstacle	C_1	DG1
	C_2	DG2

.

3.1 Avoidance Obstacle Module

In the next section, we describe different stages of the fuzzy control for avoidance obstacle (Fuzzification, fuzzy inference and Deffyzzification).

- **Inputs fuzzification**

We first proceed to the fuzzification of the distances: DF1, DF2, DG1, DG2, DD1, and DD2. The same fuzzy model used for the different measured distances.

These distances are represented by ***proxy_horizontal*** variable which is calculated by Aseba tool.

The membership functions of these distances near and far have been fixed after several tests and experiments.

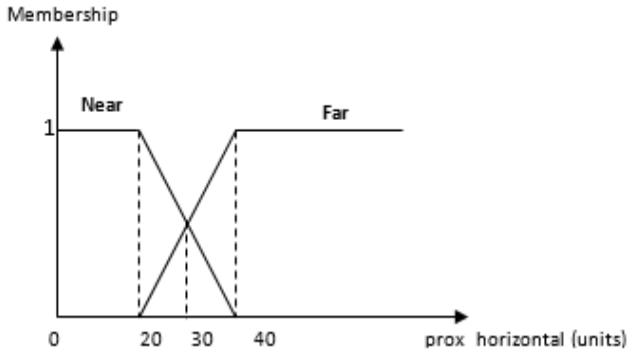


Figure 4. The fuzzy sets of proxy_horizontal variable

The **prox_horizontal** variable is calculated by Aseba tool.

- **Outputs fuzzyfication**

The speed fuzzification of the right and left wheels: **motor_right_target** and **motor_left_target** *is presented in figure 5***.**

The membership functions for this variable are: Dacc (deceleration), Zacc (no speed change) and Acce (acceleration).

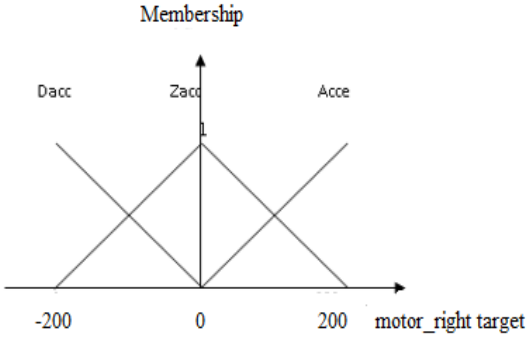


Figure 5. Fuzzy sets of motor_right_target and motor_left_target variable.

- **Fuzzy inference**

This step presents the inference rules elaboration to determine the robot behavior according to its intrinsic parameters. It will be seen later that the number of potential rules of reaction increases directly with the number of labels of the variables.

According to the membership degree of the different distances measured by the robot sensors, each values combination of these input variables, a command presented by the outputs variables (motor_right_target and motor_left_target) is associated with it.

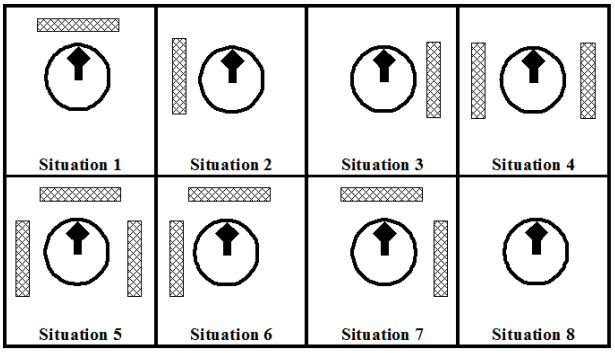


Figure 6. The eight situations of avoidance obstacle

The inference rules definition are based on the following situations (figure 6) which shows the different cases of avoidance obstacles that the robot can undergo during its navigation.

Each situation is defined by a set of inference rules depending on the distances of obstacles-robot. As example, we give the set rules of situation 1: **«Front close"**.

The robot stops and performs randomly one of the following maneuvering:

- Deviation in a right.
- Deviation in a left.
- A half-rotation.

TABLE 2.
RULE INFERENCE OF SITUATION 1 "FRONT-CLOSE".

	Distance sensor	**Fuzzy Value**		**Action**	
If	DF1	Near	**Then**	Deviation in right	
	DF2	Near			
	DD1	Far		Motor_left_target	**200**
	DD2	Far			
	DG1	Far		Motor_right_target	**-200**
	DG2	Far			

2.2.3 Defuzzification

In this section, we present a defuzzification process used in our fuzzy system. The defuzzification step allows the transformation of the output fuzzy values into the corresponding physical values.

We opted for the defuzzification method called "average weighted method". This choice is usually conditioned by a compromise between implementation ease and computing performance [16].

2.2.2 Aggregation Module

The objective of this module is to stop the robot when it is on the environment edges.

To define these edges, we used another color (black) different in the light intensity with the inside area color (white) where the ground sensors of Thymio robot can detect it easily.

The variables measured by these soil sensors are:

- ✓ ***prox.ground.ambiant***: ambient light intensity on the ground, varies between 0 (no light) and 1023 (maximum light).
- ✓ ***prox.ground.reflected***: quantity of light received while the sensor emits infrared, ranging from 0 (no reflection) and 1023 (high light).
- ✓ ***prox.ground.delta***: difference between reflected light and ambient light, related to the distance and color of the soil.

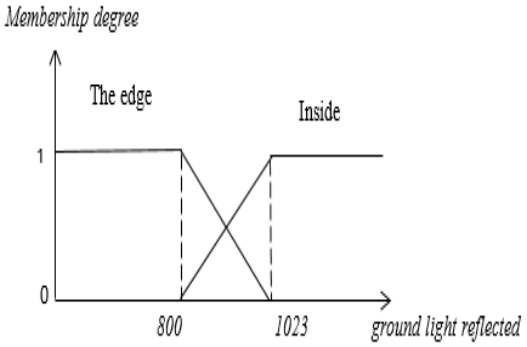


Figure 7. Fuzzy sets of ground light reflected

The variation of the light reflected from the ground is described by two subsets: at edges and Inside.

IV SIMULATION AND RESULTS

In this section, we evaluate the proposed approach efficiency by a set of experimentations on Thymio II robots group.

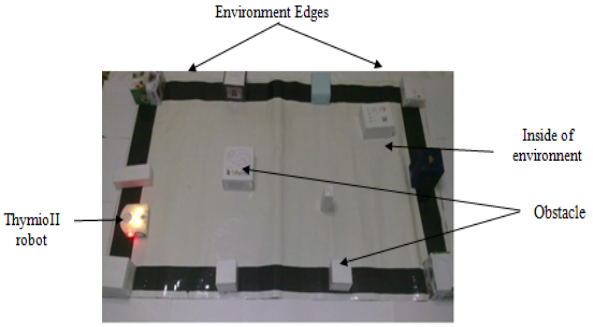
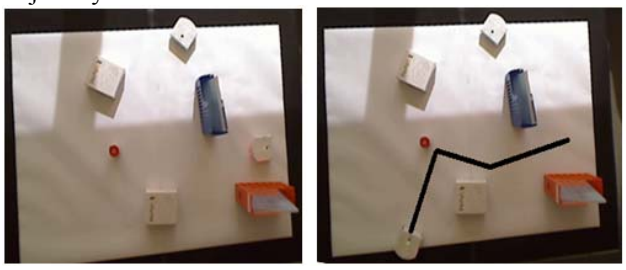


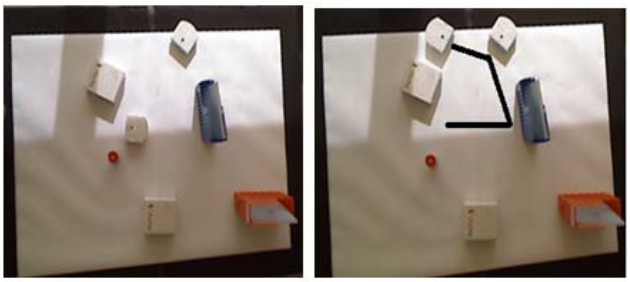
Figure 8. Experiment environment for Thymio II robot.

The experimentation scenes are timed and filmed by several cameras. In Figures 9-10, we show several cases of environments simples and complexes to validate the proposed approach. The bold line represents the trajectory of the robot.



(a) First position (b) Final position

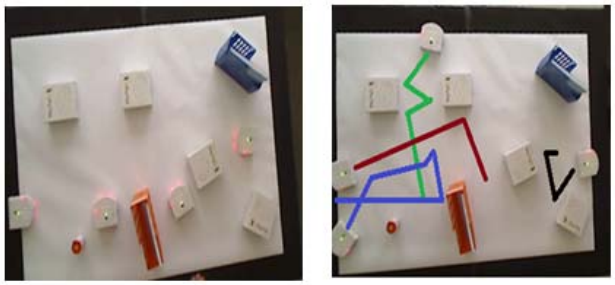
Figure 9. Navigation and avoidance obstacle in the environment 1.



(a) First position (b) Final position

Figure 10. Navigation and avoidance obstacle in the environment 2

.



(a) First position (b) Final position

Figure 11. Navigation and avoidance obstacle in the environment 3.

In these various cases of test we can see that the proposed approach is safe and efficient.

Even if the environment is complex and dynamic (several obstacles and active robots), the ThymioII robot is able to avoiding different static and dynamic obstacles encountered with a flexible navigation and stop at the environment edges.

CONCLUSION

In this article, a proposed solution has been presented to the problem of navigation and obstacle avoidance by developing a fuzzy navigation controller dedicated to ThymioII robots under Aseba environment. The experimental results are satisfactory and validate the proposed approach. The robot navigates autonomously and safe despite the complexity of the environment.

REFERENCES

[1] Adouane, L, and N. LeFort-Piat, 2004, Hybrid Behavioural Control Architecture for the Cooperation of Minimalist Mobile Robots, In Proceedings of the International Conference on Robotics and Automation ICRA04, pp. 3735-3740, New Orleans-USA).

[2] Khatib, O, 1986, “Real time obstacle avoidance for manipulators and mobile robot”, international journal robotics research 5(1) 90-80.

[3] Kren, Y, and J. Borensrein, 1991, “Potential Field Methods and their inherent limitations for mobile robot navigation”, IEEE conference on robotics and automation. 1398- 1404.

[4] Ulrich, B, 1998, “Reliable obstacle avoidance for fast mobile robots”, in IEEE int conf on robotics and automation.Leuven, Belgium.

[5] Meyer, J.A, and D. Filliat, 2003, “Map-based navigation in mobile robots a review of map- learning and path-planning strategies”, Journal of Cognitive Systems Research.4 (4), 283317.

[6] Adouane, L, 2009, “Orbital obstacle avoidance algorithm for reliable and on-line mobile robot navigation”, In Proceedings of the In 9th Conference on Autonomous Robot Systems and Competitions.

[7] Adouane, L, and A. Benzerrouk, and P. Martinet, 2011, “Mobile Robot Navigation in cluttered Environment using Reactive Elliptic Trajectories”, In Proceedings of the 18th IFAC World Congress, Milano-Italy.

[8] Hoang, T and T. Nguyen, 2007, “Obstacle Avoidance for Power Wheelchair Using Bayesian Neural Network”, Proceedings of the 29th Annual International, Conference of the IEEE EMBS, Cit International, Lyon, FranceIn.

[9] Duguleana, M, and F. Barbuceanua, and A.Teirelbarb, and G.Mogana, 2012, “Obstacleavoidance of redundant manipulators usingneuralnetworks based reinforcement learning”, Robotics and Computer-Integrated Manufacturing, Volume 28, Issue 2, Pages 132146

[10] Guy, S, and J.Chhugani and C.Kim and N.Satish and M.Lin and D. Manocha and P.Dubey, 2009, “ClearFootpath: Highly Parallel Collision Avoidance for Multi-Agent Simulation”, Eurographics/ ACM SIGGRAPH Symposium on Computer Animation, E. Grinspun and J. Hodgins (Editors).

[11] Hosseinzadeh, A and H.Izadkhah, 2010, “Evolutionary Approach for Mobile Robot PathPlanning in Complex environment”, IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 4, No 8.

[12] Couceiro, M.S, and R.P. Rocha and N.M.F. Ferreira, 2011, “A novel multi-robot exploration approach based on Particle Swarm Optimization algorithm Safety, Security, and Rescue Robotics (SSRR)”, IEEE International Symposium on.

[13] MagnenatS., RiedoF., Bonani M. and MondadaF, 2012.,“A Programming Workshop using the Robot "Thymio II": The Effect on the Understanding by Children”, The IEEE International Workshop on Advanced Robotics and its Social Impacts, Munich, Germany.

[14] Cherroun. L, r. mechgoug, m. boumehraz. 2011, “path following behavior for an autonomous mobile robot using fuzzy logic and neural networks. courrier du savoir. n°12, octobre, pp.63-70.

[15] Melin.p, o. mendoza, o. Castillo, 2011, “face recognition with an improved interval type-2 fuzzy logic sugeno integral and modular neural networks. IEEE transactions on systems, man, and cybernetics, part a 41(5): 1001-1012.

[16] Lee, and Wang, 1994, “collision avoidance by fuzzy logic control for automated guided vehicle navigation”, journal of robotics system, vol. 11 no.8, pp 743-760

A Survey on Prevention Approaches for Denial of Sleep Attacks in Wireless Networks

G.Mahalakshmi[1], Dr.P.Subathra[2]
[1]Assistant Professor, Department of Information Technology, NPR College of Engineering And Technology, TamilNadu, India.
[2]Professor, Department of Computer Science &Engineering, Kamaraj College of Engineering & Technology, TamilNadu, India.

***Abstract*—Wireless sensor network is vulnerable to various attacks due to the deployment in hostile environment. Among various types of security threats, low power sensor nodes are affected by the attacks that cause random drainage of the energy level of sensors. It leads to the death of nodes. The denial of sleep attack is the most dangerous type of attack in this category. Most of the existing approaches to detect denial of sleep attack involve lot of overhead, which lead to poor throughput. In this survey, different approaches for the detection and prevention of denial of sleep attacks in wireless sensor networks are described.**

***Index Terms*—Anomaly detection, denial of sleep attack, frequency hopping, sleep deprivation.**

I. Introduction

Security of Wireless sensor network (WSN) becomes a significant issue with the fast development of WSN that is exposed to a wide range of attacks due to the limited resources and its deployment in the hostile environment. Intrusion detection system is one of the main and effective defensive approach against attacks in WSN. A particular devastating attack is denial of sleep attack (DS attack). It is a specific type of DoS (denial-of-service) attack. The denial of sleep attacks can be categorized into six categories depending on the attack strategy as

- Sleep deprivation attack
- Barrage attack
- Synchronization attack
- Replay attack
- Collision attack
- Broadcast attack.

In DS attack, the main aim of the intruder is to increase the power consumption of the sensor node. Its battery life is also decreased. Due to this attack, the lifetime of WSN is decreased. The attack accomplishes this by making the sensor node busy and preventing it from going into low power sleep mode. So the energy of the sensor nodes is wasted. Large amount of energy consumption is imposed upon the limited power sensor nodes by this attack. This leads to denial of service through denial of sleep.

The sleep deprivation attack is a severe attack in WSN as replacing or recharging node batteries in WSN is not possible. Barrage attack causes its victims to spend more energy by attacking them with legitimate requests. In both barrage attack and sleep deprivation attack, the victim will never enter its low power sleep mode. The aim of the synchronization attack is to cause relative time synchronization problems at the MAC layer. The synchronization attack is hard to detect because it stays within the confines of the protocol. A replay attack is a breach of security in which data is stored without authorization and then retransmitted in order to trick the receiver into energy exhaust operations. In the broadcast attack, the unauthenticated traffic is broadcasted into the network by the attacker node to reduce the lifetime of sensor nodes.

This type of attack is hard to detect as it does not affect legitimate throughput, which might signify an ongoing network attack. The collision attack can be launched by a compromised node that does not follow the medium access control protocol. Collision is caused with neighbor transmissions by sending a short noise packet. Many existing detection approaches depend on deterministic model. But in sensor network, the status of affected nodes changes with time.

The paper is organized as follows. Section 2 describes prevention approaches for denial of sleep attacks in wireless sensor networks. Section 3 presents results and discussion. Section 4 provides conclusion and future enhancement

II Prevention Approaches for Denial of Sleep Attack

2.1 Absorbing Markov Chain (AMC) Model

A mathematical model based on Absorbing Markov Chain (AMC) [1] was used to detect denial of sleep attack in sensor network. The probabilistic nature of sensor nodes was detected by the AMC model. In this approach, denial of sleep attack is detected by considering the expected death time of sensor network under common scenario. The method evaluate the behavior of compromised sensor nodes depending on the Markov chain with an absorbing state. The Absorbing Markov Chain was used to model the behavior of each and every sensor node. Instead of concentrating on single

doi:10.4304/jetwi.6.1.106-110

sensor node's behavior, the network flow is monitored by the approach to detect intrusion.

In this approach, the expected absorption time of senor network is examined which denotes the network lifetime. If the network state tends to death fast when compared to common death time of sensor network, then the network is affected by denial of sleep attack. The method was more accurate than the deterministic model.

2.2 Hierarchical Framework based on Distributed Collaborative Approach

A hierarchical framework based on distributed collaborative approach [2] was used for the detection of sleep deprivation attack in WSN. The method utilized anomaly detection approach in two steps to minimize the probability of false intrusion and to provide a reliable and energy effective heterogeneous WSN. The values are compared with predefined parameters specified in normal profile to detect the anomaly. The responsibility of each node changes dynamically to minimize the burden of a single node. In order to reduce the attack, the method physically prevents malicious nodes from the network and denies fake packets. Cross layer energy-efficient security mechanism [3] was used for the protection of network from denial of sleep attack. Simulation showed that the approach achieved significant performance in avoiding network nodes from distinct denial of sleep attacks. The cross layer interaction concept was used to prevent sensor nodes from energy exhaust attacks.

The cross layer information i.e. one hop routing table from the network layer was used by MAC layer to identify attackers. If the sender does not belong to the routing path of the receiver node, all received RTS packets are rejected. The Received Signal Strength Indication (RSSIs) of received packets are then computed and compared with RSSI of neighborhood routing node in order to prevent network nodes from malignant denial of sleep attacks such as replaying attacks. The approach is well suited for resource constrained WSNs as it requires low additional cost.

2.3 Lightweight hierarchical model for HWSNET

Heterogeneous wireless sensor network (HWSNET) is more suitable for real life applications when compared to the homogeneous counterpart. Security of HWSNET becomes a significant issue with the fast development of HWSNET. A lightweight, hierarchical model [4] was used for the insomnia detection of sensor nodes affected by sleep deprivation attack in HWSNET. The approach used cluster based method in an energy effective manner in order to build a five layer hierarchical network to increase the network scalability and lifetime. Here the sensor network is partitioned into clusters which are again divided into sectors.

Partitioning the sensor field preserves communication bandwidth and prevents redundant exchange of messages among senor nodes. In this approach, energy efficiency was accomplished by keeping a minimum number of sensors active. A dynamic model was designed to overcome the sudden death of intrusion detection system (IDS) enabled sensor nodes that are responsible for all detection tasks, due to power exhaustion. Anomaly detection technique was used in the approach in such a way that the phantom intrusion detection was avoided.

2.4 Swarm based Defense Approach

A swarm based defense approach [5] for denial of sleep attack utilizes an anomaly detection model to determine the affected traffic between the nodes. Depending on this, frequency hopping approach was initiated. Ant agents of swarm intelligence are then applied to gather the frequency hopping time and communication frequency. The faulty channel is identified depending upon the frequency hopping time. And when the administer node gets this data, it deletes the faulty channel. The simulation results showed that the approach is effective in faulty channel detection. Less energy is consumed as the data about all the attackers can be known by utilizing ants. A framework for preventing denial of sleep attack [6] consists of four key components

- Strong link-layer authentication
- Anti-replay protection
- Jamming identification and mitigation
- Broadcast attack defense

Strong link-layer authentication is the most significant and first component of denial-of-sleep defense and must be included into any WSN that might be exposed to attack.

2.5 Secure Topology Maintenance Protocol (Sec-TMP)

The scalable Secure Topology Maintenance Protocol (Sec-TMP) [7] was resilient to sleep deprivation attacks. Sec-TMP does not require underlying routing and pair-wise node confidentiality. It was highly scalable as the newly deployed nodes were involved in the TMP (topology maintenance protocol) by the pre-existing nodes in the network. It utilizes one-hop communications. A novel approach for detection of sleep deprivation attacks [8] depends on wireless sensor network (WSN) clustering. It includes recursive clustering of sensors till a required granularity is achieved. The approach is applied with two distinct clustering algorithms. Fast and Flexible Unsupervised Clustering Algorithm (FFUCA) was used. To launch the sleep deprivation attack, the adversary nodes become cluster heads.

2.6 Random Vote, Round Robin and Hash-based Scheme

Three clustering approaches for mitigating sleep deprivation attack: the random vote, round robin and hash-based scheme were analyzed [9]. These approaches prevent the adversary from becoming a cluster head and also minimizes the impact of sleep deprivation attack. The random vote scheme randomizes the cluster head selection. The round robin scheme was used to overcome the lack of scalability problem in random vote clustering algorithm. In round robin scheme, clusters were maintained for long periods of time. The round robin scheme consists of two phases

- Bootstrapping phase
- Maintenance phase

Initial clusters were formed in bootstrapping phase. In the maintenance phase, the precise membership of each

cluster is updated due to addition of new nodes, removal of nodes from network and node mobility. The round robin scheme requires only a single iteration for selecting cluster head. But, a list indicating nodes in cluster at all times must be maintained at each sensor node in the round robin scheme. The excessive overhead inherent in the round robin scheme is overcome by the hash-based cluster head selection scheme. Dynamic clustering in an attack and fault tolerant manner is performed without excessive overhead.

Many denial of sleep attacks don't require a constant signal. So it is difficult to recognize the traffic as malicious and to identify the attacking node through its emitted transmissions [10]. The denial of sleep attack focuses MAC protocols. A clever denial of sleep attack keeps the radios of sensor nodes on and their batteries are drained in few days. The denial of sleep attack is mitigated by a framework including authentication at link layer, protection of broadcast attack and tamper resistance.

2.7 Isolation Table Intrusion Detection System (ITIDS)

Isolation tables and routing tables [11] are combined to detect anomalies. Isolation Table Intrusion Detection System (ITIDS) detects malicious nodes depending on attack behaviors. The malicious node is detected through its unusual behavior. The sensor node behaviors are compared with the attack behaviors for determining anomalous information. If the node is anomalous, it is isolated and recorded in isolation table. In ITIDS, sensor nodes of all kinds are concerned by monitoring task and control each other to detect denial of sleep attack. There are four characteristics of ITIDS

- Base station (BS)
- One Primary Cluster Head (PCH)
- Several Secondary Cluster Heads (SCHs)
- The remaining senor nodes are MNs (member nodes)

The approach consists of four stages. Initially, the system predefines IDS. MNs are monitored by SCH. Then PCH is monitored by SCHs and MNs. Lastly, IDS backups the isolation table in base station.

2.8 Ant-based Routing Algorithm

An ant-based routing algorithm [12] was used to detect denial of sleep attacks in WSN. The denial of sleep attacks are detected by using age, energy and reliability as parameters. The impact of distributed denial of service attacks on the performance of WSN is evaluated by using OPNET modeler. Packet authentication is used to prevent the denial of sleep attack [13]. Continuous resetting of sleep timers and link-layer authentication is used to protect from denial of sleep attack.

2.9 Secure Wake up Scheme

A secure wake up scheme [14] activates a sensor node by a secure wake up radio from a sleep state only if messages from legitimate and authenticated nodes are pending. The approach uses a lightweight security verification scheme which can be performed without requiring the change of node to its active state. The network can be protected from the sleep deprivation attack by moving the authentication from application level to physical level. Time synchronized one-time-password scheme [15] provide a secure wake up authentication.

The scheme under denial of sleep attack consume less power than the self-discharge of the batteries. The main idea is to keep the node always in the sleep mode and wake it up only as and when communication is essential. An additional receiver is used that remains in idle mode and uses only small amounts of energy. The communication requests are captured by this receiver. It also wake up the parts of node that are currently in rest in order to receive data.

All incoming requests to the wake up receiver on the physical layer are authenticated to prevent the adversaries targeting to drain the energy. The request that wireless sensor nodes utilize to wake up each other is known as token. The unnecessary traffic is reduced by computing the tokens instead of exchanging them. The counter-synchronized one time password is used for token generation.

2.10 Storm Control Mechanism

Storm control mechanism [16] was used to mitigate flooding and denial of sleep attacks. The frequency of the received packets is tracked by the system. An alert is triggered when it goes beyond a configured limit. The node sends alert to the base station and shuts its wireless transceiver for a predefined period of time. The storm control mechanism was implemented in TinyOS as a security layer incorporated in the communication stack. TOSSIM is used to test the implementation.

2.11 Clustered Adaptive Rate Limiting (CARL) Approach

Clustered Adaptive Rate Limiting (CARL) [17] approach depends on current host-based intrusion detection methods to prevent denial of sleep attacks. It is a rate limiting approach. In this adaptive rate limiting approach, network traffic is restricted when enough malicious packets are sensed in order to suspect that there is presence of attack. It can be also utilized to maintain better throughput and network lifetimes at a time even during sleep deprivation attack. A fake schedule switch scheme with RSSI measurement [18] defends denial of sleep attack. The scheme is implemented in S-MAC protocol. a quickest intrusion detection scheme, markov decision process (MDP) [19] keeps a minimal number of sensors active. The approach ensures energy expenditure for sensing, communication and computation is reduced. It also ensures that the network lifetime is reduced. A secure intrusion detection system [20] was used for denial of sleep attack prevention in WSN.

III RESULTS AND DISCUSSION

The results of the survey are shown in table 1. Various approaches for the prevention of denial of sleep attacks in wireless networks are depicted. AMC model evaluates the behavior of compromised sensor nodes to detect denial of sleep attack. Hierarchical framework based on

distributed collaborative approach uses anomaly detection approach to detect sleep deprivation attack. Swarm based defense approach uses ant agents of swarm intelligence. The lack of scalability problem in the random vote scheme and excessive overhead inherent in round robin scheme is overcome by hash based scheme. FFUCA algorithm was used in recursive clustering of sensors approach. Routing and isolation tables are used in ITIDS technique.

TABLE 1:

TECHNIQUES FOR PREVENTION OF DENIAL OF SLEEP ATTACKS

Authors	Year and reference	Technique	performance
Bhattasali and Chaki	2012 [1]	Absorbing Markov Chain (AMC) model	The behavior of compromised sensor nodes is evaluated based on Markov chain
Bhattasali et al.	2012 [2]	Hierarchical framework based on distributed collaborative approach	Uses anomaly detection approach to detect sleep deprivation attack.
Boubiche et al.	2012 [3]	Cross layer energy-efficient security mechanism	Uses cross layer interaction concept to prevent sensor nodes from energy exhaust attacks.
Bhattasali and Chaki	2011 [4]	Lightweight hierarchical model for HWSNET	Detects insomnia of sensor nodes affected by sleep deprivation attack
Periyanayagi and Sumathy	2013 [5]	Swarm based defense approach	Uses ant agents of swarm intelligence
Gabrielli et al.	2009 [7]	Scalable Secure Topology Maintenance Protocol (Sec-TMP)	Highly scalable and uses one-hop communication
Fouchal et al.	2013 [8]	Recursive clustering of sensors	Uses Fast and Flexible Unsupervised Clustering Algorithm (FFUCA)
Pirretti et al.	2006 [9]	Hash-based scheme	Overcome lack of scalability problem and excessive overhead inherent
Chen et al.	2010 [11]	Isolation Intrusion Detection System for hierarchical WSN	Routing tables and isolation tables are combined to detect anomalies
Juneja et al.	2010 [12]	Ant-based routing algorithm	Uses age, energy and reliability as parameters
Rughinis and Gheorghe	2010 [16]	Storm control mechanism	The storm control mechanism is implemented in TinyOS

CONCLUSION

The paper comprises the survey results in prevention approaches for denial of sleep attack in wireless sensor networks. The survey describes various methods and techniques such as AMC model, cross layer security mechanism, swarm based defense approach, ITIDS, storm control mechanism, ant-based routing algorithm. The results of the survey show that the solutions, which require large scale alterations are unrealistic. As a future work, the prevention of denial of sleep attacks can be performed with minimum changes, cost and resources

REFERENCES

[1] T. Bhattasali and R. Chaki, "AMC Model for Denial of Sleep Attack Detection," *Journal of Recent Research Trends (JRRT),* pp. 1-4, 2012.

[2] T. Bhattasali, R. Chaki, and S. Sanyal, "Sleep Deprivation Attack Detection in Wireless Sensor Network," *arXiv preprint arXiv:1203.0231,* 2012.

[3] D. E. Boubiche, A. Bilami, and S. Athmani, "A Cross Layer Energy Efficient Security Mechanism for Denial of Sleep Attacks on Wireless Sensor Network," in *Networked Digital Technologies*, ed: Springer, 2012, pp. 151-164.

[4] T. Bhattasali and R. Chaki, "Lightweight hierarchical model for HWSNET," *Networking and Internet Architecture,* pp. 1-14, 2011.

[5] S. Periyanayagi and V. Sumathy, "Swarm Based Defense Technique for Denial-of-Sleep Attacks in Wireless Sensor Networks," *International Review on Computers & Software,* vol. 8, 2013.

[6] D. R. Raymond, R. C. Marchany, M. I. Brownfield, and S. F. Midkiff, "Effects of denial-of-sleep attacks on wireless sensor network MAC protocols," *Vehicular Technology, IEEE Transactions on,* vol. 58, pp. 367-380, 2009.

[7] A. Gabrielli, M. Conti, R. Di Pietro, and L. V. Mancini, "Sec-tmp: a secure topology maintenance protocol for event delivery enforcement in wsn," in *Security and Privacy in Communication Networks*, ed: Springer, 2009, pp. 265-284.

[8] S. Fouchal, D. Mansouri, L. Mokdad, and M. Iouallalen, "Recursive - clustering - based approach for denial of service (DoS) attacks in wireless sensors networks," *International Journal of Communication Systems,* 2013.

[9] M. Pirretti, S. Zhu, N. Vijaykrishnan, P. McDaniel, M. Kandemir, and R. Brooks, "The sleep deprivation attack in sensor networks: Analysis and methods of defense," *International Journal of Distributed Sensor Networks,* vol. 2, pp. 267-287, 2006.

[10] D. R. Raymond and S. F. Midkiff, "Denial-of-service in wireless sensor networks: Attacks and defenses," *Pervasive Computing, IEEE,* vol. 7, pp. 74-81, 2008.
[11] R.-C. Chen, C.-F. Hsieh, and Y.-F. Huang, "An isolation intrusion detection system for hierarchical wireless sensor networks," *Journal of Networks,* vol. 5, pp. 335-342, 2010.
[12] D. Juneja, N. Arora, and S. Bansal, "An Ant-Based Routing Algorithm for Detecting Attacks in Wireless Sensor Networks," *International Journal of Computational Intelligence Research,* vol. 6, 2010.
[13] S. Saha, "ZIGBEE OPNET Modeller: An Efficient Performance Analyzer for Wireless Sensor Networks."
[14] R. Falk and H.-J. Hof, "Fighting insomnia: a secure wake-up scheme for wireless sensor networks," in *Emerging Security Information, Systems and Technologies, 2009. SECURWARE'09. Third International Conference on*, 2009, pp. 191-196.
[15] A. Kavoukis and S. Aljareh, "Efficient time synchronized one-time password scheme to provide secure wake-up authentication on wireless sensor networks," *arXiv preprint arXiv:1302.1756,* 2013.
[16] R. Rughinis and L. Gheorghe, "Storm control mechanism in wireless sensor networks," in *Roedunet International Conference (RoEduNet), 2010 9th*, 2010, pp. 430-435.
[17] D. R. Raymond and S. F. Midkiff, "Clustered adaptive rate limiting: Defeating denial-of-sleep attacks in wireless sensor networks," in *Military Communications Conference, 2007. MILCOM 2007. IEEE*, 2007, pp. 1-7.
[18] C. Chen, L. Hui, Q. Pei, L. Ning, and P. Qingquan, "An Effective Scheme for Defending Denial-of-Sleep Attack in Wireless Sensor Networks," in *Information Assurance and Security, 2009. IAS'09. Fifth International Conference on*, 2009, pp. 446-449.
[19] K. Premkumar and A. Kumar, "Optimal sleep-wake scheduling for quickest intrusion detection using sensor networks," in *Proc. IEEE Infocom*, 2008.
[20] P. Sharma, N. Sharma, and R. Singh, "A Secure Intrusion detection system against DDOS attack in Wireless Mobile Ad-hoc Network," *International Journal of Computer Applications,* vol. 41, pp. 16-21, 2012.

A Comprehensive Study of Finding Copy-and-Paste Clones from Program Source Codes

Kamran Khan,
Shaheed Zulfiqar Ali Bhutto
Institute of Science and Technology
Islamabad, Pakistan
Kamran_3388@yahoo.com

Saif Ur Rehman, Kamran Aziz
CORDE: Center of Research in Data
Engineering, MAJU University,
Islamabad, Pakisan
Saifi.ur.rehman@gmail.com,
kamrandik@gmail.com

Simon Fong
Department of Computer and
Information Science
University of Macau
Macau SAR
ccfong@umac.mo

***Abstract*—In any programming language source code, the code that is repeated is called the clone. The clone detections have got much attention in the recent years. In literature there are a number of clone detection techniques have been proposed. These techniques includes CP-Miner, CC-Finder etc. each of these techniques attempts to detect the clone from the source code of various programming languages. In this study, we will provide comprehensive details of the various clone detection techniques proposed so far. These techniques have been critically evaluated based on a no of efficiency measure parameters. In our future work we will propose our own clone detection technique that will more efficient and accurate in terms of code clone detection from multiple programming languages.**

I. INTRODUCTION

In software development, programmer often uses copy-and-paste technique. The aim of copy-and-paste technique is save efforts in manually typing over the codes again in computer program software. It is mandatory to detect and remove such clones. In the past, many techniques have been proposed. For example, [1], [2] use copy-and-paste detection tool for detecting code clones.The main issue associated to clones in programming languages is that copy-and-paste introduces bugs in programming code due to forgetting to change identifiers each time right through the code that is pasted from some somewhere else [2].

There are many issues associated with copy-and-paste source code when the size of the code get bigger, furthermore handling these issues are even greater challenge. A bug in one module is reproduced in every copy [3]. As many of the copy-and-paste codes are not documented which provides that which part of the code repeated in which parts, it is extremely hard to find and fix such programming bugs. These bugs are the main source of issues related to maintenance of existing software and removing such bugs are more complex and costly. Moreover, understanding and reusing such code is also a challenge for programmers which reduce the level of abstraction and adding new functionality to the code [3].

Different research studies have already been done to identify the duplicates in software applications [12]. However, these techniques have limitations regarding the support for certain programming languages. In literature different tests have been performed on known tools and techniques for clone detection but the results reveals that there is none good approach that produce efficient and optimal output.

In software engineering the topic of clone detection has received much attention in last decades. In literature several methods for clone detection have been proposed. These techniques are widely used in software domain[8] the existing clone detection techniques focuses on finding similar codes in source code, known as clone, which resulted in reduced update issues and application size. These gains, however, can be improved by evaluating the level of clone analysis [4]. Previous studies showed that these gains can be detect design level similarities which can aid to the software design in terms of code optimization and understanding the design of software.

The rest of the paper is organized as follows: After presenting some basic definitions about clones in Section 2, we present some summaries and strong points and weak points of clone detection techniques. In Section 3, we present a different comparison of these clone detection techniques have been critically evaluated.

II. LITERATURE REVIEW

In [1] Kamiya *et al.* have proposed a technique for detecting clones in large source codes. Their technique is called CC-Finder. The CC-Finder is based on the following elements: (1)-transformation rules; (2) a token-based comparison and (3) optimization techniques. These three elements attempt to improve the performance and efficiency of clone detection in the source code. The CC-Finder works as follow. In first step, the source code is divided into tokens. Afterwards, all these tokens are concatenated into a single token file. Then clones detection is performed on this single file. During the token analysis process, the white spaces are removed between the tokens and these characters are sent to the formatting step. In the second step, the token sequence is transformed using transformation rules. In this step the

doi:10.4304/jetwi.6.1.111-118

special characters, operators and name spaces are discarded. In the third step, following the second step transformation the same pairs are detected as clone pair [1]. In the fourth and last step the proposed method locates the copy-and-paste and converted them into the line numbers on the original source code. Afterword [1] is implemented in C++ language. The results showed that [1] can extract the clone from different languages source code including C, C++, Java and COBOL. Although, CC-Finder is very efficient for clone detection but this technique is failed to detect the clones that come from two different programming language source codes.

In [2] Li *et al.* have proposed another useful tool for clone detection. The method proposed in [2] applies techniques of data mining to identify copy-and-pasted code in a huge source code. The proposed technique is capable to find operating system associated bugs [2]. The proposed approach has two main functionalities (1). Detecting copy-and-paste code segments (2) finding copy-and-paste related bugs. It works as follow: In step reduce the development time. A software developer in his development career frequently uses copy-and-paste and they use the same code again and again. Copy-and-paste method reduces programming effort and time therefore programmer uses copy-and-paste rather than writing new code from scratch. In literature a large number of techniques have been developed for detecting duplicated code in

No.1 the proposed approach first change the problem into frequent subsequence mining problem [2]. The author used CloSpan algorithm to detect the basic copy-and-paste segments. In this step the proposed approach detects copy-and-paste very competently. To further accelerate the process the [2] eliminate unnecessary comparison by using frequent subsequence mining method. In second step, the proposed approach found code clones which were the main source of bugs. Furthermore to evaluate the efficiency of the [2] different experiments were performed using [2] the results showed that the execution time of the proposed technique is optimal, proposed approach took 11-12 minutes to find 101,699-198,605 code clones in a source code segment in a Linux. Further [2] was compared with [1], the comparisons established that execution of [1] is similar to [2] but the [2] is more efficient and detects much more copy-and-paste segments. So there work is supported by very good explanatory examples. The currently developed tool is only work for programs written in C or C++ and their tool detects only simple cases of errors which is not complicated.

In [3] Wahler *et al.* have proposed a new method for detecting clones. Their technique is based on frequent itemset. The proposed technique works as follow; in first step, it takes source code from different compilers as an input, the [3] then creates a consistent code from the input using corresponding parser. In the next step, it uses the popular frequent itemset finding technique to generate most frequent itemset from the XML file. In the last step, it removes or detects the itemset as the duplicate code. By using JDK, the authors have implemented their approach. Although, the experimental results shows that it can efficient and accurate to detect the clones from the source code; but this detection is limited to clone of type 2.

In [4] Basit *et al.* have proposed a method for detecting or finding the clones in a source file. Their technique is based on token-base technique for finding (a) simple clones (b) finds co-occurring clones in a different files by using a frequent itemset mining and it perform file clustering to find clusters that are similar. The execution process of [4] is as follow; In first step, the input file is converted into tokens and the efficient suffix array based algorithm is used to find the repeated tokens. The first step provides a data in a suitable format for second step. In the second step, [4] detect clones that occur together and frequently in a different source by using a frequent itemset technique. In this step they found some un-significant files coverage, for these un-significant files coverage they used third step (clustering highly clone method). The proposed approach is implemented using the C++ compiler for efficiency measure; they used the Java source file as an input to find the clones from it. Although, this technique can find clone but this detection is limited to just one programming language.

In [5] hummel *et al.* have proposed a new technique for finding clones. Their method is known as “Novel Index Based”. Their technique is used for both incremental and scalable to a very large source code. The incremental based method consists of three steps. In a first step the source code is read from a disk and converted into tokens. So the result of this step is the list of normalized statements for each file. In the second step, it finds the global statements list for equal sub strings. The result of this step is cloning information on the level of statement sequences. In the last step this technique creates cloning information on the level of code regions from cloning information on the level of normalized statements. Their clone index approach is similar to the inverted index used in document retrieval system. This method is not only used for the retrieval of all the clones enclosed in a source file but it is also useful for the efficient retrieval of clones from the source file. They use their method in a distributed environment across different machines for the creation of index and retrieval of clones. In a distributed environment they experiment on Java, C and C++ source codes. They have showed the efficiency of their work experimentally.

In [6] Baxter *et al.* proposes the Abstract Syntax Tree (AST) to detect the clones in the source code. Their proposed approach is the simple one as compared to the rest of the clone detecting techniques available in the literature. This technique is efficient and can detect the clones accurately from the code as compare to other clone detection techniques which only focus on either the string matches or near misses only on the body of the underlying functions. The technique proposed in [6] first of all the source code is parsed and from this parsed code an Abstract Syntax tree is produced. Then a set of three algorithms have been applied on Sub-Tree Clones. In this step the similarity is measured between the sub trees. This similarity is measured using the formula given

below in Eq.1. This AST is used to find the clones from the source code. The first algorithm is called Finding

$$Similarity = 2 \times S / (2 \times S + L + R) - \quad Eq.1 \quad [1]$$

Where, S is the number of shared nodes [6], L is the number of different nodes in sub-tree 1; R is the number of different nodes in sub-tree 2 [6]. The [6] applied their proposed technique on real world software systems, which confirms that this new technique [6] can generated more accurate clone detection results as compared to the other clone detection techniques. This technique has following strong points. (1) This technique is straightforward for detecting the clone in the source code. (2) It is more efficient as compare to it is proved form the experimental results [1]. (3) It will defiantly open the new direction for the detection of the clone code in the source code of the program. Although [6] has given new dimensions to the clone detection mechanism but the technique it uses, AST uses a thresholds value which eliminates small trees comparisons but the proposed method failed to find the close clones such that in which one clone instance is small, and the other is large [6].

In [7] Falke *et al.* have proposed a method called "Abstract syntax Suffix Tree". Their method is used to find clones in based on [6]. The approach proposed in [7] is used to find syntactic code clones in an optimal manner. Their method is very efficient especially in token base clone detection as suffix tree token base clone detection is very fast. Basically suffix tree is originally used for efficient string searching. The suffix tree represents a string where every suffix is shown through a path from root to a leaf and the edges are labeled with substrings [7]. Comparison between token base ad AST base shows that suffix-tree-based study offers a lot of advantages over other methods. Their comparison shows that token base clone detector is familiar to a new language in a very small time. The algorithm of linear time is consists of the following steps; In a first step, the source code is parsed and from this parsed code an Abstract Syntax tree is produced. Then in the 2nd step they serialize the AST nodes by a preorder traversal [7]. In the 3rd step each AST node represents a token and suffix tree clone detection is based on token. In this step the actual value of string is not disturbed when they are represented as a node. The output of this step is set of clone classes. The clone classes are consisting of AST node sequences and these sequences may or may not be syntactic clones [7]. Therefore in last and 4th step these sequences are decomposed into syntactic clones. They compare their method with 9 other techniques which is a plus point of their technique. Their method is good to find syntactic clones in a source but their technique is less efficient.

In [8] Jablonski *et al.* have proposed a new approach for detecting a copy-and-paste clones and changing the identifiers name in an integrated development environment. This technique is called a CReN. CReN finds the code clones which is occurred during the copy-and-paste in the integrated development environment (IDE) and the proposed method in [8] uses set of rules which are based on the identifiers relationship in the code fragments. Their tool CReN is implemented as an eclipse plug-in in Java [8]. They have performed experiment on the source code which is written in a Java language. Then they apply the proposed tool to find the copy-and-paste in the input file, the proposed tool is a set of identifiers placed in a code fragment and map the identifiers pairs which are placed in a same location. Their tool uses AST API of eclipse JDK framework [8] and this AST allows the proposed tool is used to create connection in clone code. The [8] tool is also used for renaming the identifiers just within copy-and-paste fragments. Experimentally their tool is so good for clone detection but it is used only for Java source code.

In [9] Uchida *et al.* have presented, the broad analysis of a code clones and for this purpose they use 125 packages of open source which is written in a C language. For analysis they use a CC-Finder technique to determine the code clones and evaluate them statistically. They also use a clone warrior tool for code clone identification for identifying and classifying the code clones and to examine the causes for their production. For token base clone detection they use CC-Finder method which have an industrial potency and applicable to a million lines size of source c ode. For detecting a code clones using a CC-Finder method they follow the following steps; In the first step, the input files are converted into tokens based on the lexical analyzer rules of the programming language [9]. In this step, the proposed method prunes white spaces and remarks or comments. In the next step, data types, variables and constants are replaced by the same respective tokens [9]. This replacement is useful for identifying a code clones as pair of code lines where only the variables names are differ. In the last step, all the substrings which are transferred as token sequences; a pair of identical substring is detected as clone pair. So, for visualization of a code clone they used a clone warrior tool which consists of a graphical user interface (GUI). Using clone warrior tool, first they specify the criteria for clone detection i.e. the smallest amount of a fragments. In next they specify the input files which have to be analyzed and after all the clones are detected from the input file and stored in code clone storage and provide the view in three forms i.e. the file list view, the code clone list view and the source code editor. This work is supported by very good explanatory examples. So their experiment works fine but it is applicable only to a source file which is written in a C Language.

In [10] Jia *et al.* have proposed a new technique for the precise and efficient clone identification. This approach is called Kclone. It combines lexical and local dependencies analysis in order to generate precise clones from the data source without affecting the clone deduction speed. This technique works in three steps; first of all it converts the code into an internal representation. After transformation it identifies those parts that denote clones and finally it combines the clone pairs into clone classes. For further details readers are referred to read [10]. Efficiency of this technique is supported by experiment which performs

clones deduction in various programming code including C, C++ and Java. The author have concluded that it is fast and requires less memory as compare to other clone detection techniques. Although it can find type 3 clones from the source code but it is less efficient for type 2 clones.

Shinobi [11] is another useful and novel approach for the clone's detection and then the modification of the code clones. This technique was proposed by Kawaguchi *et al.* in 2oo9. The main features of this technique are to (1). Identify the clones from the source code and (2). Highlight these identified clones segments. This technique is very tightly integrated with Microsoft visual studio. It works in such a simple way that as it detects the code clones, it warns the developers about the clones, so that the developer can get read of it. Shinobi is a token-base detection method. It takes the source code automatically and then detects the clones. The source code may be from a CSV file or source file directory. This approach is very useful for clone detection while a developer is working on the software maintenance. The detail procedure of Shinobi is available in [11]. This work is supported by very strong arguments. These are (1) It can be implemented as add-in in Microsoft visual studio. (2) Its clone detection process is very fast and automatic. (3) It is very useful in clone detection while a developer is working in a software maintenance phase. Although, this technique is simple, fast and accurate it does not provide any kind of compatibility with other development environment like Java Netbeans, Oracle edition etc.

There are two types of code defects that may exist in any of the source code file. These two issues are rule violation defects and copy-and-paste problems. Zhang [12] *et al.* have suggested a model that copes with these two types of source defects. The authors have used the data mining technique for their approach. This technique is frequent pattern mining. It is not an easy task to detect the programming defects. The proposed approach is tested with a C source file having 4 million lines of code. This approach is implemented in C++ and Python. They perform various experiments. The results of these experiments clearly show the efficiency and accuracy of this technique. Since this approach can detects maximum programming defects in one phase that is why testing time is very much optimized with this method. The authors have implemented the data mining techniques effectively to identify and modify the programming defects. Although this technique works fine, they have not compared their experiment results with other techniques, including [2] and [1] etc.

III. Critical Evaluation

This section primarily reflects the comparison and contrast of the above reviewed literature regarding the different clone detection techniques. It identifies the similarities and differences among the various research works on the clone detection algorithms. The critical review is given in the Table-1(A), Table-1(B) and Table-1(C), below. This will help for the future research in the clone identification and deletion of the clone data.

The clone detection technique in [1] provided sufficient experimental details of the methods with implementation in C++ but the method [1] can detect clones only in one language. In contrary the work [2] also provided detailed experiments and with C++ implementation but the method [2] can detect clones in C and C++ source codes. The clone detection approach for java based source code is provided in. The authors in [4] and [8] both have provided with sufficient details but [8] is more accurate as it implements consist renaming approach in detected clone.

Researchers in [6], [7], [8], [9] and [10] proposed approaches for detection clones written in C language, these have provided sufficient details and the results are presented in tabular as well as graphically format except [8] and [9]. The author [6] in his future work planned to provide method which is more efficient and removal of detected clones. In contrary authors from [8] and [9] suggested in their future work to extend their approaches for detecting clones in source code written in more than one language, but the researcher in [10] has not discussed any future work. To detect clones written in .Net language the work [11] provided sufficient details but the future work is not discussed. The work [12] is Python based detection tool.

IV. Conclusion and Future Work

In this study, we have presented the summary information of the different clone detection techniques. These clone detection techniques are based on the CC-Finder, CP-Miner, Abstract Syntax Tree, and Frequent Itemset Technique. In addition, we also have highlighted the research contributions and found out some limitations in different research works. Consequently, this work also depicts the critical evaluation in which comparison and contrast have been taken out to show the similarities and differences among different author's works. The spatiality of this work is that it reveals the literature review of different clone detection techniques and provides a vast amount of information under a single paper. In our future work, we have planned to propose our own technique based clone detection technique, and provide its implementation and compare its results with the different existing clone detection algorithms.

References

[1] T. Kamiya, S. Kusumoto and K. Inoue, "CCFinder: A Multilinguistic Token-Based Code Clone Detection System for Large Scale Source Code", IEEE, 2002.

[2] Z. Li, S. Lu, S. Myagmar and Y. Zhou,"CP-Miner: Finding Copy-and-paste and Related Bugs in Large-Scale Software Code", IEEE, 2006.

[3] V. Wahler, D. Seipel, J. Wolff, V. Gudenberg and G. Fischer,"Clone Detection in Source Code by Frequent Itemset Technique".

[4] H. A. Basit and S. Jarzabek,"Detecting Higher-level Similarity Patterns in Programs", ACM, 2005.

[5] B. Hummel, E. Juergens, L. Heinemann and M. Conradt,"Index-Based Code Clone Detection: Incremental, Distributed, Scalable".

[6] I. Baxter, A.Yahin, L. moura, M. S. Anna and L. Bier,"Clone Detection Using Abstract Syntax Trees"ICSM, 2005, p3.
[7] R. Falke, P. Frenzel and R. Koschke, "Clone Detection Using Abstract Syntax suffix Trees".
[8] P. Jablonski and D. Hou," CReN*: A Tool for Tracking Copy-and-Paste Code Clones and Renaming Identifiers Consistently in the IDE", ACM, 2007.
[9] S. Uchida,T. Kamiya,A. Monden,K. Matsumoto,N. Ohsugi and H. Kudo,"SOFTWARE ANALYSIS BY CODE CLONES IN OPEN SOURCE SOFTWARE",2005.
[10] Y. Jia, D. Binkley, M. Harman, J. Krinke and Makoto Matsushita, "KClone: A Proposed Approach to Fast Precise Code Clone Detection", 2009.
[11] S. Kawaguchi, T. Yamashina, H. Uwano, K. Fushida, Y. Kamei, M. Nagura and H. Lida, "SHINOBI: A Tool for Automatic Code Clone Detection in the IDE", 2009.
[12] Y. Zhang, Y. Liu and L. Zhang, "A data mining based method: Detecting software defects in source code", IEEE, 2010.

TABLE-1
CRITICAL EVALUATION OF VARIOUS CLONE DETECTION TECHNIQUES

Authors	Features	Experiment Detail	Methodology	Complexity	Future Work
Kamiya [1]	• It combines lexical and dependencies analysis in order to generate precise clones from the data.	• Sufficient experimental details are provided • They used data from multiple domains with various data size	Implementation (in C++)		They are trying to expand their tool to detect clone detection in multiple languages.
Z. Li [2] et al	Their approach uses data mining techniques to competently detect copy-and-pasted code in large software and CP-Miner is very efficient as compare to other clone detection methods like CC-Finder etc.	• They provide sufficient experiment details. • They used data in different operating systems.	• Implementation (in C and C++)		Not discussed
Wahler[3] et al	Their new approach is used to detect clones of types 1 and 2.	• They provide sufficient experiment detail • They used different packages of Java for clone detection.	• Implementation (in Java)		They are tried to extend their algorithm to detect clones of type 3.
Basit[4] *et al.*	• Their technique is based on token-base methodology for finding (a) simple clones (b) Finds co-occurring clones in a different files by using a frequent itemset mining method.	• • They provide experiment results. • They used java source code for analysis.	• Implementation (in C++)	•	• They discuss about a demanding feature that would be able to provide multiple language ability within the same system.

hummel[5] *et al.*	• In this method source code is read from a disk and converted into tokens. • This method is useful for the efficient retrieval of clones from the source file.	• They provide their results experimentally.	Implementation(in C++)		They plan to develop algorithms that detect the clone of type 3 and also plan to extend their index base tool for plagiarism detection in text documents.
Baxter[6] *et al*	Their proposed approach AST is the simple one as compared to the rest of the clone detecting techniques and efficient and can detect the accurate clones from the code as compare other clone detection techniques	• Experimental details are given • These experimental details are provided in tabular as well as in graphically	implementation(in C)		To automate the removal of code clones detected in a source file [6]. Increasing performance of the proposed method.
Falke[7] *et al*	Their method is used to find clones in abstract syntax trees. Their method is very efficient and fast especially in token base clone detection for type-2.	• Experimental details are given • These experimental details are provided in tabular as well as in graphically	Implementation (in C)		To improve and uses a token counts instead of lines during the measurement of clone size.
Jablonski[8] *et al*	Their proposed tool is implemented as an eclipse plug-in in Java and their tool is also used for renaming the identifiers just within copy-and-paste fragments	• Experimental details are given in tabular form	Implementation (in Java)		To generalize the proposed tool and provide consistent renaming support.
Uchida [9] *et al*	They use [1] tool for detection of code clones and for identification they use a clone warrior tool.	• Experimental details are given in tabular form.	Implementation (in C)		To perform analysis on multiple languages.

Jia[10] *et al*	The proposed approach is fast and requires less memory as compare to other code clones detecting tools.	• Experimental details are given • These experimental details are provided in tabular as well as in graphically	Implementation (in C)		Not Discussed
Kawaguchi [11] *et al*	The proposed tool is token base clone detection and used for identifying and Highlight these clones. This method integrated in Microsoft visual studio.	Experimental details are provided graphically	Implementation (in .net)		Not Discussed
Zhang [12] *et al*	The proposed approach is used to detect maximum programming defects in one phase that is why testing time is very much optimized with this method.	• Experimental details are given • These experimental details are provided in tabular as well as in graphically	Implementation (in C++ and Python)		To compare the proposed approach with other code clones detection tools like [1] and [2] etc.

An Optimized Parallel Computing Paradigm for Mobile Grids Based on DSPOM

Aghila Rajagopal*, M.A. Maluk Mohamed**
Software System Group, MAM College of Engineering, Affiliated to Anna University Chennai.
*E-mail: ssg_akila@mamce.org
**E-mail: ssg_malukmd@mamce.org

***Abstract*—Parallel computing methods decrease the processing time in mobile distributed systems compared to the conventional sequential computing techniques. But as they are developed from smaller mobile clusters to extensive mobile grids, they are prone to issues like high latency/jitter, processing speed, communication overhead, and low data transfer rate. So, an efficient and optimized parallel computing paradigm known as Distributed Shared Proxy Object Model (DSPOM) is developed based on Surrogate Object Model (SOM) integrated with Distributed Shared Object (DSO) for mobile grid. SOM is chosen to enhance the resource sharing of mobile grid computing, while DSO is chosen to reduce the computational complexity. The unused computing determinant is utilized by SOM to save the processing time. The transparency of the DSO model in terms of distribution and heterogeneity reduces the computational complexity. DSO also enhances the load adaptability and fault-tolerance to parallel programs on the mobile grid. The DSPO model performs better in terms of query time, query latency, packet loss, load adaptability, and fault-tolerance.**

***Index Terms*—Control Object, Distributed Shared Proxy Object (DSPO), Mobile Host (MH), Peer-to-Peer (P2P), Proxy Object (PO), and Surrogate Object Model (SOM).**

I. Introduction

PARALLEL computing is an important aspect in mobile distributed systems to reduce fault-tolerance, computational time and overhead. It assigns the computations in an adaptive manner according to the work load. When parallel computing techniques are developed from smaller mobile clusters to extensive mobile grids, they are prone to issues like high latency/jitter, processing speed, communication overhead, and low data transfer rate [1].

Parallel and distributed mobile computing fully utilizes the computing power in mobile systems. These systems must ideally offer flexible communication, unlimited computing capacity and higher availability of the mobile distributed information [2]. The computing resources of a distributed mobile system are combined to work in a common system, forming mobile clusters. A group of mobile clusters forms a mobile grid, which can be defined further as a mobile cluster of mobile clusters.

When MHs form a part of the mobile grid, they can be represented as both 'consumer' and 'provider' of services and other data resources. In mobile grids the tasks are computationally complex and need to be distributed across multiple hosts. This requires collaborative processes between distributed tasks. When MHs are provider of services and resources, the user needs to be confirmed that the service can complete the assigned task. Parallel computing on mobile distributed systems decreases the computation time, but also brings forth some issues like load variation, limited node availability, and heterogeneity in processor speed, network speed, operating system (OS), and architecture.

A mobile distributed system consisting of interconnected mobile clusters handles a vast load fluctuation on individual mobile nodes. The unused network computing capacity must be utilized in every way for maximum efficiency. The programs executed on the mobile grid should be *load adaptive*. Conventionally, the communication processes are executed as a collection of processes (COP). Using the COP model, a programmer will not be able to recognize the varying network load patterns. The initiator node gets heavily loaded during the program execution, which decreases the performance of the system.

Generally, mobile distributed systems involve varying numbers of active and inactive mobile nodes over a period of time. This should not affect the computation time, but there would be a significant change with larger differences in the number of mobile nodes [3]. When two different program instances use different number of mobile nodes, the occurrence of node failures is less relative to a single program execution. When a mobile node or a connection crashes during a subtask computation, the mobile node may be activated before the completion of the program execution.

The variations in the network speeds and processor speeds results in the higher communication overhead of the mobile distributed systems. This makes them suitable only for common parallelism. When the communication overhead is increased, the speedup of parallel processes is decreased. The processors in a heterogeneous cluster have varying speeds and the effective capacity may vary due to severe loading. Thus, appropriate grain sizes must be selected during the runtime as an effect of load variation. So, the programming paradigm must be designed for flexible task grain size.

The interconnected mobile clusters possess heterogeneity in system architectures and operating clusters. This issue is fixed by mobile distributed operating systems [4] and mobile distributed file systems [5]. The

doi:10.4304/jetwi.6.1.119-132

heterogeneity in system architectures has more impact than that of operating systems.

Some of the existing methods to solve the issues in parallel computing are discussed. Parallel and distributed computing has been performed using methods like DP [6], Moset [7], ADM [8], Sprite [9], EMPS [10], Condor [11], Piranha [12], Batrun [13], Comet [14], Imapreduce [15] and NOW [16]. Mobile telemedicine is a recent application of parallel and distributed computing [17]. A general telemedicine system consists of a small group of hospitals which can provide remote healthcare services. But, in developing nations there is a requirement for larger Internet-based telemedicine systems due to the large population is in the rural areas. A variable Internet-based P2P architecture is applied in these telemedicine networks. The principle of this system is based on a distributed context-aware scheduler and a store and forward model.

A transparent programming model can be used for the parallel communication in a grid [18]. The GDP (Distributed Pipes with grid abstraction) model performs the inter-process communication between machines. This permits the random migration of parallel tasks corresponding to the grid dynamics. It can support both sequential load and parallel load. But this model does not provide support for mobile systems.

The mobility of the mobile hosts (MHs) may lead to message loss in a distributed mobile computing environment [19]. The battery power can be greatly saved when the message delivery is guaranteed exactly once. The limited resources of MHs involved in the mobile grid need an efficient communication protocol for delivering the messages exactly once and avoiding further retransmissions. So, an exactly once multicast protocol (EOMP) is used to enhance the power efficiency. An unreliable wireless MAC layer multicast is used to transmit the messages to the MHs. This protocol tolerates the system failures at the mobile support station (MSS) by switching the MSS as stateless.

Computational mobile grid can also be regarded as an integration of mobile clusters [20]. Mobile cluster computing can also be designed using IPv6 protocol [21]. An iterative grid-based application for parallel and distributed computing is designed in a distributed solution environment [22]. The environment is based on mobile agent systems.

The process interactions in distributed computations are studied in [23]. They are programs whose communications depend on the transmission of messages. These programs generally execute on network architectures such as distributed parallel machines or NOW (Network of Workstations) [24]. Few example models for process interaction in distributed computations are heartbeat algorithms, network of filters, decentralized servers, broadcast algorithms, bag of tasks, token-passing algorithms, and echo/probe algorithms. These models consist of modules such as, computing network topology, parallel sorting, and termination detection.

Distributed computing can also be tested using tools such as *OptimalGrid* which is a middleware pattern for the computation of larger problems in distributed computing applications [25]. OptimalGrid comprises of an automatic problem partitioning, runtime management, problem deployment, and dynamic redeployment. CADP 2011 tool is another tool used for the construction and analysis of distributed processes [26]. *Grumbach and Wang* designed a rule-based language for distributed programming [27].

A cost-effective computing was proposed for the cloud infrastructure in [28]. This method utilized dynamic resource management to provide the advantages to the cloud infrastructure services. An integrated architecture was proposed to enable migration of virtual machines and live resource scaling. *Shanmugam* and *Mohamed* proposed a data management scheme for the mobile cloud using surrogate object [29]. A Surrogate Object based Cloud Caching (SOCC) procedure was proposed to provide self-healing ability. The data required for the transfers between several surrogate objects over the cloud were saved and the customer requirements were analyzed. Gang scheduling algorithm was used to schedule the parallel tasks on the cloud [30]. The surrogate model was evolved for a transaction management scheme in mobile cloud [31]. The mobile nodes in the cloud transmit the transaction request to its proximate surrogate object. The surrogate objects enhanced the network lifetime by checking its data cache for the execution of the transactions. The fault tolerance and reliability were enhanced by transferring the transaction request to surrogate object, while the network lifetime was increased by transferring the surrogate objects to the least loaded base stations.

An efficient and optimized mobile grid parallel computing paradigm known as Distributed Shared Proxy Object Model (DSPOM) is developed based on Surrogate Object Model (SOM) [32] and Distributed Shared Object (DSO) [33]. Grid computing is integrated with service composition procedure to enhance the accessibility and computational capability of mobile distributed systems. The object model comprises of a combination of middleware solutions, resource-sharing solutions, and wireless resource access. The mobile distributed system involving the computational and data resources is modeled as a Peer-to-Peer (P2P) model [34]. Service composition is enabled by virtualizing the data resources as services.

Each mobile host (MH) in the wired network is represented by a proxy object that uses application defined data structures and methods. These MHs have limited resources under high mobility. The proxy objects contain a data cache in the MH and decrease the wireless data transfers [32]. The location management issue is solved by providing a unique agent for MH location details.

The optimized parallel computing on the mobile grid is performed by the distributing and sharing the states of the local objects [33]. Here, the local objects are the proxy objects. The unused computing ability in the network is utilized by this technique. The communication processes is executed as a single element comprising of many loosely connected distributed shared proxy objects.

By integrating SOM and DSO, the following benefits can be achieved:

- DSO is used to increase the information processing capacity, service sharing by providing context and

location sensitive information, while the issues due to the asymmetry of the mobile distributed systems in network connectivity, mobility, and computing power are solved by SOM.

- The heterogeneity in operating systems, system architectures, and load variations are solved in a fair manner by the DSO technique, while the unused computing determinant is utilized by SOM to save the processing time.
- The transparency of the DSO model in terms of distribution and heterogeneity reduces the computational complexity, while SOM is chosen to enhance the resource sharing of mobile grid computing.
- DSO also enhances the load adaptability and fault-tolerance to parallel programs on the mobile grid.

The remaining part of the paper is organized as follows: Section II involves the detailed description of the DSPOM based mobile grid. Section III involves the design issues observed during the implementation of the DSPOM. Section IV involves the performance evaluation and comparison of the DSPOM with SOM and DSO. The paper is concluded in Section V.

II. DSPOM Based Mobile Grid

The proposed mobile grid is defined as a cluster of mobile clusters [32]. Each cluster is a combination of MHs and static hosts (SHs) grouped logically/virtually. The MHs are served by a conventional mobile support station (MSS). Each cluster is managed by a SH designated as cluster head (CH). The role of a CH is to manage the services and resources within its cluster. MSS and CH can be assigned to the same host when the MSS load is less. The communication between the CHs of the mobile grid is performed by a P2P overlay [35].

A. Visualization of Mobile Grid

An example mobile grid structure in terms of distributed shared proxy objects is shown in Fig. 1. The mobile grid consists of Mobile Clusters (MCs), Cells, Mobile Support Stations (MSSs), proxy objects, Cluster Head (CH) which is also a Static Host (SH), P2P overlay for the communication between the CHs, and Mobile Hosts (MHs). The states of the distributed proxy objects alone are shared.

Each MH is visualized by a proxy object (PO) in the wired network. This ensures the conservation of transparency to the imbalance in wireless communication. An active PO monitors the information pertaining to the current state. When a MH comes into a specific cell it transmits a control signal to the MSS of the cell, which contains the address of the MH's former MSS. When a MH enters the cluster for the first time, the address of the former MSS is specified as NULL, and a PO is created for complete encapsulation of the mobile device. The object pointer to a new PO is transferred to its MH for its further communication with the PO. A unified object model can be accomplished by visualizing the other MHs and SHs as independent objects. The multiple interface of an object represents the multiple nodal services. The mobile nodes involved in the computations are optimized into distributed mobile objects and the mobile grid is converted from a nodal collection into a distributed mobile objects. These mobile objects form the basic blocks of the mobile grid.

The abstraction of MH into PO addresses most of the critical issues affiliated with the MH [36]. Some of the relevant aspects of proxy objects are as follows:

- It gives a solution to the mobile asymmetry problem due to difference in wired and wireless bandwidths.
- It avoids the location management issue by defining a storage point for MH location details.
- It reduces the query response time and avoids data loss by caching the host specific data and buffering the user requests temporarily when MH is disconnected.
- It ensures the optimal utilization of wireless bandwidth.

The distributed and decentralized nature of the proxy objects enhance the resource sharing in the mobile grid. The proxy objects represent the active hosts and the wired network represents them. The mobility of the MHs does not degrade the provision of the services because of the full-state maintenance of hosts. The proxy objects are fully dynamic, secure, and autonomous i.e., the operational capabilities of POs even after the MHs are not reachable and disconnected. The PO may be related to MHs of three cases:

1. MH outside the local but inside the same MC.
2. MH outside any MC.
3. Foreign MH.

The dynamicity of the mobile grid architecture is enabled by permitting the alteration of new services and extending the wired network. This aids the adaptability of the services based on the necessities of the mobile grid users. PO possesses the properties and characteristics of a host. The characteristics of a PO are represented as attributes, sub-objects, and methods. The attributes of a PO comprise the computational capability, bandwidth, and memory consumptions [32]. The sub-objects and the methods contain the services and other data resources offered by the host. The integration of the agreement and security policy by PO for all services specifies the operation mode of services.

SOM operates in wireless connectivity mode when the surrogate objects are transferred to the wired network from the MHs. The state information is used to maintain the services of PO, even when the MH is disconnected. The data and services results can be delivered once the MH reconnects. This programming principle (or) paradigm enhances security by integrating authority schemes within the POs to operate the objects and services.

MHs are also information service providers [36]. The energy and bandwidth consumption can be reduced by contacting the PO for information rather than contacting the MHs. Congestion can also be avoided by replicating the POs, thereby enhancing the system scalability.

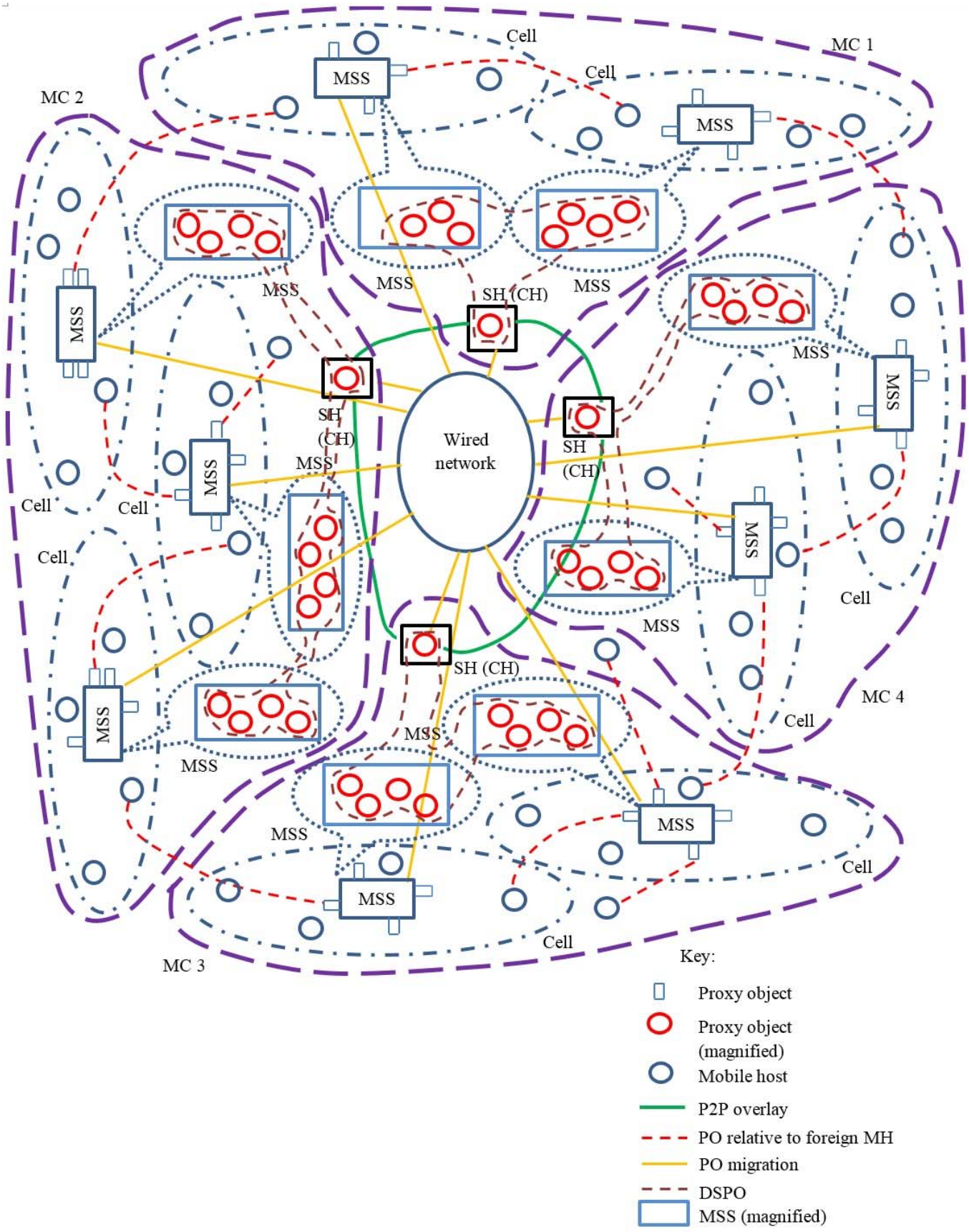


Fig. 1. Proposed mobile grid architecture based on DSPOM.

B. *Management of Services*

The mobile grid comprises of services as the basic blocks for large-scale computations [32]. The main issues in the mobile grid are service detection and cooperation, i.e. to synchronize a group of services executing on heterogeneous resources under various controls in order to solve a unified problem. The computational resources and their functionalities are defined as composable services. Mobile grid applications can be constructed by creating these services at a higher abstraction level.

Service composition is an important function which permits various autonomous services to be formed into a new service with a unique functionality and permits development of independent and modular services. Service composition is dependent on the location of the users and focuses on management of the integrated services for task completion [32].

Mobile grid formed by MHs poses new obstacles for service composition. The service results are not displayed when a user requests a service from a host, but moves to another host with higher functionality. The result becomes invalid when the user moves across various access points for location dependent services. The infrastructure for service composition of the mobile grid is given in Fig. 2.

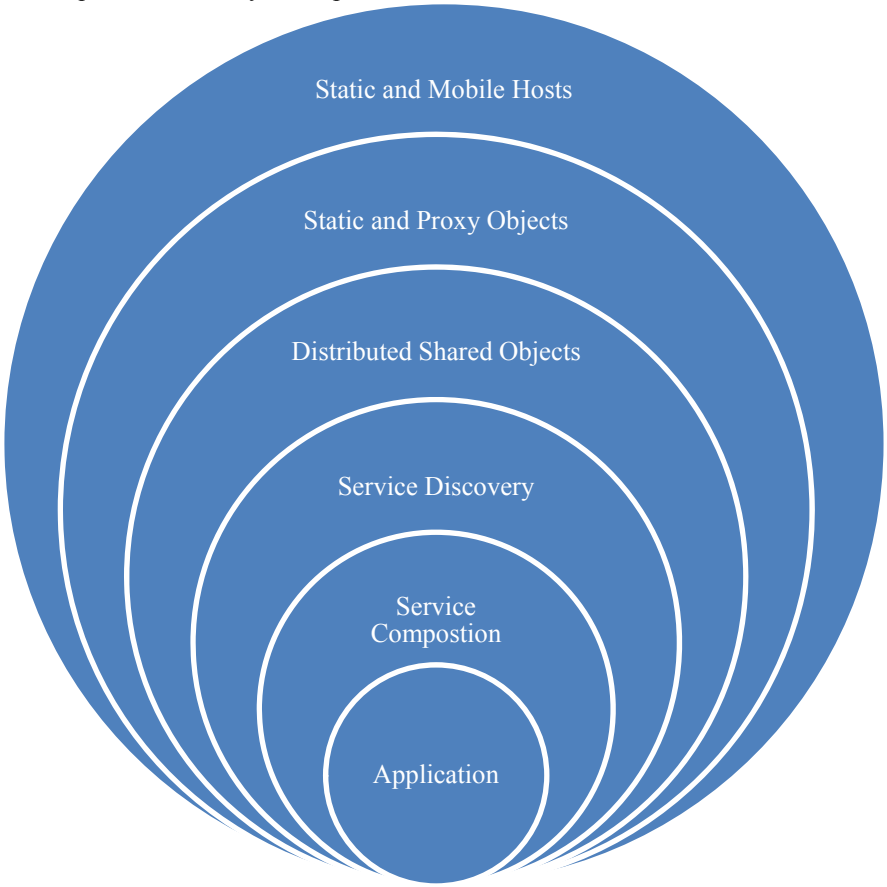


Fig. 2. Layered reference architecture for services management in mobile grid.

The distributed proxy objects structure the principles through which services are defined and used for resource sharing in the mobile grid. The services are virtualized as software functional components which involves abstraction of pre-defined functionalities. The services are announced and discovered using inspection and directories. After service detection, an appealing element can connect to the selected mobile device and begin the communication with its explicit functions via platform autonomous protocols. A vast conglomeration of composable software and hardware resources can be obtained by integrating the services with virtualized hardware resources.

Service detection is an important step, as the system must find a service before it could utilize it. The mobile grid should be capable of supporting variable service announcement and discovery. The service detection is based on the characteristics, location, and functionality of the services [36]. Runtime connection enhances reliability and load balancing of the system. It also supports a wide range of application compatibility in terms of network configurations and platforms. When the task is completed the services are returned back to the conglomeration for assignment to other users.

1) *Service characterization*

The service providers need to issue the constraints and features of the services. These descriptions are sent to CH for registration. A discovery protocol is required to map the services to the application queries. This service model

has the following advantages:

- No particular schema is needed for the system to work in a heterogeneous environment.
- Freedom to express conditions on the services the model is ready to serve.

The characterization of services involves attributes and conditions [32]. The attributes of the service model comprise of the resource characteristics such as CPU usage, location, and free memory. Dynamic characteristics of the services can be obtained by DPSOs running on the data resource. Conditions include the limitation expression given by the service provider for service allocation.

The services are maintained by an exchange service repository. The logical mobile clusters contain a service directory in the CH, which is the center for services registry. A MH that desires to give its services avails the information in the service directory.

2) Service detection

The services must be identified in order to coordinate with other mobile users and utilize them. The present methods for service detection constrain the interoperability and lead to expressive power loss during component characterization. The DPSOM gives a clear differentiation between global service management and local service management. Global service management groups the data from the cells into large mobile clusters, thus availing the data available to the users [36]. It also comprises global tracking services and search methods for clients. Local service management controls the services within a logical mobile cluster or a domain like cell. The directory service regards the location of two processes in a similar cell also as same.

Local service management occurs in CH of each logical mobile cluster. A service is affiliated with the Proxy Object Identifier (POID) depending on the client's needs. The basic requirements for this management are execution monitor, information database, and scheduler.

- The information related to the data resources and services are stored in the information database.
- The scheduler computes an association of objects with services based on the information database.
- This connection is used to contact the objects and affirm the schedule. All these processes are managed by the execution monitor.

3) Service composition

The combination of high and low performance devices interconnected to each other makes the integration and execution of heterogeneous tasks a tough task [32]. Service composition is used to construct complex services from basic services, thus resulting flexibility of new service creations. It can be shown as a variable integration of multiple services in the mobile grid in response to a client request.

The critical issue in service composition involves the management of disconnections during the execution of services and formulating context dependent service execution. The execution of service composition is performed based on optimal computing resources under conditions like data resource reliability and execution cost. A user initializing a service composite request can also indulge in another service composition. The P2P overlay within the CHs ensures the fault-tolerance of the system.

C. Distribution and Sharing of Proxy Objects

A DSO makes use of multiple interfaces, each composing of a group of methods [33]. The proxy objects act as local objects in a DSPO. The mobile objects in this model are passive, while the user threads use these mobile objects by executing the code for their methods. A single mobile object can be accessed by multiple processes simultaneously. The modifications to a mobile object's state by a process are visible by the other processes. The distributive nature of the mobile shared objects enables the active copies of a mobile object's state to be stored simultaneously on multiple machines. But, the communication protocols, distribution/migration of states, and replication methodologies are embedded in the interface.

A significant difference between DPSOs and remote objects is that there is no priori differentiation between users and servers. The processes that communicate via method invocation combine in its object implementation.

1) Merits of DPSOs

Some of the advantages of DPSOs are:

- A distributed proxy object enables well-structured interfaces to its applications. The user is separated from the communication and replication processes.
- Complete encapsulation of communication and persistence in a distributed proxy object. This implies that the implementation of a distributed system is not bounded to a small group of consistency algorithms or communication protocols.
- Ability to load object implementations at runtime.
- A process consists of a local implementation of the distributed proxy object's interface in its own address space. A DSPO is visualized only as a local object in the perspective of a process.

2) Architecture of DSPO

The distributed proxy objects are a group of local proxy objects that communicate and furnish the user of the object with the delusion of the shared state [33]. This characteristic is advanced over a remote object model since it is not bounded to a small set of predefined communication patterns. The example architecture of a DSPO is shown in Fig. 3.

The distributed proxy object is used in four address spaces, where each address space comprises of a proxy object. The group of proxy objects forms the distributed proxy object. The proxy objects utilize the communication services of a network to operate on the distributed proxy object and maintain the distributed proxy object.

The proxy object implementation is assorted from an application via an explicit interface table comprising of method pointers [33]. The interface table is triggered when the process connects to the object. The contents of the interface table are variable over time. This enables the dynamic adaption of the distributed object's local implementation. The adaption process does not affect the

interface to the application that triggers its methods.

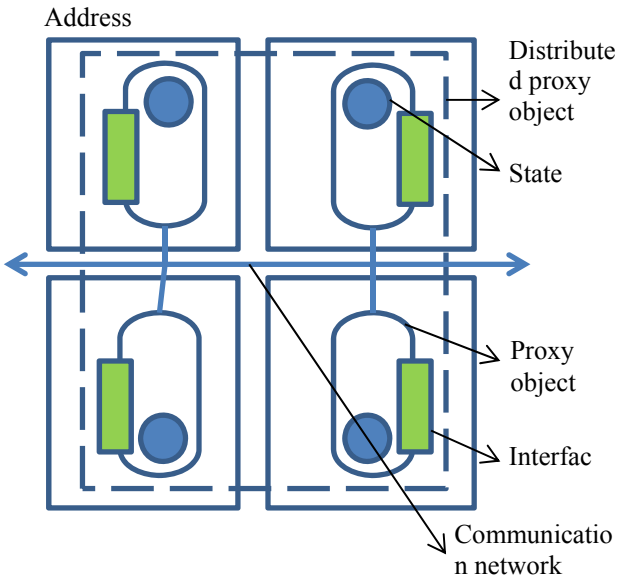


Fig. 3. Example architecture of a DSPO.

The implementation of a distributed proxy object can utilize random communication patterns while communicating with local objects. The communication process can also integrate data placement and replication. This scheme is applicable for efficient implementations of various communication paradigms. There are no limitations on the predefined operations since the interfaces are completely user-defined.

D. Transparent Communication

The developer of the distributed proxy object requires being isolated from the data placement and replication [33]. So, a standard hierarchy is developed for the implementation of a distributed proxy object. The hierarchy of a local object is given in Fig. 4.

The distributed proxy object's developer is segregated from communication, consistency management, and replication by using a *communication object* and a *replication object*. The object developer implements the *semantics object* which enables the actual functionality of the distributed proxy object, while the communication and replication objects are chosen from a library. A *control object* is used to manage the interactions between the replication objects and the sematic objects as a consequence of method invocation by an application. The control object can generated automatically.

The proxy object is capable of exporting methods that can execute on internal state. A control object is produced based on the interface to the semantics object. The access of the control object is synchronized with that of the distributed proxy object by serializing permissions to the semantics object. This prohibits the race conditions by triggering the replication object to stabilize the state of the distributed proxy object. The interface exported by the control object is similar to that of the semantics object.

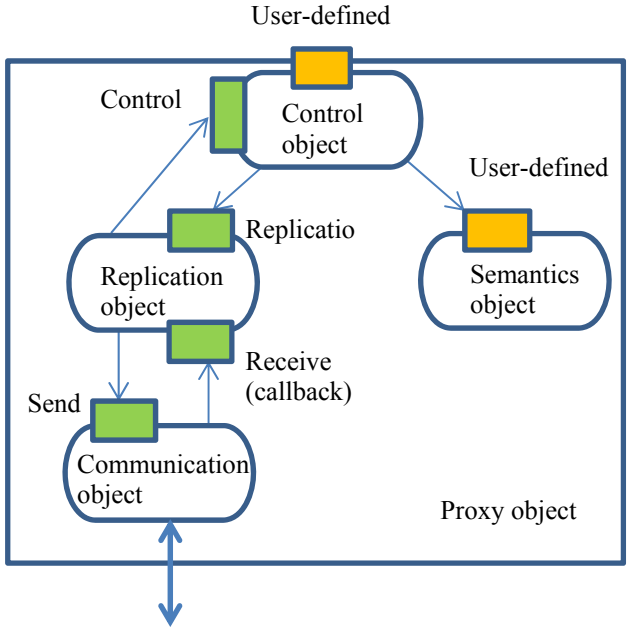


Fig. 4. Hierarchy of a proxy object.

The implementation of a method invocation by the control object is performed via three consecutive steps:

1. The *start* method controls the execution of global state functions. During remote execution, the control object passes the mobilized arguments of the method invocation to the replication object. The execution of the replication object occurs according to a specific replication protocol and returns the mobilized results to the control object according to the remote method invocation.
2. During local execution, the control object triggers the related method on the semantics object. During active replication with a local copy, the control object offers the replication object with the mobilized arguments of the method invocation. Then, the replication object executes the protocol to transfer the arguments to all replicas. This enables the synchronization with the other replicas.
3. Finally, the control object triggers the corresponding method on the semantics object.

The control object triggers the *finish* method on the replication object, which yields the replication object an opportunity to update the remote replicas. Two extensions are needed in this model to enhance its practical performance.

1. The control object and replication object needs to distinguish various operation types. It is also required to differentiate operations that alter only a part of the global state, which may occur in the case of nested or segmented objects.
2. Few extensions are required on specific criteria to handle synchronization since the semantic operations are serialized and not permitted to block for a long time.

The operations can be secured by providing blocks on a conditional basis. The status information is returned to the control object after the possible alterations are made. The

control object will delay the execution of the operation until the next state alteration.

This model of shared state comprising of operations results in passive objects, where the activity is given by the threads executing in the processes. For seamless integration of communication in this model, *pop-up threads* are instantiated with the incoming messages [33]. So, the communication object will open a new thread to handle an incoming message. The communication object inside the new thread triggers a method on the replication object's callback interface. The replication object requests the control object's callback interface with the mobilized arguments of the request.

III. Design Issues

The performance of the DSPO scheme depends on the mobility of the nodes, tolerance to network traffic, and node density. The number of mobile nodes entering a group of cells should be managed properly. The parallel computing on mobile grid depends upon several key issues such as, mobility of nodes, fault-tolerance, connectivity of mobile nodes, uneven nodal distribution, transmission time, heterogeneity of nodal performances, and uneven load in the network.

A. Load Balancing

The optimum number of implementation processes and their granularities can be computed from the details respective to the availability of nodes and load in the network. The various mobile nodes contain partitioned computation sub-domains. The granularity of each subdomain depends upon the load ratio on each mobile node. The heterogeneous load conditions are handled by issuing load indices (threshold values). The load can be balanced by nullifying the mobile nodes for which the load indices cross a specific value. The processing power of individual elements also contributes to the load balancing in a heterogeneous group of mobile nodes.

B. Bandwidth

The nodes possess high variations in the network bandwidth, depending on whether it is a static node or a mobile node, and on the type of connection to the present cell. The DSPO model distinguishes the type of connectivity and provides flexibility in terms of network bandwidth and task size.

The proxy objects does not support flexible wireless bandwidth, which is overcome in the DSPO model by differentiating the type of connectivity in a cell of the mobile cluster.

C. Handoff Process

When a MH moves from a cell to another, the transition process is termed as *handoff*. The channel resources should be monitored to preserve the connectivity of the network during handoff. The channel used in the previous cell may not be reusable in the present cell because of co-channel interference or adjacent channel interference or low signal strength. So the transiting mobile device gets separated from the rest of the mobile cluster. When a fresh channel has not been allotted to the mobile node within a limited time span, the messages transmitted in the network are delayed, resulting in a retransmission of data.

The transmission time can be limited and the retransmission of data can be averted by using various topologies based management schemes. Two types of topology of nodes are used such as, *tree topology* and *ring topology*. The tree topology consists of an organizational structure of nodal levels forming a tree. The number of nodes in the lower level of the tree is greater than that of the previous level of the tree. Each mobile node in the ring topology consists of exactly two neighbors forming a planar structure. This structure rearranges the MHs arranged in a mobile grid of rows and columns. A fragmented topology is reframed prior to the deletion of the pre-defined topology during the traversing of a mobile node. The reconstruction of the fragmented topology is based on an optional node to the migrated node. The reconstruction time should be least for the satisfactory performance of the mobile grid computing.

The handoff issue was not properly addressed by the proxy objects, but it is efficiently handled by the topology scheme introduced in the DSPO model.

D. Disconnectivity of Mobile Nodes

Timers are maintained for the detection of mobile node disconnectivity from its affiliated cell. A mobile host when it does not return to its cell within the time set in the timer and then the sub-process is retransmitted to some other mobile node. The disconnectivity issue was not properly addressed by the proxy objects, but it is efficiently handled by the timer scheme introduced in the DSPO model.

E. Tolerance to Network Traffic

The network traffic generated during the migration of PO is managed by a real-time queue model. A robust queue estimation algorithm can be applied for managing the dynamic network traffic. Reliable and accurate queue information controls the network traffic of mobile grids adaptively to the real traffic queue sizes. A primitive conservation model is used to estimate the queue size (system state) with the flow-in and flow-out readings. These results are updated with a measurement equation based on the time occupancies from link-entrance and mid-link loop detectors. This model also consists of a single point correction technique which resets the estimated results to avoid the counting errors over time.

The network traffic issue was not properly addressed by individual proxy objects, but it is efficiently handled by the queue introduced in the DSPO model.

IV. Performance Analysis

The DSPO is compared with SOM [32] and ARC [37] in terms of the following parameters, query time and query latency with respect to packet loss and migration frequencies, number of messages exchanged, speedup, and processing time. *Speedup* is the ratio of sequential execution time to the parallel execution time. The network architecture consists of a heterogeneous combination of 30 mobile and static nodes in a simulated area of 100 * 100 m.

There are 5 mobile clusters with a cluster head for each cluster. The nodal velocity is varied from 5 to 30 m/s.

A section of the simulation field is shown in Fig. 5. The yellow colored links between the CHs represent the P2P overlay. A CH and their respective nodes are linked by a DSPO as highlighted by black colored links. The radio of each node is given as green colored circles. The routing path of a particular node n11 is highlighted in red colored lines as shown in Fig. 6.

The frequency of the presence of the mobile nodes in the respective mobile clusters is computed. Each mobile cluster is represented by their cluster node (or) home node. The messages to be communicated in packets are created at consequent time intervals in the corresponding source nodes. The messages are relayed to intermediate nodes and then received by the final destination nodes.

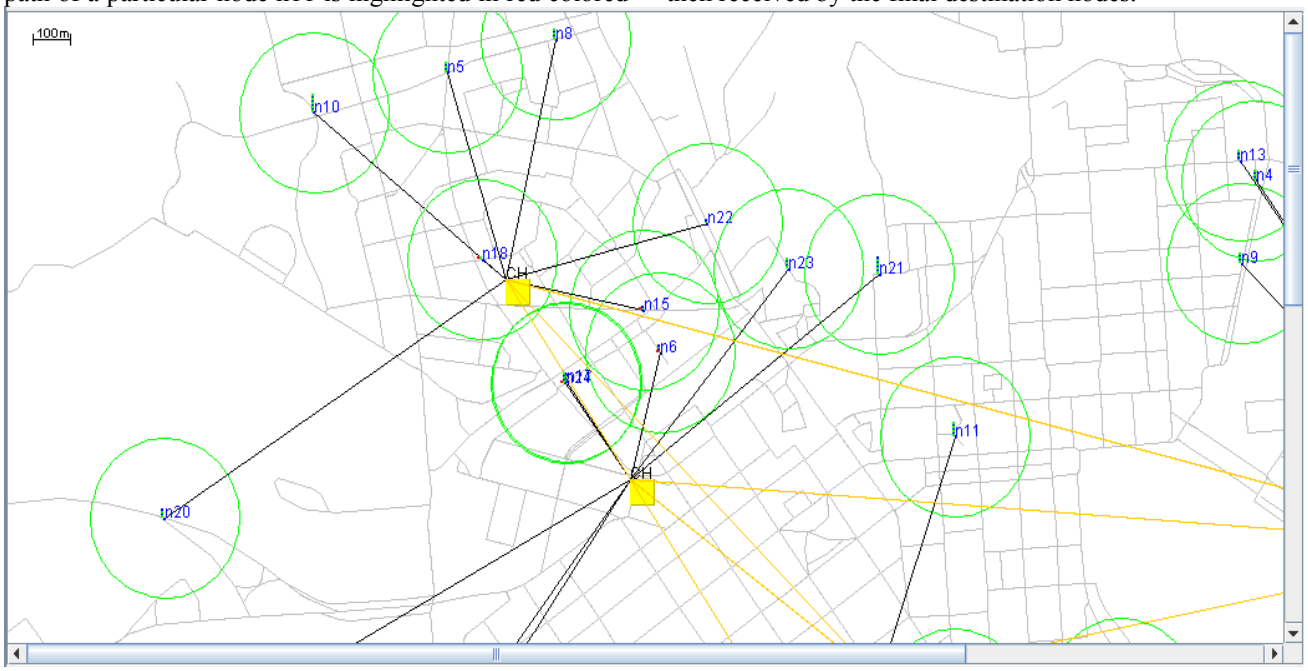


Fig. 5. Section of the simulation field.

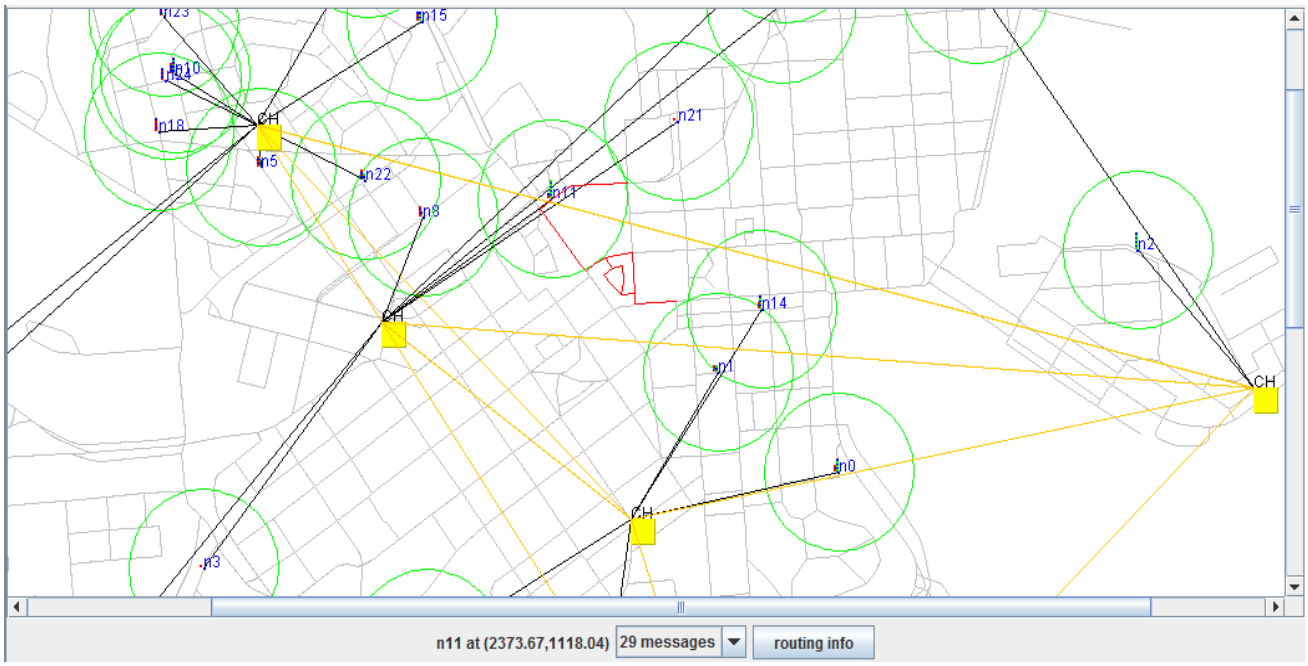


Fig. 6. The routing path of node n11 is highlighted in red colored lines.

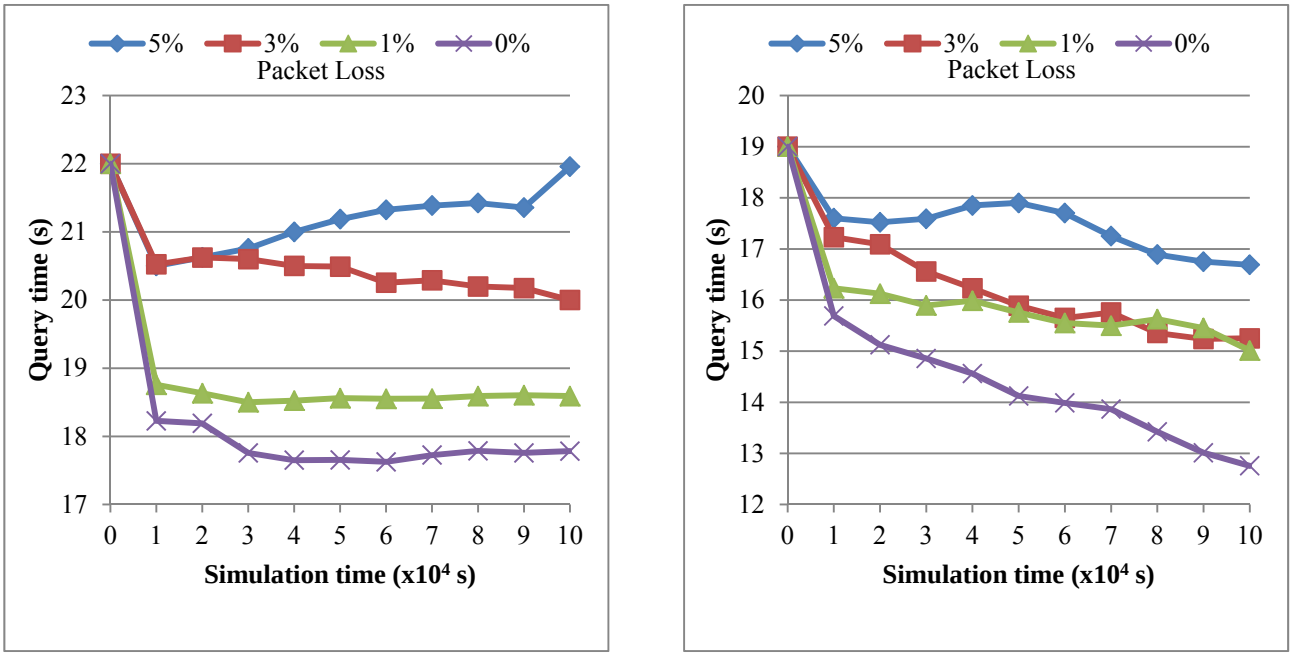


Fig. 7. Comparison of query latencies for various packet losses in: (a) SOM, and (b) DSPOM.

A. *Query Time and Latency with Respect to Packet Loss*

The query latencies are analyzed in terms of query time for a caching application using SOM and DSPO. The comparative analysis is performed against simulation time for four packet loss probabilities of 0%, 1%, 3%, and 5% as shown in Fig. 7.

The results show that the query time for DSPO model decreases almost gradually, but the decrease for SOM in the cases of packet loss probabilities of 0%, 1%, and 3% is smaller compared to that of the former model. For a packet loss probability of 5% in SOM there is an increase in query time after simulation time greater than 1 x 10^4 s. The query time for the DSPO model is about 14% lesser than SOM with respect to four packet loss probabilities of 0%, 1%, 3%, and 5%.

The query time increases as the packet loss probability also increases in SOM. However, in DSPOM this increase is not much greater compared to the same in SOM. It can also be noted that the maximum query time with null packet loss probability is higher in SOM, compared to the maximum query time in DSPOM, even with 5% packet loss probability. The percentage increase in query time for three packet loss probabilities is observed in Fig. 8.

In the case of packet loss probability equal to 0.3, the difference in increase of task time between SOM and DSPOM is about 16%, with SOM possessing higher increase in query time. The maximum increase in query time is higher in SOM (19%) compared to the DSPO model (18%).

B. *Query Time and Latency with Respect to Migration Frequency*

The migration of PO to any node in the wired network

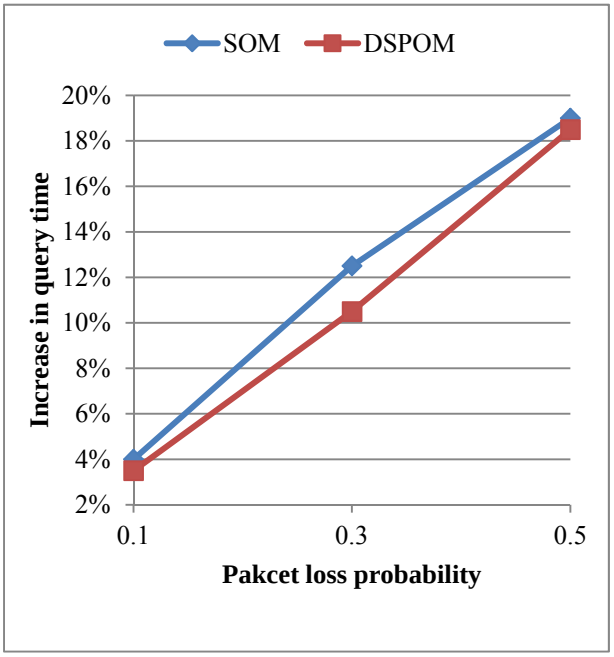


Fig. 8. Increase in query time over increasing packet loss probability.

and the distributed and shared computing contributes an advantage over the conventional SOM and DSO methods. This enhances the load balancing, failure recovery, and network latency reduction of the system. The effect of PO migration on the query latency for different migration frequencies in SOM and DSPOM is analyzed in Fig. 9.

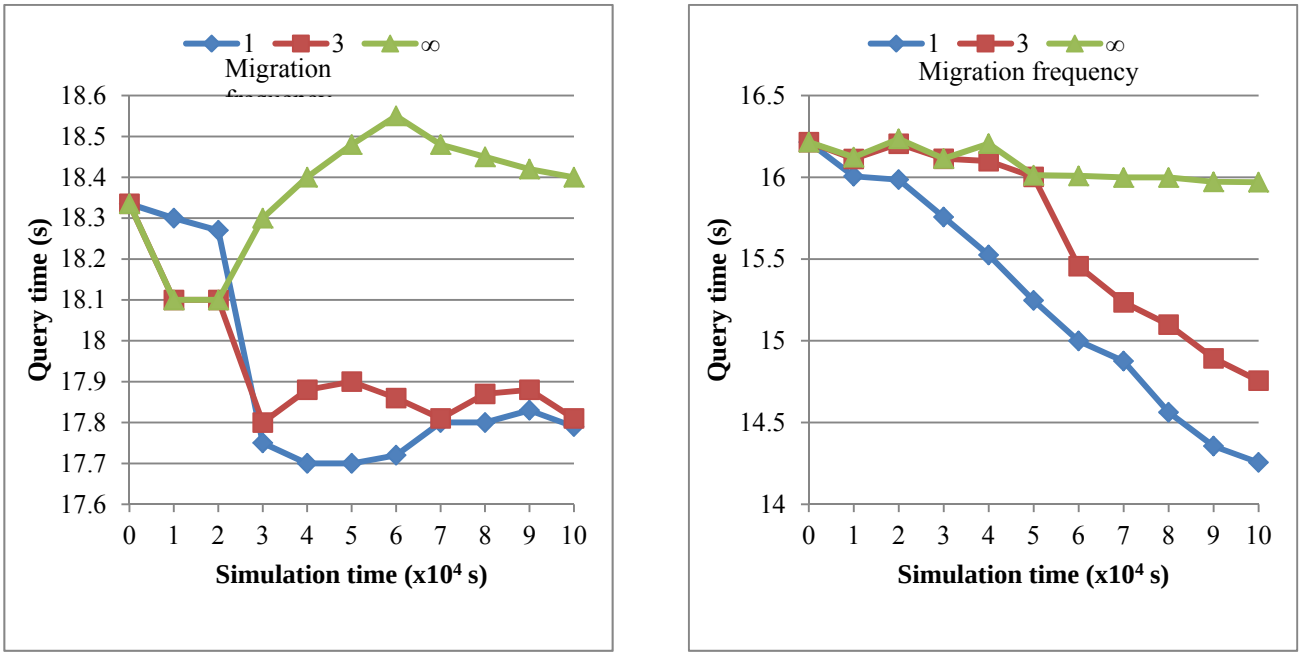


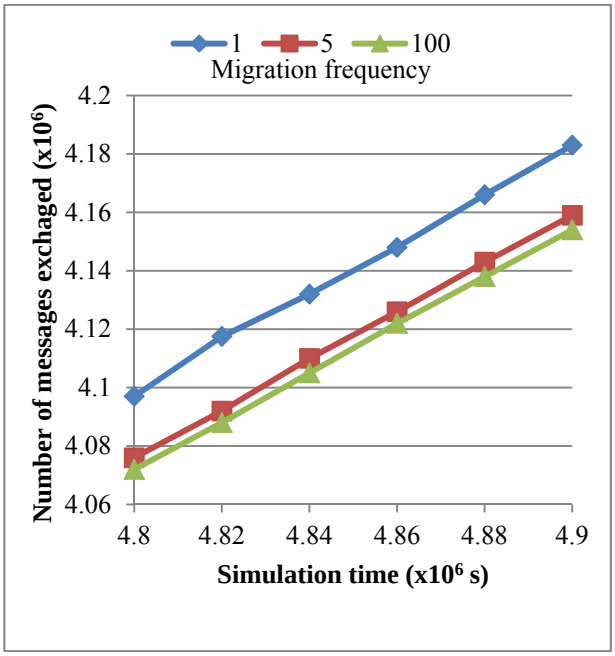
Fig. 9. Comparison of query times in: (a) SOM, and (b) DSPOM.

The results show that the query time for DSPO model decreases almost gradually in the cases of migration frequencies of 1 and 3 and decreases only a bit in the case of migration frequency of infinity. The query time for SOM decreases randomly in the cases of migration frequencies of 1 and 3. It increases randomly in the case of migration frequency of infinity after a simulation time of 2 x 10^4 s. The query time for the DSPO model is about 11.56% lesser than SOM with respect to three migration frequencies of 1, 3, and infinity.

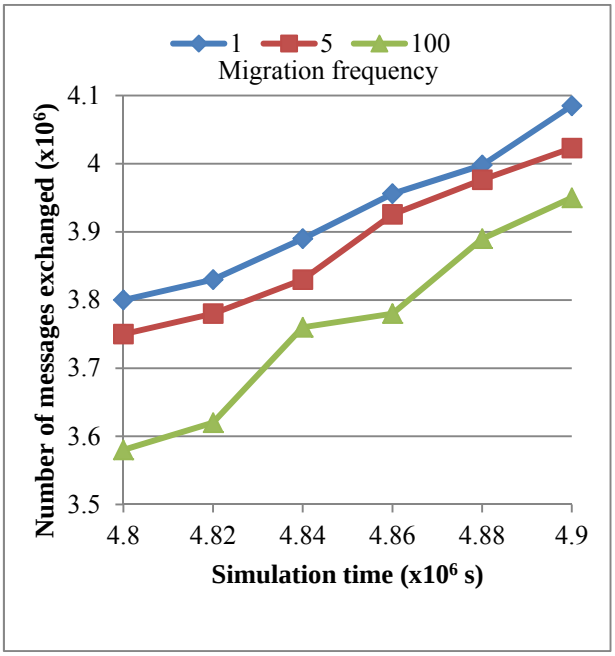
The query latencies become high when the PO is migrated for every movement of MH. This is due to the higher amount of time for migration. But, when the migration frequency is decreased, the query time is increased because the MH moves to various cells but the PO moves only once. This implies that the query time remains constant when the PO is static and MH moves many times. This is due to the fact that the query response time is dependent on the PO nearness to the initial MSS. Thus, queries are randomly generated from various parts of the network.

C. Network Traffic

The network traffic generated due to PO migration is analyzed for the determination of the freedom of PO migration. The increase in the number of messages exchanged relative to simulation time at different migration frequencies is shown in Fig. 10.



(a)



(b)

Fig. 10. Comparison of number of messages exchanged in: (a) SOM, and (b) DSPOM.

The results show that the number of messages exchanged increases randomly in DSPO model and linearly in SOM for migration frequencies of 1, 5, and 100. It is observed that the number of messages exchanged in the DSPO model is about 12.1% higher than SOM with respect to the migration frequency of 100. The maximum number of messages exchanged is higher in SOM (4.183 x 10^6) compared to the DSPO model (4.635 x 10^6).

It is observed that the magnitude of the traffic is higher when the PO moves for every movement of MH. But as the migration frequency decreases, the generation of traffic is much lesser. The migration frequency value other than unity does not influence the network traffic. This proves that it is better not to move the PO whenever the MH moves.

D. Speedup and Processing Time

The mobile grid architecture is realized using a cellular model with seven nodes as MSSs, and the remaining nodes are designated as MHs of the mobile grid. The MHs are partitioned into seven groups, where communication occurs between the MH of a group with other nodes through the relative MSSs. Each MH consists of a client and a DSPO. The clients are involved in the task submission, while the MHs are involved in the distributed mobile processing. The performance analysis displays the speedup attained by the parallel execution of the problem. TABLE I compares the speedup achieved using the mobile grid under a different number of participating MHs with mobility and without mobility conditions.

TABLE I
COMPARISON OF SPEEDUP IN SOM AND DSPOM

No. of MHs	Without Mobility – Speedup		With Mobility – Speedup	
	SOM	DSPOM	SOM	DSPOM
4	2.745	2.956	2.6	2.754
8	3.889	3.957	3.37	4.245
16	6.778	7.256	5.453	6.235
32	9.948	10.687	8.443	9.256
42	15.632	16.425	12.665	13.687

TABLE I shows that the speedup in DSPO model is higher than that of SOM with and without mobility conditions. The average speedup for the DSPO model is about 5.38% higher than that of SOM without mobility considering MH quantities equal to 4, 8, 16, 32, and 42. The average speedup for the DSPO model is about 11% higher than that of SOM with mobility considering MH quantities equal to 4, 8, 16, 32, and 42. It is observed that in the case of with mobility, the speedup decreases relatively in both the cases due to higher processing time taken by moving devices.

TABLE II
COMPARISON OF PROCESSING TIME IN SOM AND DSPOM

No. of MHs	Without Mobility – Processing Time (s)		With Mobility – Processing Time (s)	
	SOM	DSPOM	SOM	DSPOM
4	279	256	302	292
8	197	185	233	225
16	113	105	144	138
32	77	64	93	85
42	49	32	62	59

TABLE II compares the processing time taken for SOM and DSPOM under a different number of participating MHs with mobility and without mobility conditions. TABLE II shows that the processing time in DSPO model is lesser than that of SOM with and without mobility conditions. The average processing time for the DSPO model is about 14.6% lesser than that of SOM without mobility considering MH quantities equal to 4, 8, 16, 32, and 42. The average processing time for the DSPO model is about 4.86% lesser than that of SOM with mobility considering MH quantities equal to 4, 8, 16, 32, and 42.

A program for solving double precision matrix multiplication problem is loaded onto the DSO model and DSPO model. Their overheads are compared in terms of speedup at no load condition. TABLE III compares the speedup values achieved for different task sizes.

TABLE III
COMPARISON OF SPEEDUP IN DSO AND DSPOM

Task size	Nodes	Speedup – DSO	Speedup – DSPOM
50 x 50	2	1.46	1.65
100 x 100	2	1.75	1.89
	4	3.2	3.67
	5	3.8	4.05
150 x 150	3	2.63	2.89
	6	4.75	4.86
200 x 200	4	3.48	3.78
	5	3.48	3.85
	8	6.24	6.92

TABLE III shows that the speedup in DSPO model is higher than that of SOM for various task sizes and numbers of nodes equal to 2, 3, 4, 5, 6, and 8. The average speedup for the DSPO model is about 8.5% higher than that of SOM considering various MH quantities and task sizes equal to “50 x 50”, “100 x 100”, “150 x 150”, and “200 x 200”. It is seen that DSPOM possesses slightly lesser overhead than DSO technique, with retained features such as fault-tolerance, heterogeneity support, and load adaptability.

V. CONCLUSION

Parallel computing methods are preferred over sequential computing techniques to decrease the processing time in mobile distributed systems. But they are subject to many issues such as, high latency/jitter, processing speed, communication overhead, and low data transfer rate, when they are developed from smaller mobile clusters to extensive mobile grids. So, an efficient and optimized parallel computing paradigm known as Distributed Shared Proxy Object Model (DSPOM) is developed based on Surrogate Object Model (SOM) and Distributed Shared Object (DSO) for mobile grid.

By integrating SOM and DSO, the following benefits can be achieved: 1) DSO is used to increase the information processing capacity, service sharing by providing context and location sensitive information, while the issues due to the asymmetry of the mobile distributed systems in network connectivity, mobility, and computing power are solved by SOM, 2) The heterogeneity in operating systems, system architectures, and load variations are solved in a fair manner by the DSO technique, while the unused computing determinant is utilized by SOM to save the processing time, 3) The transparency of the DSO model in terms of distribution and heterogeneity reduces the computational complexity, while SOM is chosen to enhance the resource sharing of mobile grid computing, and 4) DSO also enhances the load adaptability and fault-tolerance to parallel programs on the mobile grid.

POM consists of a grid computing technique and service composition technology which combines the universal resource access and middleware solution for mobile computing to enhance the resource sharing solutions of mobile grid computing. The unused computing determinant is utilized to save the processing time. The proxy objects are distributed and shared to avail user-defined functions through distribution and replication of states. The transparency of the program model in terms of distribution and heterogeneity reduces the computational complexity. It also enhances the load adaptability and fault-tolerance to parallel programs on the mobile grid. The approach of DSPOM performs better in terms of query time, query latency, packet loss, load adaptability, and fault-tolerance.

REFERENCES

[1] S. W. Keckler, *et al.*, "GPUs and the Future of Parallel Computing," *Micro, IEEE*, 2011, vol. 31, no. 5, pp. 7-17.
[2] C. Delporte-Gallet, *et al.*, "The disagreement power of an adversary," *Distributed Computing*, 2011, vol. 24, no. 3-4, pp. 137-147.
[3] P. Kumar and R. Garg, "Soft-Checkpointing Based Coordinated Checkpointing Protocol for Mobile Distributed Systems," *International Journal of Computer Science Issues*, 2010, vol. 7, no. 3, pp. 40-46.
[4] D. Dib, *et al.*, "Towards Multi-level Adaptation for Distributed Operating Systems and Applications," in *Algorithms and Architectures for Parallel Processing*. vol. 7440, Y. Xiang, I. Stojmenovic, B. Apduhan, G. Wang, K. Nakano, and A. Zomaya, Eds., ed: Springer Berlin Heidelberg, 2012, pp. 100-109.
[5] P. Donnelly and D. Thain, "Fine-Grained Access Control in the Chirp Distributed File System," in *Cluster, Cloud and Grid Computing (CCGrid), 2012 12th IEEE/ACM International Symposium on*, 2012, pp. 33-40.
[6] B. K. Johnson, *et al.*, "DP: a paradigm for anonymous remote, computation and communication for cluster computing," *Parallel and Distributed Systems, IEEE Transactions on*, 2001, vol. 12, no. 10, pp. 1052-1065.

[7] M. A. M. Mohamed, *et al.*, "Moset: An anonymous remote mobile cluster computing paradigm," *Journal of Parallel and Distributed Computing*, 2005, vol. 65, no. 10, pp. 1212-1222.
[8] J. Casas, *et al.*, "Adaptive load migration systems for PVM," presented at the *Proceedings of the 1994 ACM/IEEE conference on Supercomputing*, Washington, D.C., 1994.
[9] F. Douglis and J. Ousterhout, "Transparent process migration: Design alternatives and the sprite implementation," *Software: Practice and Experience*, 1991, vol. 21, no. 8, pp. 757-785.
[10] G. J. W. van Dijk and M. J. van Gils, "Efficient process migration in the EMPS multiprocessor system," in *Parallel Processing Symposium, 1992. Proceedings., Sixth International*, 1992, pp. 58-66.
[11] M. Litzkow and M. Solomon, "Supporting checkpointing and process migration outside the UNIX kernel," in *Mobility*, M. Dejan, cacute, D. Frederick, and W. Richard, Eds., ed: ACM Press/Addison-Wesley Publishing Co., 1999, pp. 154-162.
[12] D. Gelernter and D. Kaminsky, "Supercomputing out of recycled garbage: preliminary experience with Piranha," presented at the *Proceedings of the 6th international conference on Supercomputing*, Washington, D. C., USA, 1992.
[13] F. Tandiary, *et al.*, "Batrun: utilizing idle workstations for large scale computing," *Parallel & Distributed Technology: Systems & Applications, IEEE*, 1996, vol. 4, no. 2, pp. 41-48.
[14] B. He, *et al.*, "Comet: batched stream processing for data intensive distributed computing," presented at the *Proceedings of the 1st ACM symposium on Cloud computing*, Indianapolis, Indiana, USA, 2010.
[15] Y. Zhang, *et al.*, "iMapReduce: A Distributed Computing Framework for Iterative Computation," *Journal of Grid Computing*, 2012, vol. 10, no. 1, pp. 47-68.
[16] T. E. Anderson, *et al.*, "A case for NOW (Networks of Workstations)," *Micro, IEEE*, 1995, vol. 15, no. 1, pp. 54-64.
[17] S. Kailasam, *et al.*, "Arogyasree: an enhanced grid-based approach to mobile telemedicine," *Int. J. Telemedicine Appl.*, 2010, vol. 2010, pp. 1-11.
[18] D. Janakiram, *et al.*, "GDP: A Paradigm for Intertask Communication in Grid Computing Through Distributed Pipes," in *Distributed Computing and Internet Technology*. vol. 3816, G. Chakraborty, Ed., ed: Springer Berlin Heidelberg, 2005, pp. 235-241.
[19] M. A. M. Mohamed, *et al.*, "EOMP: an exactly once multicast protocol for distributed mobile systems," *International Journal of Parallel, Emergent and Distributed Systems*, 2010, vol. 25, no. 3, pp. 183-207.
[20] M. Mohamed, "Communication and Computing Paradigm for Distributed Mobile Systems," *Journal on Information Sciences and Computing*, 2007, vol. 1, no. 1, pp. 33-41.
[21] A. Basit and C.-C. Chang, "Mobile cluster computing using IPV6," in *Ottawa Linux Symposium*, 2002, pp. 31-39.
[22] K. Hairong, *et al.*, "Iterative grid-based computing using mobile agents," in *Parallel Processing, 2002. Proceedings. International Conference on*, 2002, pp. 109-117.
[23] G. R. Andrews, "Paradigms for process interaction in distributed programs," *ACM Comput. Surv.*, 1991, vol. 23, no. 1, pp. 49-90.
[24] N. Lopes and A. Rybalchenko, "Distributed and Predictable Software Model Checking," in *Verification, Model Checking, and Abstract Interpretation*. vol. 6538, R. Jhala and D. Schmidt, Eds., ed: Springer Berlin Heidelberg, 2011, pp. 340-355.
[25] T. J. Lehman and J. H. Kaufman, "OptimalGrid: middleware for automatic deployment of distributed FEM problems on an Internet-based computing grid," in *Cluster Computing, 2003. Proceedings. 2003 IEEE International Conference on*, 2003, pp. 164-171.
[26] H. Garavel, *et al.*, "CADP 2011: a toolbox for the construction and analysis of distributed processes," *International Journal on Software Tools for Technology Transfer*, 2013, vol. 15, no. 2, pp. 89-107.
[27] S. Grumbach and F. Wang, "Netlog, a Rule-Based Language for Distributed Programming," in *Practical Aspects of Declarative Languages*. vol. 5937, M. Carro and R. Peña, Eds., ed: Springer Berlin Heidelberg, 2010, pp. 88-103.
[28] M. Razavi and K. Zamanifar, "Cost-Effective Computing in the Cloud Infrastructure," *International Journal of Future Computer and Communication*, 2013, vol. 2, no. 3, pp. 184-187.
[29] R. Shanmugam and M. A. M. Mohamed, "Data Management in the Mobile Cloud Using Surrogate Object," *International Journal of Future Computer and Communication*, 2012, vol. 1, no. 2, pp. 187-192.
[30] R. Srivastava, "Evaluation of Response Time Using Gang Scheduling Algorithm for B2C Electronic Commerce Architecture Implemented in Cloud Computing Environment by Queuing Models," *International Journal of Future Computer and Communication*, 2013, vol. 2, no. 2, pp. 71-75.
[31] R. Shanmugam, *et al.*, "Evolving a Surrogate Model of Transaction Managementfor Mobile Cloud," *International Journal of Future Computer and Communication*, 2012, vol. 1, no. 1, pp. 62-66.
[32] M. Mohamed, "An object based paradigm for integration of mobile hosts into grid," *International Journal of Next-Generation Computing*, 2011, vol. 2, pp. 1-23.
[33] K. Ørbekk, "Distributed Shared Objects for Mobile Multiplayer Games and Applications," Norwegian University of Science and Technology, 2012.
[34] R. Rodrigues and P. Druschel, "Peer-to-peer systems," *Commun. ACM*, 2010, vol. 53, no. 10, pp. 72-82.
[35] H. Liu, *et al.*, "Neighbor Selection in Peer-to-Peer Overlay Networks: A Swarm Intelligence Approach," in *Pervasive Computing*, A.-E. Hassanien, J. H. Abawajy, A. Abraham, and H. Hagras, Eds., ed: Springer London, 2010, pp. 405-431.
[36] M. M. Mohamed, *et al.*, "Surrogate Object Model: A New Paradigm for Distributed Mobile Systems," in *Proceedings of the 4th International Conference on Information Systems Technology and its Applications (ISTA'2005), May*, 2005, pp. 124-138.
[37] R. K. Joshi and D. J. Ram, "Anonymous remote computing: a paradigm for parallel programming on interconnected workstations," *Software Engineering, IEEE Transactions on*, 1999, vol. 25, no. 1, pp. 75-90.

DWDE-IR: An Efficient Deep Web Data Extraction for Information Retrieval on Web Mining

Aysha Banu[1], M. Chitra[2]
[1]Research Scholar, Anna University Chennai, ayshahusain11@gmail.com
[2]Professor, Department of Information Technology, Sona College of Technology, Salem,

Abstract— **Deep Web is a widely unexplored data source, is becoming an important research topic. Retrieving structured data from deep web pages is the challenging problem due to their complex structure. In this paper, Information extracts on the Deep Web pages based on the Deep Web Data Extraction technique (DWDR-IR). Search engines usually return a large number of pages in response to the user queries. To help the users to navigate in the result list, ranking methods are activated on the search results. In this paper, a page ranking mechanism called Coherence Ratio based Page (CRP) ranking algorithm is used. To retrieve the information accurately, an approach called WordNet is used. WordNet checks the similarity of data records and find the correct data region with higher precision using the semantic properties of data records. This concept is very important to display the valuable results occur on the top of the result list on the basis of browsing behavior of the user, it reduces the search space and provides high accuracy. This approach handles the visual features on the deep web data extraction, including data item extraction, data record extraction and visual wrapper generation. The proposed work removes all noise such as header, footer, irrelevant advertisement and irrelevant content using NoiSe Filter (NSFilter) algorithm. The proposed method retrieves perfect extraction of relevant results from the deep web pages. DWDE-IR results higher precision, recall and filter accuracy than the existing method ViDE.**

Index Terms— **Data item extraction, Data record extraction, Deep web data extraction, Ranking algorithm, Visual wrapper generation, WordNet.**

I. Introduction

Deep web is the web that is dynamically generated from the data sources such as file systems or databases. On surface web, data are available through URLs; whereas in deep web data are guarded by a search interface. Crawling deep web is the process of collecting hidden data by issuing queries through various search interfaces. It includes web services, HTML forms and programmable web APIs crawls deep web data sources is essential for several reasons, such as indexing deep web data sources or backing up data.

Most of the information available on the internet cannot be directly accessed via the static link; Users must type some keywords before getting the information hidden in the web database. Deep web contains 400-500 times more information and 15% larger visit capacity than that of surface web. In addition quality of data is also relatively higher. All the web databases make up the deep web; it may be hidden web or invisible web). Generally the query which is entered by the user is enclosed in web pages in the form of data records. These web pages are difficult to index by the traditional crawler based search engines like Google and Yahoo. Data records on a query result page display uniformity in their content, appearance and structure. They show structural and visual similarities. The displayed data in the primary database, and the data items for each record are retrieved from the database in response to the user's query, the same template is used to present the record.

The lexical database for the English is WordNet. It is used to check and find the meaning of words in their contents using the semantic relations of the words. Usually WordNet is used for information retrieval and this approach is also applicable to data extraction on deep web pages. Most of the existing work deals with the extraction of data records are based on the theme of identifying repeated patterns. They identify repeated patterns in the HTML source code of multiple training pages from a data source in order to infer a common structure or template and use it to extract structured data from new pages from the same source. As most of the web pages are written in HTML, it is Web page programming language dependent i.e. HTML dependent. Moreover HTML is no longer the exclusive web page programming language; the other languages have been introduced such as XHTML and XML. Simultaneously to attract human users' consumption of knowledge retrieved from search engines; best template designers of deep web pages always arrange the data records and the data items with visual uniformity to meet the reading habits of human beings. The Document Object Model (DOM) is a cross-platform and language independent convention for representing and interacting with objects in HTML, XHTML and XML documents.

In this paper, an efficient Deep Web Data Extraction for Information Retrieval technique (DWDE-IR) is proposed. Here Visual Block Tree is used to extract the semantic structure of a web page based on the visual presentation. Visual features are used to identify the special information present on the deep web pages. Then data record

doi:10.4304/jetwi.6.1.133-141

extraction phase includes the Filtering, Clustering and Regrouping process. Then Data Item Extraction phase consists of Record Segmentation and Item Alignment steps. Finally Visual Wrapper Generation includes Data Record wrapper and Data Item Wrapper process. The proposed work removes all the noise blocks present on the entire web page and also it reduces the time with high accuracy.

The remaining part of the paper is organized as follows: Section II involves the works related to the Deep Web Data Extraction. Section III involves the existing method ViDE description and problems of ViDE. Section IV involves the description of the proposed method DWDE-IR. Section V includes the performance analysis and explanation of results. The paper is concluded in Section VI.

II. Related Work

Wei et al proposed a vision-based approach that was Web-page-programming-language-independent. This method utilizes the visual features on the deep Web pages to implement deep Web data extraction including data record extraction and data item extraction. The evaluation measure was used to capture the amount of human effort needed to produce the extraction [1]. *Hong* proposed a method WordNet to check the similarity of data records and detect the correct data region with higher precision using the semantic properties of these data records. The proposed method extracts three types of data records namely single-section data records, multiple-section data records and loosely structured data records [2]. *Lin et al* proposed an algorithm for discerning the common path based on hierarchical DOM. Based on the common path and predefined regular expression the target data of the Deep Web was extracted [3]. *Han et al* proposed a Deep Web data extraction and service system based on the principle of cloud technology. A multi-node parallel computing system structure was adopted and a task scheduling algorithm was designed in the data extraction process. In this paper, the lask load was balanced among the nodes to accomplish the data extraction rapidly [4].

Srikantaiah et al proposed a similarity based web data extraction and integration system (WDES and WDICS) to extract search result pages from the web and integrate its contents to enable the user to perform intended analysis. The system provides for local replication of search result pages. This system results better precision and recall than Data Extraction based Partial Tree Alignment (DEPTA) [5]. *Li et al* proposed a Web data scheme and a domain data model. It also puts forward the web table positioning and web table records extracting based on web data schema and an integration algorithm based on the main data model [6]. *Huang et al* proposed an approach to minimize the communication cost to select the appropriate query. A set covering model was used to indicate the web database. An incremental harvest model was learnt by the machine learning method to select the appropriate query automatically [7]. *Hong et al* proposed an approach for information extraction based on fast heuristics techniques. Filtering rules was used to detect and filter out irrelevant data records. A tree matchnig algorithm was also used to increase the speed of data extraction. A data alignment algorithm was proposed to align iterative and disjunctive data items [8].

Kayed et al proposed an approach to measure the relevancy of retrieved web sites to user query concepts and rank them accordingly. a relevancy measure called ontology concepts was proposed. It hepls to re-rank the retrieved documents according to their relevancy to the search query [9]. WSD have been implemented in Lucene using query expansion with thesaurus and relevance feedback. The extended Lesk algorithm was re-implemented to disambiguate the query using WordNet. Expansion terms were limited up to 20 words chosen from expansion term candidates from disambiguated query's senses information, co-occurrence terms and most frequent terms using Kullback-Leibler Distance. The process gets iterated to find the best number of expansion iteration. This method provides better understanding of WSD in information retrieval system performance.

Du and Hai proposed an extension similarity and an intension similarity that analyzed a user's browsing patterns and their hyperlinks. Also the information content similarity between two nouns were compared automatically by examining their ISA and Part-Of hierarchy and using a user's web log. A method for computing the semantic similarity between two concepts in two different lattices and finding the semantic ranking of web pages is proposed. This method proved that the semantic ranking of web pages is useful and efficient for making a web crawlers choice of a web page [10]. *Palekar, et al* utilized the visual features of the deep Web pages to implement deep Web data extraction including data record extraction and data item extraction [11]. *Das and Kumar* designed a technique called Hidden Web Query Technique (HEET) for modelling and query the hidden web. Also a new approach to modelling of consecutive forms and concept of comprehensive form have been presented [12]. *Kayed and Chia Hui* proposed an unsupervised page-level data extraction approach to deduce the schema and templates for each individual deep websites, which contains either singleton or multiple data records in one web page. FiVaTech handles tree matching, mining techniques and tree alignment to achieve the challenging task. [13].

Sreekrishna et al proposed a semantic ontology based deep web data classification method (SODWEB). This method is used to classify the data in the deep web automatically. The URL is choosen for the process of analyzing the semantic association among the concepts. Then the URL of deep web search source was mapped to the category hierarchy obtained [14]. *Patel et al* proposed an effective Deep web data integration approach based on Schema and Attributes Extraction of Query Inetrfaces. This approach avoided the incorrect subsets while grouping attributes and is highly effective on schema extraction of source query interfaces on the invisible web [15]. *Anderson and Hong* proposed a novel approach for extracting the data records from deep web pages [16]. Based on structural regularity, visual and content

similarity between data records displayed on the query result page. This approach used to identify each data record individually while ignoring noise items such as navigation bars and advertisements. This approach resulted higher accuracy.

Miao et al proposed a method for record extraction that captures a list of objects in a more robust way based on a holistic analysis of a web page. It focused on the repetition of tag path appears on the DOM tree of the web document. This paper introduced a similarity measure that captures the visual signal. Tag paths are clustered based on the similarity measure and sets of tag path that form the structure od data records and extracted [17]. *Mukherjee, et al* classified users based on their internet usage patterns and for each class, maintain a cache of web documents. Searching the contents is based on term set analysis and direction cosine distance approach. Finally a more actualize and relevant result set were generated for the query [18]. *Bronzi, et al* proposed an innovative approach that aims at pushing further the level of automation of existing wrapper generation systems by leveraging the redundancy of data on the web. The result of this approach shown a relevant improvement in the precision of the extracted data without a significant loss in the result [19]. *Li and Xie* proposed a keyword-based user interface system EasyUI for achieving web-scale data integration and easy to use for ordinary users. Also the following challenging areas were addressed: indexing schemata terms, data values and domain features of the Deep Web, processing user input, mapping the user query to the domain that the user query most likely corresponds to, and translating the user query into candidate structured queries for the domain. [20].

III. Existing Method- ViDE

A. Description of ViDE

ViDE is the Vision Based Approach for Deep Web Data Extraction, that is Web-page-programming-language-independent methodology [1]. This approach uses the Visual Block Tree to extract the semantic structure of a web page based on the visual presentation. Visual features are used to identify the special information present on the deep web pages. It includes the Data Record Extraction process such as Filtering, Clustering and Regrouping. Filtering removes the top and bottom records of the web page. Clustering the blocks is based on block similarity. The regrouping process uses the regrouping algorithm to regain the blocks. Data Item Extraction phase includes Record Segmentation and Item Alignment process. Record segmentation uses the heuristic rules to segment the records. Data Item alignment focuses on the problem of how to align the data items of the same semantic together and also keep the order of the data items in each data record. Finally, the Visual Wrapper Generation process includes data record wrapper and the data item wrapper. The data record wrapper locates the data region in the Visual Block tree and then extracts the data records from the child blocks of the data region. The data item wrapper takes the attributes $\{a_1, a_2,\ldots,a_n\}$ which are obtained from the sample page and a sequence of data items $\{item_1, item_2, \ldots, item_m\}$ obtained from a new data record, the wrapper process the data items in order to decide which attribute the current data item can be matched. This process is much faster than the wrapper generation process.

B. Problems in ViDE

Extracting structured data from deep web pages is a challenging problem due to their complex structure. Until now, there are a large number of techniques have been proposed to overcome this problem. But all of them have some limitations because they are Web-page-programming-language dependent. ViDE system can only process deep web pages containing only one data region not in multidata-region. It removes the noise blocks on the top and bottom of the web pages and does not guarantee the removal of all the noise blocks. Record segmentation process is based on Visual Block tree. ViDE system takes a long time and low efficiency in extraction.

IV. DWDE-IR: Deep Web Data Extraction for Information Retrieval

A. Basic Concepts about Web Page and Layout

The web pages consist of different kind of information in the form of texts, images, flash, video etc. The Web Page Layout is represented as a coordinate system. The origin locates at the top left corner of the web page. The x-axis is represented as left - right horizontal position and the y-axis are represented as top - bottom vertical position. Suppose each text or image is incorporated in a minimum bounding rectangle with their sides are parallel to the axes, then a text or image has an exact coordinate *(x , y)* on the web page. Here *x* refers horizontal distance from the origin to the left side of its corresponding rectangle and *y* refers vertical distance from the origin to the upper side of its corresponding rectangle. The x and y coordinates and sizes of texts or images on the Web page from the Web Page Layout.

B. Visual Block Tree and Visual Features

Vision-based Page Segmentation Algorithm (VIPS) [1] aims to extract the semantic structure of a web page based on its visual presentation. Such semantic structure is usually called as a tree structure. Each node in the tree correlates to a block. Each node will be assigned a value called the Degree of Coherence. It is used to indicate how coherent of the content in the block based on visual perception. The algorithm extracts all the suitable blocks from the HTML DOM tree and then it finds the separators between these blocks. Separators denote the horizontal or vertical lines in a web page that visually cross with no blocks. Based on these separators the semantic tree of the web page is constructed so that the web page can be represented as a set of blocks. Noisy information such as navigation, advertisement and other decoration can be easily removed because they are often placed in certain positions on a page. Contents with different topics are categorized as separate blocks.

1) Visual Features

Usually web pages are used to broadcast information to users like other kinds of media such as newspaper and TV. Visual features are very important to identifying special information on web pages. Deep web pages contain data records which are retrieved from the web databases.

The visual features are recognized based on its four features:

1. Position features (PF)
2. Layout features (LF)
3. Appearance features (AF)
4. Content features (CF)

Position features demonstrate the location of the data region on deep web page**.** Data regions are always placed centered horizontally. Layout features show that how the data records in the data region is commonly arranged. Appearance feature catches the visual features within data records. Content features trace the regularity of the contents in data records.

Fig.1. (a) Example for Deep Web Page

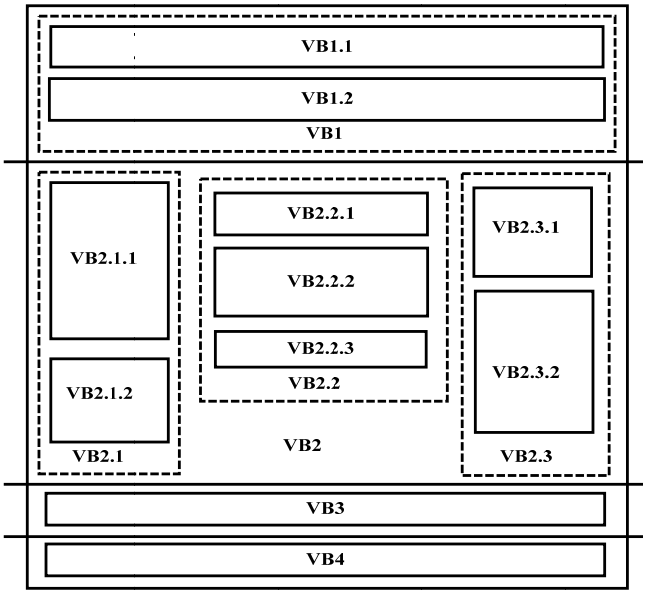


Fig.1. (b) Visual Presentation Structure

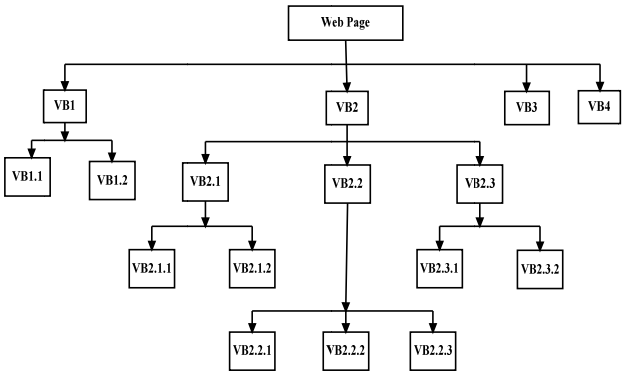


Fig.1. (c) Visual Block Tree

C. Data Record Extraction

Data record extraction aims to locate the boundary of data records and extract them from the deep web pages. So the following three techniques are incorporated:

1) Filtering – NSFilter Algorithm

Using NSFilter algorithm, all the noise blocks located on the web page are removed. Noise blocks are present on anywhere on the web page body. Existing system removes only the noise blocks present on the top and bottom of the webpage. To overcome this limitation proposed system introduced the NSFilter algorithm.

---

for each LF_i in *LF*
  If LF_i = *center*
    Add into *FB;* *//removes header, footer, right left noises*
end for

for each *FB* in new
  if *FB = Error*
  FB = FB - FB$_i$ *// removes advertisements and unwanted contents*
end for

Here *FB* is the filtered block list. Error contains the unwanted information blocks which are to be removed. Data regions are always placed centred horizontally. NSFilter removes all the noises such as header, footer, advertisements and other unwanted information's present on the webpage.

2) Clustering- Aggregate Clustering Approach

The filtered blocks are clustered using an Aggregate cluster approach based on the block and value similarity. In the existing system, similarity can be calculated using how many blocks present in the same type of the web page. It depends based upon the block count, not a value presented on the web page. To overcome this problem proposed system clusters the block based on both block similarity and value similarity. The equation for computing the appearance similarity between two blocks b_1 and b_2 is given below:

$$Sim(b_{s1}, b_{s2}) = w_i * simImage(b_{s1}, b_{s2}) + w_{pt} * simPText(b_{s1}, b_{s2}) + w_{lt} * simLText(b_{s1}, b_{s2}) \quad (1)$$

where *simImage*(b_{s1}, b_{s2}), *simPText* (b_{s1}, b_{s2}), and *simLText* (b_{s1}, b_{s2}) are the similarities established on image size, plain text font and link text font. w_i , w_{pt} , and w_{lt} are the weights of these similarities.

The equation for computing the appearance similarity between two values v_1 and v_2 of two blocks (b_{s1}, b_{s2}) is given below:

$$Sim\ (v_1, v_2) = w_t * semsim\ (v_1, v_2) \quad (2)$$

where *simImage* (v_1, v_2), *simPText* (v_1, v_2) and *simLText* (v_1, v_2) are the similarities based on Image tag, Plain text content and Link text content. w_i , w_{pt} , and w_{lt} are the weights of these similarities.

$$SimImage(b_{s1}, b_{s2}) = \frac{\min\{Ta_i(b_{s1}), Ta_i(b_{s2})\}}{\max\{Ta_i(b_{s1}), Ta_i(b_{s2})\}} \quad (3)$$

$$w_i = \frac{Ta_i(b_{s1}) + Ta_i(b_{s2})}{Ta_b(b_{s1}) + Ta_b(bs_2)} \quad (4)$$

$$SimPText(b_{s1}, b_{s2}) = \frac{\min\{Tf_{pt}(b_{s1}), Tf_{pt}(bs_2)\}}{\max\{Tf_{pt}(b_{s1}), Tf_{pt}(b_{s2})\}} \quad (5)$$

$$w_{pt} = \frac{Ta_{pt}(b_{s1}) + Ta_{pt}(b_{s2})}{Ta_b(b_{s1}) + Ta_b(b_{s2})} \quad (6)$$

$$SimLText(b_{s1}, b_{s2}) = \frac{\min\{Tf_{lt}(b_{s1}), Tf_{lt}(b_{s2})\}}{\max\{Tf_{lt}(b_{s1}), Tf_{lt}(b_{s2})\}} \quad (7)$$

$$w_{lt} = \frac{Ta_{lt}(b_{s1}) + Ta_{lt}(b_{s2})}{Ta_b(b_{s1}) + Ta_b(b_{s2})} \quad (8)$$

$$w_t = \frac{Tot\ text\ v_1 + Tot\ text\ v_2}{Tot\ b_{s1} + Tot\ b_{s2}} \quad (9)$$

Here $Ta_i(b)$ represents the total area of images available in block b. $Ta_b(b)$ denotes the total area of block b. $Tf_{pt}(b)$ denotes the total number of fonts of the plain text available in block b. $Ta_{pt}(b)$ is the total area of plain text in block b. $Tf_{lt}(b)$ denotes the total number of fonts of the link text available in block b. $Ta_{lt}(b)$ is the total area of link text in block b.

Clustering the blocks based on blocks and value similarities:

$$Sim(b_1, b_2) = sim\ (bs_1, bs_2) + sim\ (v_1, v_2) \quad (10)$$

Where $Sim(b_1, b_2)$ denotes the similarity measure between the blocks and value. The nearest value among the blocks are clustered based on the eqn(10)

3) Regrouping- MERGE Method

The clustered blocks are needed to regroup, such that the blocks belonging to the same data record form a group. Existing system uses the manual process of regrouping according to their position. The result of the existing system is not so accurate for regrouping. So in this work MERGE method is used to merge the blocks based on their tags, value and data block.

Algorithm: Regrouping- MERGE METHOD

Input: $k_1, k_2 \ldots k_m$: group of clusters generated by blocks clustering from a given deep web page *A*.
Output: $d_1, d_2 \ldots d_m$: each of them corresponds to data record on *A*.
Begin
//Step1: Sort the given blocks in k_i according to their position
(based on top to bottom; then left to right)
1: **for** each cluster k_i do
2: **for** any two blocks $b_{i,j}$ and $b_{i,k}$ in k_i
//$1 \le j \le k \le |k_i|$
3: **if** $b_{i,j}$ and $b_{i,k}$ are in different lines on *A*, and $b_{i,k}$ is above $b_{i,j}$
4: $b_{i,j} \leftrightarrow b_{i,k}$; //exchange their orders in k_i
5: **else if** $b_{i,j}$ and $b_{i,k}$ are in the same line on *A*, and $b_{i,k}$ is in front of $b_{i,j}$
6: $b_{i,j} \leftrightarrow b_{i,k}$;
7: **else**
8: if sim ($b_{i,j}$, $b_{i,k}$) > threshold
9: merge ($b_{i,j}$, $b_{i,k}$);
10: **end** until no exchange occurs;
11: form the minimum-bounding rectangle Rct_i for k_i;

//Step2: Initialize g groups and g is the number of data records on *A*
12: $k_{max=} \{k_i \mid |k_i| = \max\{|k_1|, |k_2| \ldots |k_m|\}\}$; //$g = |k_{max}|$
13: for each block $b_{max,i}$ in k_{max}
14: initialize group V_i;
15: put $b_{max,i}$ into V_i;

//Step 3: Insert the blocks into right groups and each group corresponds to a data record
16: **for** each cluster k_i
17: **if** Rct_i overlaps with Rct_{max} on *A*
18: **if** Rct_i is ahead of (behind) Rct_{max}
19: **for** each block $b_{i,j}$ in k_i
20: find the nearest block $b_{max,k}$ in k_{max} that is behind $b_{i,j}$ on the web page;
21: place $b_{i,j}$ into group V_k;
End

D. Data Item Extraction

In this paper, the problem of segmenting the data records into a sequence of data items and aligning the data

items of the same semantics together are focused. To address this issue, two techniques are used.

1) Data Record Segmentation- MSEGMENT Algorithm

The regrouped records are taken as inputs to segment the data records. The records within a data area are identified as sub-trees. The sub-trees rooted through children of the data area root. The segmentation process depends on record separators i.e sub-trees interleaved with data records. The root node of a record separator is a child of the data area root and it does not contain any text or URL. Since each of the record contains only one instance of matching node (MN), the segmentation process heuristically identifies the areas belonging to a single record.

Steps for MSegment Algorithm

Input: Regrouped records
Output: Segmented records

```
1   Consider CR
2   Expand CR
3   Consider S of CR
4   If S ϵ RS
5       Consider CR as PR
6   Else
7       Compute dist between two CR
8       Consider all 2 x (dist-1) to left and right ST
9       Repeat step 8 to all CR and compute similarity
    between the identified expansion
10      If several expansion have same similarity
11          Choose the highest similarity among R
```

Here CR denotes the candidate record, RS denotes record segmentation, S is the siblings of the CR, PR denotes proper record and R denotes records.

In the first step of the algorithm, each sub tree in the DOM containing a single MN and rooted at a direct child of the data area root is considered as a CR. A successive step tries to expand the CR to adjacent sub trees in the DOM. Therefore consider the S of CR. If they are RS, consider each CR as a PR; otherwise compute the distance between CR as the number of S between their root nodes. Then consider all the 2 x (dist-1) expansions of a record to left and right adjacent sub trees. Apply the same expansion to all the CR and compute the similarity. Choose the highest structural similarity among R.

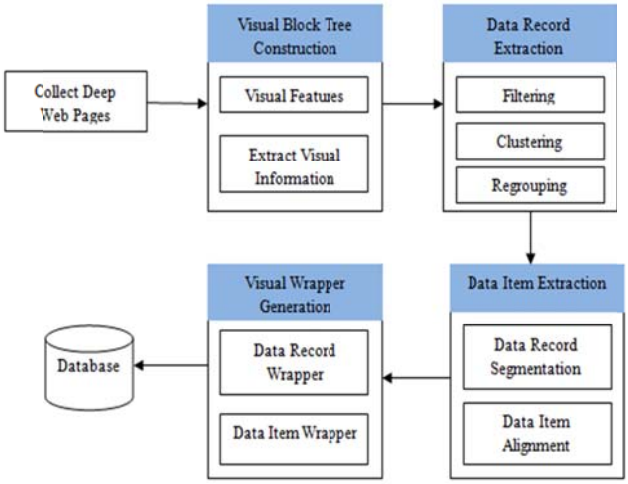


Fig.2. Extracting information from the web database

2) Data Item Alignment- Nested Structure

The data records are represented as $\{r_1, r_2 \ldots r_n\}$ and each data record r_i is denoted as a sequence of data items $\{item_1, item_2 \ldots item_m\}$. The entire data item has a unique position in its corresponding sequence according to the semantic order.

Data item alignment focuses on the problem of how to align the data items of the same semantic together and also keep the order of the data items in each data record. Proposed work uses the nested structure for data item alignment.

Given a column c_1 which contains n data values, the intra-column similarity $similarity_{intra}$ be the average data value similarity within each column in c_1

$$Similarity_{intra} = 2\,\frac{\sum_{q=1}^{m-1}\sum_{p=q+1}^{m} L_{pq}}{m(m-1)} \qquad (11)$$

Here L_{pq} is the data value similarity between pth and qth data values of c_1.

For two columns c_1 and c_2 which has f and g data values respectively, the inter-column similarity $similarity_{inter}$ is defined to be the average data value similarity of every pair of data values in c_1 and c_2.

$$Similarity_{inter} = \frac{\sum_{q=1}^{g}\sum_{p=1}^{f} L_{pq}}{fg} \qquad (12)$$

Here L_{pq} is the data value similarity between pth data values of c_1 and qth data values of c_2 using the data value similarity. After $similarity_{intra}$ and $similarity_{inter}$ are calculated for identified columns c_p , if $similarity_{inter}$ / $similarity_{intra} > S_{nest}$ is a threshold that is set to 0.5, c_p is assumed to be a nested column set, which means that the data values in it are generated from a nested structure.

E. Visual Wrapper Generation-MViDE

Wrapper Generation includes data record wrapper and the data item wrapper. They are the programs which execute the data record extraction and data item extraction with a set of parameter extracted from sample pages. Each Web database a normal deep Web page containing the

maximum number of data records are used to generate the wrappers.

1) Data Record Wrapper

For each of the record, visual data record wrapper find the first block of each record and the last block of the last data record (b_{LDR}). To achieve this objective, the visual information of the first block of each data record extracted from the sample page is saved and the distance (t) between two data records are also saved. For the child blocks of the data region in a new page, find the first block of each data record by the visual similarity with the saved visual information. Then b_{LDR} on the new page needs to be placed. The vertical distance between any two neighboring blocks in one data record is always smaller than t and the vertical distance between b_{LDR} and its next block is not smaller than t. Hence the first block is recognized whose distance with its next block is larger than t as b_{LDR}.

2) Data Item Wrapper

The data alignment algorithm groups data items from different data records into columns or attributes such that data items under the same column have the same semantics. Given a sequence of attributes $\{t_1,t_2,..,t_n\}$ obtained from the sample web page and a sequence of data items $\{d_1,d_2,\ldots,d_m\}$ which is obtained from a new data record. The data items are wrapped in order to decide which of the attribute matches the current data item. For d_i and t_j, if they are same as *T,B* and *A* then their match is recognized. The wrapper then checks whether d_{i+1} and t_{j+1} if not, it checks d_i and t_{j+1}. Repeat the process until all the data items are matched to their corresponding attributes.

T is the font used by the data items, *B* denotes Boolean, and *A* is the image, text, number, date etc.

F. Query Processing

The data record list and item are extracted from the database. Then the extracted result of the database is compared with the user query. The co-occurrence ratio is calculated for the user query and the extracted data record. Co-occurrence ratio1 is calculated using query and the key term. Co-occurrence ratio2 is calculated using query and the extracted record list. When the information is queried through a search engine, the result will be presented as a list, and the result which best satisfies the user demand will appear at the top of the list. The web pages are displayed based on the Page Ranking Algorithm.

Co-occurrence Ratio1

$$R1 = \sum_{i=1}^{m} Sim(q, t_i) \quad (13)$$

where *Sim* is the similarity measure between query q and key term t_i.

Co-occurrence Ratio2

$$R2 = \sum_{i=1}^{m} SEMSim(q, t_i) \quad (14)$$

where *SEMsim* is the semantic similarity between query q and t_i is the extracted record list based on WordNet [2].

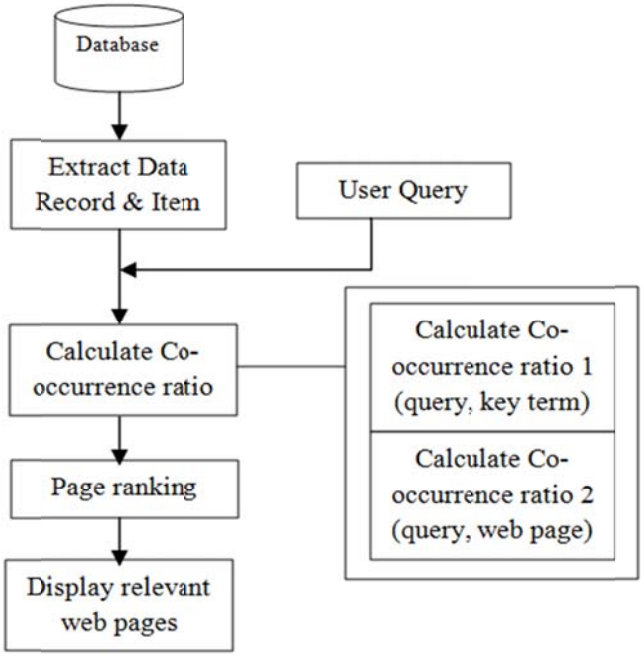


Fig.3. Flow for generating accurate results

Algorithm-CRP Ranking

Input: Query, Extracted record list q;
Output: Score List
Convert q into t1, t2… t_m;
for (j=1; j<=n; j++)
  Calculate R1 using eqn (1)
  Calculate R2 using eqn (2)
C[j] = k_1 * R1 + k_2 * R2;
end for

Select the top k results from the result set, select (R, K);
Here k_1 and k_2 are the weighting coefficient, $k_1 + k_2=1$.

Based on this algorithm the similarity measures are calculated and the web pages are displayed based on the best score among the matched web pages.

V. PERFORMANCE ANALYSIS

In this paper the performance quality is measured using Precision, Recall, Filter Accuracy and search time between the proposed system DWDE-IR and ViDE. The implementation is carried out by taking the two datasets like GDS and SDS.

A. Precision

Precision is the fraction of retrieved instances that are relevant. Precision is calculated correctly based on the extracted data items and records with the total number of extracted data records and items.

$$\textit{Precision for data record extraction} = \frac{Rec_c}{Rec_e} \quad (15)$$

where Rec_c is the total number of correctly extracted data records and Rec_e is the total number of data records extracted.

$$\textit{Precision for data item extraction} = \frac{Item_c}{Item_e} \quad (16)$$

where $Item_c$ is the total number of correctly extracted data items and $Item_e$ is the total number of data items extracted.

Fig.4. and Fig5 are the precision output for the existing technique ViDE and the proposed method DWDE-IR for record and item extraction.

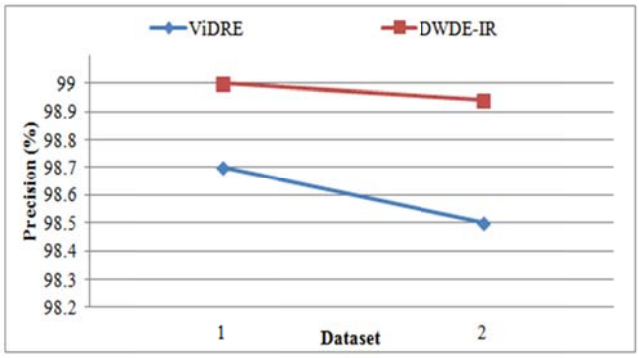


Fig.4. Precision between ViDE and DWDE-IR for record extraction

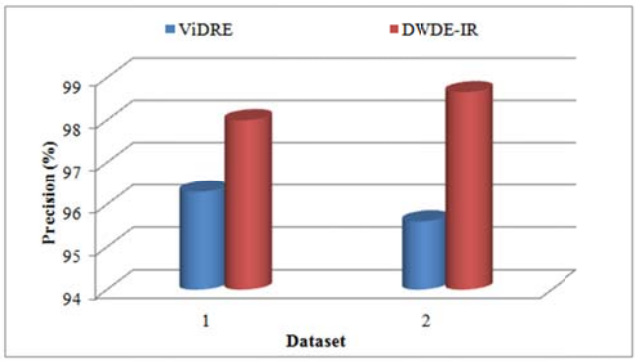


Fig.5. Precision between ViDE and DWDE-IR for item extraction

B. Recall

Recall is the fraction of relevant instances that are retrieved. Recall is calculated correctly based on the extracted data items and records with the total number of data records and items.

$$\textit{Recall for data record extraction} = \frac{Rec_c}{Rec_d} \quad (17)$$

where Rec_c is the total number of correctly extracted data records and Rec_d is the total number of data records.

$$\textit{Recall for data item extraction} = \frac{Item_c}{Item_d} \quad (18)$$

where $Item_c$ is the total number of correctly extracted data items and $Item_d$ is the total number of data items.

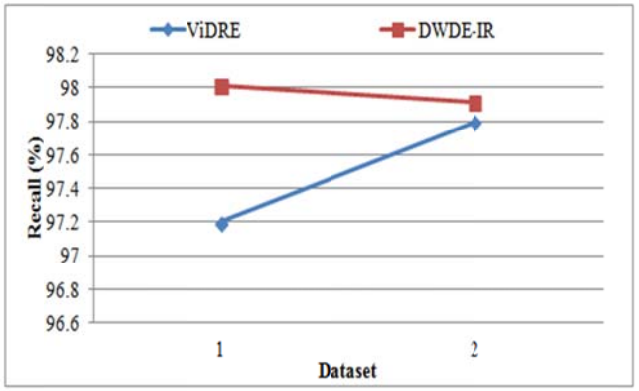


Fig.6. Recall between ViDE and DWDE-IR for record extraction

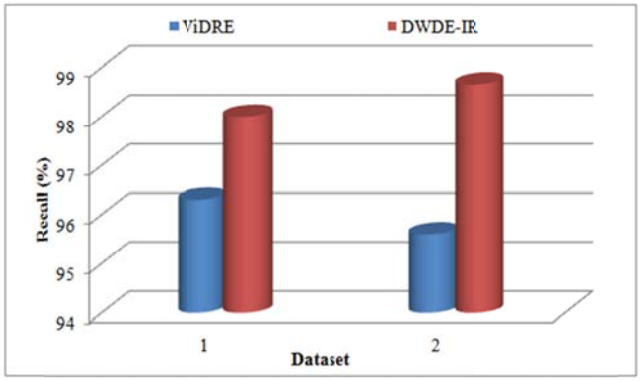


Fig.7. Recall between ViDE and DWDE-IR for item extraction

Fig.6. and Fig.7 shows the comparison of recall calculated for record and item extraction between ViDE and DWDE-IR. The result shows that the proposed system DWDE-IR generates better relevant instances than the existing system ViDE.

C. Filter Accuracy

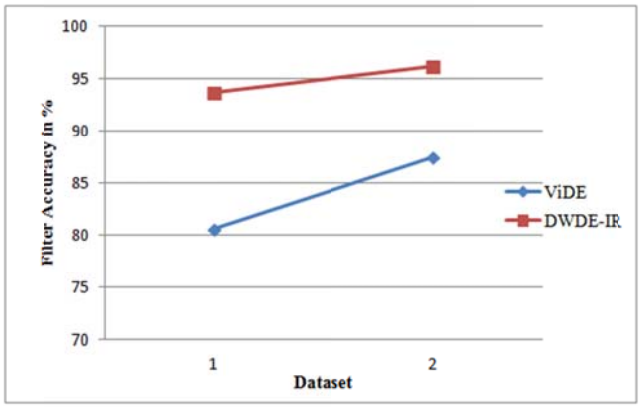


Fig.8. Filter Accuracy between ViDE and DWDE-IR

DWDE-IR generates more accurate result than the existing ViDE approach. DWDE-IR removes the noise present on anywhere on the webpage; whereas ViDE removes only the noise present on the top and bottom of the webpage.

D. Search Time

The searching time for the proposed DWDE-IR consumes less time when compared with the existing ViDE. It is shown in Fig.9.

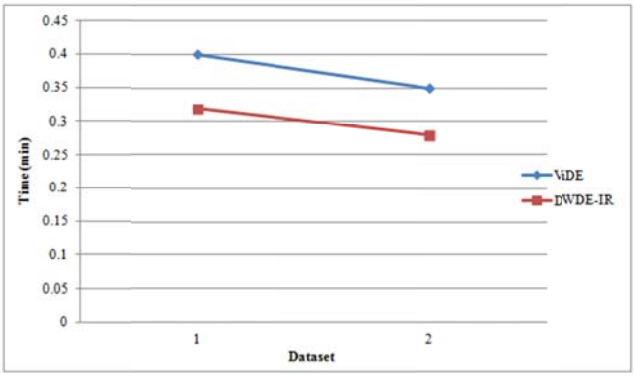


Fig.9. Search time between ViDE and DWDE-IR

VI. Conclusion And Future Work

The problem of Deep Web Data extraction has received a lot of problems in recent years and most of the proposed solutions are based on HTML source code. In this paper, an efficient Deep Web Data Extraction for Information Retrieval technique (DWDE-IR) is proposed. The proposed system is html independent and can even process deep web pages containing multiple data region. DWDE-IR removes all the noise blocks such as header, footer, irrelevant advertisement and irrelevant content from the webpage. It uses the page ranking algorithm to display the best suited result at the top of the result list. The processing time is quite less and also produces high accuracy results. The proposed system results higher precision, recall and filter accuracy than the existing approaches.

In future, the different types of wrapper generation method are applied to test better data record wrapping and data item wrapping. Then best clustering technique is incorporated to attain an accurate information retrieval.

References

[1] L. Wei, M. Xiaofeng, and M. Weiyi, "ViDE: A Vision-Based Approach for Deep Web Data Extraction," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 22, pp. 447-460, 2010.

[2] J. L. Hong, "Data extraction for deep web using wordnet," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on,* vol. 41, pp. 854-868, 2011.

[3] T. Lin, B. H. Qiang, S. Long, and H. Qian, "Deep Web Data Extraction Based on Regular Expression," *Advanced Materials Research,* vol. 718, pp. 2242-2247, 2013.

[4] Z. Y. Han, F. Y. Wang, P. Sun, and Z. Y. Li, "A Deep Web Data Extraction and Application System Based on Cloud Technology," *Advanced Materials Research,* vol. 756, pp. 2583-2587, 2013.

[5] K. Srikantaiah, M. Suraj, K. Venugopal, S. Iyengar, and L. Patnaik, "Similarity Based Web Data Extraction and Integration System for Web Content Mining," in *Advances in Communication, Network, and Computing*, ed: Springer, 2012, pp. 269-274.

[6] G. Li, Z. Y. Han, Z. X. Chen, Z. Y. Li, and P. Sun, "Web Data Extraction and Integration in Domain," *Advanced Materials Research,* vol. 756, pp. 1585-1589, 2013.

[7] Q. Huang, Q. Li, H. Li, and Z. Yan, "An Approach to Incremental Deep Web Crawling Based on Incremental Harvest Model," *Procedia Engineering,* vol. 29, pp. 1081-1087, // 2012.

[8] J. L. Hong, E.-G. Siew, and S. Egerton, "Information extraction for search engines using fast heuristic techniques," *Data & Knowledge Engineering,* vol. 69, pp. 169-196, 2// 2010.

[9] A. Kayed, E. El-Qawasmeh, and Z. Qawaqneh, "Ranking web sites using domain ontology concepts," *Information & Management,* vol. 47, pp. 350-355, 12// 2010.

[10] Y. Du and Y. Hai, "Semantic ranking of web pages based on formal concept analysis," *Journal of Systems and Software,* 2012.

[11] V. Palekar, M. Ali, and R. Meghe, "Deep Web Data Extraction using Web-Programming-Language-Independent Approach," *Journal of Data Mining and Knowledge Discovery,* vol. 3, 2012.

[12] N. N. Das and E. Kumar, "Hidden Web Query Technique for Extracting the Data From Deep Web Data Base," in *Proceedings of the World Congress on Engineering and Computer Science*, 2012.

[13] M. Kayed and C. Chia Hui, "FiVaTech: Page-Level Web Data Extraction from Template Pages," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 22, pp. 249-263, 2010.

[14] M. Sreekrishna, P. Sundaramoorthy, and T. Rajendran, "A Novel Approach for Semantic Based Deep Web Classification," *International Journal of Advanced and Innovative Research (IJAIR),* vol. Vol.2 pp. 394-402, 2013.

[15] G. Patel, A. S. Rajawat, and S. Vyas, "Deep web Data Integration Approach Based on Schema and Attributes Extraction of Query Interfaces," *International Journal of Managment, IT and Engineering,* vol. 2, pp. 272-284, 2012.

[16] N. Anderson and J. Hong, "Visually extracting data records from the deep web," in *Proceedings of the 22nd international conference on World Wide Web companion*, 2013, pp. 1233-1238.

[17] G. Miao, J. Tatemura, W.-P. Hsiung, A. Sawires, and L. E. Moser, "Extracting data records from the web using tag path clustering," in *Proceedings of the 18th international conference on World wide web*, 2009, pp. 981-990.

[18] I. Mukherjee, V. Bhattacharya, S. Banerjee, P. Gupta, and P. Mahanti, "Efficient web information retrieval based on usage mining," in *Recent Advances in Information Technology (RAIT), 2012 1st International Conference on*, 2012, pp. 591-595.

[19] M. Bronzi, V. Crescenzi, P. Merialdo, and P. Papotti, "Wrapper generation for overlapping web sources," in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on*, 2011, pp. 32-35.

[20] Y. Li and C. Xie, "A easy user interface of IR system over large scale deep web," in *Mechatronic Science, Electric Engineering and Computer (MEC), 2011 International Conference on*, 2011, pp. 250-253.

An Adaptive Weight Calculation based Bandwidth Allocation Scheme for IEEE 802.16 Networks

Akash Deep
UIET, Panjab University, Chandigarh, India
akashdeep@pu.ac.in

KS Kahlon
Department of CSE, GNDU Amritsar, India
kskahlon@yahoo.com

***Abstract*— WiMAX standard also known as IEEE 802.16 is a wireless MAN standard driven by WiMAX Forum. WiMAX forum defines specification for QoS provisioning for manufacturing of WiMAX equipment by different vendors but the amount of bandwidth allocated to specific service in WIMAX is still an open issue. A number of algorithms have been presented to satisfy the diverse QoS requirements for traffic in WiMAX networks. The focus of all these algorithms is towards real time traffic. Non real time traffic like nrtPS and BE is quite often neglected. This paper is an attempt towards fair allocation of resources to non real time traffic simultaneously taking care of real time traffic also. The approach uses uncertainty principles of fuzzy logic to calculate new weights for different flows and makes allocation based on these new weights. Simulation results show that the proposed scheme was able to meet required performance levels for both kinds of traffic.**

***Index Terms*—Scheduling, IEEE 802.16, fuzzy, weight**

I. INTRODUCTION

Ever increasing demand for multimedia application has shifted the focus of wireless fraternity to IEEE 802.16 standard, also known asWiMAX,. It is designed to provide both wired and wireless broadband access in Metropolitan areas [1] [2] guaranteeing minimum quality of service (QoS) for both real and non real types of applications. In order to make resource allocation effective WiMAX support scheduling based channel access in which different resources like time, frequency are allocated in proportion to service priorities calculated on the basis of bandwidth request made by different SS.

Support for varying applications is provided by associating each connection between BS and SS to five different service classes as defined by the standard. UGS

Unsolicited grant service (UGS) applications are fixed size real time applications with streaming on a periodic basis, e.g. VoIP without silence suppression. Real time polling service (rtPS) provides real time applications with variable-size data packets streaming on a periodic basis, e.g. MPEG video. Non-real time polling service (nrtPS) is for non real time Variable Bit Rate applications that are delay tolerant and require minimum bandwidth guarantee e.g. FTP. Best Effort service(BE) is used for best-effort traffic such as HTTP or www. A new service flow called extended real-time Polling was added in IEEE 802.16e standard to incorporate the advantages of both UGS and rtPS. The signaling mechanism in WiMAX permits SS to predict its bandwidth requirements and request for it from BS and BS makes the allocation on the basis of bandwidth requested and QoS parameters for the connection.

Each frame in WiMAX is divided into two parts : downlink and uplink subframe. Downlink subframe is used by BS to send control and data information to different SS while uplink subframe is utilized by SS for data transmission. Time slots are granted to a connection depending on service flow of connections.IEEE 802.16 standard defines QoS service framework for different services as shown in figure 1. but how much amount is to be allocated for each service has been left as open issue to be explored and implemented by equipment manufacturers. This paper is an attempt in this direction. The proposed scheme uses the dynamism provided by fuzzy logic and can be easily implemented in base stations. This paper is organized as follows. Section-II is devoted to studies available in this direction followed by proposed scheme in the next section. Simulations and results are given in next section and finally conclusion is presented.

II. RELATED WORK

Problem of bandwidth allocation and scheduling of resources together form one of the major challenges in a packet switched network like WiMAX and literature is rich enough with a number of studies in this direction. For the design of new approach authors has studied a number of papers few of these are listed in this sections.

Shreedhar andVarghese[3] presented a changed form of round robin method and named it deficit round

doi:10.4304/jetwi.6.1.142-147

robin(DRR) whch they hope will provide O(1) complexity if specific constraints are met. The scheduler has disadvantage as only one packet can be served from the queue which may lead to real time packet missing their deadlines.

Wongthavarawat et al.[4][5] utilized multiple algorithms for scheduling of different service classes. After making fixed allocations to UGS, rtPS and nrtPS was scheduled using Earliest Deadline First (EDF) and Weighted Fair Queuing (WFQ) respectively. RR was used for BE traffic. The inter-class scheduling was done using fixed

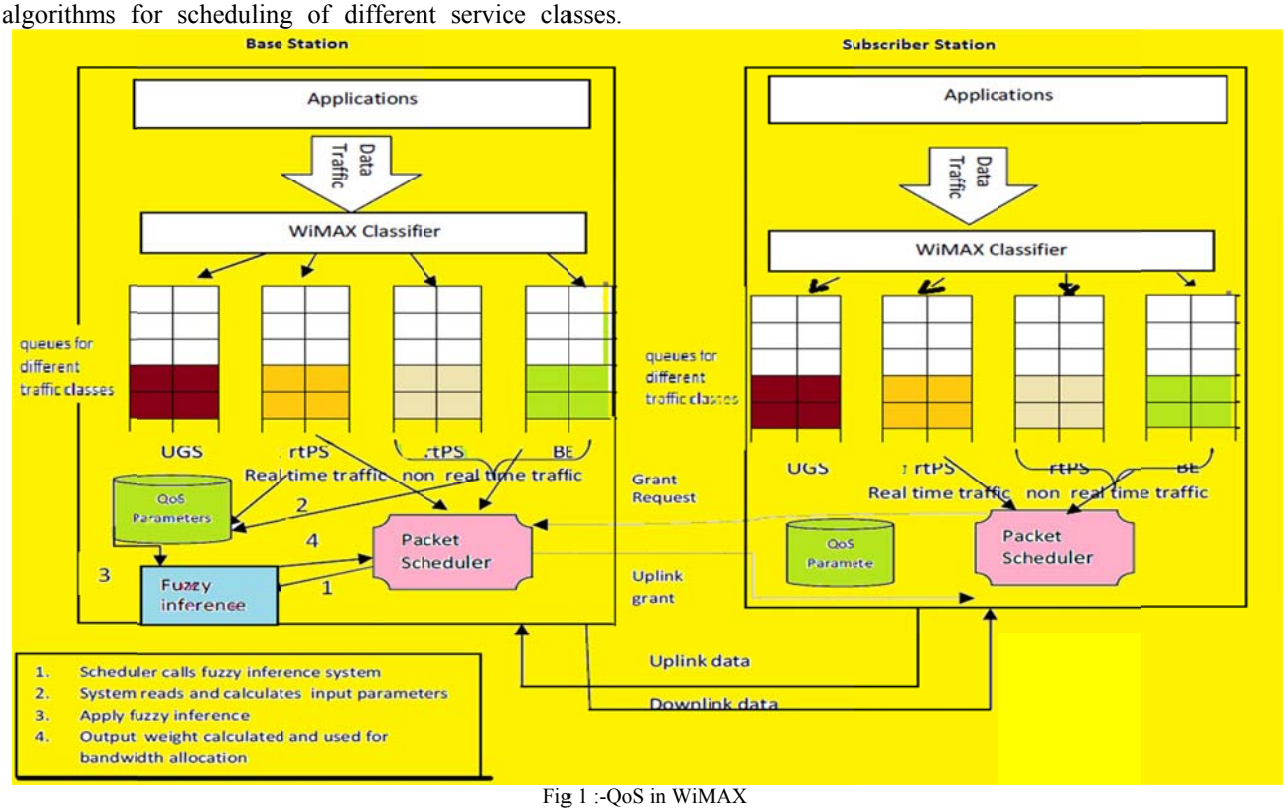


Fig 1 :-QoS in WiMAX

priority in the order of UGS>>rtPS>>nrtPS>>BE. A connection admission mechanism using token bucket has also been proposed.

Ball et al. [6] proposes a RR scheduler instead of assigning priority to real time applications tries to allocate resources to all connection in a round robin fashion. The results indicate that although BE service performs good but performance for real time application is not good enough however algorithm simplifies the complexity as no decision time was required.

Authors of [7] and [8] has proposed weighted round-robin (WRR) and earliest deadline first (EDF) schemes. The transmission in EDF is as per their associated deadlines while WRR provides allocation in round robin manner but different flows or connections are associated weights. Proportional fairness (PF) scheme was implemented by etl al in[9]. Allocation in PF is to the connection with highest priority which may lead to starving of low priority flows. Other schemes like Weighted fair queuing (WFQ) in [10] and modified largest weighted delay first (M-LWDF) in [11] scheme are also available.

J. Sun *et al*[12] proposed two different schedulers for BS and SS. Priorities to UGS and bandwidth requests opportunities for rtPS and nrtPS connections were assigned at the time of connection setup. Data grants for rtPS, nrtPS classes were scheduled taking into account bandwidth request information and their minimum requirements. The residual bandwidth was distributed in accordance to pre-assigned weights. Fixed priority scheme was implemented at SS assigning priorities of 1,2,3 and 4 for for BE, nrtPS, rtPS and UGS service classes respectively. UGS is allocated guaranteed bandwidth at the first instant, deadlines for rtPS packets were then calculated based on arrival time and tolerated delay and were scheduled on the basis of approaching deadlines. For nrtPS packets each packet is associated with virtual time that is calculated to guarantee minimum reserved bandwidth and all BE class were scheduled in FIFO manner.

N Liu *et al* [13] presented another scheme that uses combination of three different schedulers to meet Qos requirements of various classes. Scheduler I serves time sensitive traffic streams like UGS, rtPS and nrtPS and uses EDF algorithm. WFQ was used to schedule minimum bandwidth reserving flows like nrtPS where weights were in proportion to bandwidth requirements while BE also employs WFQ scheduling technique which was implemented by scheduler 3. Weights in scheduler 3 correspond to traffic priorities of each BE connection and these schedulers were served in fixed priority order with scheduler I being assigned highest priority. The proposed scheme used to plan contention and reserved transmission opportunities in accordance to bandwidth availability.

Juliana Freitag *et al*.[14] used the concept of high, intermediate and low priority queues to handle varying types of traffic. High priority queue is used to handle flows that must be scheduled in next frame which includes UGS packets and uni-cast request opportunities for rtPs and nrtPS flows. Intermediate and low priority queues were used to handle rtPS, nrtPS and BE flows respectively. Queues were served using strict priority however starvation was handled as request whose deadline is going to expire is migrated to high priority queue.

One of the most recent work in the field of dynamic scheduling has been done by M. Fathi *et al*[42] where a joint scheduling and CAC method is proposed. WFQ was used initially to assign weights to different classes in accordance to priorities. Bandwidth allocation is done on the basis of packet dropping probability.

The papers discussed in this section indicate that non-real time traffic is always sacrificed to real time traffic which tends to degrade the performance for non-real time applications. In actuality non-real time traffic holds a very large share of traffic on a network. This means that non-real time traffic cannot be neglected and a suitable scheme is necessary for making fair allocations to it.

III. PROPOSED SCHEME

The scheduling problem in WiMAX is quite novel as it has to satisfy a number of applications and these applications tend to have different QoS satisfaction levels. WiMAX has dynamic QoS and SS request resources from BS for transmission on the basis of its incoming traffic. Requirement of incoming traffic can change momentarily in multimedia application and therefore choosing a single and fixed bandwidth allocation mechanism for all sorts of application is not good idea as it may deteriorate QoS levels and real time applications can starve non-real time application like BE and nrtPS. In such situations a bandwidth allocation mechanism that can adapt itself to the changing requirement of QoS will be preferred. Authors in this paper tend to solve scheduling problem in WiMAX using a dynamic approach that is based on the principles of fuzzy logic. The indecisive theories of fuzzy logic are very helpful in providing solution to the problem where designing of exact mathematical models is very difficult. The main aim of the scheduler is to provide fairness in terms of bandwidth allocation to neglected traffic classes of nrtPS and BE.

The traffic in WiMAX consists of combination of real and non-real time applications categorized into 5 different QoS service classes namely Unsolicited Grant Service(UGS), ertPS(extended real time polling service), rtPS (real time polling service), nrtPS(non-real time polling services) and BE(Best Effort). Out of these first three belong to real time traffic class while the last two fall to non real time traffic class. Both these traffic have different QoS parameters ,real time traffic requires a guaranteed delivery of packets within stringent time constraints while maximum sustainable time requirement of non-real time traffic shall be satisfied by the implemented scheduler. Therefore the designed scheduler shall be able to adhere to these requirements and shall provide a fair share of allocation to both these classes of traffic.

The fuzzy approach proposed in this study takes as input two variables, latency requirements for real time traffic and throughput for non-real time traffic. The output of the fuzzy inference system is weight, whose value will be used to make bandwidth allocations to both these classes. The respective membership functions for all three variables are shown in figures( ). The fuzzy system is built over five different linguistic variables for all the input and output variables .The membership function are defined as NB(Negative Big), NS(negative small), Z(Zero), PS (Positive small) and PB(positive big). The range of all these variables is from 0 to 1. The rule base consists of 5 x 5=25 rules which is considered to be sufficiently large for two input fuzzy inference system is used for inference database. The rule base has been defined considering the nature and dynamism of input traffic.

The default weight for any flow in WFQ is calculated from the following equation

$$w_i = \frac{R_{min(i)}}{\sum_{i=0}^{n} R_{\min(i)}} \quad (1)$$

Where And $R_{min(i)}$ is the minimum reserved rate for flow(i) and all flows shall satisfy the constraint of equation 2.

$$\sum_{i=0}^{n} w_i = 1 \qquad 0.001 \leq w_i \leq 1 \quad (2)$$

The fuzzy system consists of three steps:-fuzzification where the system reads in system input variables ie throughput and latency. Fuzzy reasoning where input state variables read in previous step are manipulated as per the rule base and provides an output value. Last step defuzzification, employs centre of gravity method to calculate a crisp value for our output variable weight. The outputted value is taken as the weight for real time traffic and weight of non-real time traffic is calculated by subtracting from 1 since the total weights for all queues shall satisfy the constraint defined in 2. The bandwidth allocation to different queues is made on the basis of weight assigned to that queue using the bandwidth allocation formula

$$Bandwidth = R_{max} \times \frac{w_i}{\sum_{i=0}^{n} w_i} \quad (3)$$

IV. RESULTS AND DISCUSSION

In order to implement the concept of adaptive weight calculation for scheduling, a wimax network consisting of one BS and a number of SS was created in a simulator written in C++.The mail aim of the simulation is to check whether the proposed scheduler is successful in providing fairness to nrtPS and BE classes. The performance of these traffic classes is observed with varying number of

real time connections. Performance is studied on the basis of parameters like delay and throughput. Various simulation parameters used are defined in table 1.Simulation is aimed at making sure that the proposed scheme is able to provide a relative good QoS levels or not. This is done by increasing the number of UGS connections in the scenario while keeping number of rtPS and nrtPS fixed. Scenario consists of one BS and 110 SSs. For experimentation 10 ertPS, 10 rtPS connections,

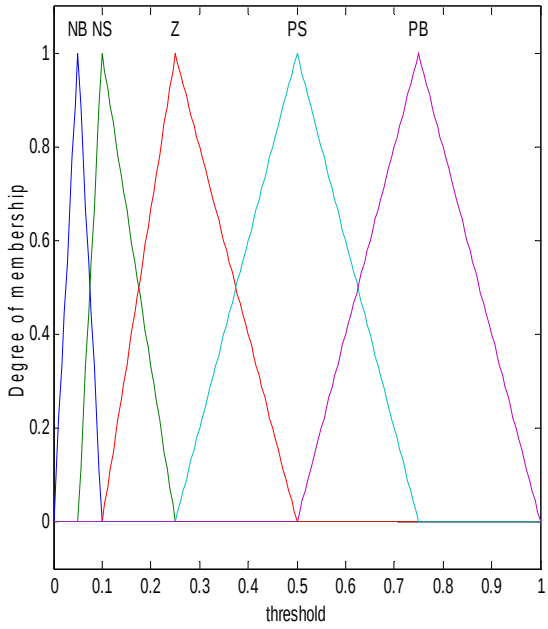


Fig 2 :- Fuzzy membership diagram for input variable threshold

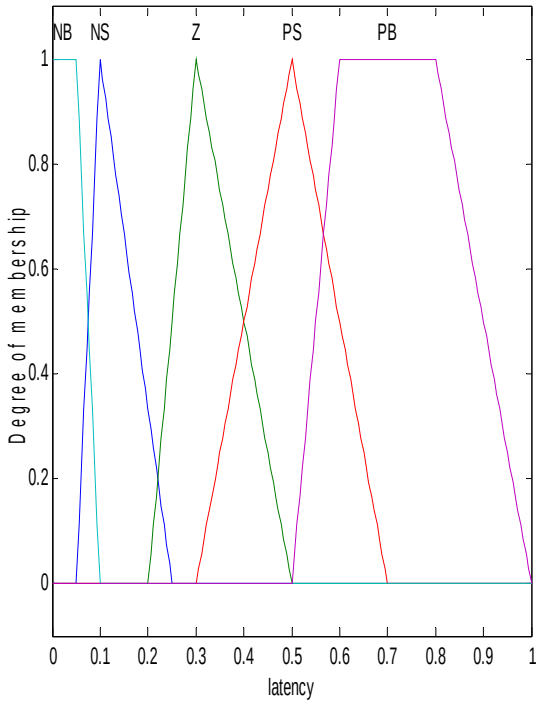


Fig 3 :- Fuzzy membership diagram for input variable latency

25 nrtPS connections and 25 BE connections were established, and the number of active UGS connections were varied from varies from 10 to 40. The performance is verified by measuring the delay and throughput of real and non real time traffic with an increase in number of UGS connections.

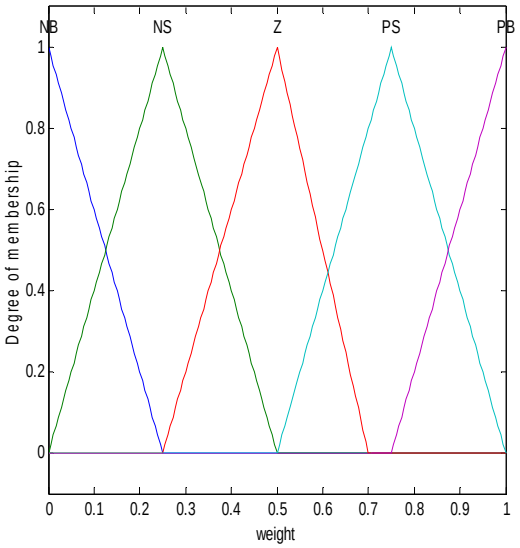


Fig 4 :- Fuzzy membership diagram for output variable weight

TABLE 1:
SIMULATION PARAMETERS FOR SCENARIO

Simulation parameters	
Radio type	802.16 Radio
Operating frequency	2.4GHz
Transmission Power	30(dbm)
Channel Bandwidth	20MHz
FFT Size	2048
Antenna Model	Omni directional
Cell Radius	1500 m
Frame duration	10ms
Number of nodes	110
Base Station	1
Node placement	Unifrom
Mobility Model	Random way point
Pathloss model	Two-ray
Latency for VoIP	<150ms

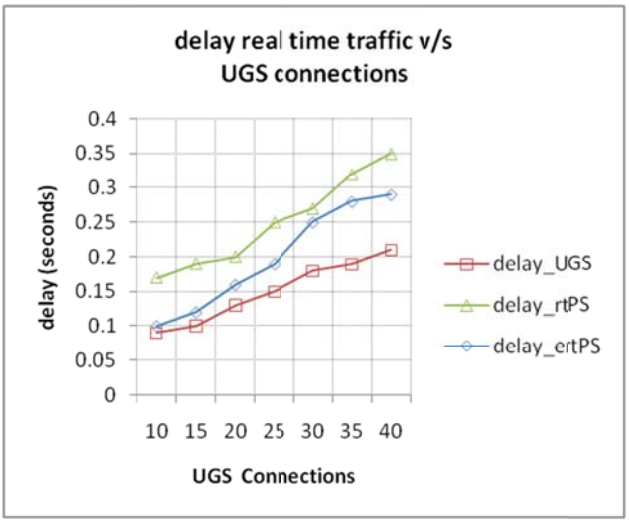


Fig 5:- Delay of UGS, ertPS and rtPS

Figure 2and 3 shows plot of queuing delay incurred by real time traffic consisting of UGS, ertPS and rtPS traffic classes and non real time classes consisting of nrtPS, BE with an increase in number of UGS connections. The variation in the queuing delay of UGS connections is minimal since they have the highest priority. Delay for ertPS and rtPS classes shows an increase when number of UGS connections approaches above 30 which is understandable as more resources are used to satisfy high priority traffic. The delay for nrtPS and BE traffic classes shows the maximum variations. The delay for both these classes shows steep rise when amount of traffic increased in the network, this was expected since these classes have minimum priority amongst all traffic classes. Inspite of it the delay values are considerably manageable and scheduler was able to provide acceptable levels of delay for all classes.

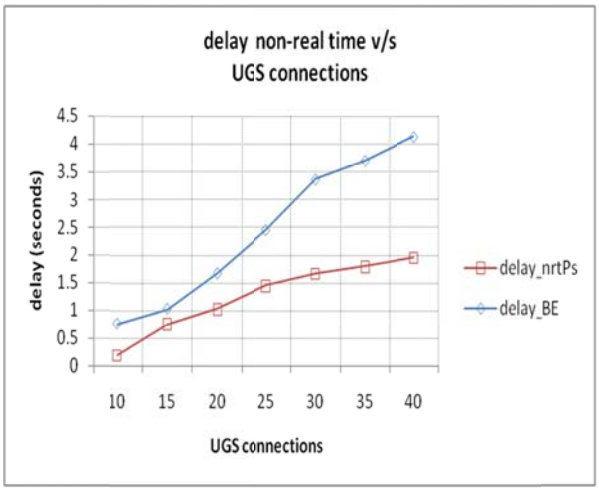


Fig 6:- Delay of nrtPS and BE classes

Figure 7and 8 shows the plot of throughput for real and non real time classes with variations in UGS connections. The graph shows that throughput obtained for UGS is maximum and is almost fixed and does not show any variations as per IEEE 802.16 WiMAX standard. Throughput for ertPS and rtPS was quite good until

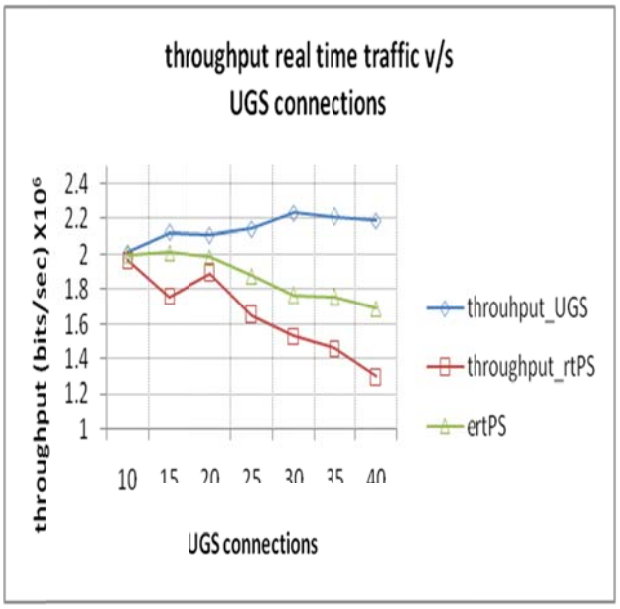


Fig 7:- throughput of UGS, ertPS and rtPS

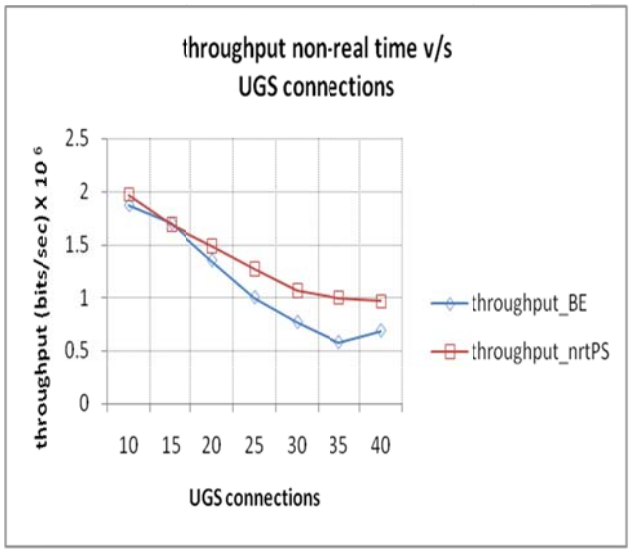


Fig 8:- throughput of nrtPS and BE

number of UGS connection increased beyond a limit after which the throughput for nrtPS started to decrease but the throughput for ertPS is still competitive enough.

Fig. 8 shows the throughput of the nrtPS and of the BE connections. nrtPS throughput shows small oscillations and scheduler provided with minimum bandwidth to achieve minimum threshold levels for nrtPS class. The BE service enjoys the same throughput as enjoyed by nrtPS service for small number of active connections, since the scheduler could have allocated the residual bandwidth to this service. When number of connections is high, the throughput for BE decreases

V. CONCLUSION

The above study proposed an adaptive scheduler for solving bandwidth allocation problem for WiMAX networks using the concepts of fuzzy logic. The proposed method supported different traffic types as specified by the standard and takes into account the QoS requirements

for making bandwidth allocations. The fuzzy system makes its scheduling decision on the basis of two input variables and gives queue weight as its output. Traffic in WIMAX networks consist of mixture of real and non-real time traffic and simulation shows that proposed scheduler was able to handle both these classes fairly. The scheduler was able to satisfy the latency requirements for real time and minimum throughput requirement for non real time traffic efficiently.

REFERENCES

[1] IEEE, Draft IEEE Standard for Local and metropolitan area networks Corrigendum to IEEE standard for Local and Metropolitan Area Networks—Part 16:Air Interface for Fixed Broadband Wireless Access Systems(Corrigendum to IEEE Std 802.16- 2004). IEEE Std P80216/Cor1/D2. 2005.

[2] IEEE, Draft IEEE Standard for Local and metropolitan area networks Corrigendum to IEEE Standard for Local and Metropolitan Area Networks- Advanced Air Interface. IEEE P80216m/D10, November 2010. 2010: 1-1132.

[3] M. Shreedhar, G. Varghese, "Efficient fair queuing using deficit round-robin" IEEE ACM Transactions on Networking 1996 vol 4, pp 375–85.

[4] K. Wongthavarawat and A. Ganz, "Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems," *International Journal of Communication System*, vol. 16, issue 1, pp. 81–96, Feb. 2003.

[5] K. Wongthavarawat and A. Ganz, "IEEE 802.16 based last mile broadband wireless military networks with quality of service support," *IEEE Military Communication Conference, 2003. MILCOM 2003*, vol. 2, no. 1, pp. 779–784, Oct. 2003.

[6] C.F. Ball, F. Treml, X. Gaube, and A. Klein, "Performance Analysis of Temporary Removal Scheduling applied to mobile WiMAX Scenarios in Tight Frequency Reuse", *16th Annual IEEE International Symposium on Personal Indoor and Mobile Radio Communications, PIMRC'2005, Berlin*, vol 2,pp 888-894,11 – 14, September 2011.

[7] Cicconetti, C., Erta, A., Lenzini, L., & Mingozzi, E. " Performance evaluation of the IEEE 802.16 MAC for QoS support" *IEEE Transactions on Mobile Computing*, *6*(1), 26–38 2007.

[8] V. Sagar, & D Das, "Modified EDF algorithm and WiMAX architecture to ensure end-to-end delay in multihop network" In *IEEE TENCON* 2008

[9] Jalali, A., Padovani, R., & Pankaj, R. "Data throughput of CDMA-HDR a high efficiency high data rate personal communication wireless system", In *IEEE VTC* 2000.

[10] K S Demers and S Shenker, "Analysis and simulation of a fair queuing algorithm", In *SIGCOMM* 1989

[11] M Andrews, Kumaran, K., Ramanan, K., Stolyar, A., & Whiting, P, "Providing quality of service over a shared wireless link" *IEEE Communications Magazine*, *32*(9), 150–154.2001.

[12] J. Sun, Y. Yao, and H. Zhu, "Quality of service scheduling for 802.16 broadband wireless access systems," in *IEEE 63rd Vehicular Technology Conference, 2006. VTC 2006-Spring, vol. 3, 7-10, May 2006, pp. 1221–1225.*

[13] N. Liu, X. Li, C. Pei, and B. Yang, "Delay character of a novel architecture for IEEE 802.16 systems," in *Sixth International Conf. Parallel Distributed Comput., Appl. Technol., 2005. PDCAT 2005*, Dec. 2005, pp. 293–296.

[14] Juliana Freitag, Nelson L.S. da Fonseca, "Uplink Scheduling with Quality of Service in IEEE 802.16 Networks" *in IEEE Global Telecommunications Conference, GLOBECOM 2007*, 26-30 Nov 2010, pp 2503 -2508.

[15] Mohammad Fathi , Samaneh Rashidi-Bajgan, Alireza Khalilzadeh , Hassan Taheri, " A dynamic joint scheduling and call admission control scheme for 802.16 networks", *Telecommun Syst (2013) 52:195–202 DOI 10.1007/s11235-011-9555-8 http://www.springerlink.com/ content h81w80515trt0843*

Tumor Detection by Color Segmentation of PET/CT Liver Images

Neha Bangar[a], Akash Deep[a]
[a]University Institute of Engineering and Technology, Panjab University,
Chandigarh, India
{neha.bangar1@gmail.com, akashdeep@pu.ac.in}

***Abstract*— A wide range of different image modalities for medical imaging are available now-a-days which provide view of internal structures of the human body such as brain, kidney, liver etc. Among these medical image modalities, Ultrasound (US), Magnetic Resonance Imaging (MRI), Computed Tomography (CT), Positron Emission Tomography (PET) and Positron Emission Tomography combined Computed Tomography (PET/CT) imaging has gained importance in research areas. The key problems within medical image analysis techniques are segmentation and shape feature extraction that will be referred to in this paper. Segmentation of grayscale medical images can be difficult since the intensity values between healthy tissue and tumor may be very close. PET/CT provides more accurate measurements of tumor size than is possible with visual assessment alone. In this paper, segmentation method for the detection of liver tumor in PET/CT scans is proposed. The images are denoised using median filter and binary tree quantization clustering algorithm is used for segmentation. Finally ROI selection and shape feature extraction is performed on the selected cluster to quantify the size of the tumor and to check the accuracy of our method with original image and K-means Clustering method.**

***Index Terms*— Denoising, Liver segmentation, Binary Tree Quantization Algorithm, K-means Clustering Algorithm, Shape Feature Extraction.**

I. Introduction

Tumor/ Cancer are the abnormal growth of cells or tissues in the organ. Liver Cancer is a common type of cancer that affects the largest organ in the abdomen, the liver. There are two types of liver cancer namely Primary Liver Cancer and Secondary Liver Cancer. Primary liver cancer is the cancer that originates from the liver itself. This type of liver cancer is also known as hepatocellualr carcinoma (HCC) or hepatoma. It is the fifth most frequent cancer form in the world and third leading cause of cancer death. HCC cases mostly occur in Asia and Africa, but its number is increasing rapidly in U.S. and other western countries [1]. Secondary liver cancer is the cancer which originates in other organs but then spreads to liver [2]. Secondary liver cancer is also known as Metastatic liver cancer. Our interest of research is hepatocellular carcinoma, which we will further refer to as liver tumor. According to American Cancer Society [3], men suffer from liver cancer more than women. As per data from the World Health Organization (WHO), less than 30 cases per 100,000 people worldwide die as a result of liver cancer, with high rates in parts of Africa and Eastern Asia. So, it is clear that the major cause of death in human is considered to be liver tumor [4]. There are various signs and symptoms of liver tumor which include unexplained weight loss, swollen abdomen, yellowish skin, dark colored urine, pale colored stools, few weeks of appetite loss, feeling full even after a small meal, high temperature and sweating, vomiting and nausea. There are various risk factors that can lead to liver cancer such as Hepatitis B infection, Hepatitis C infection, excessive alcohol consumption, diabetes and obesity.

Analyzing medical images manually is very difficult due to large size of medical image database, so computer oriented surgery has become one of the major research subject [5] and this has led to medical imaging modalities such as X-ray, CT [6], Magnetic Resonance Imaging (MRI) [7, 8], SPECT [9], PET [10] and ultrasound [11, 12].

In this paper, problem of PET/CT based liver segmentation to detect liver tumor is addressed using Binary tree quantization clustering algorithm and feature extraction is performed to calculate area, perimeter and equiv-diameter of tumor detected by binary tree quantization. These are calculated to check the accuracy of our process and all the parameters of the output image are calculated and compared with the original input image. Binary tree quantization clustering algorithm is used as it directly works on the RGB image which eliminates the requirement of converting the grayscale image into colored image and so is well suited to be used for PET/CT images. Firstly image pre-processing is done to reduce the effect of noise and preserve liver tumor edges for successful segmentation. Then, Binary tree quantization clustering is applied for segmentation. After segmentation, ROI selection is done manually to calculate various features of tumor. Finally the method is compared with same features calculated for original image and traditional K-means clustering algorithm for checking the performance of our method with K-means method. The results show that the method successfully segmented tumor with accuracy compared to K-means method. In PET/CT imaging modality, within a single scanner PET and CT scans are combined and hence in a single scanning session this modality provide both

doi:10.4304/jetwi.6.1.148-156

functional and anatomical imaging [13] i.e. in PET/CT, in the same session, images are acquired from both the modalities sequentially which are then combined into single superposed image. CT pictures of the body structures are taken firstly. Then at the same time, areas where cells are more active than normal cells of the body are shown up by the radioactive substance. Both the information is then combined by the scanner which allows the doctor to visualize changes in the cell's activity (if any) and know where the changes are happening exactly. The most commonly radioactive substance used is 18F-FDG (fluorodeoxyglucose). For carrying out the PET/CT scan, radioactive substance is injected through a small tube which is put into the veins in the back of hand or arm. The scan is carried out only when the body has absorbed the radioactive substance (it takes at least one hour) and the bladder is emptied. The patient needs to lie down on a narrow table which moves through the scanner and detector. During the scan, X-rays emitted by the CT scanner go to the detector through the patient's body to generate its anatomical structures. Then PET scanner (which is a large machine with round doughnut shaped machine in the middle similar to CT and has multiple rings of detectors), detects the radiations which are emitted from the patient because of the radioactive substance which is used by the computer to generate the cell activity (functional activity) in tissues of the body. All the cells in the body use radioactive substance (namely F18-FDG) as an energy source. So, cancer cells grow more rapidly than healthy cells. This forms the image generation in PET/CT scans. Due to reduced scanning time in PET/CT, it is considered reliable in clinical practices [14].

II. Related Work

Variety of segmentation methods have been proposed for CT, MRI and PET images to detect tumor but there have been little attempts to work on PET/CT images. *Amir H. Foruzan et.al* [15] proposed a technique for segmentation of liver in CT data for liver initial border estimation in which they started with image simplification, then searched rib bones, connected them together to find ROI of liver. They then used split thresholding technique to segment the images. Different colors were assigned to objects present in ROI, the split-threshold step and the objects that were found in 75% of right part of the abdomen. After this a colored image was obtained in which liver had a specific color from where liver boundary was extracted. In the method of *Xing Zhang et.al* [16], automatic liver segmentation included average liver shape model localization in CT via 3D generalized Hough Transform, subspace initialization of Statistical Shape Model through intensity and gradient profile and then deforming the model to liver contour through optimal surface detection method based on graph theory. *Laszlo Rusko et.al* [17] method automatically segments the liver using region-growing facilitated by pre- and post-processing functions, which considers anatomical and multi-phase information to eliminate over and under-segmentation. *Hassan Masoumi et.al* [18] extracted features of liver region in MRI images using watershed algorithm and artificial neural network leading to automatic liver segmentation. But all these methods resulted only in liver boundary and did not detect liver tumor. *O. Lezoray et.*al [19] proposed an unsupervised clustering technique in which watershed operates on distance function to centers of class for determining the number of classes. In this method, segmentation of colored image considered pairwise color projections where each of these projections is analyzed to look for the dominant colors of 2-D histogram and to fully automate the segmentation, energy function was used to quantify the quality of the segmentation. But the difficulty with the histogram method is to identify peaks and valleys in the image. *Marisol Martinez-Escobar et.al* [20] first colorized the pixels representing tumor and healthy tissues and then used threshold method for segmentation to detect tumor to overcome the problem as faced in histogram. But these methods are either performed on CT or MRI images or the images are first colored and then segmentation is applied. Since morphological changes always precede metabolic changes and are detected through imaging modalities like CT or MRI, PET is expected in enabling an early assessment of response to treatment. 18F-FDG PET has been reported to give earlier response for tumor detection than CT [21].

Positron emission tomography (PET) with 18FFluorodeoxyglucose (18F-FDG) is widely suggested method medical imaging as numerous tumors are diagnosed very accurately which has improved the decision for therapy consideration and assessing patients having cancer at different stages in the last two decades [22, 23]. It is based on the tumor specific high intracellular accumulation of the glucose analog fluorodeoxyglucose (18F-FDG) [24]. It gives tumor's physiological information and its metabolic activities [25]. PET/CT provides functional and anatomical imaging within a single scanner in a single scanning session [26]. Though PET has been replaced by PET/CT, most of the segmentation work to detect tumor has been done on PET only [27-30].

Baardwijk et. al [31] provided the advantage of PET combined CT images for the segmentation purpose. *Potesil et. al* [32] segmented PET/CT images using initial hot spot detection and segmentation in PET for tumor structure appearance and used shape model to classify voxels in CT. *Xia et. al* [33] proposed an expectation-maximization algorithm using stimulated annealing to automatically segment brain PET/CT images. But long execution times were the main drawback of this method. *Yu et. al* [34] proposed the co-registered multimodality pattern analysis segmentation system (COMPASS) to extract texture features from PET/CT and then used decision-tree based K-nearest-neighbor classifier to label each voxel as either "normal" or "abnormal" and the performance was compared with threshold methods: SUV value and signal/background ratio. *Yu et. al* [35] evaluated the effectiveness co-registered segmentation method to distinguish tumor from healthy tissue in

regions of head and neck. Most approaches use standard uptake value (SUV) to detect tumor which is a semi-quantitative normalization of FDG uptake in PET images and gives concentration of FDG in dose/gram body mass. But these approaches are basically used to evaluate the value and application of FDG PET/CT in clinical practices to detect tumors. It is clear from the literature review that most of the work is done either on CT, MRI or PET images alone. Most of the segmentation work has been done using thresholding algorithms, region-based algorithms, edge-based algorithms, watershed algorithms, fuzzy clustering based algorithms and graph-based algorithms in PET images, CT images or combined PET/CT images. But these methods have various limitations. Thresholding algorithms depend on intensity distribution of image. In PET and PET/CT images, thresholding and graph based methods are generally based on SUV values which are sensitive to volume variation in tumor or structural/functional volumes. Edge-based algorithms produce disjoint edges. Region-based algorithms require manual initialization. Watershed algorithm leads to over-segmentation and is poor in detecting thin structures in images. Graph based algorithms provide global solutions and so are computationally expensive. Fuzzy clustering based algorithms require training of the dataset and the accuracy depends on the selected training samples and therefore, these are more tedious to use. The proposed algorithm i.e. *Binary Tree Quantization Clustering Algorithm* is conceptually simple as it is based on divide and conquer rule which recursively divides the problem (image) into sub-problems where solving the sub-problems and combining their solutions give the solution to the original problem. The method makes use of binary tree structure so it reduces the computation greatly and does not lead to over-segmentation.

III. Proposed Work

Our proposed method for PET/CT-based liver segmentation to detect tumor and perform feature extraction consists of 5 steps as shown in *fig1*. The method started with reading the input image and applying the RGB median filtering which is to be done to remove noise to increase accuracy of our process. The output of median filtering process is taken as input for segmentation. Then image clustering is performed using binary tree quantization method to segment the image. In this method on the basis of the color, the clusters of pixels are computed. Then feature extraction is done for calculating area, perimeter and equiv-diameter of tumor detected by binary tree quantization. By this procedure we came to know about accuracy of our process. For this we have compared all the parameters calculated on our output image with the original input image. Finally the method is also compared with traditional K-means method.

A. Preprocessing

Noise in PET/CT images may be caused by variability in organ/tissue which looks like random noise. Noise can be removed by smoothing the image with a two types of filters i.e. *low pass averaging filters* and *low pass median filters*. Low pass filtering means removing the high frequency content from the image. Noise is a high frequency component and hence low pass filtering is used to remove noise.

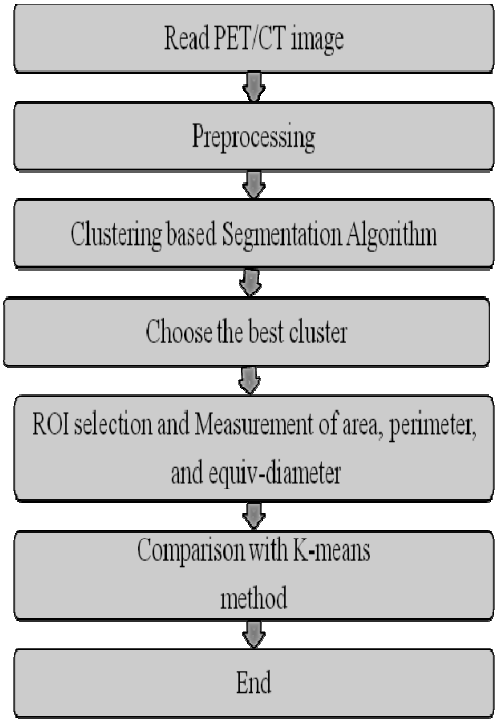


Fig 1: Flowchart of the Proposed Method

Low pass averaging filter is also known as linear filter where value of each pixel is replaced by the average of all the values of intensities in the neighborhood defined by the filter mask and image with reduced sharp transitions in intensities is produced. However, low pass averaging filter results in blurred edges. *Low pass median filter* is also known as non-linear filter as its response is based on the ordering of the pixels in an image, and replaces the value of a pixel by the median of the intensity values in the neighborhood of that. In our work to remove noise, we denoised the image using 2D-median filter in which depending upon the intensity, the pixels in the neighborhood window are ranked and the median (i.e. the middle value) becomes the output value for the centre pixel. The image is denoised using median filter rather than averaging filter as it removes noise without distorting the edges. This is because in median filtering, the output value is from the neighboring values, new unknown values are not created near the edges, so median filtering is more effective when our main aim is to simultaneously reduce noise and preserve the edges. To apply median filter on colored images, number of colors is firstly determined in the image and then the filter is applied on each color separately. All the denoised components are combined to get the final denoised image i.e. image is firstly converted into red, green and blue planes or constituents, median filtering is then performed on red, green and blue constituents separately and then all

the denoised RGB constituents are combined to get the final denoised image.

The output of median filtering process i.e. the denoised image is taken as input for segmentation step. Steps of median filtering are shown in *fig 2:*

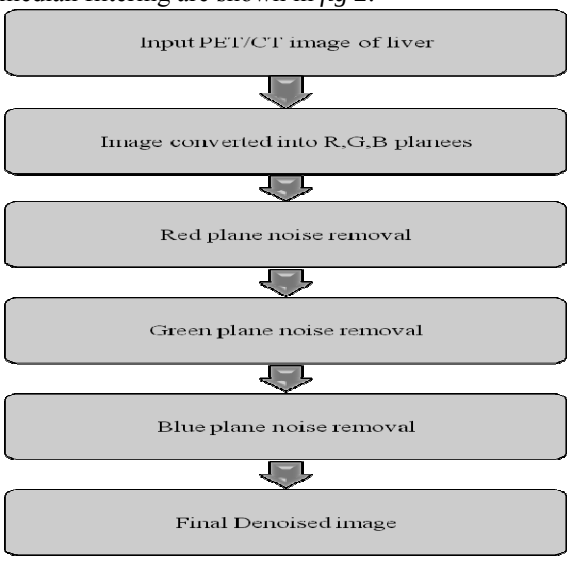


Fig 2: Flowchart of median filtering in all planes of RGB image

B. Segmentation

The denoised image is taken as input for segmentation. The proposed segmentation algorithm is based on binary tree quantization algorithm performed on PET/CT images. In this method on the basis of the color, the clusters of pixels are computed. The Orchard and Bauman method of binary tree is the base of this algorithm. Orchard and Bauman proposed a divisive palette generation algorithm, which uses a local optimization strategy in selecting the splitting axis (i.e. RGB co-ordinate axis of the color plane). In this method, partition with the largest eigen value is chosen to be split. Eigenvalue is the scalar which is associated with eigenvector where eigenvector is a vector that changes its magnitude but not its direction. The splitting axis is the one that passes through the centroid of all the colors in that region and must be perpendicular to the direction of the maximum total squared variation which is derived from the eigenvalue and eigenvector of the covariance matrix of the data in that particular region. The splitting of a node into two nodes is equivalent to selecting a plane which best splits the cluster's colors which can be found with principal component analysis (PCA) and it is called the principal axis or the eigenvector with largest eigenvalue. Mainly variance of the split node is reduced along the direction of principle eigenvector, so decrease in total squared error is considered to be proportional to total squared variation along the principle eigenvector. Therefore, it is best to split the node with the largest principal eigen vector. To conclude:

All the pixels are placed in the same cluster in the beginning. The cluster is then split around its mean value projection on the first principle component. The parameters to the clustering algorithm are the number of clusters, K, to use which is set/input by the user, and the first cluster C_1.

1. Initialize the first cluster, C_1
2. Calculate the mean, μ_1, and the covariance matrix Σ_1 of the cluster C_1
3. For i=2 to K do
4. Find the cluster, C_n with the largest eigenvalue and its associated eigenvector e_n.
5. Split C_n in two sets along the mean value projections on the eigenvector, $C_l = \{x \in C_n : e_n^T x_n \leq e_n^T \mu_n\}$ and update the original cluster with other half $C_n^* = C_n - C_l$.
6. Compute mean and covariance matrix of the two halves obtained in step2 as μ_n^*, Σ_n^*, μ_l and Σ_l.

C. ROI Selection and Shape Feature Extraction

After segmentation, Region of interest (ROI) is selected using GUI based polygon method which selects a polygonal region of interest within an image [36]. Then shape feature extraction is performed on the selected ROI to calculate area, equiv-diameter, and perimeter of the tumor. *Area* used to calculate the actual number of pixels in the region. *Equiv-diameter* is the scalar which specifies the diameter of the region of interest. *Perimeter* gives the distance between each adjoining pair of pixels around the border of the region.

D. K-means Clustering Method

K-means method is a data-clustering algorithm which groups together the objects with similar characteristics to form clusters. The number of clusters (k-clusters) to be formed to classify the dataset is fixed apriori and is given as input parameter. Each object is assigned to the cluster which is close to it. The mean of the cluster is recomputed and the process is repeated again. It works as follows:

1. Cluster seeds are chosen at random initially i.e. value of k which represent “mean” of the clusters.
2. Each object is assigned to the cluster close to it based upon the Euclidean distance between the object and the cluster center. The object is assigned to the cluster with the smallest squared Euclidean distance.
3. For each cluster, new cluster center is recomputed as average of the objects/points in that cluster.
4. Step 2 and 3 are repeated until no more objects are left for assignment to the cluster i.e. until the clusters converge.

K-means method was implemented so as to compare its results with our proposed method to check accuracy.

E. Comparison of Proposed Method with K-means Clustering Method

The proposed method is compared with k-means method to check for accuracy of segmentation. This was

done by comparing the shape features of the proposed method with the features extracted from original image and k-means clustering method. The shape features used for comparison are area, perimeter and equiv-diameter. These shape features have been used as a measure to quantify the size of tumor which is one of the prime concerns for medical practitioners and patients.

IV. Visual Results of Proposed Method and K-means Clustering Method

The binary tree quantization method was evaluated on PET/CT liver images obtained from multi-specialist hospital. We have tested our algorithm on twelve PET/CT images. The images of liver tumor were provided by a reputed hospital and no detail of the patient history was provided. Accuracy of our method is compared with traditional K-means clustering approach. Output resulting images of the proposed method on first pateint chosen at random are shown from *figures 3-7* and output images of K-means method are shown in *figure 8.* A Visual comparison of the best image produced by the proposed method and k-mean clustering is also presented in *figure 9.*

1. *Fig 3* shows the Original PET/CT scan image of liver tumor

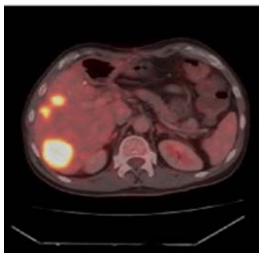
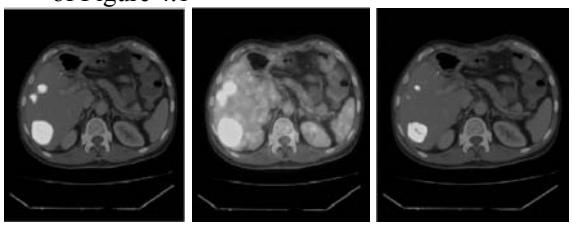


Fig 3 : Original PET/CT scan image of Patient-1

2. *Fig 4* shows the Red, Green and Blue Constituents of Figure 4.1



(a) (b) (c)

Fig 5: (a)Red Constituent; (b)Green Constituent; (c)Blue Constituent

3. *Fig 6* shows the Denoised Image after Median Filtering

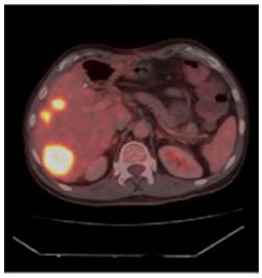
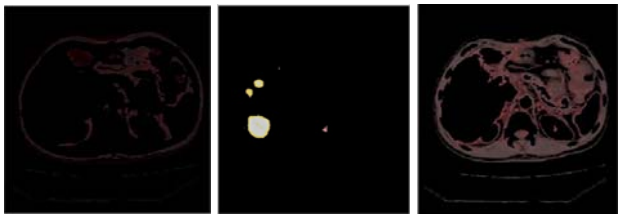
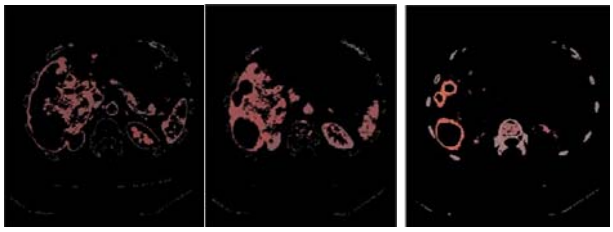


Fig 6 : Figure 4.1 after filtering

4. *Fig 7* shows Output of Binary Tree Quantization Clustering Method



(a) (b) best cluster (c)

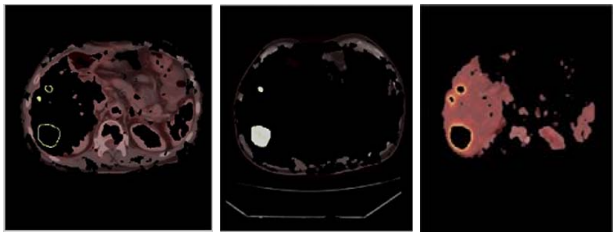


(d) (e) (f)

Fig 7: Clusters of Figure 4.3 after Binary Tree Quantization Clustering

Fig. 7 shows various clusters generated by binary tree quantization clustering method where (a) cluster1 roughly shows the organs excpet the liver; (b) cluster2 shows the brightened tumor detected by the proposed method and so, is selected as the best cluster for feature extraction; (c) cluster3 shows the organs other than liver; (d)-(e) cluster4 and 5 show liver area; (f) cluster6 shows the rough boundary of the detected tumor in liver.

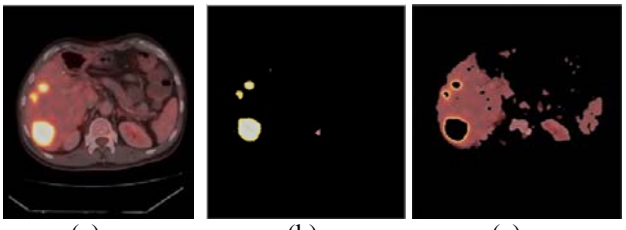
5. *Fig 8* shows the Output of K-means Method



(a)Cluster 1 (b) Cluster 2 (c) Cluster 3 (best cluster)

Fig. 8 shows various clusters generated by k-means clustering method where (a)-(b) cluster 1 and 2 do not show the entire tumor present in the liver; (c) cluster 3 shows the entire tumor present in the liver and so is taken as the best cluster for feature extraction.

6. *Fig 9* show the visual comparison of the proposed method with k-means method



(a) (b) (c)

Fig 9 Visual Comparison of proposed method with k-mean clustering on Patient-1(a) Original Image (b)Best image of proposed method (c) Best image produced by k-means method.

Fig 9 draws a visual comparison of the best clusters produced by our method and k-mean clustering method. The output shows that proposed method performs better than k-mean clustering as it had detected the tumor present in the image very clearly. The size and shape of the tumour is very much similar to original PET/CT

image.Although some tiny spots other than tumor are detected but they are negligible.

From the visual results of binary tree quantization method and k-means clustering method it is clear that clusters generated by the proposed method are more accurate, clear and brighter than the clusters generated by k-means clustering method. Tumor detected in the best cluster by the proposed method is close to the tumor in original image whereas k-means method was not able to detect tumor completely or tumor detected is not clear.

V. Quantitative Comparison of Binary Tree Clustering Method and K-means and Original Image

To check accuracy of the proposed method, the method is compared with traditional k-means clustering method and original image by calculating shape features on the selected cluster. Shape features calculated on original image, best cluster generated by binary tree quantization method and k-means clustering are shown in section A, B and C respectively.

A. Shape Feature Results on Original Input Image

All the shape features calculated on original input image are shown in *table I:*

TABLE I.
RESULTS ON ORIGINAL IMAGE

Images	Area	Perimeter	Equiv-diameter
Image 1	3333	514.5412	65.1437
Image 2	2527	191.9239	56.7228
Image 3	950	154.4092	34.7790
Image 4	955	132.8528	34.8704
Image 5	223	56.4264	16.8503
Image 6	493	97.3970	25.0541
Image 7	1551	203.2376	44.4387
Image 8	2822	60.526	59.9423
Image 9	3181	551.0437	63.6410
Image 10	255	60.8701	18.0188
Image 11	3770	239.9239	69.2828
Image 12	2690	325.0193	58.5236

B. Shape Feature Results on Best Cluster Generated by Binary Tree Quantization Clustering Method

After selecting the best cluster, shape features i.e. area, perimeter and equiv-diameter are calculated on ROI. This result is given in *table II* as:

TABLE II:
RESULTS OF BINARY TREE QUANTIZATION CLUSTERING

Images	Area	Perimeter	Equiv-diameter
Image 1	3195	215.3381	63.7809
Image 2	2290	166.7107	53.9974
Image 3	873	143.9828	33.3397
Image 4	839	92.4264	32.6841
Image 5	133	33.6985	13.0131
Image 6	415	92.4264	22.9868
Image 7	1329	123.0538	41.1356
Image 8	2683	56.2843	58.4474
Image 9	2863	361.6051	60.3762
Image 10	243	59.2132	17.5897
Image 11	3743	238.9361	69.0343
Image 12	2632	304.5341	57.8893

C. Shape Feature Results on Best Cluster Generated by K-means Clustering Method

After selecting the best cluster, shape features i.e. area, perimeter and equiv-diameter are calculated on ROI. This result is given in table III as:

TABLE III:
RESULTS OF K-MEANS METHOD

Images	Area	Perimeter	Equiv-diameter
Image 1	747	47.6985	30.8401
Image 2	974	70.5269	35.2156
Image 3	633	136.6102	28.3894
Image 4	750	111.2965	30.9019
Image 5	113	22.6274	11.9948
Image 6	383	81.6985	22.0828
Image 7	1095	193.8234	37.3390
Image 8	1783	23.8995	47.6465
Image 9	582	116.8112	25.9282
Image 10	146	48.1421	13.6343
Image 11	3642	232.1076	68.0965
Image 12	1546	269.8650	44.3670

D. Comparison with Original Image and K-means Method

Quantitative results obtained from binary tree quantization are compared results obtained from k-means method to check accuracy of the results with the same shape features calculated on original image. Relative accuracy of each calculated shape feature is calculated as modulus of calculated value minus original value divided by original value. This is shown in *table IV:*

TABLE IV:
COMPARISON OF PROPOSED METHOD WITH K-MEANS CLUSTERING METHOD AND ORIGINAL IMAGE BASED ON SHAPE FEATURES

S.No	Images	Area	Relative Accuracy	Perimeter	Relative Accuracy	Equiv-Diameter	Relative Accuracy
1	Original	3333		514.5412		65.1437	
	Proposed	3195	0.041	215.3381	0.581	63.7809	0.020
	K-means	747	0.75	47.6985	0.907	30.8401	0.526
2	Original	2527		191.9239		56.7228	
	Proposed	2290	0.093	166.7107	0.131	53.9974	0.048
	K-means	974	0.614	70.5269	0.632	35.2156	0.379
3	Original	950		154.4092		34.7790	
	Proposed	873	0.081	143.9828	0.067	33.3397	0.041
	K-means	633	0.333	136.6102	0.115	28.3894	0.183
4	Original	955		132.8528		34.8704	
	Proposed	839	0.121	92.4264	0.304	32.6841	0.062
	K-means	750	0.214	111.2965	0.162	30.9019	0.113
5	Original	223		56.4264		16.8503	
	Proposed	133	0.403	33.6985	0.402	13.0131	0.228
	K-means	113	0.493	22.6274	0.598	11.9948	0.288
6	Original	493		97.3970		25.0541	
	Proposed	415	0.158	92.4264	0.051	22.9868	0.082
	K-means	383	0.223	81.6985	0.161	22.0828	0.118
7	Original	1551		203.2376		44.4387	
	Proposed	1329	0.143	123.0538	0.394	41.1356	0.074
	K-means	1095	0.294	193.8234	0.046	37.3390	0.159
8	Original	2822		60.5269		59.9423	
	Proposed	2683	0.049	56.2843	0.070	58.4474	0.024
	K-means	1783	0.368	23.8995	0.605	47.6465	0.205
9	Original	3181		551.0437		63.6410	
	Proposed	2863	0.099	61.6051	0.888	60.3762	0.051
	K-means	582	0.817	116.8112	0.788	25.9282	0.592
10	Original	255		60.8701		18.0188	
	Proposed	243	0.047	59.2132	0.027	17.5897	0.023
	K-means	146	0.427	48.1421	0.209	13.6343	0.243
11	Original	3770		239.9239		69.2828	
	Proposed	3743	0.007	238.9361	0.004	69.0343	0.003
	K-means	3642	0.033	232.1076	0.032	68.0965	0.017
12	Original	2690		325.0193		58.5236	
	Proposed	2632	0.021	304.5341	0.063	57.8893	0.010
	K-means	1546	0.425	269.8650	0.169	44.3670	0.236

In this paper, binary tree quantization method was implemented for segmenting PET/CT liver images so as to detect tumor. It is clear from the resulting images that the proposed method works well on segmentation of PET/CT images for tumor detection and shape feature extraction. Binary tree quantization was able to detect tumor perfectly as compared to k-means method. Shape features were calculated to calculate area, perimeter and equiv-diamter of the tumor detected by the proposed method and k-means method. The calculated shape features were compared with features extracted from original image and k-mean clustering method to check the accuracy of the proposed method. From *table IV*, it is clear that the accuracy of the proposed method is far better than the k-means method and is very near to the actual image.

VI. CONCLUSIONS

This work has tried to segment PET/CT scan images of liver using Binary Tree Quantization clustering method for detection of tumor in the form of best cluster. Combination of filtering technique with clustering techniques had known to be beneficial for the process of segmentation of medical images as pre-processing removed the unwanted noise from the images. Feature extraction has been done to parameterize the efficiency of the process. Our method has proved that it is better than other clustering methods in detecting tumor more accurately and precisely. Our method is semi automatic as after tumor detection for the selection of region of interest we have used manual approach. This work is significant as it can help medical practitioner in focussing to the area of tumor and making it easy for classifying tumor as benign or malignant.

The previous work has focussed on segmenattion of PET/CT images although the proposed method is generating good results for liver images the same can be extended to detect tumors in other parts of the body. Other improvements can be done in the direction of making this process fully automated which can decrease the processing time of cancer detection as well as

efficiency could be increased and be able to detect tumors in large and different image datasets.

ACKNOWLEDGMENT

The authors would like to thank multi-specialist hospital for providing PET/CT images that helped to complete the work successfully.

REFERENCES

[1] CanLive, The Hepatobiliary, Cancers Foundation, http://canliv.org/livercancer?gclid=CPWPiIuKqrkCFa0l4godZzIAsA(last accessed on August 30, 2013)
[2] National Cancer Institute, http://www.cancer.gov/cancertopics/types/liver last accessed on August 30, 2013.
[3] American Cancer Society, http://www.cancer.org/cancer/livercancer/detailedguide/liver-cancer-what-is-key-statistics last accessed on August 30, 2013.
[4] Cancer Research UK, http://www.cancerresearchuk.org/cancerhelp/type/livercancer/about/symptoms-of-liver-cancer last accessed on August 30, 2013.
[5] Masuda, Y., Tateyama, T., Xiong, W., Zhou, J., Wakamiya, M., Kanasaki, S., Furukawa, A., and Chen, Y.W.: Liver Tumor Detection in CT images by Adaptive Contrast Enhancement and the EM/MPM Algorithm. In 18th IEEE Conference on Image Processing, pp. 1421-1424, September 2011.
[6] Hounsfield, G.N.: Computerized Transverse Axial scanning Tomography: Part 1, Description of the System. British Journal of Radiology 46, 1016-1022 (1973).
[7] Lipinski, B., Herzog, H., Kops, E.R., Oberschelp, W., Muleer-Gartner, H.W.: Expectation Maximization Reconstruction of Positron Emission Tomography Images using Anatomical magnetic Resonance Information", IEEE Transaction on Medical Imaging, Vol. 16, pp. 129-136, April 1997.
[8] Bazille, A., Guttman, M.A., Mcveigh, E.R., Zerhouni, E.A.: Impact of Semi-automated versus Manual Image Segmentation Errors on Myocardial Strain Calculation by Magnetic Resonance Tagging. Investigative radiology 29, 427-433 (1994).
[9] Anger, H.: Use of Gamma-Ray Pinhole Camera for viva studies. A Nature Conference on Nuclear Reprogramming and the Cancer Genome 170, pp. 200-204, August 1952.
[10] Ouyang, X., Wang, W.H., Johnson, V.E., Hu, X., Chen, C.T.: Incorporation of Correlated Structural Images in PET Image Reconstruction. IEEE Transactions on Medical Imaging, Vol. 13, pp. 627-640, August 2002.
[11] Akgul, Y.S., Kambhamettu, C., Stone, M.: Extraction and Tracking of the Tongue Surface from Ultrasound Image Sequences. In 1998 IEEE Computer Society Conference on Computer Vision Pattern Recognition, pp. 298-303, June 1998.
[12] Abeyratne, U.R., Petropulu, A.P., Reid, J. M.: On modeling the Tissue Response from Ultrasonic B-scan. IEEE Transactions on Medical Imaging, Vol. 2, pp. 479-490, August 1996.
[13] Thomas Beyer, David W. Townsend, Tony Brun, Paul E. Kinahan, Martin Charron, Raymond Roddy, Jeff Jerin, John Young, Larry Byars and Rnald Nutt, "A Combined PET/CT Scanner for Clinical Oncology", Journal of Nuclear Medicine, Vol. 41, pp. 1369-1379, August 2000.
[14] Bingsheng Huang, Martin Wai-Ming Law and Pek-Lan Khong, "Whole-Body PET/CT Scanning: Estimation of Radiation Dose and Cancer Risk", Radiology, Medical Physics, Vol. 251, pp. 166-174, April 2009.
[15] Foruzan, A.H., Zoroofi, R.A., Hori, M., and Sato, Y.: A Knowledge-based Technique for Liver Segmentation in CT Data. Computerized Medical Imaging and Graphics, Elsevier, Vol.33, pp. 567–587, March 2009
[16] Zhang, X., Tian, J., Deng, K., Yongfang, W., and Xiuli, I.: Automatic Liver Segmentation Using a Statistical Shape Model With Optimal Surface Detection. IEEE Transactions on Biomedical Engineering, Vol. 57, pp. 2622-2626, October 2010.
[17] Rusko, L., Bekes, G., and Fidrich, M.A.: Automatic Segmentation of the Liver from Multi- and Single-Phase Contrast-Enhanced CT. Medical Image Analysis, Elsevier, Vol. 13, pp. 871–882, July 2009
[18] Masoumi, H., Behrad, A., Pourmina, M.A., and Roosta, A.: Automatic liver segmentation in MRI Images using an Iterative Watershed Algorithm and Artificial Neural Network. Biomedical Signal Processing and Control, Elsevier, Vol. 7, pp. 429–437, February 2012
[19] Lezoray, O. and Charrier, C.: Color Image Segmentation using Morphological Clustering and Fusion with Automatic Scale Selection. Pattern Recognition Letters, Elsevier, Vol. 30, pp. 397-406, March 2009
[20] Escobar, M.M., Foo, J.L., and Winer, E.: Colorization of CT images to Improve Tissue Contrast for Tumor Segmentation. Computers in Biology and Medicine, Elsevier, Vol. 42, pp. 1170–1178, September 2012.
[21] Necib, H., Garcia, C., Wagner, A., Vanderleinden, B., Emonts, P., Hendlisz, A., Flamen, P. and Buvat, I.: Detection and Characterization of Tumor Changes in 18F-FDG Patient Monitoring using Parametric Imaging. J. of Nucl. Med. 52, 354-361 (2011).
[22] Lartizien, C., Francisco, S.M. and Prost, R.: Automatic Detection of Lung and Liver Lesions in 3-D Positron Emission Tomography Images: A Pilot Study. IEEE Transactions on Nuclear Science, Vol. 59, pp. 102-112, February 2012.
[23] Changyang, L., Wanga, X., Xiaa, Y., Eberlb, S., Yinc, Y. and Feng, D.D.: Automated PET-guided Liver Segmentation from Low-Contrast CT Volumes using Probabilistic Atlas. Computer Methods and Programs in Biomedicine, Elsevier, Vol. 107, pp. 164-174, July 2011.
[24] Blechacz, B., and Gores, G.J.: PET scan for Hepatic Mass. Hepatology, Vol. 52, pp. 2186–2191, December 2010.
[25] Ming, X., Feng, Y., Guo, Y. and Yang, C.: A New Automatic Segmentation Method for Lung Tumor Based on SUV threshold on 18F-FDG PET images. In 2012 IEEE International Conference on Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS), pp. 5-8, July 2012.
[26] Yong, X., Stefan, E., Lingfeng, W., Michael, F., and David, D.D.: Dual-Modality brain PET-CT image segmentation based on adaptive use of functional and anatomical information. Computerized Medical Imaging and Graphics, Elsevier, Vol. 36, pp. 47-53, June 2011.
[27] Belhassen, S., and Zaidi, H.: A Novel Fuzzy C-means Algorithm for Unsupervised Heterogeneous Tumor Quantification in PET. Medical Physics, Vol. 37, pp. 1309-1324, March 2010.
[28] Geets, X., Lee, J.A., Bol, A., Lonneux, M. and Gregoire, V.: A gradient-based method for segmenting FDG-PET images: methodology and validation. Eur. J. of Nucl. Med. Mol Imaging, Springer 34, 1427–1438 (2007).

[29] Hatt, M., Rest, C.L., Turzo, A., Roux, C., and Visvikis, D.: Fuzzy Logically Adaptive Bayesian Segmentation Approach for Volume Determination in PET. IEEE Transactions on Medical Imaging, Vol. 28, pp. 881-893, January 2009.

[30] Li, H., Thorstad, W.L., Biehl, K.J., Laforest, R., Su, Y., Shoghi, K.I., Donnelly, E.D., Low, D.A. and Lu, W.: A Novel PET Tumor Delineation Method based on Adaptive region-Growing and Dual-Front Active Contours. Medical Physics, Vol. 35, pp. 3711-3721, August 2008.

[31] Baardwijk, A., Bosmans, G., Boersma, L.: PET-CT based Auto-contouring in Non-Small-Cell Lung Cancer correlates with Pathology and reduces Interobserver Variability in the Delineation of the Primary Tumor and involved Nodal Volumes. International Journal of Radiation and Oncology, Biology and Physics, Elsevier, Vol. 68, pp. 771-778 July 2007.

[32] Potesil, V., Huang, X., Zhou, X.: Automated Tumor Delineation using Joint PET/CT information. In: Proc. SPIE International Symposium on Medical Imaging: Computer-Aided Diagnosis, Vol. 65142, March 2007.

[33] Xia, Y., Wen, L., Eberl, S., Fulham, M., Fend, D.: Segmentation of Dual Modality Brain PET/CT images using MAP-MRF model. In 2008 IEEE 10th Workshop on Multimedia Signal Processing, pp. 107-110, October 2008.

[34] Yu, H., Caldwell, C., Mah, K.: Automated Radiation targeting in head-and-neck cancer using Region-based Texture Analysis of PET and CT images. International Journal of Radiation and Oncology, Biology and Physics, Elsevier, Vol. 75, pp. 618-625, October2009.

[35] Yu, H., Caldwell, C., Mah, K. Mozeg, D.: Coregistered FDG PET/CT-based Textural Characterization of Head and Neck Cancer for Radiation Treatment Planning. IEEE Transactions on Medical Imaging, Vol. 28, pp. 374-383, February 2009.

[36] Baisa L. Gunjal, and Suresh N. Mali, "ROI Based Embedded Watermarking of Medical Images for Secured Communication in Telemedicine", International Journal of Computer and Communication Engineering, Vol. 6, pp. 815-820 May 2012.

A Model for Implementing Security at Application Level in Service Oriented Architecture

Said Nabi, Saif Ur Rehman
Department of Computer Science
Shaheed Zulifikar Ali Bhutto Institute of Science and Technology (SZABIST)
Saidnabi115,saifi.ur.rehman{@gmail.com}

Simon Fong
Department of Computer and Information Science
University of Macau Taipa, Macau SAR
ccfong@umac.mo

Kamran Aziz
Abasyn University, Islamabad Campus
Islamabad, Pakistan
kamrandik@gmail.com

***Abstract*— Securing the communication channels only, cannot guaranty end-to-end security in SOA based systems. To provide complete security, there is need to provide security at application level for SOA based systems. But it is a great challenge for the developer of the web services to implement the security during development of the web services. In this paper we have proposed a model, which will automate the generation of security policies for web services. This system will facilitate and enable the developers of the web services, to generate and implement security policies during the development of the web services, without having intensive knowledge of the security domain and the underlying system. The proposed system will also make the application level experts independent of the security experts for the generation and implementation of the security policy for the development of the web services.**

***Index Terms*—Web Services Architecture (WSA), Service Oriented Software (SOS), Service Oriented Computing (SOC), Service Oriented Architecture (SOA), Application Level Security in SOA, SOA security policy.**

I. INTRODUCTION

Requirement engineering is considered as one of the most critical phases in software engineering, specifically, in software design and development. If errors are introduced at the requirement stage, then they remain undetected till the later stages of software development process and [1]. Requirement engineering addresses the issues of requirement collection to design and develop the desired software. Requirement engineering has a direct impact on all the stages of software development including software design, architecture, implementation, testing and deployment. Software architecture deals with design and development of the abstract level structure of the software. It consists of a number of architectural elements like components and connectors, which are assembled in such a way to satisfy the functional and performance requirements [2]. There are a number of architectural styles or patterns used by the software architects including layered systems, event-based, object-oriented, data-abstraction and implicit invocation, etc. Although these styles provide sufficient space of architectural choices to the architects, but alongside pose challenge for the architects to realize the tradeoffs while selecting the best suitable style in a particular situation and environment [3].

There is a remarkable difference among the development of traditional software and SOA based application, especially at design and analysis phase. SOA application requires undergoing analysis and design phases. Analysis phase produces candidate services from the business requirements. Business analysts and service architects emphasize on the usage of standards to refine the candidate services. Therefore, the formal definition of business processes is very important in SOA. The testing and development phase of SOA, however, are similar to the traditional development processes [4].

SOA is a set of services which are the collections of software component and carry out business process independently. Services must have the properties like loose-coupling, self-containment and should have well defined independent interfaces. Self-containment of the services mean it would be able perform their functionality independent of the other services. Loose-coupling mean services communicate with each other through sending messages and are not aware of technical details of other collaborator services [5]. The flexible and modular designs of software applications have a negative impact on the security of software applications [6]. SOA design

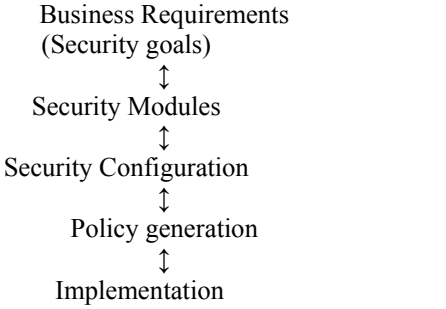
doi:10.4304/jetwi.6.1.157-163

principles provide guidelines for developing interoperable and agile service logic, which includes composition, discoverability, statelessness, independence, loose-coupling, abstraction, and reusability and service contract [7].

Although the information security standards like identification, authentication, authorization, confidentiality, integrity and availability have same connotation in SOA based applications, but the agile implementation of SOA makes it difficult to ensure secure SOA implementation within the organizations. There is a lesser possibility for successful implementation of secure SOA without specific information security guidelines [8]. There is a need for web service architecture that can control all the levels of web service stack which may include: describing, registering, managing, monitoring, deploying, wrapping, and discovering fundamental software components to self-repeatedly composing software [11]. The QoS attributes like performance, scalability, accounting and security plays a vital role for web services of different business domains [14]. The reputation and trust of data intensive services in cloud computing have generated important issues e.g., trust evaluation in collaboration services becomes a challenge. To represent service collaboration for data intensive services, the concept of service collaboration graph has been used in [15].

Securing the communication channels cannot guaranty end to end security in SOA based systems. Application level security plays a vital or key role to enable end to end security implementation in SOA. To deal with SOA security at application level, there is a need for expertise of two domains like security domain expertise and application domain expertise. There is a very rare possibility that a single person is equipped with both expertise in security domain and application domain. It is a great challenge for the developer of the web services to implement the security modules during development of the web services. To handle this issue, different researchers has proposed various techniques to facilitate the application domain experts to implement security standard for web services during its development.

Next section provides an abstract view that business and security requirements in the form of security goals are mapped into security policies, which will be implemented by the web services developers.



The above figure provides some basic idea to convert security relevant business requirements into security polices, which can be easily implemented by the web services developers. First the security goals for the business are identified, then these requirements are categorized into different security modules like identity management, confidentiality, Integrity, authentication and authorization. Next step is the configuration of above modules by applying different approaches like pattern oriented etc. further these patterns will be used by the application to generate the security policies in an automated way. Finally generated polices will be implemented by application developer for web services.

In this paper we propose a model, which will automate the generation of security policies for web services. This system will facilitate and enable the developer of the web services, to generate and implement security policies during the development of the web services, without having intensive knowledge of the security domain and the underlying system. The proposed system will make the application level experts independent of the security experts for the generation and implementation of the security policy for the development of the web services.

Previously some of the researchers has worked on the automatic generation of security polices and provided some basic information to enable the developer of the web services, to develop and implement security policies without having intensive knowledge of the security domain and the underlying system. These approaches have a number of limitations and are not mature enough to be implemented. [12] has proposed a model driven technique to generate the security policies for web services which can be easily implemented by developers of web services. But their proposed system still need for a security expert to provide some basic information about the security domain and help in the enforcement of a particular security policy to achieve the security goals for the web services. Similarly architecture for the security advisor has been proposed by [13]. It consists of an algorithm, which selects a pattern and generates an enforceable policy for given security requirements. In other word it facilitates and enables the developers of the web services to choose the security goals, generate and implement security policies with having complete knowledge of the underlying system. But proposed architecture is still depended on the security experts to capture the basic security knowledge into security pattern and application needs to be preconfigured by the security expert. To generate the security policies, it uses the Apache axis2 as application server for web services which has some limitations like complexity from user point of view and not full support for JAX-WS. But comparatively we will use the Apache CXF for enforcing standards for policy generation. Usage of Apache CXF has the following advantages over the Apache axis2.

The rest of this paper is organized as follows: the literature survey on application level security in SOA is provided in section II. A proposed system is discussed in section III. Evaluation of the proposed system is provided in section IV. Finally, we conclude in section VI along with potential future directions to this research.

II. Literature Review

According to [9] Security requirements models are used by Model driven security (MDS) approach at abstract level to generate security policies automatically. These policies are used for security services configuration. The MDS solutions focus on a single security pattern for each security requirement. This is not enough because the current Cloud and SOA services are distributed among different heterogeneous security domains and continuously changing infrastructure. It need multiple security pattern support for single security requirement. To present such security services whose configuration support for different security patterns is still a challenging task. To solve this problem author has presented a framework that adds and integrates a new layer "security pattern refinement layer" to existing MDS layers. The security pattern refinement layer helps to configure a single security service with various patterns. This framework has been properly validated by applying on healthcare system. But the proposed system is still dependable on the security domain experts and not clearly mentioned how to automate the generation of security artifacts/policies.

Menzel et al. [10] has found that although to represent security requirements as Security policies can insure the flawless usage of services. But as specification of the web services are complex, so it is a tricky and error prone task to practically implement the security policies. The proposed system will facilitate the mapping of architectural models representing simple security goals into security policies by Model driven technique. Security configuration patterns of the web services play major role in mapping process. The SOA meta model of the security configuration patterns provide a foundation to describe the objects and their association at the modelling layer. This system will also provide a formal pattern structure and to specify these pattern(s), a domain specific language was used.

Imamura et al. [8] has stated that the current tools used for the configuration of security assets of the web services presents a technology aspect, where user must fill the gap between configuration and security needs manually. This leads to misconfiguration problem and extra configuration costs. Author introduces the SOA and MDA and presented a framework for polishing the user's security requirements by using the idea of SOA and MDA. In this framework the users describe their requirements using the given vocabulary list and then transform these requirement step wise into details requirements. After getting sufficient details of the security requirements, the users select the countermeasures and transform them into a required level of details. Using a policy language the users illustrate these details and uses best practice patterns for linking the gap between these levels. The future work includes extending the tool and framework and also developing an easy to use way to select an appropriate pattern from those currently available.

Some of researchers have presented techniques used to automate the security policies generation e.g., [12] has proposed an idea of meta model driven security mechanism to ease the task of web services developers to implement security in the web services. Security meta model give a base for exchange of information and model interactions. These also illustrate the fundamental objects, associations and related roles in service oriented architecture. But it is still dependent on the security domain experts to generate and configure all the security patterns and there is no way to dynamically add and configure new security patterns.

Architecture for the security advisor has been proposed in [13]. It consists of an idea and workflow for generating security policy automatically for web services, which selects a pattern and generates an enforceable policy for given security requirements. In other word it facilitates and enables the developers of the web services to choose the security goals and configure the security module without having complete knowledge of the underlying system. But proposed architecture is still depended on the security experts to capture the basic security knowledge into security pattern and application needs to be preconfigured by the security expert. To generate the security policies, it uses the Apache axis2 as application server for web services which has some limitations like complexity from user point of view and not full support for JAX-WS. Usage of Apache CXF has the following advantages over the Apache axis2.

i) CXF uses standard API's while axis2 in general uses proprietary things.
ii) CXF is more responsive to the users issues and ensure the quick availability (release fixpacks every month or after 2 months) of the "fixpacks" to the user, while axi2 has least compatibility for older version and slow (user have to wait from 9 or 10 months to get complete patches) response to the user issues .
iii) Apache CXF is well advised for those who use Spring framework because CXF have a better integration for spring framework. CXF is considered as more embeddable into other applications.
iv) Performance wise both apache axis2 and apache CXF are comparable but in case, when you use JAXB API's then CXF is faster than axis2.

III. Proposed System

In this section we propose a model, which will help to automate the generation of security policies for web services. This system will facilitate and enable the developer of the web services, to implement security policies during the development of the web services, without having intensive knowledge of the security domain and the underlying system. The proposed system will make the application level experts independent of the security experts for the generation and implementation of the security policy for the development of the web services.

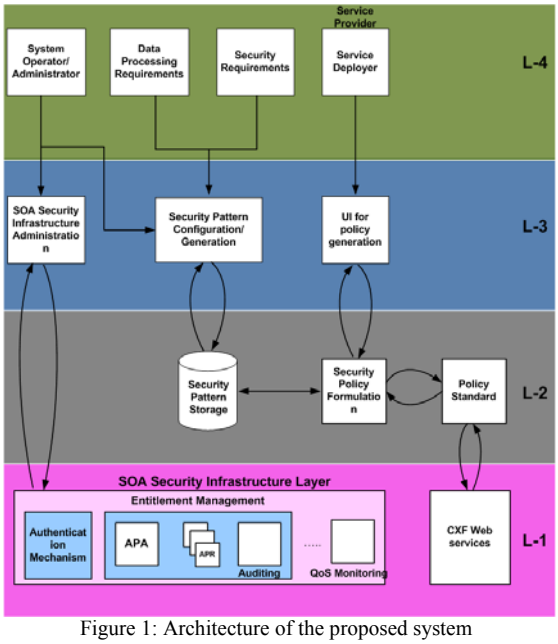


Figure 1: Architecture of the proposed system

The above figure shows the basic architecture of our proposed system, which is partitioned into four different layers including, security infrastructure layer, Business Logic Layer, Processing Layer and Interaction or application layer.

A. Infrastructure Layer

SOA infrastructure security has vital role in providing security to the web services. Infrastructure layer includes authentication, access-policy administration, access policy resolution and auditing etc. Enforcement of access policy is unluckily access policy resolution, auditing and administration are owned by different people other than developers of the business logic, changes at different times and different rates. The main benefit of separating infrastructure layer from the business logic layer is that, in case of any compliance or change in the security requirement, to adopt modification in access policy without need in the business logic of the web services. For example, consider the following scenario of Order Management of the Component Service, in which access policy for the above scenario (Component Order Management Component Service) are described as:

i) Order can only be entered by those people, who have a "broker" role,
ii) Order can only be updated by the owner or a manager of the owner of the order and
iii) Order can only be read by the owner, manager of the owner or a subject of the owner with the role of "reconcile".

In the given scenario the task of the order management development, the developers are required to highly focus on how to implement the order management services most efficiently. The resolution of access policy requires accessing the proper contextual information. The policy resolution shows wither to deny or permit the given request, which is enforced by the policy enforcement according to the decision on the request. To provide effective and practical SOA security infrastructure should allow distributed access-policy resolution by using various distributed objects of the resolution service, because the centralized access policy resolution has some scalability, availability, and performance related issues.

The auditing of access policy is needed for composite service and component service as well or the business process which invokes the component services and auditing is especially valuable for SOA based environment. There are very controlled and limited use of the application functions in non SOA based environment. On the other hand SOA based environments, the services will may called from very unpredictable and diverse ways. Auditing is a critical tool to predict issues, before they occur and locate the basic reason of problems when they do occur. Infrastructure layer also consists of application server for web services which is named as apache CFX. Apache CXF provide and enforces standards for policy generation.

B. Business Logic Layer

Business logic layer is the second layer of proposed system and consists of: security pattern repository, security policy formulation module and security policy standard provider and enforcer:

i) Security pattern repository is used to store different security patterns generated at the processing/operational layer. It communicates with security pattern generation and configuration module and on the other side it will communicate with security policy formulation module by providing the required security pattern(s).
ii) Security formulation module: this module plays a key role in the security policy creation and is the major focus of this research. After data processing requirements and security requirement has been captured in the security patterns and refined, the security formulation module is used to generate enforceable policy based and knowledge captured in the security patterns.
iii) Policy standard component of business logic layer provide and help in enforcement of security policy standards.

The detailed work flow of the security policy formulation module is shown in Figure 2.

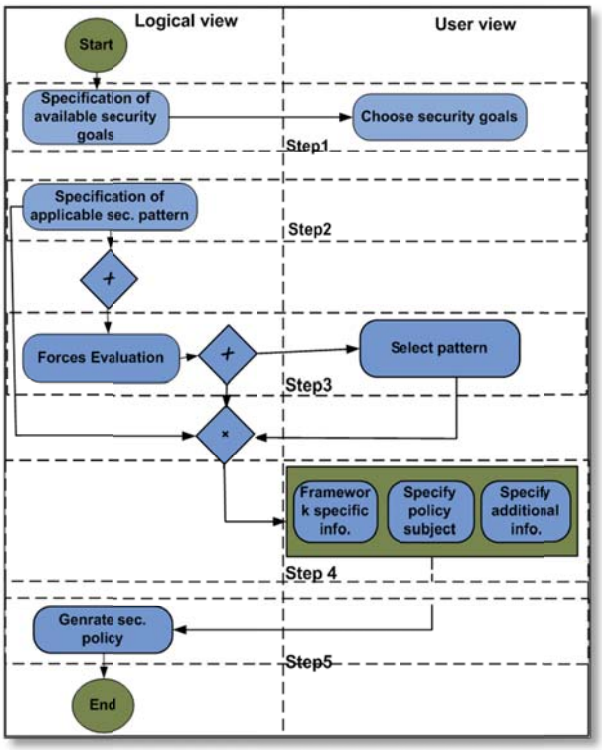


Figure2: Working details of the Security policy Formulation module

The above figure shows a model of the working flow of application for security policy formulation and is categorized into six major steps. In the first step, application will perform full search to identify the available security goals/challenges. In this step, the users of the application will select the required security goals like confidentiality, integrity, authentication etc, based on the requirements captured from business experts and requirement engineers. User can select more than one goals based on their need for the corresponding web services development. In the second step, based on the user selection of the required goals, the application will identify all those security pattern(s), which can achieve or helpful in achieving the required goal(s). If there were more than one relevant security patterns applicable to a particular security goal, then this process will enter to the forth step, otherwise it will be forwarded for policy generation. Third step is performed only when there are several approaches (in terms of security patterns) are available, then the security policy formulation module will be use the forces attributes (constraints) to select the most suitable security pattern for the security policy generation. After selecting the most relevant and highly applicable security patterns, next step of the security policy formulation module mostly concern with users of the module or application and before going to allow the security policy formulation application to generate the security policy for the selected security goals, the users are required to enter some additional information including, specification of the corresponding framework, provide information about policy subject, and some other information to further refine the policy generation process. The availability of the specific information about the underlying framework, where the generated policy will be implemented, is very important. The reason is that, it is not necessary and even not possible that each framework will be able to support the same security patterns i.e to provide confidentiality for communicating messages, using encryption techniques and different security pattern define separate encryption methods like a one security patterns uses AES-128 encryption standard and other uses security pattern uses simple DES technique for encrypting messages. Similarly it is not possible for all frameworks to support AES-128 encryption techniques. In the last step the user will generate security policy.

C. *Processing/Operational Layer*

Processing/operational layer is the third layer of our proposed system. This layer acts as an interface between interaction layer and Business logic layer. It has three major Components including, SOA security infrastructure administration view, security pattern generation and configuration and User Interface for security policy generation. The SOA security infrastructure administration view is used by the local domain administrator to provide relevant security information of the underlying infrastructure and to recognize and specify all relevant information to develop the enforceable security policies. Security pattern generation and configuration part of processing/operational layer is used by the system operator and application domain expert to generate and configure security patterns based on the knowledge captured from business process experts and requirement engineers. Security pattern generation and configuration module also interact with security pattern storage to confirm the existence of a particular security pattern. Before generation and configuration of a new security pattern, it will first interact and scan completely the security pattern storage to insure that wither the required pattern is already exist or any pattern can reused to generate and configure new pattern(s). After proper configuration of security patterns and underlying security infrastructure for achieving the required security goal the developer of the web services will use security policy formulation User Interface to generate the required security policy.

D. *Interaction Layer*

Interaction layer is the top most layer of our proposed system, where developers or system operators and service deployer interact with application. In this layer, data processing requirements are collected from business analyst and business process experts and security requirements provided and analysed by the requirement engineers for further use.

IV. Comparison-based Evaluation of Proposed System

To generate the security policies for web services, here we present different approaches initiated by other researchers. These research initiatives mostly concentrate on meta model, model driven and pattern oriented techniques to generate security policies. Few of these

researcher has focus on idea, to automate the policy generation process and has provided some basic information and future plans.

Menzel and Meinel, has proposed a meta-model based technique to generate the security policies for web services which can be easily implemented by developers of web services. But their proposed system still need for a security expert to provide some basic information about the security domain and help in the enforcement of a particular security policy to achieve the security goals for the web services. The proposed system does not provide complete information about how to generate and implement web services security polices with least required effort and lack of discussion of the automatic generation of security policies.

The system proposed by Menzel et al. [10] will facilitate the mapping of architectural models representing simple security goals into security policies by Model driven technique. Security configuration patterns of the web services and their Meta model provide a foundation to describe the objects and their association at the modelling layer. This system will also provide a formal pattern structure and to specify these pattern(s), a domain specific language was used. On the other hand for the proposed system, to secure web services, possibly there is a need for knowledge of the domain experts to devise a proper plan. Pattern engine is at initial stage, providing limited functionality and need further enhancement. The proposed system is unable to provide full information about how to generate and implement web services security polices and lack of discussion of the automatic generation of security policies.

Similarly architecture for the security advisor has been proposed by Schnjakin et al.[13]. The security advisor is an idea of an application which will enable automated generation of security policies. It selects a pattern(s) pre-configured by the security domain experts and generates an enforceable policy for given security requirements in an automated way. But on the other side the proposed architecture is still depended on the security experts to capture the basic security knowledge into security patterns and application needs to be preconfigured by the security experts. There is a possibility to have a new pattern(s) which is required to achieve a certain security goal or handle a security challenge(s), to generate and configure this new pattern, there is need to consult with security experts. In the proposed system a workflow for the automation of security policy has been presented without providing any proper algorithm or interface. To generate the security policies, it uses the Apache axis2 as application server for web services which has some limitations like complexity from user point of view and not full support for all plate forms like JAX-WS.

Our proposed system will not only combine most of the functionality provided by the system presented by Menzel et al.[10], Menzel and Meinel [12], and Schnjakin et al.[13], but also proposed some additional changes to enhance the SOA security policy generation and to automate this process. The proposed system will not be depended on the security domain experts for capturing security related information in to the security patterns and their configuration. New pattern can be generated and configured as per requirement easily and dynamically. Further comparatively we will use the Apache CXF for enforcing standards for policy generation. Apache CXF will solve the issues faced while using apache axis2 like multi-plateform support and will reduce the complexity from usability point of view. This system will generate security policy in both machine readable format like XML and normal text form.

V. Conclusion and Future Work

Due to the agile nature of SOA, it is difficult to ensure the secure implementation of SOA within the organizations. Most of the techniques proposed in the literature focus on the use of security patterns, which provide some basic concept and understanding about the security domain needed for the web services developers. But, it is still a challenge for the developers of the web services to implement security modules for the web services.

In this paper we have proposed a model, which will enable the developers of the web services to automatically generate the security policies for web services. This system will also facilitate the developer of the web services, to implement security policies during the development of the web services, without having thorough knowledge of the security domain and the underlying system. The proposed system will make the application level experts independent of the security experts for the generation and implementation of the security policy for the development of the web services. In future, we intend to look into the possibility of further refinement/enhancement of the automatic generation of security policy so that developers can implement security policies on the fly.

References

[1] A. Chakraborty, M. K. Baowaly, A. Arefin, A. N. Bahar. "The Role of Requirement Engineering in Software Development Life Cycle", Journal of Emerging Trends in Computing and Information Sciences, ISSN : 2079-8407 , Vol. 3, No. 5, pp: 723-729, 2012.

[2] P. Kruchten "The 4+ 1 view model of architecture. Software", IEEE, 12(6), 42-50, 1995

[3] D. Garlan, & M. Shaw, "An introduction to software architecture," 1994.

[4] N. A. Delessy, "A Pattern-Driven Process for Secure Service-Oriented", In Workshop on Security in Object-oriented Systems, Florida Atlantic University. Vol. 70, p. 79, 2008.

[5] D. T. Sanders, J. A. Hamilton Jr., & R. A. MacDonald, "Supporting a service-oriented architecture, "Proceedings of the 2008 Spring simulation multiconference. Society for Computer Simulation International, 2008

[6] N. Delessy and E. B. Fernandez. "A pattern-driven security process for SOA applications," Proceedings of the 3rd Int. Conf. on Availability, Reliability, and Security (ARES 2008).Barcelona, Spain, 2008

[7] J. Chetty, M. Coetzee, "Towards An Information Security Framework For Service-oriented Architecture," IEEE 2010.

[8] T. Imamura M. Tatsubori Y. Nakamura, C. Giblin," Web Services Security Configuration in a Service-Oriented Architecture," WWW, pp. 1120-1121.ACM, 2005
[9] G. D. M. H., Menghwar, Depar, A. A. Jalbani, W. M. Mashwani, (2012). Security modeling for service-oriented systems using security pattern refinement approach. Software & Systems Modeling, 1-24.
[10] M. Menzel, R. Warschofsky, C. Meinel, (2010, July). A pattern-driven generation of security policies for service-oriented architectures. In Web Services (ICWS), 2010 IEEE International Conference on (pp. 243-250). IEEE.
[11] Y. Baghdadi, "A metadata for Web services architecture: A framework for service- oriented software development," GCC Conference & Exhibition, 2009 5th IEEE Issue, On page(s): 1 – 6. March 2009.
[12] M. Menzel and C. Meinel. "A security meta-model for service-oriented architectures," In Proc. SCC, 2009.
[13] M. Schnjakin, M. Menzel, and C. Meinel. "A pattern-driven security advisor for service- oriented architectures," Pro 6th Workshop SWS (in conjunction with 16th ACM CCS),ACM Press, Chicago, USA, pages 13–20, 2009.
[14] A. Wahl, B. Hollunder, V. Sud and A. Al-Moayed. Quality Attributes for Web Services: A Model-based Approach for Policy Creation. International Journal On Advances in Software, 5(3 and 4), 166-178, 2012.
[15] L. Huang, S. Deng, Y. Li, J. Wu, J. Yin and G. Li. "A Trust Evaluation Mechanism for Collaboration of Data-Intensive Services in Cloud", Appl. Math, 7(1L), 121-129, 2013.

TABLE 1

COMPARISON OF OUR SECURITY MODEL FOR POLICY GENERATION WITH EXISTING SECURITY POLICY MODELS

Security Meta-Model for SOA[12]	**Pattern Driven Security advisor for Policy generation[13]**	**Pattern driven generation of security policy in SOA[10]**	**Our Security Model for policy generation**
Meta-Model Based	Pattern Driven	Model -cum-Pattern driven approach	Pattern Based
Define basic -Entities-Relationships -Roles Associated	Define security advisor(application)	Define -Transformation of architectural model into security policies	Define -security policy formulation module
Process Requirements includes Business requirements -Security requirement	-	Process Requirements includes -Security requirements in the form of simple security intensions.	Process Requirements includes -Business requirements -Security requirement
	Security domain knowledge provided by the security experts	Possibly there is a need for knowledge of the domain expert to make a proper strategy to secure web services and other resources	Security requirements and Business requirements are directly taken from -Business process experts -Requirement engineers
Capture process requirements into: Meta Models	Capture security domain knowledge into: -security patterns	Capture security domain knowledge into: -Security policy model using security modelling language like secureSOA	Capture business and security requirements into: -Security patterns
Dependent on:- Security domain Expert	Dependent on:- Security domain Expert	-	Do not dependent on: -security expert
Application pre-configuration by the security experts is needed	Application pre-configuration by the security experts is needed	Security configuration pattern provide expert knowledge on web services	-
	-	-	Directly generate/configure new security pattern(s) based on requirements
	Apache Axis2 web application server is used	-	Apache CXF web application server is used
Provide mapping to WS Policy and WS Security Policy	-	Transform simple security intentions into security policies.	-
Informal stated security patterns	Informal stated security patterns	Introduced a formalised pattern structure of security patterns.	Using pattern engine [10]
	Generate policy in XML form	-	Generate Security policy in both XML and normal Text form

Call for Papers and Special Issues

Aims and Scope

Journal of Emerging Technologies in Web Intelligence (JETWI, ISSN 1798-0461) is a peer reviewed and indexed international journal, aims at gathering the latest advances of various topics in web intelligence and reporting how organizations can gain competitive advantages by applying the different emergent techniques in the real-world scenarios. Papers and studies which couple the intelligence techniques and theories with specific web technology problems are mainly targeted. Survey and tutorial articles that emphasize the research and application of web intelligence in a particular domain are also welcomed. These areas include, but are not limited to, the following:

- Web 3.0
- Enterprise Mashup
- Ambient Intelligence (AmI)
- Situational Applications
- Emerging Web-based Systems
- Ambient Awareness
- Ambient and Ubiquitous Learning
- Ambient Assisted Living
- Telepresence
- Lifelong Integrated Learning
- Smart Environments
- Web 2.0 and Social intelligence
- Context Aware Ubiquitous Computing
- Intelligent Brokers and Mediators
- Web Mining and Farming
- Wisdom Web
- Web Security
- Web Information Filtering and Access Control Models
- Web Services and Semantic Web
- Human-Web Interaction
- Web Technologies and Protocols
- Web Agents and Agent-based Systems
- Agent Self-organization, Learning, and Adaptation
- Agent-based Knowledge Discovery
- Agent-mediated Markets
- Knowledge Grid and Grid intelligence
- Knowledge Management, Networks, and Communities
- Agent Infrastructure and Architecture
- Agent-mediated Markets
- Cooperative Problem Solving
- Distributed Intelligence and Emergent Behavior
- Information Ecology
- Mediators and Middlewares
- Granular Computing for the Web
- Ontology Engineering
- Personalization Techniques
- Semantic Web
- Web based Support Systems
- Web based Information Retrieval Support Systems
- Web Services, Services Discovery & Composition
- Ubiquitous Imaging and Multimedia
- Wearable, Wireless and Mobile e-interfacing
- E-Applications
- Cloud Computing
- Web-Oriented Architectrues

Special Issue Guidelines

Special issues feature specifically aimed and targeted topics of interest contributed by authors responding to a particular Call for Papers or by invitation, edited by guest editor(s). We encourage you to submit proposals for creating special issues in areas that are of interest to the Journal. Preference will be given to proposals that cover some unique aspect of the technology and ones that include subjects that are timely and useful to the readers of the Journal. A Special Issue is typically made of 10 to 15 papers, with each paper 8 to 12 pages of length.

The following information should be included as part of the proposal:

- Proposed title for the Special Issue
- Description of the topic area to be focused upon and justification
- Review process for the selection and rejection of papers.
- Name, contact, position, affiliation, and biography of the Guest Editor(s)
- List of potential reviewers
- Potential authors to the issue
- Tentative time-table for the call for papers and reviews

If a proposal is accepted, the guest editor will be responsible for:

- Preparing the “Call for Papers” to be included on the Journal’s Web site.
- Distribution of the Call for Papers broadly to various mailing lists and sites.
- Getting submissions, arranging review process, making decisions, and carrying out all correspondence with the authors. Authors should be informed the Instructions for Authors.
- Providing us the completed and approved final versions of the papers formatted in the Journal’s style, together with all authors’ contact information.
- Writing a one- or two-page introductory editorial to be published in the Special Issue.

Special Issue for a Conference/Workshop

A special issue for a Conference/Workshop is usually released in association with the committee members of the Conference/Workshop like general chairs and/or program chairs who are appointed as the Guest Editors of the Special Issue. Special Issue for a Conference/Workshop is typically made of 10 to 15 papers, with each paper 8 to 12 pages of length.

Guest Editors are involved in the following steps in guest-editing a Special Issue based on a Conference/Workshop:

- Selecting a Title for the Special Issue, e.g. “Special Issue: Selected Best Papers of XYZ Conference”.
- Sending us a formal “Letter of Intent” for the Special Issue.
- Creating a “Call for Papers” for the Special Issue, posting it on the conference web site, and publicizing it to the conference attendees. Information about the Journal and Academy Publisher can be included in the Call for Papers.
- Establishing criteria for paper selection/rejections. The papers can be nominated based on multiple criteria, e.g. rank in review process plus the evaluation from the Session Chairs and the feedback from the Conference attendees.
- Selecting and inviting submissions, arranging review process, making decisions, and carrying out all correspondence with the authors. Authors should be informed the Author Instructions. Usually, the Proceedings manuscripts should be expanded and enhanced.
- Providing us the completed and approved final versions of the papers formatted in the Journal’s style, together with all authors’ contact information.
- Writing a one- or two-page introductory editorial to be published in the Special Issue.

More information is available on the web site at http://www.academypublisher.com/jetwi/.

An Adaptive Weight Calculation based Bandwidth Allocation Scheme for IEEE 802.16 Networks 142
Akash Deep and KS Kahlon

Tumor Detection by Color Segmentation of PET/CT Liver Images 148
Neha Bangar and Akash Deep

A Model for Implementing Security at Application Level in Service Oriented Architecture 157
Said Nabi, Saif Ur Rehman, Simon Fong, and Kamran Aziz

(Contents Continued from Back Cover)

A WIFI Antenna Radiation Effects on Human Head in the ISM Band 45
Radouane Karli and Hassan Ammor

Development of a New Slit-Slotted Shaped Microstrip Antenna Array for Rectenna Application 49
Mohamed Adel Sennouni, Jamal Zbitou, Abboud Benaissa, Abdelwahed Tribak, Otman Elmrabet,and Mohamed Latrach

Performance Evaluation of WideMac Compared to ALOHA in term of Energy Consumption for IR-UWB based WSN 54
Anouar Darif, Rachid Saadane, and Driss Aboutajdine

Euclidean 3D Reconstruction of Unknown Objects from Multiple Images 59
Soulaiman El Hazzat, Abderrahim Saaidi, and Khalid Satori

Comparative Study of PCA, ICA, LDA using SVM Classifier 64
Anissa Bouzalmat, Jamal Kharroubi, and Arsalane Zarghili

SVM Classification of High Resolution Urban Satellites Images using Composite Kernels and Haralick Features 69
Aissam Bekkari, Soufiane Idbraim, Driss Mammass, Mostafa El Yassa, and Danielle Ducrot

3D Normalization Based on the Barycentric Coordinates 75
El Mostafa Rajaallah, Lahcen Moumoun, and Taoufiq Gadi

Optimization of the Organized KOHONEN Map by a New Model of Preprocessing Phase and Application in Clustering 80
Es-safi Abdelatif, Harchli Fidae, and Ettaouil Mohamed

Adaptive Learning 85
Soumaya El Mamoune, Marouane El Mabrouk, Loubna Cherrat, Ahmed Hadioui, Omar Kassara, and Mostafa Ezziyani

A Comparative Study between the Pseudo Zernike and Krawtchouk Invariants Moments for Printed Arabic Characters Recognition 89
Rachid Salouan, Said Safi, and Belaid Bouikhalene

People Tracking using Color Control Points and Skin Color 94
K.Laaroussi, A.Saaidi, K.Satori

Fuzzy Control System for Autonomous Navigation of Thymio II Mobile Robots 101
Fatma Boufera, Fatima Debbat, Francesco Mondada, and M. Faycal Khelfi

SURVEY PAPERS

A Survey on Prevention Approaches for Denial of Sleep Attacks in Wireless Networks 106
G.Mahalakshmi and P.Subathra

A Comprehensive Study of Finding Copy-and-Paste Clones from Program Source Codes 111
Kamran Khan, Saif Ur Rehman, Kamran Aziz, and Simon Fong

REGULAR PAPERS

An Optimized Parallel Computing Paradigm for Mobile Grids Based on DSPOM 119
Aghila Rajagopal and M.A. Maluk Mohamed

DWDE-IR: An Efficient Deep Web Data Extraction for Information Retrieval on Web Mining 133
Aysha Banu and M. Chitra