International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 1 - 4, 2014

# Software Engineering Challenges in Pervasive Computing: A review

Surabhi Soni
Engineer,
Mahindra Comviva,
Bangalore, India

Ayush Rastogi
Development Engineer,
Laurus Infosystems,
Bangalore, India

**Abstract**-Moving away from decades of machine-centric computing and making pervasive human-centric computing, the new wave of computing, a reality revolutionizes the relationship between humans and computing systems. There is a growing interest in the use of context-awareness as a technique for developing pervasive computing applications that are flexible, adaptable, and capable of acting autonomously on behalf of users The software challenges to turn such pervasive or ubiquitous computing environments into reality are enormous In this paper, we review some of the challenges of software engineering in pervasive computing.
**Keywords**: Human-centric computing, Pervasive computing, Software Engineering, Ubiquitous computing, human-machine interaction.

## 1. INTRODUCTION

Pervasive computing introduces a radically new set of design challenges when compared with traditional desktop computing. In particular, pervasive computing demands applications that are capable of Operating in highly dynamic environments and placing minimal demands on user attention. Late Michael Dertouzos, pointed out the need for pervasive computing. He stated the following:

*If computers are to live up to the promise of serving us, they will have to change drastically and never again subject us to the frustrating experiences we all have shared* [Dertouzos, M.L.2001][1].

It is envisioned that pervasive computing systems will help people to achieve more while doing less. These systems will:

- Understand us when we speak to them;
- Do much of our routine brainwork for us;
- Get us the information we want, when and where we want it;
- Help us work with other people across space and time;
- Adapt on their own to our individual needs and desires [Dertouzos, M.L.2001] [1].

In the pervasive computing era, we will not need to carry our own physical devices with us anymore. Instead, configurable generic devices, either handheld or embedded in the environment, will bring computation to us, whenever we need it and wherever we might be. As we interact with these *anonymous* devices, they will adopt our information personalities. They will respect our desires for privacy and security.

Pervasive human-centric computing systems will change how businesses, organizations and governments work with each other, as well as how individuals interact. It represents the dawn of a new era in Information Technology (IT) [Dertouzos, M.L.2001][1].

To shift the focus of computing from machines to humans, major changes are required not only in technologies and systems, but also in the approach to developing, deploying and managing technologies and systems. Weiser presented his vision for pervasive human centric computing in 1991. He further articulated his vision as follows:

*There is more information available at our fingertips during a walk in the woods than in any computer system, yet people find a walk among trees relaxing and computers frustrating. Machines that fit the hum an environment instead of forcing humans to enter theirs will make using a computer as refreshing as taking a walk in the woods* [Weiser, M.1993] [Weiser, M.1991][3][4].

## 2. MOTIVATION

Challenges which are based on natural characteristics of pervasive computing systems (i.e. mobility, dynamism and heterogeneity) can be evaluated from a more domain specific perspective,

Hence, we particularly list following basic interrelated requirements for such pervasive learning environments:

(1) *Device independence*: applications and data should be always accessible without any device dependence,

(2)*Application independence*: data should be always accessible without any application dependence,

(3)*Adaptivity and Adaptability*: learning environment and elements of this environment should dynamically adapt

according to context of learner(s) and users should be able to configure such environments such as composing/decomposing data and applications,

(4)*Collective Operation*: applications in such environments must be able to collectively operate for the benefit of users in a seamless manner. Adaptivity is long studied both in adaptive web systems and adaptive e-learning systems [P. Brusilovsky et.al 2007][10], and in such systems adaptivity is generally considered as an aspect between user and application based on user profiles and models.

Pervasive human-centric computing systems are proactive systems that are capable of sensing, measuring, monitoring, predicting and reacting to physical world conditions. To support a wide variety of human activities, pervasive human-centric computing systems must be:

Pervasive: being available everywhere and accessing the same information base through every portal

Embedded: sensing and affecting the physical world;

Nomadic: allowing users and computations to move around freely to meet the users' needs;

Adaptable: providing flexibility and spontaneity in response to changes in the user's requirements and operating conditions;

Powerful yet efficient: freeing itself from constraints imposed by bounded hardware resources, addressing system constraints imposed by user demands and available power or communication bandwidth;

Intentional: enabling people to name services and software objects by intent;

Eternal: never requiring shut down or reboot while components are added or removed in response to demands, errors or upgrades [MIT PROJECT OXYZEN][2].

In a pervasive computing environment, user and perceptual technologies will directly address human needs and consist of the following:

Collaboration Technologies: enabling the formation of spontaneous collaborative regions that accommodate the needs of mobile people and computations, and also provide support for recording and archiving speech and video fragments from a variety of sources and/or events;

Knowledge access technologies: offering vastly improved access to information and customized to the needs of users (i.e. people, applications and software systems);

Automation Technologies: offering natural, easy to-use, customizable and adaptive mechanisms for automating and tuning repetitive information and control tasks;

Perceptual Technologies like speech and vision technologies: enabling communication with devices, networks and software to extend the range of user technologies delivered to all places.

## 3. CONTEXT AWARE PERVASIVE COMPUTING

*Context aware computing aims to enable device to provide better service for people through applying available context information* [L.Han et.al2008][11].

Above a generic definition of context aware computing is given, which emphasizes the relation between user, context and computing, but how do we apply available context information? Although various categorizations for context-aware systems are already given [J.Pascoe1998][12], we prefer to re-interpret these categorizations based on adaptive systems, particularly according to adaptive web systems. This is because we defined adaptively as a key factor of intelligence and as a key relation between context and computing for context-aware computing systems. Therefore by referring to the field of adaptive web [P.Brusilovsky et.al2007][10] for categorization of context aware computing applications, below categorization have been proposed:

(1) *Context based filtering and recommendation of information and services*: examples might include finding the nearest printer, accessing the history of a nearby object etc.,

(2) *Context based presentation and access of information and services*: e.g. selecting voice when screen displays are not available (multimodal information presentation and user interfaces), dynamic user interfaces etc.,

(3) *Context based information and service searching*: e.g. location aware query rewriting for a search for available restaurants (query rewriting is a technique used in adaptive web systems for information filtering by rewriting a user query according to the user profiles) etc.,

(4) *Context adaptive navigation and task sequencing*: adaptive navigation is a technique employed in adaptive web systems. We can extend this idea in pervasive computing since a user's interaction might consist of several related sub-tasks in relation with his goals and might lead to context aware task sequencing,

(5) *Context based service and application modification/configuration:* this need mainly arises from different devices available in the environment, e.g. disabling particular features depending on the capabilities of target device,

(6) *Context based actions*: [J.Mantyjarvi et.al.2003][13] proposes three levels of context dependent automatic actions: manual, semi-automatic, and automatic. [11] Notes that fully automatic actions based on context are rarely useful, and incorrect action scan be frustrating,

(7) *Context based resource allocation*: this might include allocating physical recourses (e.g. memory, even non-hardware physical resources) for the use of other entities in the setting (e.g. applications, users etc.).

It is worth to note that, adaptive behaviors of context aware systems are not necessarily need to depend on the current context, rather such systems should also be able to adapt proactively by making use of current context or historical

context to predict future context of the setting. An example is given in [J.Coutaz et.al.2005] where a user walks through the building and submits a printing request, the selected printer should not depend on the user's current location but rather to his final destination.

## 4. SOFTWARE ENGINEERING CHALLANGES

Pervasive computing has introduced new high level system requirements that should be taken in consideration in the design and implementation process.

Interoperability is highly demanded in all the levels of pervasive systems.

Software components should be built independently of the context, this way they will be used in different computing environments and applications.

Heterogeneity is a challenge in pervasive environments; mobile users interact with the system using different hardware devices, the context and connectivity become dynamic. Meanwhile, the user also has a dynamically evolving profile. As a result the software should provide services that adapt with different screen resolutions, user-interaction methods, machine power and processing capacities.

Mobility is an important requirement. Actual mobility is the capability of an autonomous software agent to dynamically transfer its execution towards the nodes where the resources it needs to access are located. Exploiting this form of mobility will save network bandwidth and increase the execution reliability and efficiency.

This aspect can be deployed to help embedded software agents to follow mobile users wherever they go. Virtual agent mobility is the ability to be aware of the multiplicity of networked execution environments [Chug ES Et.al 2004][6].

Survivability and security provide systems with powerful capacities. Survivability is the ability of a system to fulfill its mission on time despite the presence of attacks and failures. Such functioning requires a self-healing infrastructure with improved qualities such as security, reliability, availability and robustness. An application may employ security mechanisms, such as passwords and encryption, yet may still be fragile by failing when a server or network link dies. The literature has presented two kinds of survivability in the context of software security: survival by protection (SP) allows security mechanisms like access control and encryption to ensure survivability by protecting applications from malicious attacks. The survival by adaptation (SA) gives the application the ability to survive by adapting itself to the changing dynamic conditions [Chug ES Et.al2004][6].

Continuity is a very demanding feature in ubiquitous/pervasive applications especially with the uncertainty and instability of user connectivity while he moves around.

The application should have the possibility to pause the user session in the case of sudden disconnection and continue to work later without losing information[ Weiser, M 1993][4].

Usability, according to human-machine interaction engineers and to interactive system designers, is how easy an interface design is to be understood and used, how unambiguous interactive controls are and how clear its navigational scheme is [Gschwind T 2002][9]. This feature is the final objective of pervasive systems.

## 5. CONCLUSION

To make pervasive computing a reality, a significant amount of research is still needed; much of it in software engineering. To ensure that pervasive computing systems fulfill their required and intended purpose and can trust these systems to perform as intended Software is an essential ingredient. To produce pervasive computing systems efficiently and economically we need new methods and techniques to meet the new challenges. Our techniques must be able to deal with highly dynamic, highly heterogeneous environments and scale the range of tiny to powerful computing elements.

## 6. REFERENCES

[1] Dertouzos, M.L., *The Unfinished Revolution: Human-Centered Computers and What TheyCan Do for Us*. New York: HarperCollins (2001).

[2]. MIT Project Oxygen, http://www.oxygen.lcs.mit.edu/

[3]. Weiser, M., The computer for the twenty-first Century. *Scientific American*, 265, **3**, 94-104, 1991

[4]. Weiser, M., Some computer science issues in ubiquitous computing. *Comm. ACM*, 36, **7**, 75-84,(1993).

[5] Chen E, Zhang D, Shi Y and Xu G, "Seamless Mobile Service for Pervasive Multimedia". In PCM'04, IEEE, 2004, p 194-198.

[6] Chung ES, Hong JI, Lin J, Prabaker MK, Landay, JA and Liu AL, "Development and evaluation of emerging design patterns for ubiquitous computing". In 2004 conference on Designing interactive systems, USA, p 233 – 242,2004.

[7] Davis J, Tierney A and Chang E, "A User-Adaptable User Interface Model to Support Ubiquitous User Access to EIS Style Applications". COMPSAC'05, IEEE, p 351–358,2005

[8] Duan Y and Canny J, "Protecting User Data in Ubiquitous Computing: Towards trustworthy environments". In PET 2004, Springer, p 167-185,2004.

[9] Gschwind T, Jazayeri M and Oberleitner J, "Pervasive Challenges for Software Components". In RISSE 2002, Springer, p 152-166,2002.

[10] P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), *The Adaptive Web*, Lecture Notes in Computer Science, Springer-Verlag, Berlin, Germany, 2007.

[11] L. Han, S. Jyri, J. Ma, K. Yu, "Research on Context-aware Mobile Computing", *Proceedings of Advanced Information Networking Applications Workshops*, pp 24-30, 2008.

[12] J. Pascoe, "Adding generic contextual capabilities to wearable computers", *Proceedings of 2nd International Symposium on Wearable Computers*, pp. 92-99, 1998.

[13] J. Mäntyjärvi, U. Tuomela, I. Kansala, J. Hakkila, "Context-studio–tool for personalizing context-aware applications in mobile terminals", *Proceedings of Australasian Computer Human Interaction Conference*, Addison-Wesley Longman, pp. 64-73, 2003.

[14] P. Korpipää, J. Hakkila, J. Kela, S. Ronkainen, I. Kansala, "Utilizing context ontology in mobile device application personalization", *Proceedings of Mobile and Ubiquitous Multimedia*, ACM Press, pp. 133-140, 2004.

[15] J. Coutaz, J. Crowley, S. Dobson, D. Garlan, "Context is Key", *Communications of the ACM*, 48(3), March 2005.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 5 - 9, 2014

# Clustering that depends upon Various Viewpoints Similarity Measure

V. Venkaiah,
Department of Computer Science and Engineering,
TKR College of Engineering and Technology
Hyderabad, A.P-500 097, India

A. Pramod Reddy
Department of Computer Science and Engineering,
TKR College of Engineering and Technology
Hyderabad, A.P-500 097, India

P.V.S. Srinivas
Department of Computer Science and Engineering
TKR College of Engineering and Technology
Hyderabad, A.P-500 097, India

**Abstract:** Between the data objects there must be cluster relationship which was assumed by clustering methods which they are applied on. Explicitly or implicitly will be defined based upon common things from a pair of objects. In this paper, we explored a novel based upon similarity measure which is multiviewpoint and also couple of clustering methods which are related. There is difference and also it is a major drawback in traditional dissimilarity/similarity measure to ours. The previous method only utilizes a single viewpoint, and then it is treated as origin, while our proposed system deals with a lot various viewpoints. Same cluster objects will not be assumed for measuring. Assessment of similarity will be gained by utilizing multiple viewpoints. Our system is supported with the help of conducting theoretical analysis and empirical study. For document clustering based on this new measure two criterion functions are explored. We measures the similarities and dissimilarities of them with a lot various clustering algorithms that utilizes another well known similarity measures on different document collections to check our proposal advantages.

**Keywords**: Clustering, Multiviewpoint, Similarity measure, Singleviewpoint, Two criterion functions.

## 1. INTRODUCTION

In data mining clustering takes key role when compared to other important ones. The main goal of clustering is to get intrinsic structures in information, and arrange them into perfect subgroups for future work and analysis. There are a lot clustering algorithms comes in every academic year. Those have been proved in various research areas, and created utilizing completely various techniques and approaches.

According to present researches, even after 5 decades after inventing, the simple algorithm k-means has its goal and still in top 10 algorithms list of data mining. Partitional clustering algorithm is the frequently utilized in reality. Not only that there is another one that recent survey states that k-means is the mostly utilized algorithm that users used in their practicals in the similar fields selected to utilize. There is no need to say, k-means has a little more drawbacks than the basic drawbacks, like cluster size sensitiveness initialization and its working capability is the most worst when in lot domains than rest of different state-of-the-art algorithms.

It is based on similarity measure from multi view points. Can process large amounts of data. It improves the clustering performances. Can be applied documents clustering. Can be applied on web document clustering. Can be applied on social web sites clustering. An optimization process is most commonly used approach for the clustering crisis. Optimizing invented optimal partition with the help of particular similarity function among data.

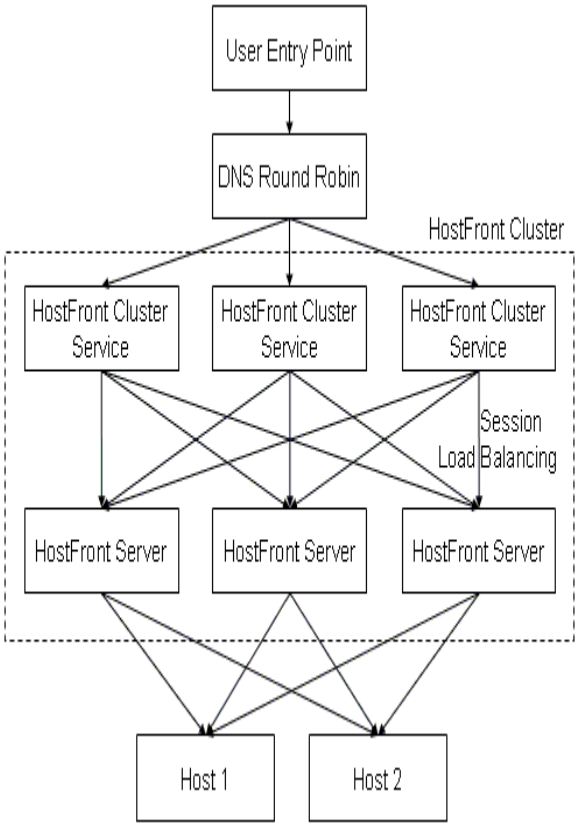



Fig: Clustering with multiple view point

Basically, there is a chance intrinsic structure of information could be accurately declared with the help of similarity formula embedded and declared in the function of clustering criterion. Based upon similarity measure on data or information the clustering algorithms effectiveness will be calculated. K-means contains objective function sum-of-squared-error that utilizes Euclidean distance. In the case of high-dimensional domains which are spherical k-means, text documents which utilizes cosine similarity (CS) as measure rather than euclidean distance cause its more efficient.

## 2. PROBLEM STATEMENT

**Existing system**:

The previous system deals with concepts like Phrase-Based clustering, Concept-Based clustering, Conceptual Tree similarity clustering. These are containing such drawbacks that make system failure.

There are some issues pertinent to most of the existing feature clustering methods.

- First, the parameter k, indicating the desired number of extracted features, has to be specified in advance. This gives a burden to the user, since trial-and-error has to be done until the appropriate number of extracted features is found.
- Second, when calculating similarities, the variance of the underlying cluster is not considered. Intuitively, the distribution of the data in a cluster is an important factor in the calculation of similarity.
- Third, all words in a cluster have the same degree of contribution to the resulting extracted feature. Sometimes, it may be better if more similar words are allowed to have bigger degrees of contribution. Our feature clustering algorithm is proposed to deal with these issues.

**Proposed System:**

Because of the nature of similarity plays an important role in clustering, there is a need of a method for calculating common data between objects. Proposed a novel clustering similarity measure based on multi view point.

## 3. SYSTEM DEVELOPMENT

**Multiviewpoint-based similarity**

The cosine similarity will be explored in the following form without changing its meaning.

$$Sim(d_i, d_j) = \cos(d_i - 0, d_j - 0) = (d_i - 0)^t (d_j - 0),$$

Where vector 0 indicates the point of origin. Based upon this formula, the calculation picks 0 as one and also point of reference. The common thing of couple of documents di and dj is defined with respect to the angle between the couple of points when the view is from origin.

**Two Clustering Criterion Functions IR and IV**

We now created our clustering criterion functions. The initial function, known as IR, is the cluster size-weighted those are sum of average of normal pair wise common things of documents which are in the one

cluster only. Initially, we explored the sum by function F in a normal form.

$$F = \sum_{r=1}^{k} n_r \left[ \frac{1}{n_r^2} \sum_{d_i, d_j \in S_r} Sim(d_i, d_j) \right].$$

**Optimization Algorithm and Complexity**

We have given clustering framework of ours by Multiviewpoint-based Similarity Clustering. At the same time, we are having MVSC-IV and MVSC-IR, those are with MVSC IR and IV criterion function, respectively. The key aim is to do clustering by optimizing on document IV and IR. Cause of this sake, the algorithm incremental k-way a sequential flow of k-means is used. Taking that the IV expression depends upon only on nr and Dr, r = 1; . . . ; k, IV will be in a general form.

$$I_V = \sum_{r=1}^{k} I_r(n_r, D_r),$$

Where Ir(nr,Dr) is the objective value of r cluster. The similar is used on IR. The incremental optimization algorithm in general form, which contains couple of important steps those are refinement and initialization, are explored.

1: **procedure** BUILDMVSMATRIX($A$)
2: **for** $r \leftarrow 1 : c$ **do**
3: $D_{S \setminus S_r} \leftarrow \sum_{d_i \notin S_r} d_i$
4: $n_{S \setminus S_r} \leftarrow |S \setminus S_r|$
5: **end for**
6: **for** $i \leftarrow 1 : n$ **do**
7: $r \leftarrow$ class of $d_i$
8: **for** $j \leftarrow 1 : n$ **do**
9: **if** $d_j \in S_r$ **then**
10: $a_{ij} \leftarrow d_i^t d_j - d_i^t \frac{D_{S \setminus S_r}}{n_{S \setminus S_r}} - d_j^t \frac{D_{S \setminus S_r}}{n_{S \setminus S_r}} + 1$
11: **else**
12: $a_{ij} \leftarrow d_i^t d_j - d_i^t \frac{D_{S \setminus S_r} - d_j}{n_{S \setminus S_r} - 1} - d_j^t \frac{D_{S \setminus S_r} - d_j}{n_{S \setminus S_r} - 1} + 1$
13: **end if**
14: **end for**
15: **end for**
16: **return** $A = \{a_{ij}\}_{n \times n}$
17: **end procedure**

Algorithm: MVS Similarity matrix construction

## 4. RELATED WORK

Hard clustering: Each document belongs to exactly one cluster. More common and easier to do Soft clustering: A document can belong to more than one cluster. Makes more sense for applications like creating browsable hierarchies.

**Cosine similarity measurement**

Assign Boolean values to a vector describing the attributes of a database element, then measure vector similarities with the Cosine Similarity Metric. Cosine similarity is a measure of similarity between two vectors by measuring the cosine of the angle between them. The result of the Cosine function is equal to 1 when the angle is 0, and it is less than 1 when the angle is of any other value. As the angle between the vectors shortens the cosine angle approaches 1, meaning that the two vectors are getting closer, meaning that the similarity of whatever is represented by the vectors increases. Assign Numeric values to non-numerical items, and then use one of the standard clustering algorithms. Then use one of the standard clustering algorithms like, hierarchical clustering, agglomerative ("bottom-up") or, divisive ("top-down"), Partitional clustering, Exclusive Clustering, Overlapping Clustering, Probabilistic Clustering , Text Clustering, Text clustering is one of the fundamental functions in text mining.

Text clustering is to divide a collection of text documents into different category groups so that documents in the same category group describe the same topic, such as classic music or history or romantic story. Efficiently and automatically grouping documents with similar content into the same cluster.

Vector space model is an algebraic model for representing text documents (and any objects, in general) as vectors of identifiers.

A text document is represented as a vector of terms <t1, t2, …, ti, …, tm>.Each term ti represents a word. A set of documents are represented as a set of vectors, that can be written as a matrix.

**Collection Reader**

Transform raw document collection into a common format, e.g., XML Use tags to mark off sections of each document, such as, <TOPIC>, <TITLE>, <ABSTRACT>,<BODY> Extract useful sections easily Example:

"Instead of direct prediction of a continuous output variable, the method discretizes the variable by kMeans clustering and solves the resultant classification problem."

Detagger

Find the special tags in document ",","." . Filter away tags. "Instead of direct prediction of a continuous output variable the method discretizes the variable by kMeans clustering and solves the resultant classification problem".

Removing Stopwords

Stopwords

Function words and connectives. Appear in a large number of documents and have little use in describing the characteristics of documents.

Example

**Removing Stopwords**

Stopwords: "of", "a", "by", "and" , "the", "instead"

Example: "direct prediction continuous output variable method discretizes variable kMeans clustering solves resultant classification problem"

- **Stemming**

Remove inflections that convey parts of speech, tense.

Techniques: Morphological analysis (e.g., Porter's algorithm), Dictionary lookup (e.g., WordNet), Stems: "prediction --->predict", "discretizes --->discretize", "kMeans ---> kMean", "clustering --> cluster", "solves ---> solve", "classification ---> classify".

Example sentence: "direct predict continuous output variable method discretize variable kMean cluster solve resultant classify problem"

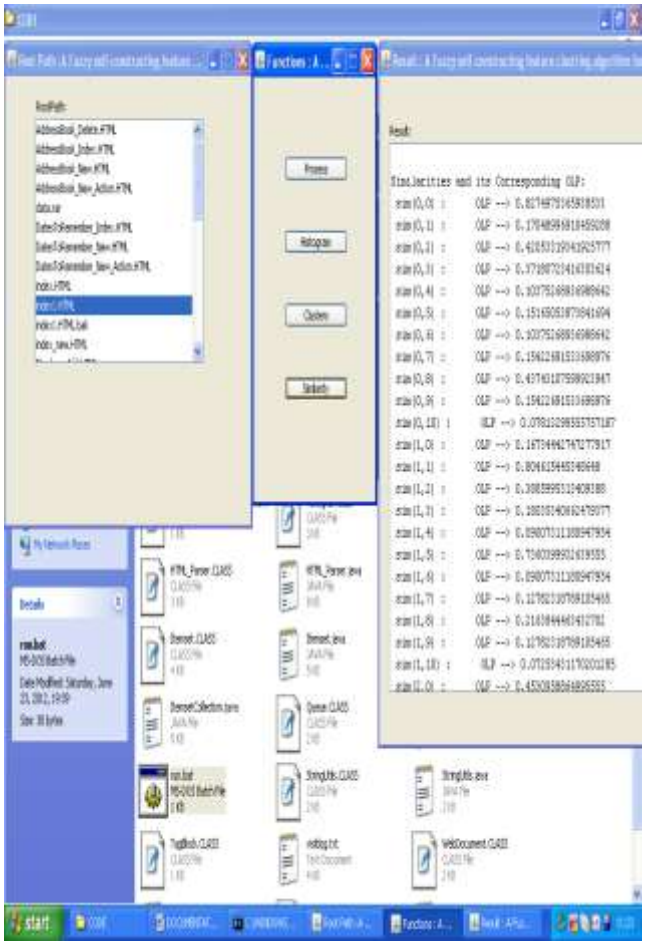
K-Means

In some case documents are taken as real-valued vectors. Select K random docs {s1, s2,… sK} as seeds. Until clustering converges or other stopping criterion: For each doc di: Assign di to the cluster cj such that dist (xi, sj) is minimal. (Update the seeds to the centroid of each cluster) For each cluster cj sj = μ(cj). A fixed number of iterations. Doc partition unchanged. Centroid positions don't change. A state in which clusters don't change. K-means is a special case of a general procedure known as the Expectation Maximization (EM) algorithm. EM is known to converge. Number of iterations could be large.

Result: Whole corpus analysis/navigation. Better user interface: search without typing. For improving recall in search applications. Better search results (like pseudo RF). For better navigation of search results. Effective "user recall" will be higher. For speeding up vector space retrieval. Cluster-based retrieval gives faster search. Cluster hypothesis - Documents in the same cluster behave similarly with respect to relevance to information needs. Therefore, to improve search recall: Cluster docs in corpus a priori. When a query matches a doc D, also return other docs in the cluster containing D. Hope if we do this: The query "car" will also return docs containing automobile. Because clustering grouped together docs containing car with those containing automobile.

## 5. CONCLUSION

In this paper, based upon similarity measuring we explored a Multiviewpoint- method, known as MVS. MVS is potentially perfect based upon theoretical analysis and empirical examples. MVS is perfect than the cosine similarity for text documents. There are couple of clustering algorithms that impacts some criterion functions, those are IR and IV.



MVSC-IR and MVSC-IV, are explored newely. State-of-the-art clustering methods utilizes various ones of similarity measure, on a huge count sets of document data and under various evaluation metrics, the explored new algorithms proven that they are capable of providing better clustering operating capability significantly. The main thing of this is the basic concept of common measure from many viewpoints. Upcoming methods capable of utilizing of the one principle, but for the relative similarity mentioned alternative forms, or wont utilize average but contains different methods to club the relative common things based upon various viewpoints.

Apart from this paper keeping interest on documents partition clustering. In upcoming days, for hierarchical clustering algorithms it may be capable to use the proposed criterion functions. At last, we have explored the MVS application and its text data clustering algorithms. It is an interesting part to show the working on various types of sparse and also on dimensional data which was high.

## 6. REFERENCES

[1]http://people.csail.mit.edu/jrennie/20Newsgroups/, 2010.

[2]Http://kdd.ics.uci.edu/databases/reuters21578/reuters21578.html. 2010.

[3] H. Kim, P. Howland, and H. Park, “Dimension Reduction in Text Classification with Support Vector Machines,” J. Machine Learning Research, vol. 6, pp. 37-53, 2005.

[4] F. Sebastiani, “Machine Learning in Automated Text Categorization,” ACM Computing Surveys, vol. 34, no. 1, pp. 1-47, 2002.

[5] B.Y. Ricardo and R.N. Berthier, Modern Information Retrieval. Addison Wesley Longman, 1999.

[6] A.L. Blum and P. Langley, “Selection of Relevant Features and Examples in Machine Learning,” Artificial Intelligence, vol. 97, nos. 1/2, pp. 245-271, 1997.

[7] E.F. Combarro, E. Montan˜ e´s, I. Dı´az, J. Ranilla, and R. Mones, “Introducing a Family of Linear Measures for Feature Selection in Text Categorization,” IEEE Trans. Knowledge and Data Eng., vol. 17, no. 9, pp. 1223-1232, Sept. 2005.

[8] K. Daphne and M. Sahami, “Toward Optimal Feature Selection,” Proc. 13th Int’l Conf. Machine Learning, pp. 284-292, 1996.

[9] R.Kohavi and G.John, “Wrappers for Feature Subset Selection,” Artificial Intelligence, vol. 97, no. 1-2, pp. 273-324, 1997.

[10] Y. Yang and J.O. Pedersen, “A Comparative Study on Feature Selection in Text Categorization,” Proc. 14th Int’l Conf. Machine Learning, pp. 412-420, 1997.

[11] D.D. Lewis, “Feature Selection and Feature Extraction for Text Categorization,” Proc. Workshop Speech and Natural Language, pp. 212-217, 1992.

[12] H. Li, T. Jiang, and K. Zang, “Efficient and Robust Feature Extraction by Maximum Margin Criterion,” T. Sebastian, S. Lawrence, and S. Bernhard eds. Advances in Neural Information Processing System, pp. 97-104, Springer, 2004.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 10 - 19, 2014

# Natural Hand Gestures Recognition System for Intelligent HCI: A Survey

Vishal Nayakwadi
Department of Computer Science
TSSM's BSCOER College of Engineering
University of Pune, India

N. B. Pokale
Department of Computer Science
TSSM's BSCOER College of Engineering
University of Pune, India

**Abstract:** Gesture recognition is to recognizing meaningful expressions of motion by a human, involving the hands, arms, face, head, and/or body. Hand Gestures have greater importance in designing an intelligent and efficient human–computer interface. The applications of gesture recognition are manifold, ranging from sign language through medical rehabilitation to virtual reality. In this paper a survey on various recent gesture recognition approaches is provided with particular emphasis on hand gestures. A review of static hand posture methods are explained with different tools and algorithms applied on gesture recognition system, including connectionist models, hidden Markov model, and fuzzy clustering. Challenges and future research directions are also highlighted.

**Keywords -** Hand gesture interface, HCI, Computer vision, Fuzzy Clustering, ANN, HMM, Orientation Histogram.

## 1. INTRODUCTION

Gestures and face expressions easily used for daily humans interactions [1] while human computer interactions still require understanding and analyzing signals to interpret the desired command that made the interaction sophisticated and unnatural [1]. Recently the designing of special input devices witnessed great attention in this field to facilitate the interaction between humans and computers [2], and to accomplish more sophisticated interaction through the computer [2]. It is worth to mention that the window manager is the earlier user interface to communicate with computers [3]. The combining of traditional devices mouse and keyboard with the new designed interaction devices such as gesture and face recognition, haptic sensors, and tracking devices provides flexibility in Tele-operating [2] [24], text editing [4], robot control [2] [50] [42], cars system control [2], gesture recognition [4], Virtual Reality (VR) [5], and multi-media interfaces [4], video games [4] [44]. Gesture considered as a natural way of communication among people especially hear-impaired [6]. Gestures can be defined as a physical movement [6] of hands, arms, or body that delivers an expressive message [6], and gesture recognition system used to interpret and explain this movement as meaningful command [6][7].

Gesture recognition has been applied in a large range of application areas such as recognizing sign language [6] [8], human computer interaction (HCI) [6] [9], robot control [6], smart surveillance [6], lie detection [6], visual environments manipulating [6], etc. Different techniques and tools have been applied for handling gesture recognition, vary between mathematical models like Hidden Markov Model (HMM) [6][10] [51] [52] and Finite State Machine (FSM) [6][11] to approaches based on software computing methods such as fuzzy clustering [12], Genetic Algorithms (GAs) [13] and Artificial Neural Network (ANN) [14]. Hand posture recognition sill an open research area [15], since the human hand is a complex articulated object with many connected joints and links, which forms the 27 degrees of freedom [16] for the hand. Typically the implementation of gesture recognition system required different kinds of devices for capturing and tracking image/ video image [6] such as camera(s), instrumented (data) gloves, and colored marker [6]. Those devices are used for modeling the communication between human and environments rather than traditional interface device such as keyboards, and mice which are inconvenient and unnatural for HCI system. Vision based technique also different according to some system environments such as number of cameras used [6], speed [6], and illumination conditions [6]. The major difficulty in gesture recognition system is how to identify a specific gesture meaning by the machines (computers/ robot) [17].

The purpose of this paper is to present a review of vision based hand gesture recognition techniques for human computer interaction, and to explain various approaches with its advantages and disadvantages. Although recent reviews [1] [6], [7] [18] [17] [19] [20] in computer vision based have explained the importance of gesture recognition system for human computer interaction (HCI), this work concentrates on vision based techniques method and it's up-to-date. With intending to point out various research developments as well as it represent good starting for interested persons in hand gesture recognition area.

## 2. HAND GESTURE TECHNOLOGY

For any system the first step is to collect the data necessary to accomplish a specific task. For hand posture and gesture recognition system different technologies are used for acquiring input data. Present technologies for recognizing gestures can be divided into vision based, instrumented (data) glove, and colored marker approaches. Figure 1 shows an example of these technologies.

### 2.1 Vision Based approaches:

In vision based methods the system requires only camera(s) to capture the image required for the natural interaction between human and these approaches are simple but a lot of gesture challenges are raised such as the complex background, lighting variation, and other skin color objects with the hand object,

besides system requirements such as velocity, recognition time, robustness, and computational efficiency [7] [17].

Fig. 1 Vision based

## 2.2 Instrumented Glove approaches:

Instrumented data glove approaches use sensor devices for capturing hand position, and motion. These approaches can easily provide exact coordinates of palm and finger's location and orientation, and hand configurations [17] [21] [22], however these approaches require the user to be connected with the computer physically [22] which obstacle the ease of interaction between users and computers, besides the price of these devices are quite expensive [22], it is inefficient for working in virtual reality [22].

Fig. 2 Data Glove

## 2.3 Colored Markers approaches:

Marked gloves or colored markers are gloves that worn by the human hand [5] with some colors to direct the process of tracking the hand and locating the palm and fingers [5], which provide the ability to extract geometric features necessary to form hand shape [5]. The color glove shape might consist of small regions with different colors or as applied in [23] where three different colors are used to represent the fingers and palms, where a wool glove was used. The amenity of this technology is its simplicity in use, and cost low price comparing with instrumented data glove [23]. However this technology still limits the naturalness level for human computer interaction to interact with the computer [5].

Fig. 3 Color Makers

# 3. GESTURE RECOGNITION TECHNIQUES

The recognition of gesture involves several concepts such as pattern recognition [19], motion detection and analysis [19], and machine learning [19]. Different tools and techniques are utilized in gesture recognition systems, such as computer vision [38] [55], image processing [6], pattern recognition [6], statistical modeling [6].

## 3.1 Artificial Neural Networks

The use of neural networks for gesture recognition has been examined by many researchers. Most of the researches use ANN as a classifier in gesture recognition process, while some others use it to extract the shape of the hand, as in [25]. Tin H. [26] presents a system for hand tracking and gesture recognition using NNs to recognize Myanmar Alphabet Language (MAL). Adobe Photoshop filter is applied to find the edges of the input image and histogram of local orientation employed to extract image feature vector which would be the input to the supervised neural networks system. Manar M. [27] used two recurrent neural network architectures to recognize Arabic Sign Language (ArSL). Elman (partially) recurrent neural networks and fully recurrent neural networks have been used separately. A colored glove used for input image data, and for segmentation process, HSI color model is applied.

The segmentation divides the image into six color layers, one for the wrist and five for fingertips. 30 features are extracted and grouped to represent a single image, fifteen elements used to represent the angles between the fingertips and between them and the wrist [27], and fifteen elements to represent distances between fingertips; and between fingertips and the wrist [27]. This input feature vector is the input to both neural networks systems. 900 colored images were used as training set, and 300 colored images for system testing. Results had shown that fully recurrent neural network system (with recognition rate 95.11%) better than the Elman neural network (with 89.67% recognition rate). Kouichi M. in [28] presented Japanese sign language recognition using two different neural network systems. Firstly, back propagation algorithm was used for learning postures of Japanese alphabet. For input postures, data glove is used, and normalization operation was applied as a preprocessing step on the input image. The feature extracted from input images was 13 data items, ten for bending, and three for angles in the coordinates.

The output of the network was 42 characters. The network consists of three layers, the input layer with 13 nodes, the hidden layer with 100 nodes, and the output layer with 42 nodes which corresponds 42 recognized characters. The recognition rate for

learning 42 taught patterns was 71.4%, and for unregistered people 47.8%, while the rate improved when additional patterns added to the system, it became 98.0% for registered, and 77.0% for unregistered people. Elman Recurrent Neural Network was the second system applied for recognition gestures. The system could recognize 10 words. The data item have been taken from data glove and the same preprocessing applied for input image. Features extracted are 16 data items, 10 for bending, 3 for angles in the coordinates, and 3 for angles in the coordinates. The network consists of three layers, the input layer with 16 nodes, the hidden layer with 150 nodes, and the output layer with 10 nodes which corresponds 10 recognized words.

Some improvement in the positional data and filtering data space are added to the system [28]. Integration of these two neural networks, in a way, that after receiving data from data glove, a determination of the start sampling time and if the data item considered a gesture or a posture is sent to the next network, for checking the sampling data and the system hold a history, which decide the end of sign language. The final recognition rate with the encoding methods was 96%. Stergiopoulou E. [25] recognized static hand gestures using Self-Growing and Self-Organized Neural Gas (SGONG) network. A camera used for acquiring the input image, and YCbCr color space is applied to detect hand region, some thresholding technique used to detect skin color. SGONG network use competitive Hebbian learning algorithm for learning process, the learning start with only two neurons and continuous growing till a grid of neurons are constructed and cover the hand object which will capture the shape of the hand. From the resultant hand shape three geometric features was extracted, two angles based on hand slope and the distance from the palm center was determined, where these features used to determine the number of the raised fingers.

For recognizing fingertip, Gaussian distribution model used by classifying the fingers into five classes and compute the features for each class. The system recognized 31 predefined gestures with recognition rate 90.45%, in processing time 1.5 second. Shweta K. in [29] introduced gesture recognition system using Neural Networks. Web-cam used for capturing input image at slow rate samples between 15-25 frames per second. Some preprocessing have been made on the input image which convert the input image into sequence of (x, y) coordinates using MATLAB, then passed into neural classifier, in which it will classify the gesture into one of several classed predefined classes which can be identified by the system. Sidney and Geoffiey [30],[31] used neural networks to map hand gestures to speech synthesizer using Glove-Talk system that translated gestures to speech through adaptive interface which is an important class of neural networks applications [31].

## 3.2 Histogram Based Feature

Many researchers have been applied based the histogram, where the orientation histogram is used as a feature vector [32]. The first implementation of the orientation histogram in gesture recognition system and real time was done by William F. and Michal R. [32]; they presented a method for recognizing gestures based on pattern recognition using orientation histogram. For digitized input image, black and white input video was used, some transformations were made on the image to compute the histogram of local orientation of each image, then a filter applied to blur the histogram, and plot it in polar coordinates. The system consists of two phases; training phase, and running phase. In the training phase, for different input gestures the training set is stored with their histograms. In running phase an input image is presented to the computer and the feature vector for the new image is formed, Then comparison performed between the feature vector of the input image with the feature vector (oriented histogram) of all images of the training phase, using Euclidean distance metric and the less error between the two compared histograms will be selected. The total process time was 100 msec per frame.

Hanning Z., et al. [33] presented hand gesture recognition system based on local orientation histogram feature distribution model. Skin color based segmentation algorithm were used to find a mask for the hand region, where the input RGB image converted into HSI color space, and then map the HSI image H to a likelihood ratio image L the hand region is segmented by thresholding value, 128 elements in the local orientation histogram feature were used. The augmented of the local orientation histogram feature vector implemented by adding the image coordinates of the sub-window. To compact features representation, k-means clustering has been applied the augmented local orientation histogram vectors. In Recognition stage, Euclidean distance used to compute the exact matching score between the input image and stored posture. Then Locality Sensitive Hashing (LSH) used to find the approximate nearest neighbors, and reduce computational cost for image retrieval. Wysoski et al. [34] presented a rotation invariant static-gesture recognition approach using boundary histograms. Skin color detection filter was used, followed by performing erosion, dilation as preprocessing operation, and clustering process to find the groups in the image. For each group the boundary was extracted using an ordinary contour-tracking algorithm.

The image Divided into grids, and normalized the boundary in size, which give the system invariance distance between the camera and hand. Homogeneous background was applied, and the boundary is represented as chord's size chain. The image was divided into a number of regions N. And the regions were divided in a radial form [34], according to a specific angle as shown in the Figure. The histogram of boundary chord's size was calculated. So the whole feature vector consists of a sequential chain of histograms. Multilayer perceptron (MLP) Neural Networks and Dynamic Programming (DP) matching were used as classifiers. 26 static postures from American Sign Language, for every posture, 40 pictures were taken, 20 pictures for training and 20 for test. Different number of histograms were used varies from 8 to 36 increasing by two, with different histogram resolutions.

## 3.3 Fuzzy Clustering Algorithm

Clustering algorithms is a general term comprises all methods that partitioning the given set of sample data into subsets or clusters [35] based on some measures between grouped elements [12]. According to this measure the pattern that share the same characteristics are grouped together to form a cluster [12]. Clustering Algorithms have been widely spread because of their ability of grouping complicated data collections into regularly clusters [35]. In fuzzy clustering, the partitioning of sample data into groups in a fuzzy way are the main difference between fuzzy clustering and other clustering algorithm [12], where the single data pattern might belong to different data groups [12].

Xingyan L. In [12] presented fuzzy c-means clustering algorithm to recognize hand gestures in a mobile remote.

A camera was used for acquire input raw images, the input RGB images are converted into HSV color model, and the hand extracted after some preprocessing operations to remove noise and unwanted objects, and thresholding used to segment the hand shape. 13 elements were used as feature vector, first one for aspect ratio of the hand's bounding box, and the rest 12 parameters represent grid cell of the image, and each cell represents the mean gray level in the 3 by 4 blocks partition of the image, where the mean value of each cell represents the average brightness of those pixels in the image, Then FCM algorithm used for classification gestures. Various environments are used in the system such as complex background and invariant lighting conditions. 6 hand gestures used with 20 samples for each gesture in the vocabulary to create the training set, with recognition accuracy 85.83%.

## 3.4 Hidden Markov Model (HMM)

Many researches were applied in the field of gesture recognition using HMM. HMM is a stochastic process [6] [52], with a finite number of states of Markov chain, and a number of random functions so that each state has a random function [6]. HMM system topology is represented by one state for the initial state, a set of output symbols [6] [22], and a set of transitions state [22] [8]. HMM contained a lot of mathematical structures and has proved its efficiency for modeling spatio–temporal information data [6]. Sign language recognition, are one of the most applications of HMM [8], and speech recognition [10]. In [9] Keskiin C., et. al. presented HCI interface based on real time hand tracking and 3D gesture recognition using hidden Markov models (HMM) [54]. Two colored cameras for 3D construction are used. To overcome the problem of using skin color for hand detection because of hand overlapping with other body parts, markers are used to reduce the complexity in hand detection process [9] [52]. Markers used to segment the hand from complex backgrounds under invariant lighting conditions.

The markers are distinguished using marker detection utility, and connected components algorithm was applied to find marker regions using double thresholding. For fingertip detection, simple descriptors were used, where the bounding box and four outmost points of the hand that defining the box is determined [9]. The bounding box in some cases needs to be elongate to determine the mode of the hand, and the points used to predict the fingertip location in different modes of the hand. Kalman filter was used for filtering trajectory of the hand motion. For 3D reconstruction of finger coordinates, calibration utility was implemented for specific calibration object [9]. Least square approach used to generate fingertip coordinates, and kalman filter applied for smoothing the trajectory of 3D reconstructed coordinates. To eliminate coordinate system dependency, the 3D coordinates are converted into sequences of quantized velocity vectors. HMM interprets these sequences [9], which are directional characterizing of the motion [9]. The system designed for game and painting programs application. Hand tracking is utilized to imitate the movements of the mouse for drawing, and the gesture recognition system used for selecting commands. Eight gestures have been used for system training, and 160 for testing, with 98.75% recognition performance.

**Table 1. Comparisons between various gesture recognition systems**

| **Method** | **Type Of Input Device** | **Segmentation Type** | **Features (Geometric Or Non)** | **Feature Vector Representation** | **Classification Algorithm** | | **Recognition Rate** |
|---|---|---|---|---|---|---|---|
| Tin H. [26] | Digital camera | threshold | Non geometric | Orientation histogram | supervised neural network | | 90% |
| Manar M. [27] | Colored glove, and Digital camera | HSI color model | N/A | Available Features from resource | Two neural network system | Elman recurrent network | 89.66% |
| | | | | | | Fully recurrent network | 95.11% |
| Kouichi M. [28] | Data glove | threshold | Non geometric | 13 data item (10 for bending, 3 for coordinate angles) | Two neural network system | back propagation network | 71.4% |
| | | | | 16 data item (10 for bending, 3 for coordinate angles, 3 for positional data) | | Elman recurrent network | 96% |

| | | | | | | |
|---|---|---|---|---|---|---|
| Stergiopoulou E. [25] | Digital camera | YCbCr color space | geometric | Two angles of the hand shape, compute palm distance | Gaussian distribution | 90.45% |
| Shweta in [29] | Web-cam | | Non geometric | | supervised neural network | N/A |
| William F. and Michal R. [32] | Digital camera | | Non geometric | Orientation histogram | Euclidean distance metric | N/A |
| Hanning Z., et al. [33] | Digital camera | threshold | Non geometric | augmented of the local orientation histogram | Euclidean distance metric | 92.3% |
| Wysoski, et al. [34] | Digital camera | skin color detection filter | Geometric | the histogram of radial boundary | MLP+ DP matching | 98.7% |
| Xingyan L. [12] | Digital camera | threshold | Non geometric | One dimensional array of 13 element | Fuzzy C-Means algorithm | 85.83% |
| Keskiin C., et al. [39] | Two colored cameras and marker | connected components algorithm with double thresholding | Geometric | sequences of quantized velocity vectors | HMM | 98.75% |

## 4. RECOGNITION SYSTEM METHODOLOGY

Many researchers have been suggested on gesture recognition system for different applications, with different recognition phases but they all agree with the main structure of the gesture recognition system. These phases are segmentation, features detection and extraction, and finally the classification or recognition phase. One of these structures illustrated in Figure 4.

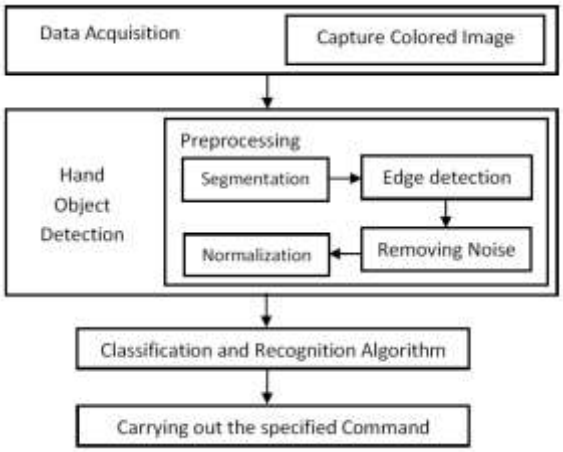



Fig. 4 The flow of gesture recognition system [73]

### 4.1 Segmentation

Segmentation phase plays an important role in the system recognition process. Perfect segmentation effects on the accuracy of the recognition system [74]. For any segmentation process, some image processing operations are required for hand shape detection [7] [74].

Segmentation image algorithms can be classified into two types according to image gray level properties as explained in [78]:

#### *4.1.1 Discontinuity:*

Which tries to find a mass change in the contrast?

#### *4.1.2 Similarity:*

Which computes the similarity between neighbor pixels?

When the input gesture acquired form colored camera, instrumented glove device or colored glove as shown in Figure 1. The first step is segmentation, to extract the hand region from the input image and isolate it from the background [74]. There are two main methods for object segmentation, first method depends on the color model that can be extracted from the existence RGB color model which could be HSV color model [74] [75] [77] or YCbCr color space [25] which deals with the pigment of the skin of the human hand [74], the significant property of this color space is that the human different ethics group can be recognized according to their pigment concentration which can be distinguished according to some skin color saturation [74]. Then, the hand area is isolated from the input gesture with some threshold value. Some normalization for the segmented image might require for obtaining the gestures database which should be invariant against different perturbations like translation, scaling and rotation [74]. The database created with many samples per single gesture, the relation between the number of samples and the accuracy is directly proportional, and between number of samples and the speed is inversely proportional [74].

Hasan [74] used HSV color model to extract the skin-like hand region by estimating the parameter values for skin pigment, and used Laplacian filter for detection the edges. Stergiopoulou [25] used YCbCr color model to segment the hand. Maraqa (2008) used color glove for input gestures and HSI color space for the segmentation process. Ghobadi (2008) treated the segmentation process as clustering method by grouping the image pixels among image objects. Lamberti [23] used HSI color model to segment the hand object. Table 3 shows some applications of the segmentation methods used in the discussed method.

## 4.2 Features Detection and Extraction

The features are the useful information that can be extracted from the segmented hand object by which the machine can understand the meaning of that posture. The numerical representation of these features can be obtained from the vision perspective of the segmented hand object which forms the feature extraction phase [76]. Many researchers have been applied to form this feature vector which takes different sizes as well as meanings. Hasan [74] extracted the features vector by dividing the segmented hand object into fixed block size 5x5 brightness value moments; this produce 625 features vector size and only 98 are stored as actual features vector. Stergiopoulou [25] applied Self-Growing and Self-Organized Neural Gas (SGONG) network to extract the exact shape of the hand region and determine three characteristics as the features; Palm region, Palm center, and Hand slope. Compute angle between the finger root and the hand center named RC Angle, and the joints the fingertip and the hand center named TC, and angle and distance from the palm center. Li [12] defined a grid of fixed size with 12 blocks gray scale features vector, and each grid cell represents the mean value of the average brightness of the pixels in the block. Lamberti [23] defined the distance $d$ from the palm to the fingers $di$(i = 1, ..., 5), and computed the angle $\beta$ between the line connecting the centroids of the palm and The fingers, which produces four angles $\beta i$(i = 1, ..., 4), so the hand represented by nine numerical features vector [23]. Table 4 demonstrates features vector representation of these methods.

## 4.3 Recognition

Recognition or classification of hand gestures is the last phase of the recognition system. Hand gestures can be classified using two approaches as mentioned in [7]

### *4.3.1 Rule based Approaches:*

Which represents the input features as manually encoded rule, and the winner gesture is the one that matched with the encoded rules after his features has been extracted. The main problem of this technique is that the human ability to encode the rules limits the successfulness of the recognition process [7]

### *4.3.2 Machine Learning based Approaches:*

The most common approaches that considered the gesture as result of some stochastic processes [7]. Most of the problems that based on machine learning have been addressed based on the statistical modeling [16], such as PCA [79], FSM [80]. Hidden Markov Models (HMMs) [9] [54] have been paid attention by many researchers [7], Kalman filtering [77], Artificial Neural networks (ANNs) [27] [28] [30] [31] which have been utilized in gesture recognition as well. Some researchers used Gaussian distribution for gestures classification [25] and Euclidian distance metric [74].

# 5. APPLICATIONS

Lately there has been a great emphasis on Human-Computer Interaction (HCI) research to create easy-to-use interfaces by facilitating natural communication and manipulation skills of humans. Among different human body parts, the hand is the most effective interaction tool because of its dexterity. Adopting hand gesture as an interface in HCI will not only allow the deployment of a wide range of applications in sophisticated computing environments such as virtual reality systems and interactive gaming platforms, but also benefit our daily life such as providing aids for the hearing impaired, and maintaining absolute sterility in health care environments using touch less interfaces via gestures [38].Gesture recognition has wide-ranging applications [49] such as the following:

## 5.1 Virtual Reality:

Gestures for virtual and augmented reality applications have experienced one of the greatest levels of uptake in computing. Virtual reality interactions use gestures to enable realistic manipulations of virtual objects using ones hands, for 3D display interactions [56] or 2D displays that simulate 3D interactions [57].

## 5.2 Robotics and Telepresence:

Telepresence and telerobotic applications are typically situated within the domain of space exploration and military-based research projects [47] [24]. The gestures used to interact with and control robots are similar to fully-immersed virtual reality interactions, however the worlds are often real, presenting the operator with video feed from cameras located on the robot [58]. Here, gestures can control a robots hand and arm movements to reach for and manipulate actual objects, as well its movement through the world.

## 5.3 Desktop and Tablet PC Applications:

In desktop computing applications, gestures can provide an alternative interaction to the mouse and keyboard [59] [43]. Many gestures for desktop computing tasks involve manipulating graphics, or annotating and editing documents using pen-based gestures [60].

## 5.4 Games:

When, we look at gestures for computer games. Freeman et al. [61] tracked a player's hand or body position to control movement and orientation of interactive game objects such as cars [44]. Konrad et al. [62] used gestures to control the movement of avatars in a virtual world, and Play Station 2 has introduced the Eye Toy, a camera that tracks hand movements for interactive games [63].

## 5.5 Sign Language:

Sign language is an important case of communicative gestures [15]. Since sign languages are highly structural, they are very suitable as test beds for vision algorithms [64]. At the same time, they can also be a good way to help the disabled to interact with computers. Sign language for the deaf (e.g. American Sign Language) is an example that has received significant attention in the gesture literature [65] [66] [67] [68].

## 5.6 Vehicle Monitoring:

Another important application area is that of vehicle interfaces. A number of hand gesture recognition techniques for human

vehicle interface have been proposed time to time [69] [70]. The primary motivation of research into the use of hand gestures for in-vehicle secondary controls is broadly based on the premise that taking the eyes off the road to operate conventional secondary controls can be reduced by using hand gestures.

### 5.7 Healthcare & Medical Assistance:

The healthcare area has also not been left untouched by this technological wave. Wachs et al. [71] developed a gesture based tool for sterile browsing of radiology images. Jinhua Zeng, Yaoru Sun, Fang Wang developed a wheelchair with intelligent HCI [40] [45].

### 5.8 Daily Information Retrieval:

Sheng-Yu Peng implemented an approach that provides daily information retrieved from Internet, where users can operate this system with his hands' movements [41] [42].

### 5.9 Education:

Zeng, Bobo, Wang, Guijin presented a system using hand gestures to control poerpoint presentations [48].

### 5.10 Television Control:

Last application for hand postures and gestures is controlling Television devices [22]. Freeman [72] developed a system to control a television set by hand gestures. Using an open hand and the user can change the channel, turn the television on and off, increase and decrease the volume, and mute the sound.

## 6. IMPLEMENTATION TOOLS

Many implementation hardware and software tools have been utilized for recognizing gestures depending on the application fields used.

### 6.1 Hardware Implementation Tools

Input devices used for gesture recognition systems are various and different according to the system and application used for recognition process. Single camera can be used for postures recognition since this environment might be inconvenient for other types of image-based recognition [24]. Stereo camera which consists of two lenses with an isolated sensor for each lens [24], which imitates human visual system therefore, the 3D effect of views is created [24]. Stereo camera can be used to make 3D pictures for movies [24], or for range imaging [24]. Tracking device such as instrumented data gloves measure finger movements through various types of sensors technology [21], [22]. It can provide accurate information about the position and orientation of the hands using magnetic or inertial tracking devices [24]. For more details about various types of glove-based input device refer to [21], [22].

Controller-based gestures, in this type of input gestures, controllers represent a complement of the human, so that when the body moves to perform some gestures [24], these motions are captured using some software [24]. Mouse gestures are an example of such controllers [24]. Other systems based on accelerometers to measure hand movements [36] [37].

### 6.2 Software Implementation Tools

Software tools are the programming language and windows system used for implementing the gesture recognition system. Some researches applied programming languages like C, C++, and Java language. To simplify the work especially when image processing operations are needed, MATLAB® with image processing toolbox is used. Tin H. [26] used MATLAB® for hand tracking and gesture recognition. Manar M. [27] use MATLAB6 and C language, MATLAB6 used for image segmentation while C language for Hand Gesture Recognition system. Kouichi [28] use SUN/4 workstation for Japanese Character and word recognition. Also Stergiopoulou [25] used Delphi language with 3 GHs CPU to implement hand gesture recognition system using SGONG network. Shweta [29] used MATLAB® for hand recognition. Freeman and Michal Roth [32] used HP 735 workstation was used for implementing the system. Hanning Zhou et. al. [33] used C++ implementation costs only 1/5 second for the whole preprocessing, feature extraction and recognition, when running on a 1.3G Intel Pentium laptop processor with 512MB memory.

## 7. CONCLUSION

Building an efficient human-machine interaction is an important goal of gesture recognition system. Many applications of gesture recognition system ranging from virtual reality to sign language recognition and robot control. In this paper a survey on tools and techniques of gesture recognition system have been provided with emphasis on hand gesture expressions. The major tools surveyed include HMMs, ANN, and fuzzy clustering have been reviewed and analyzed. Most researchers are using colored images for achieving better results. Comparison between various gesture recognition systems have been presented with explaining the important parameters needed for any recognition system which include: the segmentation process, features extraction, and the classification algorithm. In this paper a literature review on gesture recognition has been reviewed and analyzed; the major tools for classification process include FSM, PCA, HMMs, and ANNs are discussed. Descriptions of recognition system framework also presented with a demonstration of the main three phases of the recognition system by detection the hand, extraction the features, and recognition the gesture.

## 8. REFERENCES


[1] John Daugman, Face and Gesture Recognition: Overview, IEEE transaction on pattern analysis and machine intelligence, vol. 19(7).

[2] Sanjay Meena, A Study on Hand Gesture Recognition Technique, Master thesis, Department of Electronics and Communication Engineering, National Institute of Technology, India. 2011

[3] Myers, B.A., A Taxonomy of User Interfaces for Window Managers, IEEE Transaction in Computer Graphics and Applications, 8(5), pp. 65-84.Doi; 10.1109/38.7762 , 1988

[4] Myers B. A., A Brief History of Human Computer Interaction Technology, ACM interactions. pp. 44-54, Vol. 5(2). Doi: 10.1145/274430.274436, 1998

[5] Mokhtar M. Hasan, and Pramod K. Mishra, Hand Gesture Modeling and Recognition using Geometric Features: A Review, Canadian Journal on Image Processing and Computer Vision 2012 Vol. 3, No.1.

[6] S. Mitra, and T. Acharya, Gesture Recognition: A Survey, IEEE Transactions on systems, Man and Cybernetics, Part C: Applications and reviews, vol. 37 (3), pp. 311-324, doi: 10.1109/TSMCC.2007.893280.

[7] G. R. S. Murthy & R. S. Jadon, A Review of Vision Based Hand Gestures Recognition, International Journal of

Information Technology and Knowledge Management, 2009 vol. 2(2), pp. 405-410.

[8] Thad Starner and Alex Pentland, Real-Time American Sign Language Recognition from Video Using Hidden Markov Models, AAAI Technical Report FS-96-05, The Media Laboratory Massachusetts Institute of Technology. 1996.

[9] C. Keskin, A. Erkan, L. Akarun, Real Time Hand Tracking and 3D Gesture Recognition for Interactive Interfaces using HMM, In Proceedings of International Conference on Artificial Neural Networks 2003.

[10] Lawrence R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, Proceedings of the IEEE, vol. 77 (2), pp. 257 – 286. 1989.

[11] Pengyu H., Matthew T., Thomas S. Huang, Constructing Finite State Machines for Fast Gesture Recognition, IEEE Proceedings, 15th International Conference on Pattern Recognition (ICPR 2000), vol. 3 ,pp. 3691-694, 2000, doi:10.1109/ICPR.2000.903639.

[12] Xingyan Li, Gesture Recognition based on Fuzzy C-Means Clustering Algorithm, Department of Computer Science. The University of Tennessee. Knoxville. 2003.

[13] David E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Pearson (2002) Edition 1

[14] Ben Krose, and Patrick van der Smagtan, An Introduction to Neural Networks, the University of Amsterdam, eighth edition. 1996.

[15] Sara Bilal, RiniAkmeliawati, Momoh J. El Salami, Amir A. Shafie, 2011. Vision-Based Hand Posture Detection and Recognition for sign Language - A study, IEEE 4th international conference on Mechatronics (ICOM 2011), pp. 1-6.

[16] Vladimir I. Pavlovic, Rajeev Sharma, and Thomas S. Huang, Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review, IEEE Transactions On Pattern Analysis And Machine Intelligence, vol. 19(7), pp. 677- 695. 1997.

[17] Pragati Garg, Naveen Aggarwal and Sanjeev Sofat, Vision Based Hand Gesture Recognition, World Academy of Science, Engineering and Technology 49, pp. 972-977, 2009.

[18] Thomas B. Moeslund and Erik Granum, A Survey of Computer Vision-Based Human Motion Capture, Elsevier, Computer Vision and Image Understanding 81, Ideal, pp. 231–268, 2001.

[19] Ying Wu, Thomas S. Huang, Vision-Based Gesture Recognition: A Review, Lecture Notes in Computer Science, Gesture Workshop, proceedings of the international Gesture Workshop on Gesture-Based communication in Human-Computer interaction, vol.(1739), pp. 103-115, 1999.

[20] Ali Erol, George Bebis, Mircea Nicolescu, Richard D. Boyle, Xander Twombly, Vision-based hand poses estimation: A review, Elsevier Computer Vision and Image Understanding 108, pp. 52–73, 2007.

[21] Laura Dipietro, Angelo M. Sabatini, and Paolo Dario, A Survey of Glove-Based Systems and their applications, IEEE Transactions on systems, Man and Cybernetics, Part C: Applications and reviews, vol. 38(4), pp. 461-482, doi: 10.1109/TSMCC.2008.923862, 2008.

[22] Joseph J. LaViola Jr. A Survey of Hand Posture and Gesture Recognition Techniques and Technology, Master Thesis, NSF Science and Technology Center for Computer Graphics and Scientific Visualization, USA, 1999.

[23] Luigi Lamberti & Francesco Camastra, Real-Time Hand Gesture Recognition Using a Color Glove, Springer 16th international conference on Image analysis and processing: Part I (ICIAP'11), pp. 365-373, 2011.

[24] MacLean J, Herpers R, Pantofaru C, Wood L,Derpanis K, Topalovic D, Tsotsos J, Fast Hand Gesture Recognition for Real-Time Teleconferencing Applications, IEEE Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, Digital Object Identifier: 10.1109/RATFG.2001.938922, Publication Year: 2001 , Page(s): 133 – 140

[25] E. Stergiopoulou, N. Papamarkos, Hand gesture recognition using a neural network shape fitting technique, Elsevier Engineering Applications of Artificial Intelligence 22, pp. 1141–1158, 2009.

[26] Tin Hninn H. Maung, Real-Time Hand Tracking and Gesture Recognition System Using Neural Networks, World Academy of Science, Engineering and Technology 50, pp. 466- 470, 2009.

[27] Manar Maraqa, Raed Abu-Zaiter, Recognition of Arabic Sign Language (ArSL) Using Recurrent Neural Networks, IEEE First International Conference on the Applications of Digital Information and Web Technologies, ICADIWT, Aug. 2008, pp.478-48, doi: 10.1109/ICADIWT.2008.4664396

[28] Kouichi Murakami and Hitomi Taguchi, Gesture Recognition using Recurrent Neural Networks, ACM, pp. 237-242., 1999.

[29] Shweta K. Yewale, Artificial Neural Network Approach For Hand Gesture Recognition, International Journal of engineering Science and Technology (IJEST), vol. 3(4), 2011.

[30] S. Sidney Fels, Geoffrey E. Hinton, Glove-Talk: A Neural Network Interface Between a Data-Glove and a Speech Synthesizer, IEEE transaction on Neural Networks, vol. 4(1), pp. 2-8, doi: 10.1109/72.182690

[31] S. Sidney Fels, Geoffiey E. Hinton, Glove-Talk II—A Neural-Network Interface which Maps Gestures to Parallel Formant Speech Synthesizer Controls, IEEE transactions on neural networks, vol. 9(1), pp. 205-212, doi: 10.1109/72.655042

[32] William T. Freeman and Michal Roth, Orientation Histograms for Hand Gesture Recognition, IEEE International Workshop on Automatic Face and Gesture Recognition, Zurich.

[33] Hanning Zhou, Dennis J. Lin and Thomas S. Huang, Static Hand Gesture Recognition based on Local Orientation Histogram Feature Distribution Model, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04). 2004.

[34] Simei G. Wysoski, Marcus V. Lamar, Susumu Kuroyanagi , Akira Iwata, A rotation invariant approach on static-gesture recognition using boundary histograms and neural networks, IEEE Proceedings of the 9th International Conference on Neural Information Processing, Singapura, November 2002.

[35] James C. Bezdek, Robert Ehrlich, William Full, FCM: The Fuzzy C-Means Clustering Algorithm, Computers & Geosciences vol. 10(2-3), pp. 191-203.

[36] Damien Zufferey, Device based gesture recognition, ACM Second International Conference on Tangible and. Embedded Interaction (TEI'08), 2008.
[37] Marco Klingmann, Accelerometer-Based Gesture Recognition with the iPhone, Master Thesis in Cognitive Computing, Goldsmiths University of London, 2009.
[38] J. P. Wachs, M. Kolsch, H. Stern, and Y. Edan, Vision-based hand gesture applications, Commun. ACM, vol. 54, pp. 60–71, 2011.
[39] Haoyun Xue, Shengfeng Qin, Mobile Motion Gesture Design for Deaf People, IEEE Proceedings of the 17th International Conference on Automation & Computing, September 2011
[40] Jinhua Zeng , Yaoru Sun , Fang Wang, A Natural Hand Gesture System for Intelligent Human-Computer Interaction and Medical Assistance, Intelligent Systems (GCIS), 2012 Third Global Congress on Digital Object Identifier: 0.1109/GCIS.2012.60 Publication Year: 2012 , Page(s): 382 – 385
[41] Sheng-Yu Peng, Kanoksak, Wattanachote, Hwei-Jen Lin, Kuan-Ching Li, A Real-Time Hand Gesture Recognition System for Daily Information Retrieval from Internet, IEEE DOI 10.1109/U-MEDIA.2011.45, Publication Year: 2011
[42] Chun Zhu, Weihua Sheng, Wearable Sensor-Based Hand Gesture and Daily Activity Recognition for Robot-Assisted Living, IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, Volume: 41 , Issue: 3 Digital Object Identifier: 10.1109/TSMCA.2010.2093883, Publication Year: 2011 , Page(s): 569 - 573
[43] Stern H I, Wachs J P, Edan Y, Human Factors for Design of Hand Gesture Human -Machine Interaction , IEEE International Conference on Systems, Man and Cybernetics, SMC '06. Volume: 5 Digital Object Identifier: 10.1109/ICSMC.2006.384767 Publication Year: 2006 , Page(s): 4052 – 4056
[44] Ashwini Shivatare, Poonam wagh, Mayuri Pisal,Varsha Khedkar, Hand Gesture Recognition System for Image Process Gaming, International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 3, March - 2013ISSN: 2278-0181
[45] Yoshinori Kuno, Teruhisa Murashima, Mobutaka Shimada and Yoshiaki Shiraia, Interactive Gesture Interface for Intelligent Wheelchairs, IEEE Conference on Multimedia and Expo. ICME 2000,Vol. 2, 2000
[46] Juan Wachs, Helman Stern, Yael Edan, Michael Gillam, Craig Feied, Mark Smith, Jon Handler A Real-Time Hand Gesture Interface for Medical Visualization Applications, Advances in Intelligent and Soft Computing Springer 2006
[47] Jithin Thampi, Muhammed Nuhas, Mohammed Rafi, Aboo Shaheed, Vision based hand gesture recognization: medical and military applications, Advances in Parallel Distributed Computing Communications in Computer and Information Science Volume 203, 2011, pp 270-281
[48] Zeng, Bobo, Wang, Guijin ; Lin, Xinggang , A hand gesture based interactive presentation system utilizing heterogeneous cameras, IEEE Tsinghua Science And Technology ISSN 1007-0214 15/18, 17(3):329-336, June 2012, DOI: 10.1109/TST.2012.6216765
[49] Michael J. Lyons, Julien Budynek and Shigeru Akamatsu, Automatic classification of single facial images, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No. 12, December 1999.
[50] Asanterabi Malima, Erol Özgür, and Müjdat Çetin, A Fast Algorithm For Vision-Based Hand Gesture Recognition For Robot Control, IEEE Signal Processing and Communications Applications, 2006 IEEE 14th, DOI: 10.1109/SIU.2006.1659822, Publication Year: 2006 , Page(s): 1 - 4 Cited by: Papers (9)
[51] J. Yamato, J. Ohya, and K. Ishii, Recognizing human action in time sequential images using hidden Markov model, in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recogn., Champaign, IL, 1992, pp. 379–385.
[52] F. Samaria and S. Young, HMM-based architecture for face identification, Elsevier, Image Vision Computing, vol. 12, pp. 537–543, 1994.
[53] Stefan Reifinger, Frank Wallhoff, Markus Ablassmeier, Tony Poitschke, and Gerhard Rigoll Static and Dynamic Hand-Gesture Recognition for Augmented Reality Applications, Springer LNCS 4552, pp. 728–737, 2007.
[54] Mahmoud Elmezain, Ayoub Al-Hamadi, Jorg Appenrodt, and Bernd Michaelis, A Hidden Markov Model-Based Isolated and Meaningful Hand Gesture Recognition, International Journal of Electrical and Electronics Engineering 3:3 2009
[55] Ashutosh Samantaray, Sanjaya Kumar Nayak, Ashis Kumar Mishra , Hand Gesture Recognition using Computer Vision, International Journal of Scientific & Engineering Research, Volume 4, Issue 6, June 2013 ISSN 2229-5518
[56] Sharma, R., Huang, T. S., Pavovic, V. I., Zhao, Y., Lo, Z., Chu, S., Schulten, K., Dalke, A., Phillips, J., Zeller, M. & Humphrey, W, Speech/Gesture Interface to a Visual Computing Environment for Molecular Biologists, In: Proc. of ICPR'96 II (1996), 964-968.
[57] Gandy, M., Starner, T., Auxier, J. & Ashbrook, D, The Gesture Pendant: A Self Illuminating, Wearable, Infrared Computer Vision System for Home Automation Control and Medical Monitoring, Proc. of IEEE Int. Symposium on Wearable Computers. (2000), 87-94.
[58] Goza, S. M., Ambrose, R. O., Diftler, M. A. & Spain, I. M, Telepresence Control of the NASA/DARPA Robonaut on a Mobility Platform, In: Proceedings of the 2004 Conference on Human Factors in Computing Systems. ACM Press, (2004) 623–629.
[59] Stotts, D., Smith, J. M. & Gyllstrom, K. Facespace: Endoand Exo-Spatial Hypermedia in the Transparent Video Facetop. In: Proc. of the Fifteenth ACM Conf. on Hypertext & Hypermedia. ACM Press, (2004) 48–57.
[60] Smith, G. M. & Schraefel. M. C, The Radial Scroll Tool: Scrolling Support for Stylus-or Touch-Based Document Navigation, In Proc. 17th ACM Symposium on User Interface Software and Technology. ACM Press, (2004) 53–56.
[61] Freeman, W., Tanaka, K., Ohta, J. & Kyuma, K. Computer Vision for Computer Games, Tech. Rep. and International Conference on Automatic Face and Gesture Recognition, (1996).
[62] Konrad, T., Demirdjian, D. & Darrell, T. Gesture + Play: Full-Body Interaction for Virtual Environments. In: CHI '03 Extended Abstracts on Human Factors in Computing Systems, ACM Press, (2003) 620–621.
[63] Website: University of Chicago: Mcneill Lab for Gesture and Speech Research. Electronic Resource, (2006).

[64] Valli, C. & Lucas, C. Linguistics of American Sign Language: An Introduction, Washington, D. C.: Gallaudet University Press, (2000).

[65] Martinez, A., Wilbur, B., Shay, R. & Kak, A. Purdue RVLSLLL ASL Database for Automatic Recognition of ASL, In IEEE Int. Conf. on Multimodal Interfaces, (2002) 167–172.

[66] Starner, T., Weaver, J. & Pentland, A. Real-Time American Sign Language Recognition using Desk and Wearable Computer Based Video, PAMI, 20(12) (1998) 1371–1375.

[67] Vogler, C. & Metaxas, D. A Framework for Recognizing the Simultaneous Aspects of American Sign Language, Computer Vision and Image Understanding, 81(3) (2001) 358–384.

[68] Waldron, M. Isolated ASL Sign Recognition System for Deaf Persons, IEEE Transactions on Rehabilitation Engineering, 3(3) (1995) 261–271.

[69] Dong Guo Yonghua, Vision-Based Hand Gesture Recognition for Human-Vehicle Interaction, International Conference on Control, Automation and Computer Vision, 1998

[70] Pickering, Carl A. Burnham, Keith J. Richardson, Michael J. Jaguar , A research Study of Hand Gesture Recognition Technologies and Applications for Human Vehicle Interaction, 3rd Conference on Automotive Electronics, 2007

[71] Juan P. Wachs , Helman I. Stern, Yael Edan, Michael Gillam, Jon Handler, Craig Feied, Mark Smith, A Gesture-based Tool for Sterile Browsing of Radiology Images, Journal of the American Medical Informatics Association (2008; 15:321-323, DOI 10.1197/jamia.M24)

[72] Freeman, W. T., & Weissman, C. D, Television control by hand gestures, IEEE International Workshop on Automatic Face and Gesture Recognition. Zurich.

[73] Moni, M. A. & Ali, A. B. M. S, HMM based hand gesture recognition: A review on techniques and approaches, 2nd IEEE International Conference on Computer Scienceand Information Technology, (ICCSIT 2009).

[74] Hasan, M. M., & Mishra, P. K., HSV brightness factor matching for gesture recognition system, International Journal of Image Processing (IJIP), 2010 vol. 4(5)

[75] Hasan, M. M., & Mishra, P. K., Gesture recognition using modified HSV segmentation, IEEE International Conference on Communication Systems and Network Technologies.2011

[76] Hasan, M. M., & Mishra, P. K. , Brightness factor matching for gesture recognition system using scaled normalization, International Journal of Computer Science & Information Technology (IJCSIT), 2011 3(2).

[77] Mo, S., Cheng, S., & Xing, X, Hand gesture segmentation based on improved kalman filter and TSL skin color model, International Conference on Multimedia Technology (ICMT), Hangzhou.2011

[78] Peter, H. P. (2011). Image Segmentation. (1sted.). India. (Part 2). Image Segmentation Methods Image Segmentation through Clustering Based on Natural Computing Techniques.

[79] Kim, J., & Song, M, Three dimensional gesture recognition using PCA of stereo images and modified matching algorithm, IEEE Fifth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD '08, pp. 116 – 120

[80] Verma, R., & Dev A, Vision based hand gesture recognition using finite state machines and fuzzy logic, IEEE International Conference on Ultra Modern Telecommunications & Workshops, ICUMT '09, pp. 1-6.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 20 - 23, 2014

# Float-based Pipeline Monitoring Network

Ata Barzegar
Computer Engineering Department
Islamic Azad University, Science and Research Branch,
Khouzestan, Iran

Sattar Hashemi
Computer Engineering Department
Shiraz University
Shiraz, Iran

Gholamreza Akbarizadeh
Electrical Engineering Department
Shahid Chamran University of Ahvaz
Ahvaz, Iran

***Abstract*:** Wireless sensor networks are used to monitor water/oil pipelines. The excising schemes for pipeline monitoring are not energy efficient. Network operators have to daily travel across pipelines to change batteries of sensor nodes. In this paper, we propose a novel sensor-based pipeline monitoring protocol called FPMN. FPMN utilizes float nodes to reduce energy consumption in sensor nodes by reducing tasks of sensor nodes. FPMN tries to spread best route information along the pipelines in a way to reduce bandwidth and energy consumption. Our simulation experiments show that FPMN reduces energy consumption by 68.3% without change in packet delay and network traffic.

**Keywords:** wireless network; pipeline; sensor; forwarding; routing

## 1. INTRODUCTION

Sensor networks usually contain a large number of sensors distributed in a controlled area and a single sink node that collects sensed data from sensor nodes. Today's wireless sensors are very limited in terms of battery, processing power, and memory capacity except for the sink. For example, MICA2 contains 4 kilobytes memory [1].

During the operation of a wireless sensor network, sensor nodes frequently die because of physical damage or battery discharge. In addition, a packet transmitted between two wireless nodes may contain errors due to noise or interference [2].

In this paper, we propose a novel pipeline monitoring protocol called FPMN. FPMN tries to utilize float nodes instead of fixed nodes. FPMN contains a number of mechanisms to achieve this goal. Our simulation experiments show that FPMN reduces energy consumption in fixed nodes without change in network overhead and delay.

The rest of this paper is organized as follows. We review the related work on wireless sensor networks and pipeline monitoring designs in Section II. We propose FPMN in Section III. Section IV contains our simulation results and section V concludes the paper.

## 2. RELATED WORK

In this section, we review former researches related to our work.

### *A. Wireles Sensor Network*

There are a number of routing protocols such as AODV [3], CTP [4] that discover the network topology to find a route. Then, each node identifies the neighbor to which it has to forward the packet. In this way, packets travel the shortest path from the source to the destination which is the sink in wireless sensor network.

Now we review the existing routing algorithms for wireless sensor networks. Directed Diffusion [5] routing protocol aims at diffusing the data through the sensor nodes by using a naming scheme. Rumor Routing [6] is a variation of Directed Diffusion intended for scenarios where geographic routing is not applicable. LEACH [7] is a cluster-based routing algorithm, but uses single-hop routing and can therefore not be applied to networks deployed in large regions. GPSR [8] uses a greedy forwarding strategy to forward only local information, but it has been designed for mobile ad hoc networks and requires a location service to map locations and node identifiers. GAF [9] is an energy-aware location-based routing algorithm. Its basic idea is to set up a virtual grid based on location information. Spectra uses the cluster-based Ripple routing algorithm. When a CH requires a route to BS, it broadcasts a Route-Request message in the network. If a CH receives a Route-Request and contains a route to BS, then it sends a Route-Reply message to the requesting CH.

SenseCode [10] is a collection protocol for sensor network to employ network coding. SenseCode provides a way to gracefully introduce a configurable amount of redundant information in the network, thereby increasing reliability in the face of packet loss.

### *B. Pipeline Monitoring Networks*

PipeNet [11] is a pipeline monitoring system for collecting hydraulic and acoustic/vibration data at high sampling rates as well as algorithms for analyzing this data to detect and locate leaks.

MISE-PIPE is introduced in [12] to provide low-cost and real-time leakage detection and localization for underground pipelines.

SPAMMS [13] is an autonomous sensor-based system that combines robot agent based technologies with sensing technologies for efficiently locating health related events and allows active and corrective monitoring and maintenance of the pipelines.

Authors in [14] present a wireless sensor network system and its reliability assessment model for oil and gas pipelines condition monitoring.

SWATS [15] is a sensor network based system that aims to allow continuous monitoring of the steam flood and water flood systems with low cost, short delay, and fine granularity coverage while providing high accuracy and reliability.

In [16], for sensor nodes that are utilized to monitor oil pipelines, authors study the linear sensor placement problem with the goal of maximizing their lifetime.

In [17], an industrial wireless architecture was designed and simulated to transport data and video for a particular pipeline system.

## 3. THE FPMN PROTOCOL

In this section, we propose a pipeline monitoring protocol based on wireless sensor network. Our proposed protocol is called FPMN (Float-based Pipeline Monitoring Network).

We assume we have a number of crossing pipelines where wireless sensor nodes are placed along the pipelines (such as Fig. 1) such that the resulted wireless network is connected. A base station (sink) is placed at each end of every pipeline. Sensor nodes sometimes send monitoring data packets to the pipeline control room. All sinks are connected using a separate high speed network (such as the Internet) in a way that receiving a packet by the pipeline control room is equivalent to receiving that packet by any sink.

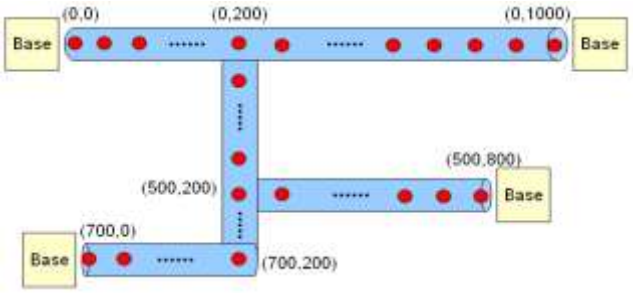



Figure 1 An pipeline network used in our simulation; Red points indicate sensor nodes; (x,y) indicates the physical cordinate of a point in the network.

The routing algorithm can be any routing algorithm that discovers the next hop to reach the sink. Each node has a fixed buffer capacity to temporarily save packets. Packet contains its generation time and expiration time.

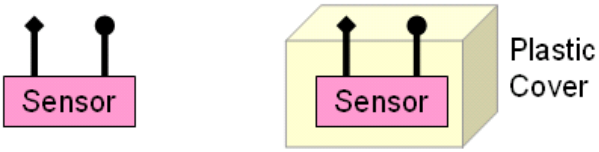
### A. *System Design*

We use two kinds of nodes:

- Fixed Node: It does both sensing, routing, and forwarding.
- Float Node: It does only packet forwarding.

The simple way for making a float node is use of a fixed node inside a water-proof plastic cover (Fig. 2). Float nodes move by the power of the streaming liquid inside the pipe. Thus, float node requires a small battery only for data processing and packet forwarding.

Figure 2 Network nodes




(a) Fixed Node (b) Float Node

Workers are assigned at each pipeline head to do the followings.

- They periodically insert float nodes inside pipelines.
- They collect outgoing float nodes from pipelines and repair them and/or change their batteries.

Fixed sensor nodes monitor the pipeline and send the sensed data toward the sink. Both fixed and float nodes are able to forward data packets. The aim is to utilize float nodes as many as possible.

Using the neighbour identification mechanism, each node is able to find out which fixed/float nodes are its neighbours in what directions.

When a node is going to forward a packet, it checks whether there is a multipipe node inside its neighbourhood. Then, it select the next hop according to the following priorities.

1. Multipipe node (highest priority)
2. Float node
3. Single-pipe node (lowest priority)

We give the highest priority to multipipe nodes, because a multipipe node determines the best route for packets. A float node is not fixed and has no information about pipe routes.

### B. *Buffer Management*

When the node receives or generates a new data packet, it tries to insert it in its buffer. If the buffer has no room, the node drops the packet.

Each node periodically checks its buffer. When the expiration time of a packet is passed, it is removed from node's buffer. This mechanism avoids delayed packet from staying in the network for a too long time.

### C. *Node Behavior*

A sensor node gets one of the following roles when forwarding a packet in FPMN:

- Source Node: A sensor node that generates a data packet.
- Single-Pipe Node: A sensor node that has only two directions of a single pipe toward the sinks. (such as node *s* in Fig. 3)
- Multi-Pipe Node: A sensor node that has more than two directions toward the sinks due to residing on the junction of multiple pipes. (such as node *t* in Fig. 3)

The node acts differently when it gets one of the roles as below.

- Fixed Source Node: Fig. 4.
- Float Node or Fixed Single-Pipe Node: When the node receives a packet from side of the pipe, it simply forwards the packet toward the other side of the pipe. If the node is fixed, it increases the packet's FixedNodeCount by one before forwarding.
- Fixed Multi-Pipe Node: Fig. 5.

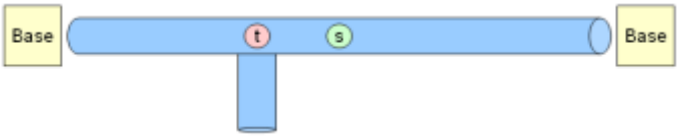



Figure 3 An example of pipelines

Algorithm. SourceSteps

Assumption: A node generated a packet *pkt* to send toward the sink.

I. Set FixedNodeCount(*pkt*) = 0

II. Mark *pkt* as a data packet.

III. Do the steps of a multi-pipe node.

Figure 4. Steps when a source generates a data packet

Algorithm. MultiPipeSteps

Assumption: The node knows the shortest path SP to the sink. The node has to forward a packet *pkt*. The packet is either generated in this node or received from another node.

I. Set FixedNodeCount(*pkt*) = FixedNodeCount(*pkt*) + 1

II. If all the branches have fixed node estimation, then

a. Find branch B that has the least number of fixed nodes among all the branches.

b. Send *pkt* on B.

III. Otherwise, if SP has fixed node estimation, then

a. Find branch B that has the least number of fixed nodes among the branches with fixed node estimation.

b. Send *pkt* on B.

IV. Otherwise, send *pkt* on SP.

V. End

Figure 5 Steps when a multi-pipe node has to forward a packet

TABLE I. SIMULATION SCENARIO I AND II

| Parameter | Scenario I | Scenario II | Scenario III |
|---|---|---|---|
| Number of fixed nodes | 125 and 250 | 125 | 125 |
| Network Traffic | A Pareto-On/Off flow from every fixed node to the sink | | |
| Average Bit Rate per traffic flow | 1 Kbps | | |
| The idle_time period per traffic flow | 2 seconds | | |
| Packet Expiration Duration | 3 seconds | | |
| Node's Buffer Capacity | 20 packets | | |
| Number of float nodes | 125 | | |
| $T_H$ | 1.5 s | 1.5 s | multiple cases |
| Float node speed | 0.2 m/s | multiple cases | 0.2 m/s |
| Simulation Duration | 1 hour | | |

## 4. SIMULATION

We implemented the FPMN protocol in the NS2 [18] network simulator. In this section, we evaluate the performance and the overhead of FPMN. To evaluate FPMN, we define the two simulation scenarios presented in Table I. Fig. 1 depicts the pipeline network topology.

Every node sends traffic to the sink. We use a Pareto-On/Off model (one of the traffic generators in NS2 [18]) for each traffic flow with idle_time equal to 2 seconds. In this way, every node sends variable-bit-rate data and all the nodes do not synchronously send data to the sink.

In Scenario I, we use different numbers of nodes in different executions. In Scenario II, we use different values for float node speed in different executions whereas the other parameters are fixed. In Scenario III, we use different Hello intervals.

We simulated the following protocols:

- FPMN
- No-FPMN: Simplified FPMN without float nodes in which each source sends generated packets to the sink along the shortest path.

### A. *Simulation Results*

End-to-end packet delay depends on node delay and path length. For packet delay, the main difference between FPMN and No-FPMN is that FPMN uses float nodes instead of some fixed nodes in data path. However, FPMN uses the least fixed node path instead of the shortest path sometimes. The least fixed node path is a short path and probably the shortest path. Therefore, there seems no delay difference between FPMN and No-FPMN. Fig. 6 shows end-to-end delay in the simulated protocols.

FPMN tries to utilize float nodes instead of fixed nodes as much as possible. Fig. 7 shows the energy consumed in fixed nodes. Energy consumption is proportional to number of nodes. In this experiment, FPMN averagely requires 68.3 percent less energy in fixed nodes compared to No-FPMN.

We define traffic overhead as follows.

Traffic Overhead = (Control Traffic) x 100 / (Data Traffic + Control Traffic)

Fig. 8 compares traffic overhead in the two simulated protocols. In this experiment, traffic overhead is averagely 20.8% in FPMN and 19.2 in No-FPMN.

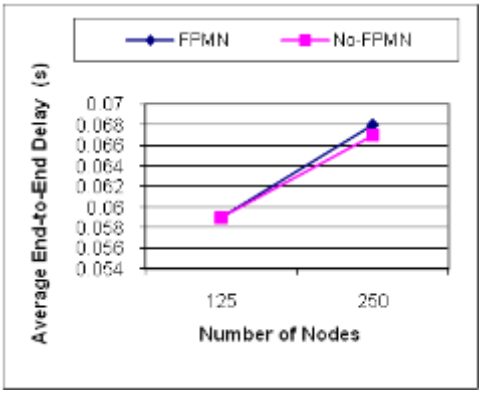



Figure 6 Average end-to-end packet delay versus number of nodes

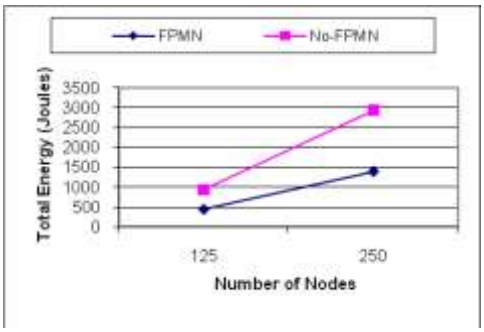



Figure 7 Total energy consumption in fixed nodes versus number of nodes

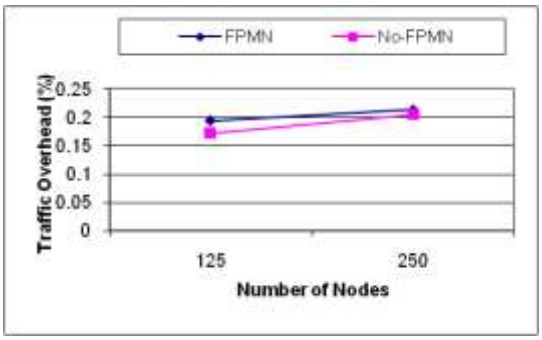



Figure 8 Traffic overhead in float and fixed nodes versus number of nodes

## 5. CONCLUSIONS

In this paper, we propose a novel pipeline monitoring protocol called FPMN. FPMN tries to efficiently spread packets along the pipelines by choosing the best path to send a packet. This path is either the shortest path or the least fixed node path. FPMN finds the least fixed node path using the fixed-node-count field in data packets. Our simulation experiments show that FPMN reduces energy consumption with acceptable overhead.

## 6. ACKNOWLEDGEMENT

The work described in this paper was supported by the Shahid Chamran University of Ahvaz (SCU) as a M.Sc. research thesis. It was financial supported with Grant treaty in 2013 with number 874095.The authors would like to thank from the Shahid Chamran University of Ahvaz (SCU) for financial support.

## 7. REFERENCES

[1] Crossbow technology inc. URL: http://www.xbow.com

[2] H. Karl, and A. Willig, Protocols and Architectures for Wireless Sensor Networks, John Wiley & Sons, 2005, ISBN: 0-470-09510-5.

[3] C. E. Perkins, "Ad hoc On Demand Distance Vector (AODV) Routing", Internet Draft, draft-ietfmanet-aodv-01.txt, 1998.

[4] "The Collection Tree Protocol (CTP)", http://www.tinyos.net/tinyos-2.x/doc/html/tep123.html.

[5] C. Intanagonwiwat, R. Govindan, and D. Estrin, "Directed diffusion: a scalable and robust communication paradigm for sensor networks", In Proceedings of the 6th Annual ACM/IEEE International Conference on Mobile Computing and Networking, August 2000, Boston, USA.

[6] D. Braginsky, and D. Estrin, "Rumor routing algorithm for sensor networks", In Proceedings of the First Workshop on Sensor Networks and Applications (WSNA), October 2002, Atlanta, GA.

[7] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks", In Proceedings of the 33rd Hawaii International Conference on System Sciences, January 2000, Hawaii. IEEE, pp. 1–10.

[8] B. Karp, and H. Kung, "GPSR: Greedy perimeter stateless routing for wireless sensor networks", In Proceedings of the 6th Annual ACM/IEEE International Conference on Mobile Computing and Networking, August 2000, Boston, MA.

[9] Y. Xu, J. Heidemann, and D. Estrin, "Geography-informed energy conservation for ad hoc routing", In Proceedings of the 7th Annual ACM/IEEE International Conference on Mobile Computing and Networking, July 2001, Rome, Italy.

[10] L. Keller, E. Atsan, K. Argyraki, and C. Fragouli, "SenseCode: Network coding for reliable sensor networks", EPFL Technical Report, October 2009.

[11] I. Stoianov, L. Nachman, and S. Madden, "PIPENET: A Wireless Sensor Network for Pipeline Monitoring", Int'l. Conf. Info. Process. Sensor Net., Apr. 2007.

[12] Z. Sun, P. Wang, M. C. Vuran, M. Al-Rodhaan, A. Al-Dhelaan, and I. F. Akyildiz, "MISE-PIPE: Magnetic induction-based wireless sensor networks for underground pipeline monitoring", Ad Hoc Networks 9(3), pp.218-227, 2011.

[13] J. Kim, G. Sharma, N. Boudriga, and S. Iyengar, "SPAMMS: A Sensor-based Pipeline Autonomous Monitoring and Maintenance System", Proceedings of The 2nd International Conference on Communication Systems and Networks (COMSNETS), Bangalore, India, pp.118-127, IEEE, January 2010.

[14] K. EL-Darymli, F. Khan, M. H. Ahmed, "Reliability Modeling of Wireless Sensor Network for Oil and Gas Pipelines Monitoring", Sensors & Transducers Journal, Vol. 106, Issue 7, pp. 6-26, July 2009.

[15] S. Yoon, W. Ye, J. S. Heidemann, B. Littlefield, and C. Shahabi, "SWATS: Wireless sensor networks for steamflood and waterflood pipeline monitoring", IEEE Network 25(1), pp.50-56, 2011.

[16] Y. Guo, F. Kong, D. Zhu, A. S. Tosun, and Q. Deng, "Sensor Placement for Lifetime Maximization in Monitoring Oil Pipelines", in the Proc. of the ACM/IEEE First International Conference on Cyber-Physical Systems (ICCPS), Stockholm, Sweden, Apr. 2010.

[17] R. Beatriz, and A. Meza, "Wireless sensors network for the cerro corona tailings pipeline system", in proc. of the Ibero-American Conference on Trends in Engineering Education and Collaboration, 2009.

[18] Information Sciences Institute, NS-2 network simulator, Software Packet, 2012, http://www.isi.edu/nsnam/ns/.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 24 - 28, 2014

# Proposed Algorithm for Surveillance Applications

Youssef Benabbassi
University of Bechar
RIIR Laboratory, Oran
Bechar, Algeria

Hafid Haffaf
University of Oran
RIIR Laboratory, Oran
Oran, Algeria

Congduc Pham
University of Pau
Pau LIUPPA, France
Pau, France

**Abstract**: Technological systems are vulnerable to faults. In many fault situations, the system operation has to be stopped to avoid damage to machinery and humans. As a consequence, the detection and the handling of faults play an increasing role in modern technology, where many highly automated components interact in a complex way such that a fault in a single component may cause the malfunction of the whole system. This work introduces the main ideas of fault diagnosis and fault-tolerant control under the optics of various research work done in this area. It presents the Arduino technology in both hardware and software sides. The purpose of this paper is to propose a diagnostic algorithm based on this technology. A case study is proposed for this setting. Moreover, we explained and discussed the result of our algorithm.

**Keywords**: Technological systems, control system, actuator faults, Arduino, sensors faults

## 1. INTRODUCTION

The manufacturing systems consist of many different machine tools, robots and transportation systems all of which have to correctly satisfy their purpose in order to ensure an efficient and high-quality production. Mobile communication provides another example where networked components interact so heavily that component faults have far reaching consequences.

In the general sense, a fault is something that changes the behavior of a system such that the system does no longer satisfy its purpose. In large systems, every component has been designed to accomplish a certain function and the overall system works satisfactorily only if all components provide the service they are designed for. Therefore, a fault in a single component usually changes the performance of the overall system. The presence of a fault detection and isolation (FDI) system is then necessary to detect the occurrence of a fault and isolate it. Normally, the task which comes after the isolation of a fault is its repairing. The problem of FDI and Fault Tolerant Control (FTC) is an important problem to deal with, since faults in sensors, actuators and components are usually associated to increasing operating costs, off-specification production, line shut-down and possible detrimental environment impact.

Overall, fault-tolerant control is a complex interdisciplinary research field that covers a diverse range of engineering disciplines, such as modeling and identification, applied mathematics, applied statistics, stochastic system theory, reliability and risk analysis, computing, communication, control, signal processing, sensors and actuators, as well as hardware and software implementation techniques.

The measures should be carried out by the control equipment, in order to avoid production deteriorations or damage to machines and humans. Their aim is to make the system fault tolerant. The control algorithm adapts to the faulty plant and the overall system satisfies its function again as it is shown in Figure 1.

In the literature, most of the motivation and research work in fault tolerant control involves solving problems encountered in safety critical systems such as aircraft.

In [1] the author proposes a solution for fault detection and isolation using system dynamics identification techniques. In [2] the authors propose a fault detection and isolation scheme for industrial systems based on multiple operating model. Rosa and al. in [3] describe an application of a new fault detection and isolation (FDI) technique based on set-valued observers (SVOs) to a linear parameter varying (LPV) longitudinal aircraft dynamic model. The authors in [4] propose parameter estimation methods for fault detection and isolation. In [5] Tharrault and al. propose fault detection and isolation with robust principal component analysis. This proposed scheme avoids the combinatorial explosion of faulty scenarios related to multiple faults to be considered. The authors in [6] designed closed-loop fault-tolerant control for uncertain nonlinear systems. This solution is based on a new algebraic estimation technique of the derivatives of a time signal. This yields good estimates of the unknown parameters and of the residuals of the fault indicators.

A general active fault-tolerant control framework is proposed in [7] for nonlinear systems with sensor faults. According to their identifiability, all sensor faults are divided into two classes: identifiable faults and non-identifiable faults.

In [8] the authors propose a new fault-tolerant control methodology using adaptive estimation and control approaches based on the learning capabilities of neural networks or fuzzy systems. On-line approximation-based stable adaptive neural/fuzzy control is studied for a class of input–output feedback linearizable time-varying nonlinear systems. The authors in [9] discussed the problem of designing fault tolerant compensators that stabilize a given system both in the nominal situation, as well as in the situation where one of the sensors or one of the actuators has failed.
The paper is organized as follows. Section 2 presents the Fault-Tolerant Control. Section 3 presents a case study. Section 4 presents our algorithm proposed for diagnosis based on Aduino programming. Finally, section 5 concludes the paper and points out open research problems.

# 2. FAULT-TOLERANT CONTROL

## 2.1 Definitions

**System:** A system is a set of interconnected components. Each of the components has been chosen or designed by the system engineer so as to achieve some function of interest. A function describes what the design engineer expects the components to perform, independently of how it is performed. A component performs some function because it has been designed so as to exploit some physical principles. Which in general are expressed by some relationships between the time evolution of some system variables. Such relationships are called constraints, and the time evolution of a variable is called its trajectory [10][11].

**Fault:** A fault in a dynamical system is a deviation of the system structure or the system parameters from the nominal situation. Examples for structural changes are the blocking of an actuator, the loss of a sensor or the disconnection of a system component.

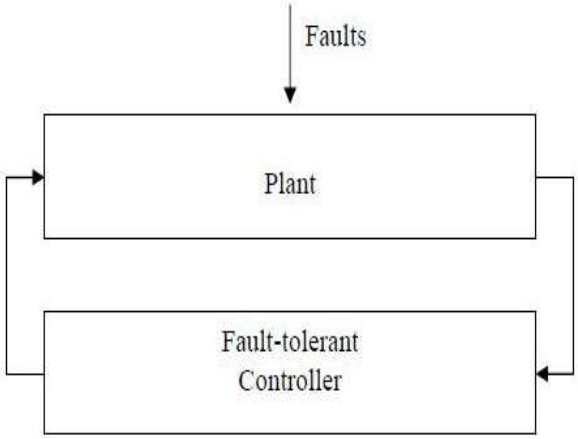



Figure 1. Fault-tolerant system

## 2.2 Classification of faults

The faults are often classified as follows:

**Plant faults:** such faults change the dynamical I/O properties of the system.

**Sensor fault:** the plant properties are not affected, but the sensor readings have substantial errors.

**Actuators faults:** the plant properties are not affected, but the influence of the controller on the plant is interrupted or modified.

## 2.3 Fault-tolerant control notions

The aims of fault-tolerant control are related to these notions, which result from different views on faulty systems:

### *2.3.1 Safety*

It describes the absence of danger. A safety system is a part of the control equipment that protects a technological system from permanent damage. A controlled shut-down brings the technological process into a safe state, if specified conditions are met. Dedicated actuators stop the process. The over-all system is then called a fail-safe system.

### *2.3.2 Reliability*

It is the probability that a system accomplishes its intended function for a specified period of time under normal conditions. Fault-tolerant control cannot change the reliability of the plant components, but it improves the reliability of the overall system, because with a fault-tolerant controller the overall system remains operational after the appearance of faults.

### *2.3.3 Availability*

It is the probability of a system to be operational when needed. Contrary to reliability it also depends on the maintenance policies, which are applied to the system components.

### *2.3.4 Dependability*

It lumps together the three properties of reliability, availability and safety. A dependable system is a fail-safe system with high availability and reliability.

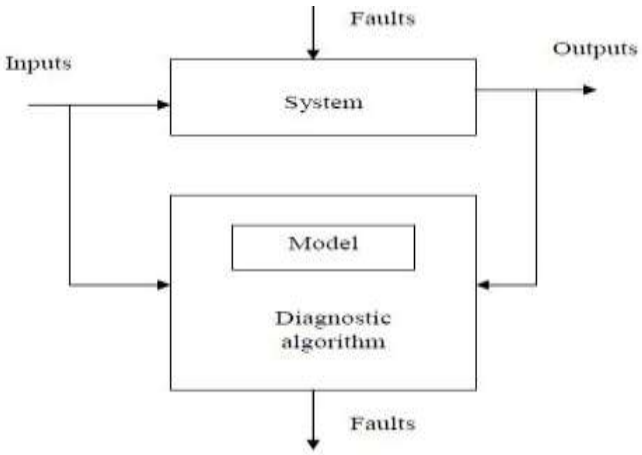



Figure 2. Fault diagnosis

## 2.4 Physical redundancy and analytical redundancy

Fault tolerance necessitates redundancies. The main advantage of fault-tolerant control over measures for fault tolerance is the fact that fault-tolerant control makes intelligent use of the redundancies included in the system and in the information about the system in order to increase the system availability. The analytical redundancy is cheaper than duplicating all vulnerable components.

Three methods for ensuring fault tolerance are:

### *2.4.1 Robust control*

A fixed controller is designed that tolerates changes of the plant dynamics. The controlled system satisfies its goals under all faulty conditions. Fault tolerance is obtained without changing the controller parameters. It is, therefore, called passive fault tolerance.

### *2.4.2 Adaptive control*

The controller parameters are adapted to changes of the plant parameters. If these changes are caused by some fault, adaptive control may provide active fault tolerance.

### *2.4.3 Diagnosis steps*

For fault-tolerant control, the location and the magnitude of the fault have to be found. Different names are used to distinguish the diagnosis steps according to their depth:

- Fault detection: Decide whether or not a fault has occurred. This step determines the time at which the system is subject to some fault.
- Fault isolation: find in which component a fault has occurred. This step determines the location of the fault.
- Fault identification and fault estimation: identify the fault and estimate its magnitude. This step determines the kind of fault and its severity.

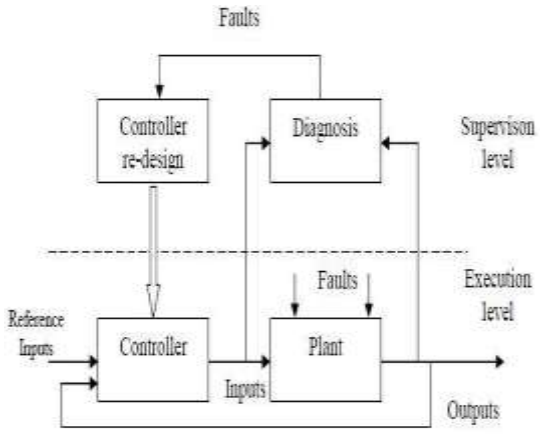



Figure 3. Architecture of fault-tolerant control

To diagnose a system by testing the consistency of the measurements with a model is a general idea, which does not depend on the kind of mode used.

Several direct consequences of this principle should be mentioned:

- Fault detection is possible without any information about the behavior of the faulty plant. Fault detection algorithms use only a model of the nominal plant. The main idea is to identify deviations of the current system behavior from the nominal behavior, which is possible without a list of all possible faults.
- Without information about the faults and about the way in which the faults affect the system, no fault isolation and identification is possible. In order to identify the fault, fault models have to be known.
- With a given measurement configuration, not all faults can be distinguished. Diagnosability considerations can be used to determine those faults that can be separately identified. Generally, the way to make a system fault-tolerant consists of two steps:

Fault diagnosis: The existence of faults has to be detected and the faults have to be identified.

Control re-designs: The controller has to be adapted to the faulty situation so that the overall system continues to satisfy its goal.

## 3. CASE STUDY

This section presents the technique to monitor our process using an Arduino-based microcontroller programmed. More details on the Arduino-Uno and programming will be given in the next section. Our process is seen as a functional system which outputs values indicating the status of its operation. This operating state is classified either correct mode operation, malfunction or faulty. The first output, indicates the temperature (T) released, and the second indicates the pressure (P) collected during operation of the process. In nominal mode, the values of these two variables are the zero state (0). So the system is safe mode. But once one of the two values of these two variables changes to state one (1) this means that the system switches to malfunction. An alarm is triggered, indicating the presence of this fault, which requires adequate operation, either cooler level, or at the compressor. In the worst case we ordered a forced shutdown of the system. As these two variables are monitored, one can have the four possible modes of operation of the system, they are given as follows in table 1:

| Temperature (T) | Pressure (P) | Operating Mode |
|---|---|---|
| 0 | 0 | Functional (safe) |
| 0 | 1 | Dysfunction (in compressor) |
| 1 | 0 | Dysfunction (in cooling) |
| 1 | 1 | Failure |

Table 1. Operating modes of the process

Our process monitoring scheme based on an Arduino microcontroller is given as follows in Figure 4:

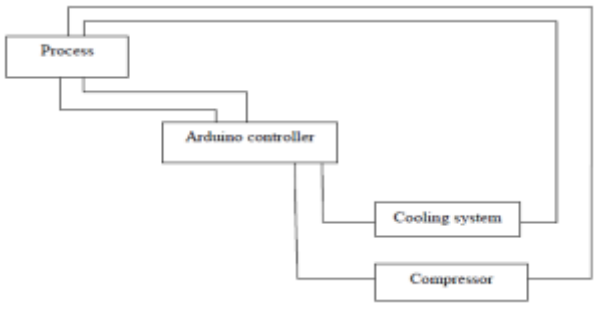



Figure 4. Process monitoring scheme based on Arduino

## 4. PROGRAMMING ARDUINO

The Arduino is a microcontroller board based on a (mini-computer) Atmel ATMEGA8 or ATMEGA168. It has in its basic version 1KB of RAM and 8K of flash memory for storing programs. It can be connected to 13 digital inputs or outputs, 3 PWM (up to 3 analog outputs: see http://fr.wikipedia.org/wiki/PWM) and 6 analog inputs

converting 10 bits. In the most common version, the communication with the computer is via a USB port as shown in Figure 5. There are several versions of the Arduino, including a miniaturized version [12]. The card has an internal software system (editable) and user programs.

Figure 5. Arduino Uno board

Arduino programs are written in C or C++. The Arduino IDE, as shown in Figure 6, comes with a software library called "Wiring" from the original Wiring project, which makes many common input/output operations much easier. Users only need define two functions to make a runnable cyclic executive program. The first function is setup (). This function is launched once at the start of a program that can initialize settings. The second function is loop (). It is called repeatedly until the board powers off.

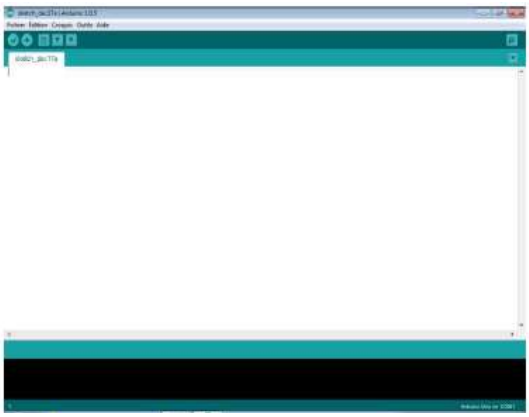


Figure 6. Arduino IDE

Our proposed diagnostic algorithm is defined as follows:

```
Diagnosis Algorithm
Input:
    Process_Temerature_Status;
    Process_Pression_Status;
Output:
    Process_Alert_Status;
    Process_Temperature_Rgulation;
    Process_Pression_Regulation;
BEGIN
Int reading_Pin_13=LOW;
Int reading_Pin_12=LOW;
Int reading_Pin_8=LOW;
Int reading_Pin_7=LOW;
Set_Method()
Begin
        PinMode(13,Input);
        PinMode(12,Input);
        PinMode(11,Output);
        PinMode(10,Output);
        PinMode(7,Output);
        PinMode(8,Output);
End
Loop_Method()
Begin
reading_Pin_13 = digitalRead(13);
reading_Pin_12 = digitalRead(12);
        if ((reading_Pin_13=HIGH) and (reading_Pin_12=HIGH)) Then
        begin
                digitalWrite(11, HIGH);
                Temerature_regulation
                digitalWrite(10, HIGH);
                Pression_regulation;
                Alert_status_Failure;
                digitalWrite(7, HIGH);
                digitalWrite(8, HIGH);
        end
        else
        begin
                if ((reading_Pin_13=HIGH) and (reading_Pin_12=LOW)) Then
                begin
                        digitalWrite(11, HIGH);
                        Temerature_regulation;
                        digitalWrite(10, LOW);
                        Alert_status_Temperature_Dysfonction;
                        digitalWrite(7, HIGH);
                        digitalWrite(8, LOW);
                end
                else
                begin
                if ((reading_Pin_13=LOW) and (reading_Pin_12=HIGH)) Then
                begin
                        digitalWrite(10, HIGH);
                        Pression_regulation
                        digitalWrite(11, LOW);
                        Alert_status_Pression_Dysfonction;
                        digitalWrite(7, LOW);
                        digitalWrite(8, HIGH);
                end
                else
                begin
                        digitalWrite(11, LOW);
                        digitalWrite(10, LOW);
                        Alert_status_Safe;
                        digitalWrite(7, LOW);
                        digitalWrite(8, LOW);
                end
                end
        end
End
END
```

The program is written on the IDE Aduino then it was compiled. The executable code is loaded to the Arduino

board. The system is connected to the Arduino. At this point, the Arduino board controls the system. The results of the implementation of our application are given as follows:

The system provides two output values. The first represents the state of the temperature and the other represents the state of the pressure, in which the operation of the system continues. Then the system generates these two values to the Arduino board, in the pin 13 we find the value indicating the status of the temperature, zero (0) means nothing to report, one (1) means that there is an anomaly to report. So, the state of pin 7 passes to state one (1). It triggers an alarm and proceeds by actuating the cooler. Pin 12 gives the value indicating the status of the pressure, zero (0) means nothing to report, one (1) means that there is an anomaly report. So, the pin 8 passes to state one (1). It triggers an alarm and proceeds by actuating the pressure regulator. In both cases the system switches to malfunction. In the event, the two anomalies occur together the system switches to state alarmed. In the worst case, we proceed by a forced shutdown of the system.

## 5. CONCLUSION

This paper has presented a brief introduction to the fields of FTC and FDI. An algorithm for monitoring a system was proposed by using Arduino technology.

As an emerging and active area of research in automatic control, fault-tolerant control has recently attracted more and more attention. When a fault occurs in a system, the main problem to be addressed is to raise an alarm, ideally diagnose what fault has occurred, and then decide how to deal with it. The problem of detecting a fault, finding the source/location and then taking appropriate action is the basis of fault tolerant control.

In our future work we are interested in neuro-fuzzy modeling and diagnosis.

## 6. REFERENCES

[1] L., Jiang.2011. Sensor fault detection and isolation using system dynamics identification techniques. PhD Thesis. University of Michigan, USA.

[2] M., Rodrigues, D. Theillol, M., Adama and D., Sauter. 2008. A fault detection and isolation scheme for industriel systems based on multiple operating model. Control Engineering Practice 16,2 (2008) 225-239.

[3] P., Rosa, C., Silvestre, J.S., Shamma, and M., Athans. 2010. Fault detection and isolation of an Aircraft using Set-Valued Observers. Georgia, USA.

[4] T., Escobet and L., Travé-Massuyès. 1995. Parameter estimation methods for fault detection and isolation. Campus de Terrassa, Barcelona, Spain.

[5] Y., Tharrault, G., Mourot, J., Ragot and D., Maquin. 2008. Fault detection and isolation with robust principal component analysis. Int.J.Appl.Math.Comput.Sci., 2008, Vol. 18, N°4, 429-442.

[6] M., Fliess, C., Join, and H., Sira-Ramirez. 2006. Closed-loop fault-tolerant control for uncertain nonlinear systems. Ecole plytechnique. Palaiseau, France.

[7] Y., Wang, D., Zhou, S., Joe Qin, and H., Wang. 2008. Active fault-tolerant control for a class of nonlinear systems with sensor faults. International journal of control, automation, and systems, vol.6 n°.3, (2008), pp. 339-350.

[8] Y., Diao, K.M., Passino. 2002. Intelligent fault-tolerant control using adaptive and learning methods. Control Engineering Practice 10 (2002) 801-817.

[9] J., Stroustrup, and V.,D., Blonded. 2004. A simultaneous stabilization approach to (passive) fault tolerant control. Department of control engineering institute of electronic systems, Denmark.

[10] M., Staroswiecki, M. Blanke, and N., Eva. 2001. Concepts and methods in fault-tolerant control. Book.

[11] M., Staroswiecki, M., Blank, M., Kinnaert, J., Lunze. 2006. Diagnosis and fault-tolerant control. Springer.

[12] J.,N., Montagné. 2006. Initiation à la mise en œuvre matérielle et logicielle d'Arduino. Livret en français, Centre de Ressources Art Sensitif.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 29 - 32, 2014

# An Improved Image Fusion Scheme Based on Markov Random Fields with Image Enhancement Scheme

Nalla Aravind Babu
Department of ECE
Vardhaman College of Engineering
Hyderabad, India

M. Veeraswamy
Department of ECE
Vardhaman College of engineering
Hyderabad, India

**Abstract:** Image fusion is an important field in many image processing and analysis tasks in which fusion image data are acquired from multiple sources. In this paper, we investigate the Image fusion of remote sensing images which are highly corrupted by salt and pepper noise. In our paper we propose an image fusion technique based Markov Random Field (MRF). MRF models are powerful tools to analyze image characteristics accurately and have been successfully applied to a large number of image processing applications like image segmentation, image restoration and enhancement, etc.,. To de-noise the corrupted image we propose a Decision based algorithm (DBA). DBA is a recent powerful algorithm to remove high-density Salt and Pepper noise using sheer sorting method is proposed. Previously many techniques have been proposed to image fusion. In this paper experimental results are shown our proposed Image fusion algorithm gives better performance than previous techniques.

**Keywords**: Markov Random Field, image fusion, image segmentation, image restoration and enhancement, Decision based algorithm.

## 1. INTRODUCTION

The applications were image data should be gathered (or collected) from many input images as source images, in such applications image fusion plays an important role. The source images contain about some information of the original scene but individually they cannot represent full scene perfectly in order to gather the original scene from the source images image fusion is used. The input images for fusion can be capture at different time and/or by an utilizing various sensors, because of this source images individually cannot represent full scene [1]. Thus we perform image fusion in order to get an accurate view at the scene, that's why it is more applicable in many image processing applications [2]. In applications like satellite image sensing image fusion is utilize to improve the resolution of the images [3], [4]. Here, we propose fusion scheme for source images having same resolution. The multispectral fusion scheme consists of a range of bands so as to improve spectral resolution [4]. We use several bands improving spectral resolution.

The available schemes of image fusion can be broadly classified into three categories: feature level, decision level and pixel level. In our paper we are using pixel level type of image fusion; so many techniques have been proposed of image fusion for various applications. In general there are two steps involved in pixel level image fusion.

a) We have to estimate whether given source image will contribute to the final fused image or not. For each and every pixel.

b) By using all the source images the intensity of the fused image can be formed.

The true previous techniques of fusion are by averaging pixel intensities of input images or by taking maximum pixel intensity out of total available source images pixel values. Gaussian noise can be effectively removed in case of averaging scheme and also it increases SNR value but fails to maintain proper contrast.

Recently, a name approach has been introduced i.e. MD-based scheme (multi scale-decomposition) and it shows better results for various applications of image processing. It consists of following three steps.

a) The input images are decomposed into n levels by using wavelet transform.

b) At, each and every level of source images the fusion approach is applied.

c) Convert the image spatial domain by applying inverse wavelet transform, in order to synthesize the resultant image.

This MD-based approach attains good performance and increases computational complexity became use of wavelet transform, so depending on application requirement one can select or skip the transformation step. The pixel level fusion approach may introduce distortion ine2 the resultant fused image, because of the drawback of pixel level fusion that the decision on a specific input image or not, is difficult.

This noteworthy to consider spatial co-relation while we are doing fusion and the users such model will increase the performance, one of the way to use this spatial co-relation is by the use of window based or region based method [5][6][8] .

Here in our paper we introduce MRF(Markov Random Field) as it has been used in image processing applications such as change detection[10], image de-noising [9][12], edge detection [11] and image restoration [13][14].

MRF can also be used for pixel level image fusion. The first algorithm focus on decision making as an MRF and the next algorithm both on decision making and exact image as an MRF and the contrast can be improved by finding the upper and lower band intensity values for the resultant image. Our scheme can be tested in the presence of salt and pepper noise and impulse noise as well and it shows better performance.

## 2. PROBLEMM FORMULATION

The main aim of imager fusion is to predict the specified scene, by considering that every source image has good view

of small part of scene [1]. Consider there are source images each can be represented as,

$$yi(r) = Hi\ x(r) + wi(r), \qquad i=1,\ldots\ldots,N \qquad (1)$$

Where r represents spatial coordinates, $y_i(r)$ is the brightness or intensity of *i*th image, $x(r)$ is the intensity of original scene or source scene at r which has to be predicted $w_i(r)$, we presents noise and sensor selectivity coefficient is represented by $H_i(r)$. The above equation (1) represents a relationship between actual scene and input image; in practice this model has some drawbacks. The images captured from various sensor gives different aspects of original scene in this case this scheme is an approximation.

The fusion problem necessarily include the prediction of *x* and $H_i$. Previous algorithms may also be represented by above model. In case of averaging $H_i=1$ for any *i* and for maximizing case $H_i=1$ if $i= \max_i\{y_i\}$;$H_i=0$ otherwise. If $H_i$ is known the gray scale values of fused image will be calculated by least square scheme [15] as,

$$\hat{x} = (H^T H)^{-1} H^T Y \qquad (2)$$

Where H is a vector having values $[H_1,H_2,\ldots..,H_N]^T$ and Y is a vector $[y_1,y_2,\ldots..,y_N]^T$. in general H will be available without any information about H to predict the values of H. the LS algorithm says that H will take the values which generates highest energy i.e.

$$\begin{aligned}\hat{H} &= \min_H \left\{(Y - H\hat{x})^T (Y - H\hat{x})\right\} \\ &= \min_H \{Y^T Y - (Y^T H)(H^T H)^{-1}(H^T Y)\} \\ &= \max_H \{(Y^T H)(H^T H)^{-1}(H^T Y)\} \end{aligned} \qquad (3)$$

$H_i \in \{0,1\}$, H has $2^N$ possible values, now the values *x* will be estimated by LS technique as

$$\hat{x} = (\hat{H}^T \hat{H})^{-1} \hat{H}^T Y \qquad (4)$$

The above scheme is sensitive to noise as H as well as intensity of resultant image *x* estimated pixel. The accuracy of estimation of H coefficients has much impact on resultant image accuracy. A simple way is to consider that the pixel values lie in a small window will be constant now choose coefficients which generates highest energy in that window [7]. The main aim of LS technique is to reduce the error $||y - \hat{y}||^2$, a famous scheme for enhancing the predicted error of LS is by incorporation of prior knowledge of H or *x* [16]. We presented to use an MRF model in order to predict the coefficients of H. it is predicted to enhance the accuracy of H coefficients and resultant image quality.

## 3. PROPOSED ALGORITHM

The problem in image fusion is to predict the original scene. The estimation of H is performed first before we go for *x*, which will decide whether the previous work we predict the value of H by using maximizing, LS, Averaging and Windowing scheme. Here in this session we represent the two schemes based on MRF which will consider the spatial coordinates as well. Therefore, the intensity value in the resultant image will be decided by intensity of source images and also by neighboring pixels as well. The first scheme (MRF_H) with MRF models the coefficients only, but in the case of second algorithm coefficient as well as fused image will be modeled via MRF (MRF_HX) few notations.

- X: total original scene;
- $H_i$: *i*th source image coefficients;
- $Y_i$: *i*th source image intensity;
- H: source images coefficients, where $H(r,i)=H_i$;
- Y: source images intensities, where $Y(r,i)=Y_i$

A. Image fusion: H coefficients will be modeled using MRF.

This approach is motivated by the fact that an input images coefficient has spatial correlation. We proceed with H coefficients via MRF. Assume a sites in an image be represented by 'S' and $\Lambda \in \{0,1,\ldots.,L\text{-}1\}$. The phase space with [14] the marginal pdf H can be expressed as below (5) with normalization constant $Z_H$ is given below.

$$P_H(H) = = \frac{1}{Z_H} exp\ [-\frac{1}{T}\sum_{c\subset S} Uc\,(H)] \qquad (5)$$

$$Z_H(H) = \sum_{H\in\Lambda} exp\left[-\frac{1}{T}\sum_{c\subset S} Uc\,(H)\right] \qquad (6)$$

The value can be express in (6) and the optimal value of H can be represented in (7). The steps involved in these algorithms are as follows,

1) First, find the initial estimate of H and x and decide initial parameters and a looping variable has to be initiated in looping variable.
2) For every iteration a new estimate of H will be obtained by Gibbs pdf and with Gibbs potential E(H).
3) Make obtaining the resultant fused image by using (4).
4) Reduce the value of looping variable and repeat above (2) and (3) steps till convergence.

B. Image fusion: H coefficients and x image will be modeled.

The resultant fused image also consists of properties of spatial coefficients. Therefore we assume that the resultant fused image will also followed Markov Random Field along with Gibbs potential Vc(X). Therefore X marginal pdf in [9].

$$P_X(X) = \frac{1}{Z_X} \exp\ [-\tfrac{1}{T}\sum_{c\subset S} Vc\,(X)] \qquad (7)$$

Here $Z_X$ is normalization constant and can be written as (6) by this consideration the optimal X given as,

$$\hat{X}^{n+1} = \arg\left\{\max_H \left[P\left(X|Y,\hat{H}^n\right)\right]\right\} \qquad (8)$$

Therefore the $\hat{H}$ estimation will reduces to,

$$\hat{X}^{n+1} = \arg\left\{\max_X (\Delta(X))\right\} \qquad (9)$$

Where,

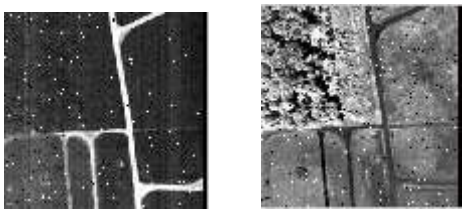
$$\Delta(X) = \frac{1}{2\sigma^2}\sum_i (Y_i - \hat{H}_i^n X)^T (Y_i - \hat{H}_i^n X) + \sum_{c\subset S} Vc\,(X). \qquad (10)$$

As compare to scheme proposed previously i.e. MRF_H in that by using LS algorithm we updated the fused image X. in this technique we use (10) to update the resultant image the algorithm is as follows.

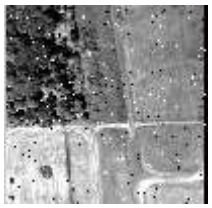
1) As a first step de-noise the input images.
2) Start with an initial estimate of H and X. Estimate the initial parameters and set the initial variable.
3) For every iteration new H vale will be obtain on the basis of its Gibbs pdf(5) with E(H) using Gibbs sampling method [9].
4) Make update the fused image by using (10).
5) Reduce the looping variable and repeating the (2 and 3) steps till convergence.
6) Finding the upper and lower band intensity values from the resultant values.

## 4. EXPERIMENTAL RESULTS

As an example we have taken three multispectral images to demonstrate our algorithm and in evaluate the fusion scheme. The fig.1 shows noisy source images, we improving quality of source images by using median filter and the enhanced images or as shown in fig.2 the different bands multispectral source images. Each image is split or decompose into four sub images by using desecrate wavelet transform method in case of windowing scheme and fusion scheme approach is employed to contribute or fuse the images and finally we have to take the inverse wavelet transform for resultant image. In fig.3 shows the fusion based results with our proposed algorithm and we present the performance resultant maximizing approach, LS approach, averaging approach, windowing based and our proposed approach. The maximizing approach is too bright resultant image. The LS approach is having drawback of showing more noise in resultant image and the resultant image produce by window based shows mosaic pattern.



(a) (b)

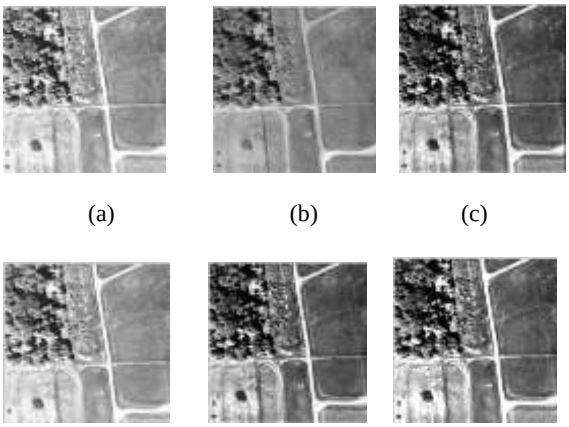


(c)

**Fig. 1: Noisy source images**

(a) (b) (c)

**Fig.2: Enhanced source images**



(a) (b) (c)

(d) (e) (f)

**Fig.3: Fusion results based on (a) maximizing algorithm (b) averaging algorithm (c) LS algorithm (d) Window algorithm (e) MRF_H algorithm (f) Proposed algorithm.**

## 5. ACKNOWLEDGMENTS

Here in this paper we have analyze the problem of image fusion on the basis of statistical model. We consider the importance of correlation and assume that it can also be modeled via MRF_H has been introduced. We presented different image fusion approaches such as LS, maximizing, averaging, windowing based and our proposed method. Which utilizes Gibbs pdf, Gibbs distribution and the results of which can be further more resultant fused image. Hear we assume a simple relation between true scene and input source image i.e. a source image may contribute or may not contribute fused image and we assume that the coefficients can take any real value in the data model. These may increase the accuracy of the image fusion.

## 6. REFERENCES

[1] Min Xu, “An image fusion approach based on markov random fields,” IEEE Trans. Geosci. Remote Sens., vol. 49, No 12, dec. 2011.

[2] R. S. Blum, “On multi-sensor image fusion performance limits from an estimation theory perspective,” Inf. Fusion, vol. 7, no. 3, pp. 250–263, Sep. 2006.

[3] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li, “A comparative analysis of image fusion methods,” IEEE Trans. Geosci. Remote Sens., vol. 43, no. 6, pp. 1391–1402, Jun. 2005.

[4] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, “Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics,” IEEE Trans. Geosci. Remote Sens., vol. 46, no. 5, pp. 1301–1312, May 2008.

[5] R. K. Sharma, T. K. Leen, and M. Pavel, “Probabilistic image sensor fusion,” in Proc. Adv. Neural Inf. Process Syst. 11, 1999, pp. 824–830.

[6] P. Burt and R. Kolczynski, “Enhanced image capture through fusion,” in Proc. 4th Int. Conf. Comput. Vis., 1993, pp.173–182.

[7] J. Yang and R. Blum, "A statistical signal processing approach to image fusion for concealed weapon detection," in Proc. IEEE Int. Conf. Image Process., 2002, pp. 513–516.

[8] E. Lallier and M. Farooq, "A real time pixel-level based image fusion via adaptive weight averaging," in Proc. 3rd Int. Conf. Inf. Fusion, 2000, pp. WEC3/3–WEC313.

[9] S. Z. Li, Markov Random Field Modeling in Computer Vision.New York: Spinger-Verlag, 2001.

[10] T. Kasetkasem and P. Varshney, "An image change detection algorithm based on Markov random field models," IEEE Trans. Geosci. Remote Sens., vol. 40, no. 8, pp. 1815–1823, Aug. 2002.

[11] Z. Tu and S. Zhu, "Image segmentation by data-driven Markov chain Monte Carlo," IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 5, pp. 657–673, May 2002.

[12] H. Chen, "Mutual information based image registration with applications," Ph.D. dissertation, Syracuse Univ., Syracuse, NY, May, 2002.

[13] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," in Readings in Uncertain Reasoning. San Francisco, CA: Morgan Kaufmann, 1990, pp. 452–472.

[14] L. Bedini, A. Tonazzini, and S. Minutoli, "Unsupervised edge-preserving image restoration via a saddle point approximation," Image Vis. Comput., vol. 17, no.11, pp.779–793, Sep.1999.

[15] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory. Upper Saddle River, NJ: Prentice-Hall, 1993.

[16] Y. C. Eldar, A. Beck, and M. Teboulle, "Bounded error estimation: A Chebyshev center approach," in Proc. 2nd IEEE Int. Workshop Compute. Adv. Multi-Sensor Adapt. Process, 2007, pp.205–208.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 33 - 38, 2014

# A NOVEL METHOD FOR THE CONSTRUCTION OF THRESHOLD MULTIPLE-SECRET VISUAL CRYPTOGRAPHIC SCHEMES BY WITHOUT PIXEL EXPANSION

P. Saranya
Sri Ramakrishna Arts and Science College for Women
Coimbatore, India

G. Sathayavathy
Sri Ramakrishna Arts and Science College for Women
Coimbatore, India

**Abstract :** The main concept of the original visual secret sharing (VSS) scheme is to encrypt a secret image into n meaningless share images. It cannot leak any information of the shared secret by any combination of the n share images except for all of images. The shared secret image can be revealed by printing the share images on transparencies and stacking the transparencies directly, so that the human visual system can recognize the shared secret image without using any devices. The visual secrets sharing scheme for multiple secrets is called multiple-secret visual cryptographic schemes (MVCSs). This paper proposed general constructions for threshold multiple-secret visual cryptographic schemes (MVCSs) that are capable of encoding s secret images. This presented MVCS schemes utilize a predefined pattern book with pixel expansion to encrypt secret images into share images. In our research, we propose a novel MVCS scheme that can share two binary secret images on two rectangular share images with no pixel expansion, but also has an excellent recovery quality for the secret images.

**Keywords:** Cryptography, Visual Cryptography, Visual Threshold, Encryption

## 1. INTRODUCTION

Visual cryptography is a cryptographic technique which allows visual information (pictures, text, etc.) to be encrypted in such a way that decryption becomes a mechanical operation that does not require a computer [1]. Visual Cryptography (VC) [2, 3, 4, 5, 6] is a variation of the conventional secret sharing scheme. In VC, instead of a numerical secret key, a secret image is shared among participants in the form of images called shares. Each participant possesses his own share which cannot reveal the secret image being alone, making it necessary to stack more than one share of a qualified participant in order to reveal the secret image. Thus in VC the stacking of shares is equivalent to the decryption process, where neither extra computations nor previous knowledge are required to reveal the secret image. Until now some important VC schemes, such as the *(k,n)*-VC scheme, the general access structure for VC and the extended VC (EVC)[7, 8 ,9, 10, 11, 12] have been proposed. Unfortunately all schemes can be cheated, if one or more participants try to generate their fake shares to force the revealed secret image to be a faked one. In this paper, we propose a cheating prevention VC scheme, in which the shares can be identified and authenticated using the EVC scheme and watermarking techniques. Using a similar idea, transparencies can be used to implement a one-time pad encryption, where one transparency is a shared random pad, and another transparency acts as the cipher text.

In the proposed VC scheme, the share of each participant can be identified by its meaningful appearance instead of noise-like image used in the conventional VC scheme. For the purpose of authentication of each share two binary watermark images are encrypted using shift operation. Before the secret image is revealed, the validation of the shares must be carried out, extracting two watermark images. If they can be extracted correctly, the revealed secret image is considered as authentic; otherwise it is determined as a faked one.

1. VCS is a kind of secret sharing scheme that focuses on sharing secret images. The idea of the visual cryptography model is to split a secret image into two random shares (printed on transparencies) which separately reveals no information about the secret image other than the size of the secret image.

2. The secret image can be reconstructed by stacking the two shares. The underlying operation of this scheme is logical operation OR.

3. In general, a traditional VCS takes a secret image as input, and outputs shares that satisfy two conditions: 1) any qualified subset of shares can recover the secret image; 2) any forbidden subset of shares cannot obtain any information of the secret image other than the size of the secret image

Visual cryptography was possible to devise a secret sharing scheme in which an image can reconstructed "visually" by superimposing two shares? Each share would consist of a transparency, made up of black and white pixels. (Note that it would be more accurate to say "transparent" rather than "white".) Examination of one share should reveal no information about the image[13, 13, 15].

Naor and Shamir devised the following scheme, illustrated in the figure below. The algorithm specifies how to encode a single pixel, and it would be applied for every pixel in the image to be shared.

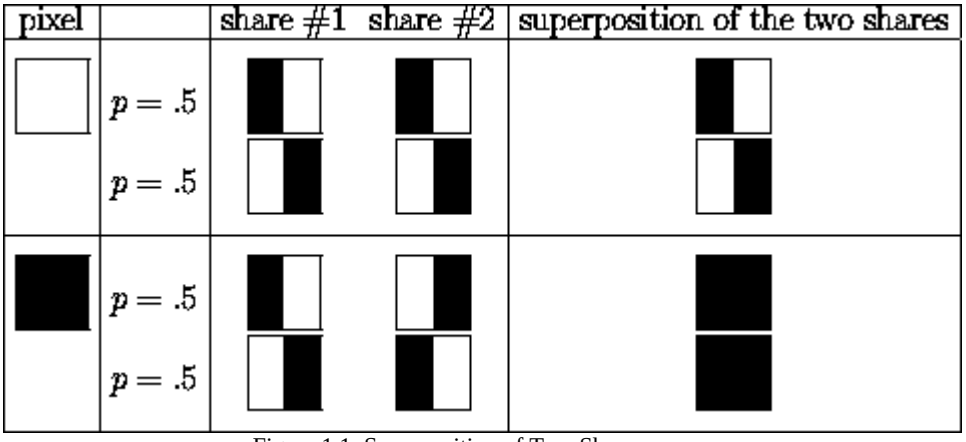



Figure 1.1: Superposition of Two Shares

A pixel P is split into two sub pixels in each of the two shares [18]. If P is white, then a coin toss is used to randomly choose one of the first two rows in the figure 1.1. If P is black, then a coin toss is used to randomly choose one of the last two rows in the figure 1.1. Then the pixel P is encrypted as two sub pixels in each of the two shares, as determined by the chosen row in the figure 1.1. Every pixel is encrypted using a new coin toss.

Suppose we look at a pixel P in the first share. One of the two sub pixels in P is black and the other is white. Moreover, each of the two possibilities "black-white" and "white-black" is equally likely to occur, independent of whether the corresponding pixel in the secret image is black or white. Thus the first share gives no clue as to whether the pixel is black or white. The same argument applies to the second share. Since all the pixels in the secret image were encrypted using independent random coin flips, there is no information to be gained by looking at any group of pixels on a share, either. This demonstrates the security of the scheme.

Now let's consider what happens when we superimpose the two shares (here we refer to the last column of the figure). Consider one pixel P in the image. If P is black, then we get two black sub pixels when we superimpose the two shares; if P is white, then we get one black subpixel and one white subpixel when we superimpose the two shares. Thus, we could say that the reconstructed pixel (consisting of two subpixels) has a grey level of 1 if P is black and a grey level of 1/2 if P is white. There will be a 50% loss of contrast in the reconstructed image, but it should still be visible.

The main focus of this research is to use the visual secret sharing as the part of the cryptography system for getting more the confidentiality, reliability and integrity. With this knowledge, the methods supports for this focus of the research has been considered and dealt.

The main objective of this research is to generate shares of secret image by using no pixel expansion strategy. Then transmit the two secret images with the use of two shares. With stacking two shares, secret image I appear and with stacking one of the shares with 90 degrees rotation in clockwise on other share appears the secret image

## 2. LITEARTURE SURVEY

Visual Cryptography is a special encryption technique to hide information in images in such a way that it can be decrypted by the human vision if the correct key image is used. The technique was proposed by Naor and Shamir in 1996[ 3, 4, 5, 6, 7,8]. Visual Cryptography uses two transparent images. One image contains random pixels and the other image contains the secret information. It is impossible to retrieve the secret information from one of the images. Both transparent images and layers are required to reveal the information. The easiest way to implement Visual Cryptography is to print the two layers onto a transparent sheet. In the overlay animation you can observe the two layers sliding over each other until they are correctly aligned and the hidden information appears.

The paper [10, 15] proposed the construction of basis matrices of visual secret sharing schemes for color images under the ($t$, $n$)-threshold access structure, where $n \geq t \geq 2$ are arbitrary integers. They treat colors as elements of a bounded semi lattice and regard stacking two colors as the join of the two corresponding elements. They generate $n$ shares from a secret image with $K$ colors by using $K$ matrices called basis matrices. The basis matrices considered in this paper belong to a class of matrices each element of which is represented by a homogeneous polynomial of degree $n$. They first clarify a condition such that the $K$ matrices corresponding to $K$ homogeneous polynomials become basis matrices. Next, they give an algebraic scheme for the construction of basis matrices. It is shown that under the ($t$, $n$)-threshold access structure they can obtain $K$ basis matrices from appropriately chosen $K - 1$ homogeneous polynomials of degree $n$ by using simple algebraic operations. In particular, they give basis matrices that are unknown so far for the cases of $t = 2$, 3 and $n - 1$.

The paper [16, 17, 18] proposed Visual cryptography was introduced by Naor and Shamir. It is a new cryptographic paradigm that enables a secret image to be split into n shares, each share being printed on a transparency. The shares are distributed among n participants of whom only some are qualified to recover the original image. The secret image is reconstructed by

stacking a certain number k (2k,n) of these transparencies from the set of qualified participants. If fewer than k transparencies are superimposed, then it is impossible to decode the original image. The resulting cryptographic scheme is called a (k;n) visual threshold scheme (VTS). Since the reconstruction is done by the human visual system, no computations are involved during decoding unlike traditional cryptographic schemes where a fair amount of computation is needed to reconstruct the plain text. It is conceivable that the encryption strategy is such that the color ratios of the different pixels in the reconstructed image are different. In this case, they could de ne the color ratio of a scheme to be the minimum value of the ratio defined, the minimum being taken over all possible different colored pixels. On the other hand, the encryption strategy could be so regular that each pixel, irrespective of its color, has the same color ratio. In this case, they need not de ne the color ratio separately for each pixel of the reconstructed image. In such a case, if the color ratio of each pixel is R, they will say that the encoding scheme attains a color ratio R.

The paper [20, 21, 22, 23] A (k; n)-threshold visual cryptography scheme ((k; n)-threshold VCS, for short) is a method to encode a secret image SI into n shadow images called shares such that any k or more shares enable the "visual" recovery of the secret image, but by inspecting less that k share one cannot gain any information on the secret image. The "visual" recovery consists of Xeroxing the shares onto transparencies, and then stacking them. Any k shares will reveal the secret image without any cryptographic computation. In this paper they analyze the contrast of the reconstructed image for (k; n) - threshold VCS. They define a canonical form for (k; n)-threshold VCS and they also provide a characterization of (k; n)-threshold VCS. They completely characterize contrast optimal (n \Gamma 1; n)-threshold VCS in canonical form. Moreover, for n 4, they provide, a contrast optimal (3; n)-threshold VCS in canonical form. They generically transform any *k*out-of-*n* threshold VCS for black-and-white images to color images. During the transformation, they use a probabilistic technique for achieving no pixel expansion. In addition, they also allow the user of the VCS to choose the number of colors that the reconstructed image will have.

## 3. PROBLEM DEFINITION

A cheating process against a VCS consists of the following two phases:

1. Fake share construction phase: the cheater generates the fake shares;
2. Image reconstruction phase: the fake image appears on the stacking of genuine shares and fake shares.

In order to cheat successfully, honest participants who present their shares for recovering the secret image should not be able to distinguish fake shares from genuine shares. A reconstructed image is perfect black if the sub pixels associated to a black pixel of the secret image are all black.

- More memory space is needed when generating shares by expanding secret image pixels
- Network traffic can be increased when sending shares of secret image over network

### 3.1 Problem Specification of (k,n,s)-MVCS

Essentially, a (k,n,s) MVCS [25] is capable of encoding secret Images P1,P2,P3…..Ps into shares S1,S2,…Sn which are distributed to the participant in P={1,2…..n} such that each group of k,k+1,….,n shares reveals P1,P2,P3…..Ps respectively, to our eyes when superimposed, but that of less than shares cannot. Here, s=n-k+1 and the cases of s≤n-k+1 can also be generated.

**Definition 1**

To reveal (or conceal) one pixel ,p∈P, either white (0) or black (1) with only two possibilities, in the superimposed result of k (or less than k) shares, the (k,n) -VCS aims at the design of a set of two basis matrices $B^0$ and $B^1$ for encoding P=0 or 1, respectively. Now, a set of *corresponding* pixels (p1,p2…ps) , either 0 or 1 for each with totally $2^s$ possibilities, should be considered in a (k,n,s) -MVCS where pixels ,p1∈ $P1, p2 \in P2$,..ps∈Ps are regarded as corresponding to each other if their positions in the secret images are all the same. To reveal (or conceal) these corresponding pixels (p1,p2…ps) in the superimposed results of k.k+1,…..n(or less than ) shares, respectively, the (k,n,s)-MVCS should rely on a set of nXm $2^s$ basis matrices $B^{00..0}$,$B^{00..1}$,$B^{11..1}$ to encode (p1,p2…ps) =(0,0,..0),(0,0,…1),…,(1,1,…1) , respectively, into m subpixels for each of the n shares.
Assume U={i1,i2….$i_u$} and V={j1,j2….$j_v$} Where U,V⊆ {1,2…n}, and $1 \leq v < k \leq u \leq n$. Let $B^{(p1,p2\ldots ps)}[U]$( $B^{(p1,p2\ldots ps)}[V]$) denote the u× $m(v \times m)$ matrix consists of rows i1,i2….$i_u$ {j1,j2….$j_v$} in $B^{(p1,p2\ldots ps)}$.

**Definiton 2**

A set of $2_s$ n× $m$ Boolean basis matrices $B^{0..00}$,$B^{0..01}$..,$B^{1\ldots11}$ , in which the result of a column permutation of $B^{(p1,p2\ldots ps)}$ defines the color of the m subpixels in each one of the shares when sharing corresponding pixels (p1,p2…ps) $\in \{0,1\}$ in p1,p2…ps, respectively, constitutes a (k,n,s)-MVCS where s=n-k+1 if the following conditions are met:
1) For each set of t+(k-1) participants U⊆ Pwhere 1<T<S And k<|U|(=t+(k-1))<n ,
2)For each set of less than k participants V⊆ P(1≤ $|V| \leq k-1$), $H(B_V^{p1,p2\ldots ps})$= $H(B_V^{p1',p2'\ldots ps'})$ , for p1,p2…ps ≠ $p1', p2' \ldots ps'$
where $p_i,p_j'$ ∈{0,1} and 1≤i,j≤s.

## 4. PROPOSED SYSTEM

In this research, we have proposed a novel MVCS scheme for sharing two binary secret images in two share images $S_1$ and $S_2$ with no pixel expansion, and have achieved an excellent recovery quality for the revealed secret images. During the encrypting process, the proposed scheme generated share images without any pre-defined pattern books. This process was different from any existing MVCS schemes. By directly stacking the two share images, $S_1$ and $S_2$ , the first secret $SE_1$ could be revealed and be recognized by the human visual

system, and the second secret $SE_2$ could be revealed by stacking one share image and the other with a rotation angle of 180 degree. Neither of the two share images leaked any information of the two secret images. Our proposed MVCS scheme has resolved the pixel expansion problem existing in MVCS schemes, whether sharing one or multiple secrets, and has increased the contrast quality of the revealed secret image by adopting the appropriate encrypting process.

## 4.2 Proposed Algorithm

Through the DSP, SP, and CMP processes, two secret images were encrypted into two share images. The two share images camouflaged by the proposed camouflaging process with maximum block density were meaningless images. From the above description, a patterns book was not adopted during the process for generating the two share images. All processes were executed block by block for the initially generated share images with h × w size. So, two secret images with h × w size were encrypted into two share images with h × w size. The size of the share image was equal to the size of the secret image. The critical pixel expansion of the visual multiple secrets sharing scheme was solved by our proposed scheme. A complete algorithm for sharing two secrets is shown as follows:

**Algorithm 1: Encrypting process of MSCV with no pixel expansion**

Input: Two h × w secret images $SE_1$, $SE_2$, block size: n × n, threshold of block density: $d_{TH}$

Output: Two h × w share images $S_1$, $S_2$

Step 1: b p;k i;j $b_{i,j}^{p,k} \leftarrow 0$; ∀i; j; k; p.

Step 2: Randomly generate 4 matrixes $C^1$, C2, C3, C4 by the conditions

$$a_{ij}^{1,k} = c_{ij}^1 + c_{ij}^2,$$
$$a_{ij}^{2,k} = c_{ij}^3 + c_{ij}^4,$$
$$|H(C^1) - H(C^2)| \leq 1 \text{ and } |H(C^3) - H(C^4)| \leq 1, \quad \forall k.$$

Step3: Let $b_{i,j}^{1,k} \leftarrow c_{ij}^1 \mathrm{v} c_{ij}^3$, $b_{i,j}^{2,k} \leftarrow b_{i,j}^{2,k} \mathrm{v} c_{ij}^2$ and $b_{n+1-i,n+1-j}^{2,k+1-k} \leftarrow b_{n+1-i,n+1-j}^{2,k+1-k} \mathrm{v} c_{ij}^4, \forall k$

Step 4: if $H(F^{p,k})/n^2 \geq dTH$

Then $d^{p,k} \leftarrow H(B^{p,k})/H(F^{p,k})$ else $d^{p,k} \leftarrow 0, \forall$ k,p

Step 5: $d_{max}^{p} \leftarrow \max d^{p,k}, \forall p,k; d \leftarrow \max d_{max}^{p}, \forall p$

Step 6: $d^{p,k} \leftarrow d_{max}^{p}, \forall k, p$

Step 7: k←1

Step 8: Repeat

Step 8.1: $r_w^{p,k} \leftarrow \lfloor \bar{d}^{p,k}(n^2 - H(F^{p,k})) \rfloor$,
$r_w^{p,K+1-k} \leftarrow \lfloor \bar{d}^{p,K+1-k}(n^2 - H(F^{p,K+1-k})) \rfloor, \forall p$

Step 8.2: $r_w \leftarrow \lfloor n^2 . \bar{d} \rfloor$

Step 8.3: Randomly generate a camouflaging matrix $C^1$with n × n by the condition

$H(C^1)=r_w$

Step 8.4: Repeat

Step 8.4.1: Randomly select one element $C_{i,j}^1 \neq 0$ from $C^1$

Step 8.4.2: if $r_w^{p,k} > 0$ and $f_{i,j}^{p,k}=0$ then $b_{i,j}^{p,k} \leftarrow 1, r_w^{p,k} \leftarrow r_w^{p,k} - 1, \forall p$

Step 8.4.3: if $r_w^{p,K+1-k}>0$ and $f_{n+1-i,n+1-j}^{p,K+1-k}=0$ then $b_{n+1-i,n+1-j}^{p,K+1-k} \leftarrow 1, r_w^{p,K+1-k} \leftarrow r_w^{p,K+1-k}-1, \forall p$

Step 8.4.4: $c_{i,j}^1 \leftarrow 0, r_w \leftarrow r_w - 1$

Step 8.4.5: until $r_w=0$

Step 8.5: k ←k +1.

Step 8.6: until k>K/2

Step 9: Repeat

Step 9.1: Randomly select one element $b_{i,j}^{p,k}=0$ if $r_w^{p,k} > 0$ and $f_{i,j}^{p,k}=0$ then $b_{i,j}^{p,k} \leftarrow 1, r_w^{p,k} \leftarrow r_w^{p,k} - 1$

Step 9.2: until $r_w^{p,k}=0 \forall$k,p

Step 10: $r_b^{p,k} \leftarrow \lfloor \bar{d}^{p,k} \cdot H(F^{p,k}) - H(B^{p,k}) \rfloor$.

Step 11: Repeat

Step 11.1: Randomly select one element $b_{i,j}^{p,k}=0$,if $r_b^{p,k} > 0$ and $f_{i,j}^{p,k}=1$

Then $b_{i,j}^{p,k} \leftarrow 1, r_b^{p,k} \leftarrow r_b^{p,k} - 1$

Step 11.2: until $r_b^{p,k} \leq 0, \forall k, p$

Step 12: Output share images $S_1$, S2 by $B^{p,k}$

# 5. RESULTS AND DISCUSSION

Applications of this technology is very vast with respect to the security for Possible applications are paper trail on electronic voting which shown by Chaum and encryption of financial documents shown by Hawkes and many other application is possible for example the computer logon by using the one share is in user USB drive and one share is with computer and when user insert the USB computer stacked the images and use image analysis to retrieve the password and allow to logon.

### 5.1 Comparison Graph Based On CPU Execution Time for Encryption of Images

The experimental result on CPU time can be obtained for proposed system over existing system is shown in the following chart:

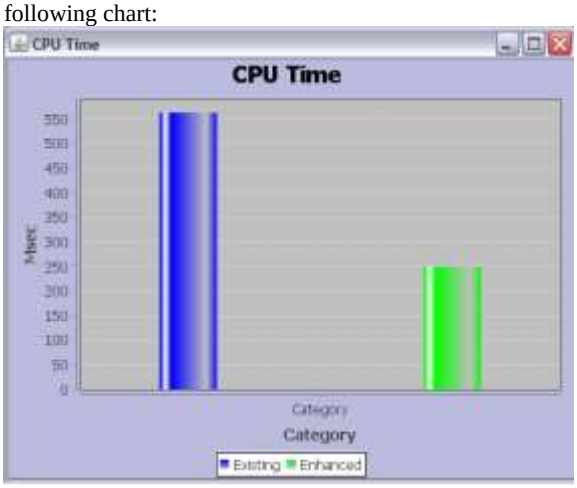


Fig 5.1 CPU Time Comparison for Existing System Vs Proposed System

### 5.2 Comparison Graph Based On Pixel Size of Encrypted Shares of Source Images

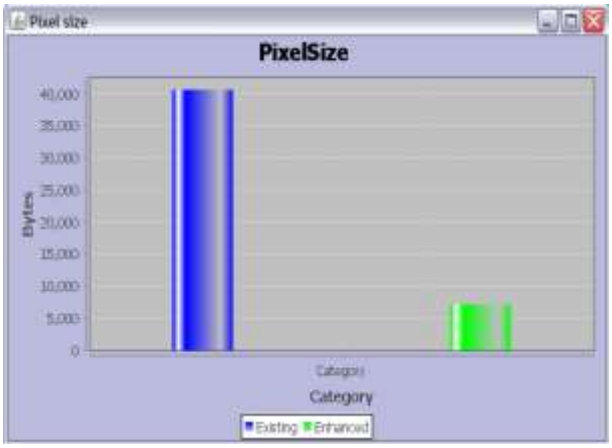



Fig 5.2 pixel size comparison for existing system Vs proposed system

The above graph shows the pixel size for existing and proposed research. The existing system uses pixel expansion so the result returns largest pixel size. Thus the proposed system works on the no pixel expansion and thus returns the smallest pixel size in the result.

## 6. CONCLUSION & FUTURE ENCHANCEMENT

Visual Cryptography is an image encryption technique used to hide the secure information in images. It allows the encryption of secret image into n number of shares and distributed into n number of participants. For example in (k, n) secret sharing problem the secret image can be visually recover by stacking together any k or more transparencies of the shares. But cannot reveal any secrete information by stacking less than k transparencies together. A novel MVCS scheme that can share two binary secret images on two rectangular share images with no pixel expansion, but also has an excellent recovery quality for the secret images.

### FUTURE WORK

In future we can propose a scheme to hide some extra confidential data in transparencies during secret image encryption in visual cryptography. The secret image is multitude into several levels first. An extended non-expansion visual secret sharing model is employed, i.e. size of transparencies is equal to that of the secret image. Thus less time and space are needed for transparencies transmission and storage.

## 7. REFERENCES

[1] Ateniese.G, Blundo.C, De. Santis.A, and Stinson.D.R, "Extended capabilities for visual cryptography," *Theoretical Computer Sci.*, vol. 250, pp. 143–161, 2001.

[2] Ateniese.G, Blundo.C, De. Santis.A, and Stinson.D.R, "Visual cryptography for general access structures," *Inf. Computat.*, vol. 129, pp. 86–106, 1996.

[3] Ateniese.G, Blundo.C, De. Santis.A, and Stinson.D.R, "Constructions and bounds for visual cryptography," *Lecture Notes Computer Sci.*, vol. 1099, pp. 416–428, 1996.

[4] Blakley.G.R, "Safeguarding cryptographic keys," in *Proc. Nat. Computer Conf.*, 1979, vol. 48, pp. 313–317.

[5] Blundo.C, De. Santis.A, and Stinson.D.R, "On the contrast in visual cryptography," *J. Cryptology*, vol. 12, pp. 261–289, 1999.

[6] Blundo.C, D'Arco.P, De. Santis.A, and Stinson.D.R, "Contrast optimal threshold visual cryptography schemes," *SIAM J. Discrete Math.*, vol. 16, pp. 224–261, 2003.

[7] Blundo.C, Cimato.S, and De Santis.A, "Visual cryptography schemes with optimal pixel expansion," *Theoretical Computer Sci.*, vol. 369, pp. 169–182, 2006.

[8] Blundo.C, DeBonis.A, and DeSantis.A,"Improvedschemesfor visual cryptography," *Designs, Codes and Cryptography*, vol. 24, pp. 255–278, 2001.

[9] Bose.M and Mukerjee.R, "Optimal visual cryptographic schemes for general ," *Designs, Codes and Cryptography*, vol. 55, pp. 19–35, 2010.

[10] Climato.S, R. D. Prisco, and A. De. Santis, "Optimal colored threshold visual cryptography schemes," *Designs, Codes and Cryptography*, vol. 35, pp. 311–335, 2005.

[11] Chen.T.H and Tsai.D.S, "Owner-customer right protection mechanism using a watermarking scheme and a watermarking protocol," Pattern Recognit., vol. 39, pp. 1530–1541, 2006.

[12] Droste.S, "New results on visual cryptography," *Advances in Cryptog- raphy-CRYPTO'96, Lecture Notes in Computer Science*, vol. 1109, pp. 401–415, 1996.

[13] Eisen.P.A and D. R. Stinson, "Threshold visual cryptography schemes with specified whiteness levels of reconstructed pixels," *Designs, Codes and Cryptography*, vol. 25, pp. 15–61, 2002.

[14] Embedded Extended Visual Cryptography Schemes F Liu… - … Forensics and Security, IEEE Transactions on, 2011 - ieeexplore.ieee.org.

[15] Free Software Foundation, lp_solve Reference Guide Menu [Online]. Available: http://lpsolve.sourceforge.net/5.5, since Feb. 1999.

[16] Hofmeister.T, M. Krause, and H. U. Simon, "Contrast-optimal Out of secret sharing schemes in visual cryptography," *Lecture Notes in Computer Sci.*, vol. 1276, pp. 176–185, 1997.

[17] Jin.D, Yan.W.Q, and Kankanhalli.M.S, "Progressive color visual cryptography," Electron.J. Imag., vol. 14, no. 3, p. 033019, 2005.

[18] Lin.C.C and Tsai.W.H, "Visual cryptography for graylevel images by dithering techniques," Pattern Recognit. Lett., vol. 24, no. 1-3, pp. 349–358, 2003.

[19] MacPherson.L.A, "Grey Level Visual Cryptography for General Access Structures,"Master Thesis, University ofWaterloo,Waterloo, ON, Canada, 2002.

[20] Nakajima.M and Yamaguchi.Y, "Extended visual cryptography for natural images," in Proc. WSCG Conf. 2002, 2002, pp. 303–412.

[21] Naor.M and Pinkas.B, "Visual authentication and identification," in Proc. CRYPTO'97, 1997, vol. 1294, pp. 322– 336, Springer-Verlag LNCS.

[22] Naor.M and Shamir.A, "Visual cryptography," in Proc. OCRYPT' 94, Berlin, Germany, 1995, vol. 950, pp. 1–12,Springer-Verlag, LNCS.

[23] Tuyls.P, Kevenaar.T, Schrijen.G.J, Staring.T, and Van Dijk.M,"Security displays enabling secure communications," in Proc. First Int. Conf. Pervasive Computing, Boppard Germany, Springer-Verlag Berlin LNCS, 2004, vol. 2802, pp. 271–284.

[24] Tsai.D.S, Chenc.T, and Horng.G, "On generating meaningful shares in visual secret sharing scheme," Imag. Sci. J., vol. 56, pp. 49–55, 2008.

[25] Wang.Z.M, Arce.G.R, and Di Crescenzo.G, "Halftone visual cryptography via error diffusion," IEEE Trans. Inf. Forensics Security, vol.4, no. 3, pp. 383–396, Sep. 2009 .

[26] Zhou.Z, Arce.Z.R, and Di Crescenzo.G, "Halftone visual cryptography,"IEEE Trans. Image Process., vol. 15, no. 8, pp. 2441–2453,Aug. 2006.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 39 - 41, 2014

# Detecting and Preventing Data Overloading Difficulty using Concept Hierarchies

Vijayakumar R
Department of Computer Science and Engineering
Muthayammal Engineering College, Namakkal
Tamilnadu, India

Rajasekar G
Department of Computer Science and Engineering
Muthayammal Engineering College, Namakkal
Tamilnadu, India

Raja Rajeswari B
Department of Computer Science and Engineering
Muthayammal Engineering College, Namakkal
Tamilnadu, India

Raja S
Department of Computer Science and Engineering
Muthayammal Engineering College, Namakkal
Tamilnadu, India

**Abstract**: Search queries on large databases, often return a large number of results, only a small subset of which is relevant to the user. When the user want to search the result for a particular query he or she find lot of difficulties when query results are large in size. To overcome the searching and navigation difficulty the following contributions are made. First, design very good user interface to search the query using front end tools like ASP.NET and it will fetch the result from database like SQL SERVER 2005. Second, Query results are organized into a tree format using tree control. Third, Ranking concept is used to display the concepts in order based on more number of times that concept is accessed. Fourth, Edge cut algorithm is used to display the query result mostly related to the user expected results in tree format. The advantage of this proposed work is minimizing navigation cost, provided good user interface to search the query and time consuming is very less. Ranking and categorization, which can also be combined, have been proposed to alleviate this information overload problem.

**Keywords**: Data mining, MeSH, Edge cut algorithm, concept hierarchy, Query results.

## 1. INTRODUCTION

Data mining is the process of extracting knowledge from large amount of data stored in database, datawarehouse or other repositories. Concept hierarchy is defined as recursively reduce the data by replacing low level concepts(such as numeric values for age) by higher level concepts(such as young, middle-aged ,or senior) [1].

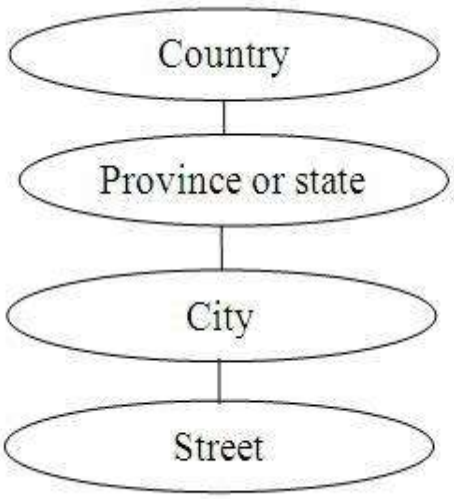



Fig: 1 Concept hierarchy

Some hierarchies can be automatically generated based on the analysis of the number of distinct values per attribute in the data set. The attribute with the most distinct values is placed at the lowest level of the hierarchy. Exceptions e.g., weekday, month, quarter, year.

### 1.1 MeSH

Medical Subject Headings (MeSH) is a comprehensive controlled vocabulary for the purpose of indexing journal articles and books in the life sciences; it can also serve as a thesaurus that facilitates searching. Created and updated by the United States National Library of Medicine (NLM), it is used by the MEDLINE/PubMed article database and by NLM's catalog of book holdings. MeSH can be browsed and downloaded free of charge on the Internet through PubMed. The yearly printed version was discontinued in 2007 and MeSH is now available online only [2] [3].Originally in English, MeSH has been translated into numerous other languages and allows retrieval of documents from different languages.

### 1.2 FEATURES

This MeSH mechanisms is contains following list of features.

1. MeSH is used on MEDLINE to index bibliographic citations and author abstracts from over 4,000 journals published in the United States and in 70 foreign countries (though mainly to English language papers). MEDLINE provides citations and, where available, astracts and links to full-text articles.

2. PubMed is a Web-based retrieval system developed by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine. It is part of NCBI's retrieval system called Entrez. MeSH vocabulary is used for indexing journal articles for Index Medicus® and MEDLINE and is also used for cataloging books.

3. MeSH terms are arranged in a hierarchy of "MeSH Tree Structures". When PubMed searches a MeSH term, it will automatically include narrower terms in the search, if applicable. This is also called "automatic explosion." NLM indexers examine articles and assign the most specific MeSH heading(s) that appropriately describes the concept(s) discussed. As many as 15 headings may be assigned Automatic Term Mapping feature to search for unqualified terms. When you click Go, PubMed will look for a match in up to four lists. It looks first for a match in the MeSH Translation Table. If it doesn't find a match, it looks in the Journals Translation Table, then in the Phrase List, and finally in the Author Index [2].

## 2. EXISTING SYSTEM

In the existing system searching is a static navigation of the database information. Here the understanding of the information is very difficult. So the categorization of the information is also a difficult task to the programmer.

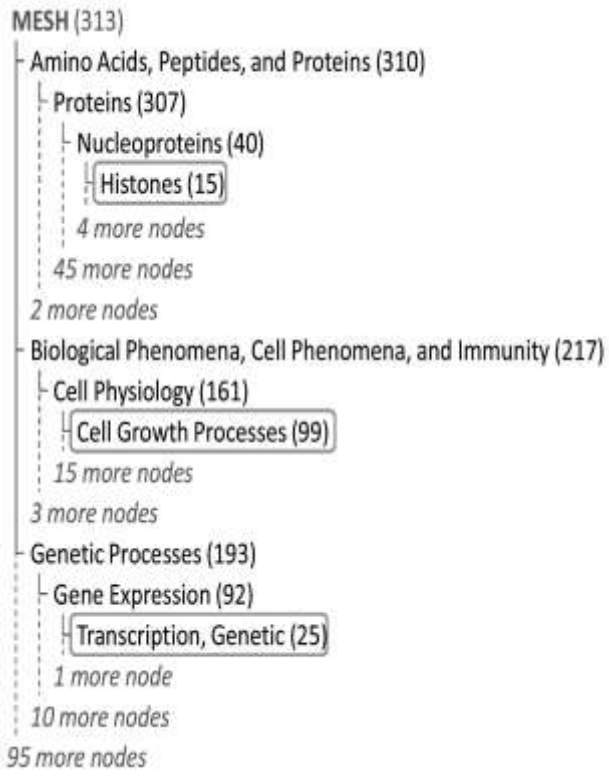



Fig 2: Static navigation on the MeSH concept hierarchy

An intuitive way to categorize the results of a query on database is by using the MeSH static concept hierarchy to build and maintain such a comprehensive structure. Each citation is associated with several MeSH concepts in two ways: 1) by being explicitly annotated with them, and 2) by mentioning those in their text.

## 3. PROPOSED SYSTEM

In addition to the static hierarchy navigation works mentioned above, there are works on dynamic categorization of query, which create unsupervised query-dependent results clusters, but do not study how the clusters should be navigated. When the user want to search the result for a particular query he or she find lot of difficulties when query results are large in size. The advantage of this proposed work is minimizing navigation cost, provided good user interface to search the query and time consuming is very less.

Search queries on databases, often return a large number of results, only a small subset of which is relevant to the user. Ranking and categorization, which can also be combined, have been proposed to alleviate this information overload problem. To overcome the searching and navigation difficulty the following contributions are made. First, design very good user interface to search the query using front end tools like ASP.NET and it will fetch the result from database like SQL SERVER 2005. Second, Query results are organized into a tree format using tree control. Third, Ranking concept is used to display the concepts in order based on more number of times that concept is accessed. Fourth, Edge cut algorithm is used to display the query result mostly related to the user expected results in tree format.

The advantage of this proposed work is minimizing navigation cost, provided good user interface to search the query and time consuming is very less. Ranking and categorization, which can also be combined, have been proposed to alleviate this information overload problem.

## 4. CONCLUSION AND FUTURE WORK

Database is created to store more number of data. Front end is designed for effective navigation of query results. Tree navigation is used to display the query result in tree format. Navigation model is used to navigate the query results based on user interests. Users can expand or backtrack, based on their interests. Ranking algorithm concept is used to display the most accessed concept in first of the tree. It is very useful for every user. Best edge cut algorithm is used to get the particular set of query results from the database. Finally, this project is designed to support more number of data.

At present, ranking is based on number of times the root node is accessed. As help of these rankings, expand more number of sub-roots will be considered at future. The challenge is to implement the databases like products, college, schools and etc. Because now I have used books and authors database only.

## 5. REFERENCES

[1] Abhijith kashyap, Vageli Hritidis, Michalis Petropoulos, and Sotiria Tavoulari. (2011)'Effective navigation of query Results based on concept hierarchies', IEEE transaction on knowledge and engineering, VOL.23, NO.4.

[2] Agrawal, J.S, Chaudhuri, S, Das, G. and Gionis, A.(2003),'Automated Ranking of Database Query Results', Proc. First Biennial Conf. Innovative Data Systems Research.

[3] Chakrabarti, K, Chaudhuri, S. and Hwang, S.W.(2004) 'Automatic Categorization of Query Results', Proc. ACM SIGMOD, pp. 755- 766.

[4] Chen, Z. and Li, T. (2007) 'Addressing Diverse User Preferences in SQLQuery- Result Navigation', Proc. ACM SIGMOD, pp. 641-652.

[5] Delfs, R, Doms, A, Kozlenkov, A. and Schroeder, M. (2004)'GoPubMed: Ontology-Based Literature Search Applied to Gene Ontology and PubMed', Proc. German Conf. Bioinformatics, pp. 169-178.

[6] Feige, U, Peleg, D. and Kortsarz, G. (2001),'The Dense k-Subgraph Problem', Algorithmica, vol. 29, pp. 410-421.

[7] Hoffman, R. and Valencia, A.(2004)'A Gene Network for Navigating the Literature', Nature Genetics, vol. 36, no. 7, p. 664.

[8] Hristidis, V. and Papakonstantinou, Y.(2002) 'DISCOVER: Keyword Search in Relational Databases', Proc. Int'l Conf. Very Large Databases (VLDB).

[9] Shatkay, H. and Feldman, R. (2003) 'Mining the Biomedical Literature in the Genomic Era: An Overview', J. Computational Biology, vol. 10,no. 6, pp. 821-855.

[10] Zhang, T, Ramakrishnan, R. and Livny, M.(1996) 'BIRCH: An Efficient Data Clustering Method for Very Large Databases', Proc. ACM SIGMOD, pp. 103-114.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 42 - 44, 2014

# Analysis of Pattern of Information Revelation and Site Use Behavior in Social Networking Sites

Sanjeev Dhawan
University Institute of Engineering and Technology
Kurukshetra University
Haryana, India

Shivi Goel
University Institute of Engineering and Technology
Kurukshetra University
Haryana, India

**Abstract**: With the emergence of online social networking sites, the rules of social interaction and communication has been changed. Most of the social networking sites motivate users to share personal information, build new relationships and increase knowledge with a perceived ease of use. But this online interaction and sharing of personal information on online sites have raised many new privacy concerns as it provide huge amount of data to third party which can be misused by malicious activities against the user's will. This research aims to develop an understanding of individual's risk taking behavior online particularly around the issue of information disclosure and to study the factors influencing the site use behavior. Human behavior concern has crucial role in development of social networking sites. Using an online questionnaire, survey data has been collected from social networking site users of different age groups and gender. It has been statistically shown that although the privacy concerns of respondents are significant but their attitude towards the risks of information disclosure is still very relaxed. People are still not aware about the actual cyber risks and therefore still happy to disregard protective advice and have risk taking behavior. Further some actual flaws and changes that user's wishes to see in online social networking sites are mentioned which are also collected from the actual online interaction with the users.

**Keywords**: Awareness, Information revelation, Privacy concern, Social networking sites (SNS), Social engineering.

## 1. INTRODUCTION

In 1954 J.A. Barnes coined the term 'social Networking' in references to a map of the relationships between individuals, indicating the ways in which they are connected through various social familiarities ranging from casual acquaintances to close familial bonds. After half a century and with the help of internet the basic concept of these social networks has completely changed. In recent years several attractive, user friendly and cost effective facilities has been introduced in online society in which social networking sites are very popular where an extensive participation of people has been seen. Boyd and Ellison defined social networking sites as web-based services that allow individuals to "(1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system" [1]. Social networking sites provide a platform to stay connected with friends, family, to develop new connections and share personal information and experience.

Social networking methods were came into focus by the end of 1990's. The new generation of social networking began to gain momentum with the surfacing of Friendster in 2002 followed by Myspace and LinkedIn a year later. Facebook launched in 2004, and now become the most popular site in the world [2]. According to rough statistics as on September 2013 [3], Facebook have 727 million daily active users on average. The phenomenon of online social interaction is more than teenagers looking to expand their friend list or to gain popularity. People of all age group and demographics are engaged in enriching their lives through the contacts they make on social networking sites. Even it provides a wide platform for marketing and advertisement agencies to promote their products.

As users can share almost anything with anybody on online social networking sites, social networking sites provide many privacy policies to control and customize what their information is available to other users. However it has been shown [4] [5] that such measures are not enough to protect one's sensitive data. Social ties represented as network data can be demoralized by the challenger to predict the value of sensitive attributes through a wide array of techniques.

The rapid growth of social networking sites has attracted a large number of researchers to explore and study its popular services. This paper focuses on the study and analysis of use behavior of these sites, user's attitude towards privacy and security in non invasive manner, to develop an understanding of individual's risk taking behavior online particularly around the issue of information revelation. Rest of the paper is organized as follows. Section 2 elaborates the theoretical background, section 3 presents the data analysis and results of the paper and section 4 concludes the work.

## 2. THEORITICAL BACKGROUND

The role of peer pressure, herding behavior, myopic privacy attitude, sense of protection offered by SNS has been clarified by Gross et al. which support the information revelation behavior of users and quantified individual's willingness to provide large amount of personal data and expose them to various physical and cyber risks [6]. In terms of perceived trust and privacy concern, subjects from Facebook and Myspace are compared in order to study the use behavior and it has been shown that the interaction of trust and privacy concern in SNS is not yet fully understood to sufficient degree by users to allow accurate modeling of behavior and activity [7]. Steinfield et al. explains the impact of self esteem on use of SNS and demonstrated that lower self esteem students appear to gain more from their use of Facebook than higher self esteemed students as SNSs mitigates fear of rejection and provide ease to facilitate initial communication [8]. Role of national cultural differences in user adoption of SNS is investigated by Veltri et al. which specially focus on Morocco and U.S. users [9]. It uses Hofstede's dimensions of national culture and diffusion of innovation. Phippen et al. demonstrated that users usually get engaged in risk taking behavior and are willing to disclose potentially harmful information to strangers [10]. Gangadharbatla et al.

investigated that internet self efficacy, need to belong and collective self esteem, all have positive effects on attitude of user towards SNS [11]. The factors that drive students to use SNS are explored by Cheung et al. [12]. They conceptualized the use of SNS as an intentional social action and demonstrated that social presence is an important factor that determines student's usage of Facebook. Even Purcell et al. demonstrated that the impact of Facebook is more than Twitter on teens [13]. Robben et al. proposed a model which can help in determining the impact of social learning technologies in workplace. It emphasizes on fundamental trust which results in knowledge productivity and enhanced social capital [14]. The impact of trust, security and privacy concern with regard to social networking websites among consumers using both reliable scale and measures are examined by Shin et al. In this paper it is suggested that SNS providers need to establish a trust relationship with consumers by developing and promoting comprehensive standards and ensuring that participants of privacy seal programs remain to those standards [15]. Lin et al. suggested that the influence on user intention to join SNS, enjoyment has stronger significant role on people's continued use of SNS [16]. It also demonstrated that gender makes a notable difference in the effect of perceived benefit indicating that men do not feel pleasure with SNS with a large number of members. Qin et al. suggested that perceived ease of use, perceived usefulness and social influence positively relate to usage intention of SNSs [17]. Tsoi et al. presented cultural comparison of the impact of privacy concern on the use behavior of SNS and revealed that under the influence of privacy concern, usage behavior on same SNS are not identical [18]. Alam et al. reported the results of an investigation of the use of Facebook by the Australian taxation office [19]. In order to understand the impact of user's privacy concern on their acceptance of SNS, Xin Tan et al. demonstrated two research models with privacy concern conceptualized either as an antecedent of acceptance intention or as a moderator of the relationship in the technology acceptance model. It is statistically shown that privacy concern did restrain the effect of perceived usefulness and perceived ease of use on user's intention to continue the use of SNS [20]. Quinn et al. demonstrates the impact of age difference on the use behavior of SNSs [21]. Dhami et al. proposed a research model which advocates that perceived security, perceived privacy and perceived trust are the factors that influence user's willingness to share information in SNSs [22]. The research model proposed by Mohamed et al. explained that self efficacy, perceived severity, perceived vulnerability and response efficacy are positively related to information revelation [23]. The research model proposed by Chen et al. advocated that social presence, ease of use and personality traits such as extroversion contributes to perceived enjoyment which positively contributes to site use. The personal beliefs of the computing environment such as internet risk perception and privacy abuse concern will impact on one's perceived risk which in turn negatively relate to site use behavior [24]. Elaborated study of personality and its effects on social media is demonstrated by Correa et al. where personality traits like extraversion, emotional stability and openness to experience is discussed [25]. User behavior is studied from four different perceptive which are connection and interaction, traffic activity, mobile social behavior and malicious behavior [26]. As SNS can be used as a tool of customer engagement, social interaction and relationship building, constantinides et al. investigated the use of SNS as a business tool [27].

## 3. DATA ANALYSIS AND RESULT

Data is obtained through the online based questionnaire. It has been statistically shown that although the privacy concerns of respondents are significant but their attitude towards the risks of information disclosure is still very relaxed. Figure 1 shows the attitude of social networking site users towards the potential risks. 30% of the total respondents do not care about the risks in spite of having the knowledge of misuse of personal information. Figure 2 shows the level of understanding of privacy settings and terms and conditions. Analysis shows that only 36% claims to have full understanding of these settings and agreements. Figure 3 shows the usage trend of social networking sites. Analysis shows that 49% respondents use these sites for sharing information. Figure 4 shows the reaction of users when asked would they stop using SNS if SNS continued to use their personal information and photographs with researchers and marketers against user's expectations. 52% users confirmed that they would definitely stop using such sites.

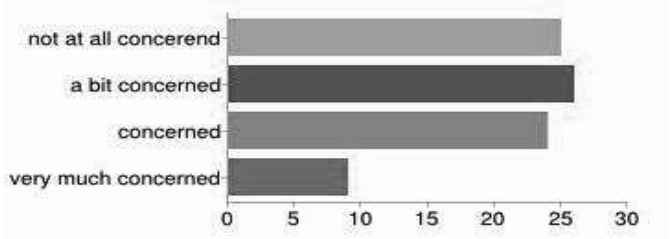



Fig. 1: attitude of users towards privacy concern

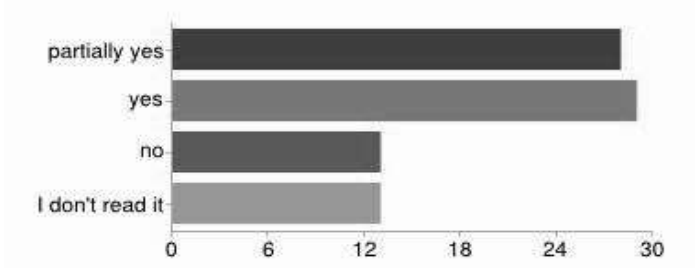



Fig. 2: level of understanding of privacy settings and terms and conditions

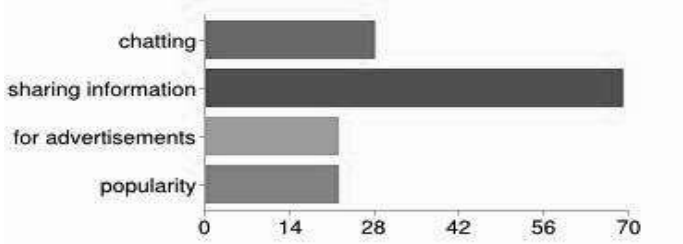



Fig. 3: usage trend

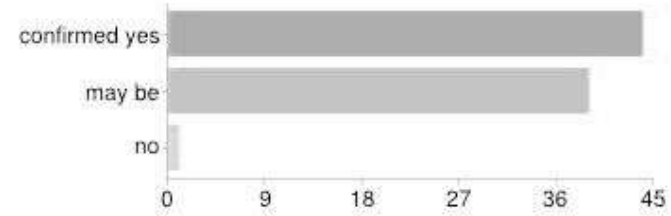



Fig. 4: user's response about using SNS under influence of unwanted information disclosure

It is analyzed that SNS user especially Indian citizen have mixed feelings on privacy issues. Their perceptions regarding privacy issues are still not clear and it is supposed that it is because of lack of awareness. As shown in above graphs that 52% users agree to stop using SNS if they found misusage of

their private information but still 30% of total respondents do not actually care about this issue.

## 4. CONCLUSION

The primary objective of this study is to examine the affect of privacy issues on site use behavior in social networking sites. Site use adds vital support to the sustainable success of networking sites and may simulate the use extent by cultivating perceived hedonic utilities. Some of the flaws and changes in terms of security and privacy as suggested by respondents of online survey are like the message receiving mechanism should be scrutinized again. There should be some control like id proof verification on fake ids. Term and condition page and privacy settings page are often too long and confusing. These pages should be aided with less text and more graphical representations to increase the understanding of the user. Also Social networking sites can provide an option of user defined security settings where user's will be allowed to adjust the privacy settings according to their needs not just according to prearranged format provided by online social networking site providers. In future detailed in depth analysis will be done to study the site use behavior in social networking and impact of privacy concern on it.

## 5. REFERENCES

[1] D.M. Boyd, N.B. Ellison, Social network sites: definition, history, and scholarship, *Journal of Computer-Mediated Communications 13 (1) (2007) 210–230.*

[2] 2014 Best Social Networking Site reviews and Comparisons at: http://social-networking-websites-review.toptenreviews.com/index.html

[3] Statistics at: http://newsroom.fb.com/Key-Facts

[4] J. He, W. Chu, and Z. Liu, "Inferring privacy information from social networks," *Intelligence and Security Informatics, pp. 154–165, 2006*

[5] E. Zheleva and L. Getoor, "To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles," in *Proc. of ACM World Wide Web, 2009, pp. 531–540.*

[6] Gross, Ralph, and Alessandro Acquisti, "Information revelation and privacy in online social networks," in *Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, ACM, 2005.

[7] Dwyer, Catherine, Starr Roxanne Hiltz, and Katia Passerini, "Trust and Privacy Concern Within Social Networking Sites: A Comparison of Facebook and MySpace" *AMCIS*. 2007.

[8] Steinfield, Charles, Nicole B. Ellison, and Cliff Lampe, "Social capital, self-esteem, and use of online social network sites: A longitudinal analysis."*Journal of Applied Developmental Psychology* 29.6 (2008): 434-445.

[9] Veltri, Natasha F. and Wafa Elgarah, "The role of national cultural differences in user adoption of social networking" in *Southern Association for Information Systems Conference, Charleston, SC*, 2009.

[10] Phippen, Andy, Richard Davey, and Steven Furnell, "Should we do it just because we can? Methodological and ethical implications for information revelation in online social networks" in *Methodological Innovations Online* 4.3 (2009): 41-55.

[11] Gangadharbatla, Harsha, "Facebook me: Collective self-esteem, need to belong, and internet self-efficacy as predictors of the iGeneration's attitudes toward social networking sites" in *Journal of interactive advertising* 8.2 (2008): 5-15.

[12] Cheung, Christy MK, Pui-Yee Chiu, and Matthew KO Lee, "Online social networks: Why do students use Facebook?" in *Computers in Human Behavior*27.4 (2011): 1337-1343.

[13] Purcell, Kristen, Aaron Smith, and Kathryn Zickuhr, "*Social media & mobile internet use among teens and young adults"* in Washington, DC: Pew internet & american life project, 2010.

[14] Robben, Joost, "A model for leveraging social learning technologies in corporate environments" in *Proceedings of Network Learning Conference*. 2010.

[15] Shin, Dong-Hee, "The effects of trust, security and privacy in social networking: A security-based approach to understand the pattern of adoption" in *Interacting with Computers* 22.5 (2010): 428-438.

[16] Lin, Kuan-Yu, and Hsi-Peng Lu, "Why people use social networking sites: An empirical study integrating network externalities and motivation theory," in *Computers in Human Behavior* 27.3 (2011): 1152-1161.

[17] Qin, Li, Yongbeom Kim, Jeffrey Hsu, and Xin Tan, "The effects of social influence on user acceptance of online social networks" in *International Journal of Human-Computer Interaction* 27, no. 9 (2011): 885-899.

[18] Tsoi, Ho Keung, and Li Chen, "From Privacy Concern to Uses of Social Network Sites: A Cultural Comparison via User Survey" in *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, pp. 457-464. IEEE, 2011

[19] Alam, Sultana Lubna, John Campbell, and Richard Lucas, "Using Social Media in Government: The Australian Taxation Office e-Tax Facebook Page."*Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference on*. IEEE, 2011

[20] Xin Tan, Li Qin, Yongbeom Kim, Jeffrey Hsu, "Impact of privacy concern in social networking web sites", Internet Research, Vol. 22 Iss: 2, 2011, pp.211 – 233

[21] Quinn, Darren, Liming Chen, and Maurice Mulvenna. "Does Age Make a Difference in the Behaviour of Online Social Network Users?" in *International Conference on Cyber, Physical and Social Computing*, IEEE, 2011.

[22] Dhami, Anil, et al. "Impact of trust, security and privacy concerns in social networking: An exploratory study to understand the pattern of information revelation in Facebook" in IEEE, 2012.

[23] Mohamed, Norshidah, and Ili Hawa Ahmad, "Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia" in *Computers in Human Behavior* (2012).

[24] Chen, Rui, "Member use of social networking sites—an empirical examination" in *Decision Support Systems* 54.3 (2013): 1219-1227.

[25] Correa, Teresa, et al, "Personality and Social Media Use", 2013.

[26] Jin, Long, Yang Chen, Tianyi Wang, Pan Hui, and Athanasios V. Vasilakos, "Understanding user behavior in online social networks: A survey." *IEEE Communications Magazine* 2013

[27] Constantinides, Efthymios, and Carlota Lorenzo-Romero, "Social Networking Sites as Business Tool: A Study of User Behavior" in *Business Process Management*, pp. 221-240, Springer Berlin Heidelberg, 2013.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 45 - 47, 2014

# Comparative Study of Particle Swarm Optimization and Fuzzy C-Means to Data Clustering

Harveen Anand
Dept. of Computer Science and Engineering
Vidyalankar Institute Of Engineering, Wadala
India

Vipul Dalal
Dept. of Computer Science and Engineering
Vidyalankar Institute of Engineering, Wadala
India

**Abstract:** Data has an important role in all aspects of human life and so analyzing this data for discovering proper knowledge is important. Data mining refers to find useful information (extracting patterns or knowledge) from large amount of data. Clustering is an important data mining technique which aims to divide the data objects into meaningful groups called as clusters. It is the process of grouping objects into clusters such that objects from the same cluster are similar and objects from different clusters is dissimilar. In data mining, data clustering has been studied for long time using different algorithms and everyday trends are proposed for better outcomes in this area. Particle swarm optimization is an evolutionary computational technique which finds optimum solution in many applications. Fuzzy C- means (FCM) algorithm is a popular algorithm in field of fuzzy clustering. In this paper, we present a comparative study of Particle swarm optimization and FCM to data clustering.

**Keywords:** Clustering**,** Particle Swarm Optimization, FCM, fuzzy clustering, swarm

## 1. INTRODUCTION

Data plays an important role in all aspects of human life and so analysing this data, extracting useful information from the data, and transforming it into understandable way is important. This is the main goal of data mining process. Clustering is one of major tasks of knowledge discovery from databases. Data clustering is the process of grouping together similar multi-dimensional data vectors into a number of clusters or bins. Data clustering is the organizing of data into specific groups, the characteristics of which are unknown. Clustering is a class of unsupervised classification method and has been applied to a wide range of problems, including exploratory data analysis, data mining, image segmentation, pattern recognition, machine learning and other diverse fields.

Data clustering has been studied for a long time by researchers with different methods. A good clustering method should produce high quality clusters. The quality of the clustering method depends on 3 components: the distance measurement used, the clustering algorithm itself (its implementation) and the ability to discover some or all the hidden patterns. Generally, clustering algorithms can be categorized into hierarchical methods, partitioning methods, Density-based methods, grid-based methods, model-based methods, constraint-based methods, frequent pattern based methods. The hierarchical clustering techniques create a hierarchical decomposition of the database .It merges some clusters in order to make a bigger cluster or divide a cluster into some clusters to make small clusters [1].On the other hand,, the partition clustering techniques partition the database into predefined number of clusters. They attempt to determine 'k' partitions that optimize a certain criterion function.

Recently, bio-inspired algorithms like particle swarm optimization (PSO), ant colony optimization (ACO) etc have found success in solving clustering problems [2].Clustering techniques based on bio-inspired algorithms have outperformed many traditional methods. This paper focuses on comparative study of Particle Swarm Optimization algorithm and Fuzzy C-Means to data clustering.

## 2. FUZZY C-MEANS CLUSTERING

Fuzzy clustering is a powerful unsupervised method for the analysis of data and construction of models. It is an extension of K-Means algorithm [3].Objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial membership[3].Fuzzy C-means (FCM) algorithm is most widely used algorithm for fuzzy clustering [3]. Bezdek [4] introduced Fuzzy C-means clustering method in 1981; extend from Hard C-mean clustering method [5]. Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. FCM employs fuzzy partitioning such that a data point can belong to all groups with different membership grades between 0 and 1[3].FCM is widely applied in agricultural engineering, astronomy, chemistry, geology, image analysis, medical diagnosis, shape analysis and target recognition[6].The working of fuzzy c-means algorithm is discussed using following steps.

Algorithm:

1. Initialize U=[$u_{ij}$] matrix, $U^{(0)}$
2. At k-step: calculate the centers vectors $C^{(k)}$=[ $c_j$ ] with $U^{(k)}$

$$c_j = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_i}{\sum_{j=1}^{n} u_{ij}^{m}}$$

3. Update $U^{(k)}$ , $U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left[ \frac{||x_i - c_j||}{||x_i - c_k||} \right]^{2/m-1}}$$

4. If $\| U^{(k+1)} - U^{(k)} \| < €$ then STOP; otherwise return to step 2.

Here *m* is any real number greater than 1,
$u_{ij}$ is the degree of membership of *xi* in the cluster *j*,
$x_i$ is the *i*th of d-dimensional measured data,
$c_j$ is the d-dimension center of the cluster,

This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data is near to the cluster center more is its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one. After each iteration, membership and cluster centers are updated according to the formula.

## 3. PARTICLE SWARM OPTIMIZATION

Particle swarm optimization algorithm (PSO) is an evolutionary computational technique based on the movement and intelligence of swarms. PSO is a population based global optimization technique proposed by Kennedy and Eberhart [7]. The algorithm uses a number of agents (particles) that constitute a swarm moving in the search space looking for the best solution. In the context of PSO, a swarm refers to a number of potential solutions to the optimization problem, where each potential solution is referred to as a particle**.** Each particle has two movement characteristics, velocity and location. Each particle's movement is influenced by its local best known position and is also guided toward the best known positions in the search-space, which are updated as better positions are found by other particles. This is expected to move the swarm toward the best solutions [8].

Equation (*1*) and Equation (*2*) are the velocity and location update functions. In each generation, all particles are updated using these equations until the maximum generation count reaches or a particle gets to our termination criteria. In all cases, the global best reached so far, is the solution of our problem.

$$V_i^{k+1} = wV_i^k + c_1\, rand_1(pbest_i - x_i^k) + c_2\, rand_2(gbest - x_i^k) \quad --(1)$$

$$x_i^{k+1} = x_i^k + V_i^{k+1} \quad --(2)$$

Here, $rand_1$ and $rand_2$ are uniformly generated random numbers in the range [0, 1]. Acceleration constants $c_1$and $c_2$ and inertia weight *w*are predefined by the user. $V_i^k$ is the current velocity and $x_i^k$ is the current location.

## 4. PSO CLUSTERING

In context of clustering, a single particle represents N cluster centroid vectors. That is, each particle $X_i$ is constructed as follows:

$$X_i = (m_{i1,}\, m_{i2,}\, .....\, m_{iNc})$$

where $m_{ij}$ refers to the j-th cluster centroid vector of the i-th particle in cluster $C_{ij}$ .[9].Therefore, a swarm represents a number of candidate clustering for the current data vectors. The aim of PSO is to find the particle position that results in best evaluation of a given fitness function. The particle swarm optimization based clustering algorithm is as follows[1]:

Algorithm:

1. Initialize each particle with random cluster centroid
2. Do
   a. For each Particle
      i. For Each Data Vector $Z_p$ in Data set
      ii. Calculate $D(Z_p, m_{ij})$ for all cluster centroids.
      iii. Assign $Z_p$ to a Cluster Ci according to this condition
   b. $D(Z_p, m_{ij}) = min_{\forall r=1,2,...N_r}\{d(Z_p, m_{ir})\}$
   c. Calculate Fitness of Particle.

$$J_c = \frac{\sum_{j=1}^{N_c}\left[\sum_{\forall Z_p \in C_{ij}} \frac{d(Z_p, m_j)}{|C_{ij}|}\right]}{N_c}$$

   d. Update the Global Best (*gbest*) and Local Best (*pbest*).
   e. Update the Particle Velocity and Location using Equation *(1)* and *(2)*.
3. While Termination Criteria is satisfied.

## 5. ADVANTAGES AND DISADVANTAGES OF PSO AND FCM

Advantages of PSO

1. PSO algorithm is based on intelligence and it is applied on both scientific research and engineering [10].
2. PSO accepts the real number code and that is decided directly by the solution. Calculation in PSO is simpler and efficient in global search [11].
3. PSO algorithm has no mutation and overlapping calculation. The search can take place by the speed of the particle .Most of optimist particle can be able to transmit the information onto other particles during the development of several generations, and the speed of researching is faster[11].
4. It is insensitive to initial conditions [10].

Disadvantages of PSO

1. Low Convergence [10].
2. Weak local Search ability.
3. PSO cannot work on the problem of non co-ordinate system like solution of energy field

and moving rules for the particles in the energy field [11].

4. The method easily suffers from the partial optimism, which causes the less exact at the regulation of its speed and direction [11].

Advantages of FCM

1. Converges [3].
2. Unsupervised [3].
3. Easy to implement.

Disadvantages of FCM

1. Depends on cluster centre initialization [3].
2. Sensitivity to the initial guess (speed, local minima) [3].
3. Sensitivity to noise and one expects low (or even no) membership degree for the outliers (noisy points).

## 6. CONCLUSION AND FUTURE WORK

From the study of PSO and FCM clustering we found that FCM which depends on initial condition may cause the algorithm to converge to suboptimal solution while on other hand, PSO is less sensitive to initial condition due to its suboptimal population-based nature. So particle swarm optimization is more likely to find near optimal solution. During the study, it was found that in future research, more improvement can be done in area of initialization of algorithm, improved fitness function, combining the advantages with other clustering algorithms to improve the efficiency and performance of clustering.

## 7. REFERENCES

[1] Ehsan Toreini, Maryam Mehrnejad "Clustering data using particle swarm optimization using a new fitness" Proceedings IEEE International Conference,2011.

[2] A.A.Esmin, D.L.pereira, F.P.A. De Araujo, "Study of different approach to clustering data by using particle swarm optimization algorithm"2008 IEEE Congress on Evolutionary Computation (CEC 2008).

[3] R.Suganya, R.Shanthi" Fuzzy C- Means Algorithm- A Review" International Journal of Scientific and Research Publications, Volume 2, Issue 11, November 2012 1 ISSN 2250-3153.

[4] J.C.Bedzek, "Pattern recognition with Fuzzy Objective Function Algorithm" ,New York, Plenum Press,1981.

[5] S. Ghosh,S. K. Dubey, "Comparative analysis of K-Means and Fuzzy C-Means" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 4, No.4, 2013.

[6] Y. Yong, Z. Chongxun and L. Pan, "A Novel Fuzzy C-Means Clustering Algorithm for Image Thresholding", Measurement Science Review, vol. 4, no.1, 2004.

[7] J. Kennedy and RC Eberhart, "Particle Swarm Optimization," Proc.IEEE International Joint Conference on Neural Networks, vol. 4, pp. 1942-1948, 1995.

[8] Yallamati Prakasarao, Arumalla Nagaraju, A.Veeraswamy, "Experimental of Data Clustering using particle swarm optimization", IJESR/Aug 2012/vol-2/Issue-8/Article No-16/864-870.

[9] Van der Merwe, D.W. and AP.Engelbrecht, "Data Clustering using particle swarm optimization", in Evolutionary Computation, 2003.

[10] Pritesh Vora, Bhavesh Oza" A Survey on K-Means and Particle Swarm Optimization" (IJISME) 2013,pp 24-26.

[11] Qinghai B," The Analysis of Particle Swarm Optimization Algorithm", in CCSE, February 2010, vol.3.

[12] Sunita Sarkar, Arindam Roy, Bipul S." Application of Particle Swarm Optimization in Data Clustering", IJCA (0975-8887), Vol 65-No 25, March 2013.

[13] Han J: Kamber, M (2001)."Data Mining: Concepts and Techniques",Morgan Kaufmann, San Francisco.

[14] Jain,A.K; Murty,M N Flynn,P.J.(1999),"Data Clustering: a review" ACM Computing Survey 31(3):264-323.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 48 - 51, 2014

# Survey on Supervised Method for Face Image Retrieval Based on Euclidean Distance

Priyanka T
Vivekanandha college of Engineering for Women
Tiruchengode, Tamilnadu
India

Kalaichelvi M
Vivekanandha college of Engineering for Women
Tiruchengode, Tamilnadu
India

Uma Maheswari C
Vivekanandha college of Engineering for Women
Tiruchengode, Tamilnadu
India

**Abstract**: Content-based image retrieval is a technique which uses visual contents to search images from large scale image databases according to users' interests. Given a query face image, content-based face image retrieval tries to find similar face images from a large image database. Initially face of the image is detected from the query image. After the removal of noise present in the image, it is separated as patches. For each patch, the Local binary pattern (LBP) is extracted which improves the detection performance. LBP is a type of feature used for classification in computer vision. The LBP operator assigns a label to every pixel of a gray level image. The label mapping to a pixel is affected by the relationship between this pixel and its eight neighbors. Support Vector Machine (SVM) is used then which will produce a model (based on the training data) that predicts the target values of the test data given only the test data attributes. When the feature values are provided to the SVM classifier, it will train about the feature. Finally it will classify about the result. SVM maps input vectors to a higher dimensional vector space where an optimal hyper plane is constructed. Among the available hyper planes, there is one hyper plane alone that maximizes the distance between itself and the nearest data vectors of each category. The Euclidean distance between the query image and database image is calculated and the index of the Euclidean distance is sorted.The indexing scheme used for this purpose provides an efficient way to search the image. Then the corresponding image from the database is retrieved based upon the index. This SVM classifier mainly improves the detection performance and the rate of accuracy.

**Keywords**: Content based face image retrieval, local binary pattern, support vector machine, Euclidean distance

## 1. INTRODUCTION

In typical content-based image retrieval systems, the visual contents in the images from the database are extracted and described by multi-dimensional feature values. The feature vectors in the images from the database form a feature database. Users provide the retrieval system with example images or sketched figures to retrieve images. The similarities /distances between the feature vectors of the query example or sketch and those of the images in the database are then calculated and retrieval is performed with the aid of an indexing scheme.

Recent retrieval systems have incorporated users' relevance feedback to modify the retrieval process in order to generate perceptually and semantically more meaningful retrieval results. Image content may include both visual and semantic features. Visual features are both general or domain specific. General visual content include color, texture, shape, etc. Domain specific visual features like the human faces, is dependent on application and may involve domain knowledge.

Semantic content is obtained either by textual annotation or by complex inference procedures based on visual content. It concentrates on general visual contents descriptions and domain specific and semantic features. A good visual feature descriptor should be invariant to the accidental variance introduced by the imaging process (e.g., the variation of the illumining of the scene). However, there is a tradeoff between the invariance and the discriminative power of visual features, since a wide range of invariance loses the ability to discriminate between essential differences.

Invariant description has been largely investigated in computer vision (like object recognition), but it is new in image retrieval. A visual feature descriptor can be local or global descriptor. A global descriptor uses the visual features from the whole image but a local descriptor uses the visual features of regions or objects to describe the image content.

Human attributes such as gender, race, hair style are high-level semantic descriptions about a person. The recent work shows automatic attribute detection has adequate quality (more than 80% accuracy) on many different human attributes. Using the described human attributes, many researchers has been achieved promising results in different applications such as verification of face, identification of face, keyword-based face image retrieval, and similar attribute search. These results indicate the power of the human attributes on face images.

Although human attributes have been shown useful on applications related to face images, it is non-trivial to apply it in content-based face image retrieval task due to several reasons. First, human attributes only contain limited dimensions. When there are too many people in the dataset, it loses discriminability because certain people might have similar attributes. Second, human attributes are represented as a vector of floating points. It does not work well with developing large-scale indexing methods, and therefore it suffers from slow response and scalability issue when the data size is huge.

## 2. LITERARY SURVEY

### 2.1 Active Shape Models (ASM)

Active shape models (ASMs) are statistical models of the shape of objects which iteratively deform to fit to an example of the object in a new image. ASM relies upon each object or image structure being represented by a set of points. The points represents the boundary of the image and internal features, etc,. The sets of points are aligned automatically to minimize the variance in distance between equivalent points.

Points are placed in the same way on each of a training set of examples belonging to the object. By the examining of statistics of the positions of the labeled points a model is derived called “Point Distribution Model (PDM)”. The model gives the average positions of the set of points, and it has a various number of parameters which control the main modes of variation found in the training set. The ASM algorithm aims to match the model to a new image.
The ASM performs by alternating the following methods:

- Generate a suggested shape by looking in the image around each point for a better position for that point. This is usually done using a model called a "profile model", which looks for strong edges to match a model template for the point.
- Conform the suggested shape to the point distribution model, commonly called a "shape model".

One of the main drawbacks of the approach is the amount of labeled training examples required to build a good model which are very time consuming to generate. They are also not appropriate for objects with widely varying shapes.

## 2.2 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables. Those variables are called the principal components. The number of principal variables is less than or equal to the number of original components. The transformation of the variables is defined in such a way that the first principal component has the largest possible variability, and each succeeding variable in turn has the highest variance possible under the constraint that it be orthogonal to (i.e., uncorrelated with) the preceding components.

The objectives of principal component analysis are to discover or to reduce the dimensionality of the database, to identify meaningful underlying variables. PCA is sensitive to the relative scaling of the original variables. PCA defines a new orthogonal coordinate system that optimally describes variance in a single dataset. The limitations include the results of PCA depend on the scaling of the variables and the applicability of PCA is limited by certain assumptions made in its derivation.

## 2.3 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a classification method that is mathematically robust and often produces models whose accuracy is as good as more complex methods. LDA is used in pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of attributes or events. The resulting combination is the linear classifier for dimensionality reduction before classification.
LDA is closely related to regression analysis, which attempts to express one dependent variable as a linear combination of other features or measurements and also to principal component analysis (PCA) and factor analysis in that they both look for linear combinations of variables which best define the data. LDA explicitly tries to model the difference between the classes of data. PCA does not take into account any difference in class. Discriminant analysis is also different from factor analysis in that it is not an interdependence technique, a distinction between independent variables and dependent variables must be made. LDA works when the measurements made on independent variables for each observation are continuous quantities.
In face recognition, each face is represented by a large number of feature values. Linear discriminant analysis is commonly used to reduce the number of features to a more manageable number before the object classification. Every new dimension is a linear combination of pixel values, which form a template. The advantages of LDA include multiple dependent variables and reduced error rates.
Drawbacks of LDA

- Linear discriminant analysis is extremely sensitive to outliers.
- The dependent variables are not correlated to the linear combination of other variables is not perfect.
- The relation between the variables are assumed to be linear in all groups.

## 2.4 Locality-constrained Linear Coding (LLC)

LLC is a simple but effective coding scheme which utilizes the locality constraints to project each descriptor into its local-coordinate coding system, and the coordinates which are projected will be integrated by max pooling to generate the final representation. LLC is easy to compute and gives superior image classification performance than any other approaches. It is a fast approximated method that proposes by first performing a K-nearest-neighbor search and then solving a constrained least square fitting problem, bearing computational difficulty. So even with very large codebooks, the system processes multiple frames per second.
LLC incorporates locality constraint instead of the scarcity constraint which leads to several favorable properties like better reconstruction, local smooth sparsity, analytical solution. Jinjun Wang and Jianchao Yang presented a promising image representation method called Locality-constrained Linear Coding (LLC).
LLC applies locality constraint to select similar basis of local image descriptors of a codebook, and learns a combination weight of these basis to reconstruct each descriptor. LLC coding criteria is used for codebook training, to improve the performance.

## 2.5 Bag-of-Words (BoW)

Automatic image annotation is the process of employing computer programs to automatically assign an unlabeled image a set of keywords or tags, each of which represents certain semantic object/concept. Bag of visual words is a sparse vector of occurrence counts of a vocabulary of local image features. To represent an image using BoW model, an image can be treated as a document. It usually requires three steps: feature, description of feature and the generation of codebook. After detecting the feature, every image is extracted by many image patches. Feature representation technique concentrates with how to represent the patches as numerical values. These values are called feature descriptors. A descriptor should be able to handle intensity variation, rotation changes and variation of affine to some level. One of the most famous descriptors is Scale-invariant feature transform (SIFT). SIFT converts each patch to some dimensional vector. After completing this step, each image will be a collection of vectors of the same dimension. The final step for the BoW model is to convert vector represented patches to "codewords", which also produces a "codebook". Each patch in an image is mapped to a certain codeword

through the clustering process and the image can be represented by the histogram of the codewords.
The disadvantages of BoW include that it ignores the spatial relationships among the patches, which is very important in image representation and the BoW model for object segmentation and localization is not well understood.
Lei Wu and Steven C. H. Hoi proposed a Semantics-Preserving Bag-of-Words (SPBoW) model, for the measurement of the semantic gap, it takes the distance between the semantically identical features and tries to train a codebook by reducing the semantic gap. They formulate the codebook generation task as a distance metric learning problem, which can be formalized as semidefinite programming (SDP). This approach overcomes the limitation of semantics lost in BoW models by bridging the semantic gap via distance metric learning method. An efficient algorithm is implemented to solve the codebook learning task and the optimization problem. The SPBoW can automatically decide the size of the codebook for each category. There are two important issues for SPC "Semantics-Preserving Codebook", such as assignment of codebook size, and generation of visual word.

## 2.6 Local Binary Pattern (LBP)

Local Binary Patterns is one of the best local feature describer operator which is nonparametric and computationally simple that is used to describe the local spatial structure of an image. LBP is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number.
The most important property of the LBP operator in real-world applications is its robustness to monotonic gray-scale changes occurred such as variation of illumination. One more important property is its simplicity regarding computational that makes the analysis of images in challenging real-time settings.
The LBP pattern encodes the textures of the facial regions while the whole shape of the face is recovered by the construction of the face feature histogram. This histogram effectively has a description of the face on three different levels of locality: the LBP labels for the histogram contain information about the patterns on the pixel-basis, the labels are concatenated over a small region to produce information on a regional level and the regional histograms are concatenated to build a global description of the face.
The idea behind using the LBP features is that the face images can be seen as composition of micro-patterns which are invariant with respect to monotonic transformations into grey scale image. Grouping these patterns, a description of the face image is obtained.
Timo Ahonen and Matti Pietik¨ainen proposed a novel approach to face recognition for the purpose of representing the facial images which considers both shape and texture information. In this approach initially the face area is divided into very small patches. From these patches, Local Binary Pattern (LBP) histograms are extracted and concatenated into a single, improved feature histogram which uniquely represents the face image. The recognition will be performed using an attribute classifier called nearest neighbor classifier in the computed feature space. It increases the robustness of the systems against different factors.
Due to its discriminative power and simplicity regarding simplicity, LBP operator is widely used approach in various applications. LBP recognizes the facial expression and also used in many other applications of biometrics like localization of eye and recognition of iris and fingerprint, palmprint recognition, and facial age classification.

## 2.7 Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are supervised learning models with associated learning algorithms that analyse data and pattern recognization. SVM is primarily used for the analysis of classification and regression. The SVM considers an input data to predict that which of two classes forms the output. SVM is a binary classifier. Given a training data, marked as belonging to one of two classes, an SVM algorithm builds a template which assigns the example data into to the belonging class. SVM model is a representation of the example data, mapped so that the examples of the separate classes are segmented by a gap that is as big as possible. The example data are then mapped into that same data and then predicted to belong to a class based on gap formed. SVM can perform linear classification as well as non-linear classification by implicitly mapping their inputs into high-dimensional feature spaces. There are three possible extensions to SVM which are listed below:

- Multiclass SVM
- Transductive SVM and
- Structured SVM

# 3. CONCLUSION

In this survey overviewed various image classifiers needed for the purpose of efficient face image retrieval. Among all of them the Support Vector Machine (SVM) is the better classifier. SVMs are helpful in text and hypertext categorization as they can significantly reduce the need for labelled training instances in both the standard inductive and transductive settings. SVMs are also useful in medical science to classify proteins with up to 90% of the compounds classified correctly. Hand-written characters can be recognized using SVM. Thus the content based face image retrieval system retrieves similar images from the large image database when a query image is provided. The retrieval is followed by initially detecting the face from the image and performing some of the preprocessing steps to remove certain noise which will be present in the image. The extracted LBP feature assigns labels to every pixel of the gray converted image. Then SVM classifier classifies and predicts about the image. This will calculate the Euclidean distance of the both the query and the dataset image. This system greatly reduces the quantization error and can achieve salient gains in face retrieval from the database. And also avoids misclassification of the images with the help of the SVM classifier.

# 4. ACKNOWLEDGMENTS

We thank the reviewers and others those who have helped us in preparing this.

# 5. REFERENCES

[1] Belhumeur.P, Boult.T, Kumar.N, and Scheirer.W, (2012) "Multi-attribute spaces: Calibration for attribute fusion and similarity search" in Proc. IEEE Conf. Computer Vision and Pattern Recognit.

[2] Cheng.W.H, Lin.H.T, Kuo.Y.H and Yang.Y.H, (2011) "Unsupervised auxiliary visual words discovery for large-scale image object retrieval," in Proc. IEEE Conf. Computer Vision and Pattern Recognit.

[3] Coates.A and Ng.A.Y, (2011) "The importance of encoding versus training with sparse coding and vector quantization," in *Proc. ICML*.

[4] Boult.T.E, Kumar.N, Ricanek.K and Scheirerand.W, (2011) "Fusing with context: A Bayesian approach to combining descriptive attributes," in Proc. Int. Joint Conf. Biometrics.

[5] Davis.L.S, Feris.R.S, and Siddiquie.B, (2011) "Image ranking and retrieval based on multi-attribute queries," in Proc. IEEE Conf. Computer Vision and Pattern Recognit.

[6] Jain.A.K and Park.U, (2010) "Face matching and retrieval using soft biometrics," IEEE Trans. Inf. Forensics Security, vol. 5, no. 3, pp. 406–415.

[7] Ho.S.C.H, Wu.L, and Yu.N, (2010) "Semantics-preserving bag-of-words models and applications," IEEE Trans. Image Process., vol. 19, no. 7, pp. 1908–1920.

[8] Cao.Z, Tang.X, Sun.J, and Yin.Q, (2010)"Face recognition with learning based descriptor," in *Proc. IEEE Conf. Computer Vision and Pattern Recognit*.

[9] Ke.Q, Shum.H.Y and Sun.J, (2010) "Scalable face image retrieval with identity-based quantization and multi reference re-ranking," in Proc. IEEE Conf. Computer Vision and Pattern Recognit.

[10] Belhumeur.P.N, Bergand.A.C, Kumar.N and Nayar.S.K, (2009) "Attribute and simile classifiers for face verification," in Proc. Int. Conf. Computer Vision.

[11] Ganesh.A, Sastry.S, Wright.J and Yang.A, (2009) "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227.

[12] S. Milborrow and F. Nicolls, (2008) "Locating facial features with an extended active shape model," in Proc. Eur. Conf. Computer Vision.

[13] Lazebnik.S, Ponce.J and Schmid.C, (2006) "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Conf. Computer Vision and Pattern Recognit*.

[14] T. Ahonen, A. Hadid, and M. Pietikainen, (2004) "Face recognition with local binary patterns," in Proc. Eur. Conf. Computer Vision.

[15] Lowe.D, (2003) "Distinctive image features from scale-invariant keypoints," Int. J. Comput. Vision.

[16] Gionis.A, Indyk.P, and Motwani.R, (1999) "Similarity search in high dimensions via hashing," in Proc. VLDB.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 52 - 55, 2014

# Modeling and Analysis of Two Node Network Model with Multiple States in Mobile Networks

Patil Shweta
Dept. of E&C
VTU Center for PG Studies
Bangalore, India

B.N. Manjunatha Reddy
Dept. of E&C
Global Academy of Technology
Bangalore, India.

**Abstract:** In this work, the decision probability of the handoff are modeled and simulated for smaller bandwidths. The smaller bandwidth is chosen just for simulation purposes and to demonstrate the applicability of the algorithm. The probability of handover and probability of incorrect decision in the handover is modeled. Two nodes of the network are modeled and the probabilities of four different states of the mobile node are also modeled. The results are presented for two cases with and without the probabilities of four different states of the mobile nodes.

**Keywords:** Wrong Decision Probability, Missing Handovers, Unnecessary handovers

## 1. INTRODUCTION

There are number of networks and technologies available today to meet the market demands and the continuous changing needs of the markets. The technologies and the topologies of the networks are designed and developed based on the requirements per the market. The networks are designed to handle enough bandwidth, signal strength, speed of handling and processing of data, voice and videos. More importantly, the networks should provide the guaranteed security so that the data transmission is safe. This is an essential requirement for the corporates using these networks in their day to day work. Another important factor one should consider when designing these networks is there should be a reliable handover of the mobile nodes when the certain condition like movement of mobile node, bandwidth availability and signal strength are considered as a condition of handover. The challenge is how smoothly one can handover the mobile node seamlessly.

Handover is performed when certain conditions in the network system are met. These conditions have been used to design multiple numbers of algorithms [1]. The conditions are based on the factors like bandwidth availability in the target network versus the present network, signal strength, movement of the mobile node, access delay. For example, when the user of the mobile node moves from one area to another area in a locomotive or an automobile, the handover happens based on the distance between the network node and the mobile node. If the distance of another network with respect to the position of the mobile node is less compared to that of the present one, handover happens. Similarly, the handover can happen if the received signal strength by the mobile node is less than certain threshold [1]. However, it must be built into the algorithm that if a mobile node is handed over based on the signal strength or movement of the mobile node, there should be availability of enough bandwidth in the target network. If bandwidth is not available, then the algorithm should also resolve such conflicts. These kinds of algorithms are already available in the literature [2].

While there is sufficient number of handover algorithms, there is lack metrics or gauge algorithm which measures the performance of the handover algorithms. The performance needs to be measured against certain criteria like number of successful handovers, number of incorrect decisions of handover etc. Hence it is required to develop a common method based on analytical models for better comparison. There are already some algorithms available which can measure the performance of the handover algorithms [3, 4]. The criteria used in these algorithms are based on received signal strength [3, 4]. There is another algorithm available which is based on the access delay [2]. The disadvantage of these algorithms is lack of robust analytical model in the structure. This gap in the measurement of performance of algorithms necessitated the concept of the Wrong decision probability model [6]. This model measures the number of in-correct decisions made by the handover algorithms irrespective of the kind of decision criteria used for the handover. In other words, the wrong decision probability model measures the performance of the handover algorithm if it is based on bandwidth, signal strength or movement of mobile node.

The performance of all the algorithms mentioned above are evaluated in the virtual environment and the measured performance is only applicable to the environment having conditions made in assumptions while developing the algorithms. The actual performance in real situation may be quite different from the simulated performance. Other algorithms can be found in references [7-20].

In this work, an improvement is made to the wrong decision model developed in ref [6]. Four different states of the mobile node [21] are considered in this work. The mobile node can stay in any of the two network nodes and in any of the four states. Semi Markov models are used to represent the actual scenario and the probabilities are considered to evaluate the performance of the two node network model.

In this work, probability models are used to determine the probabilities of unnecessary handovers, missing handovers and wrong decisions for different decision times and for two different models, namely, baseline and proposed models. Next section present the mathematical models used in the analytical formulations. Section III describes the general algorithms used and section IV explains the simulated results based on MATLAB coding. Finally important conclusions are drawn in section VI.

## 2. ANALYTICAL MODEL

There are two network nodes in the model, namely, $n1$ and $n2$ (P and Q). Assume that the mobile node is either in network node P or Q. The mobile node while in network nodes P or Q can be in any of the four states. These four states are:

**1.** Cooperative State

**2.** Malicious State

**3.** Selfish State

**4.** Failed State.

***Definitions:***

- In Cooperative state, the mobile nodes are active in route discovery and packet forwarding, but not in DOS attack launching.
- In Malicious State, the mobile nodes are active in route discovery and DOS attack launching.
- In Selfish state, the mobile nodes are active in route discovery, but not in packet forwarding and DOS attack launching.
- In failed State, the mobile nodes are not active in route discovery.

The models proposed by Chi et. al [6] have only two network nodes with just one state, i.e. cooperative state of mobile node.

The two models may be called as

1. Baseline model (Proposed by Chi et.al [6])
2. Improved model (Proposed in this work)

### A. Baseline Model:

Let there are two wireless networks. As mentioned in the baseline model [6], the probability that a mobile node continues to stay in the network node 1 is,

$$P_{n1} = \frac{P_{n1/n2}}{P_{n1/n2} + P_{n2/n1}} \quad (1)$$

And, the probability that a mobile node continues to stay in the network node 2 is,

$$P_{n2} = \frac{P_{n2/n1}}{P_{n1/n2} + P_{n2/n1}}, \quad (2)$$

Where, $n_1$ and $n_2$ are network nodes;

$B_1$ and $B_2$ are maximum available bandwidth for the two networks;

$P_{nj/ni}$ is the probability of mobile node moving from node $n_i$ to $n_j$;

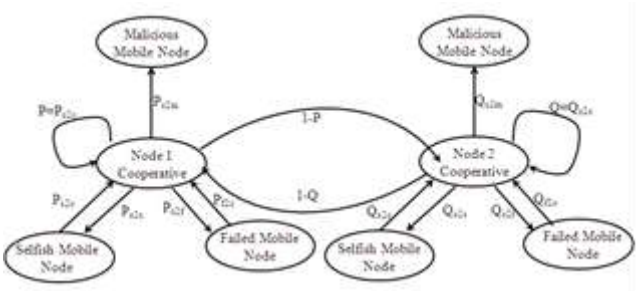



Figure 1: Semi Markov State Two Node Network Model

$P_{ni/ni}$ denotes the probability of mobile node continue to stay in $n_i$ after a time interval *D*.

### B. Proposed Model:

In the proposed model as shown in fig.1, the probability that mobile node stays in node 1 is,

$$P_{c2c-1} = \frac{P_{n1/n2}}{P_{n1/n2} + P_{n2/n1}} \quad (3)$$

and the probability that a mobile node continues to stay in the network node 1 and in cooperative state is,

$$P_{n1} = P_{c2c-1} + P_{f2c-1} + P_{s2c-1} - P_{c2s-1} - P_{c2m-1} - P_{c2f-1} \quad (4)$$

The probability that mobile node stays in node 2,

$$P_{c2c-2} = \frac{P_{n2/n1}}{P_{n1/n2} + P_{n2/n1}} \quad (5)$$

and, the probability that a mobile node continues to stay in the network node 2 and in cooperative state is,

$$P_{n2} = P_{c2c-2} + P_{f2c-2} + P_{s2c-2} - P_{c2s-2} - P_{c2m-2} - P_{c2f-2} \quad (6)$$

Where, probability of the mobile node moving from failed state to cooperative state is given by,

$$P_{f2c} = \frac{1}{T_{\text{Recover}}} \quad (7)$$

Probability of the mobile node moving from selfish state to cooperative state is given by,

$$P_{s2c} = \frac{TC_{Thr}}{TC_{Max}} \quad (8)$$

Probability of the mobile node moving from cooperative state to selfish state is given by,

$$P_{c2s} = \frac{1}{T_{Selfish}} \quad (9)$$

Probability of the mobile node moving from cooperative state to malicious state is given by,

$$P_{c2m} = q_a \frac{k_a}{N} \frac{1}{T_{attack}} \quad (10)$$

Probability of the mobile node moving from cooperative state to failed state is given by,

$$P_{c2f} = \max\left(\frac{1}{T_{Life}}, \frac{1}{T_{\text{Residence}}}\right) \quad (11)$$

More details about these equations and nomenclature can be found in ref [21]. Handover probabilities are given by,

$$HP = P_{n1}P_{n2/n1} + P_{n2}P_{n1/n2} \quad (12)$$

Wrong decision probability can be computed from the unnecessary handover probability (UHP) and missing handover probability (MHP). The exact expressions for the probabilities can be seen in reference 6. The difference lies only in the computation of the probabilities $P_{n1}$ and $P_{n2}$. Wrong decision probability is the summation of UHP and MHP i.e.

$$WDP = UHP + MHP \quad (13)$$

## 3. GENERAL ALGORITHM

Only the band width is considered as the criteria for the handover. The general algorithm that is used in this simulation is similar to the one used in the reference 6.

## 4. SIMULATION RESULTS

In this section the simulation results are presented for a maximum bandwidth of the 16 in both the network nodes 1 and 2. MATLAB is used for coding the probability models. The traffic density is varied from 1 to 15 and for six cases of the decision or life time of the mobile node. The decision times of 10 to 15 time units are modeled. The following factors are used in the simulations:

- Life time of the mobile node = Decision time
- Residence time = Decision time/4
- Attack time = Decision time/4
- Recovery time = Decision time/4
- Selfish time = Decision time/4
- probability of attack, qa=0.3
- $TC_{thr}/TC_{max}$ = 0.01
- ka/N=0.01

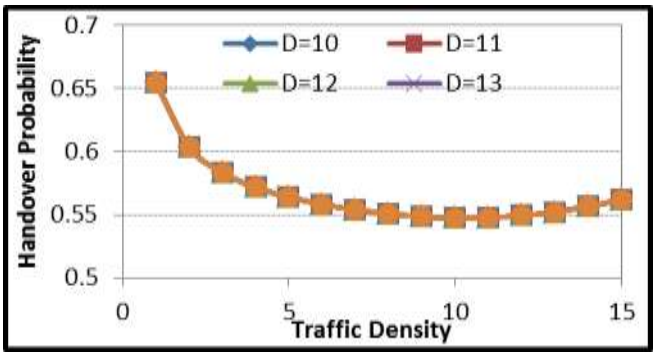



Figure.2: Handover probability versus Traffic Density for Baseline Model

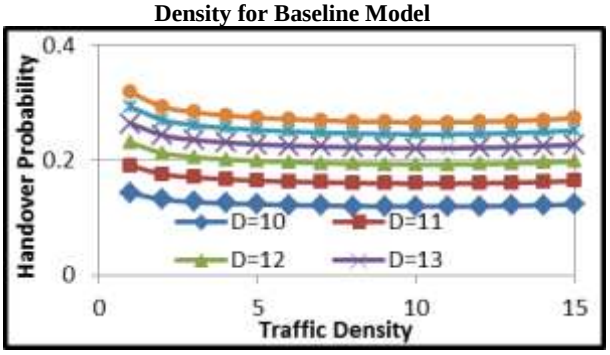



Figure.3: Handover Probability versus Traffic Density for Proposed Model

Fig. 2 shows the handover probability versus traffic density for baseline model for different decision times. It can be noticed that the handover probability is independent of the decision time. This assumption is not a valid assumption as per the baseline model proposed in reference 6. Because the conditions of the handover can change at the other networks when the decision times are large. For example, the bandwidth availability in the target network may change by the time the mobile node is actually handed over.

Fig. 3 shows the handover probability versus traffic density for proposed model. These probabilities represent the reality as the handover probability varies with respect to the decision time. Also, the the handover probability is very high in the baseline model compared to the proposed model.

Fig. 4 and Fig. 5 show the unnecessary handover probability versus traffic density for baseline model and poposed models respectively. It can be noticed that the unnecessary handover probabilities are less by around 50% in the proposed models. With the baseline models, the probability models are oversimplified and hence high values of probabilities.The probabilities start to take the reverse profile after D=10.

Fig. 6 and Fig.7 show the missing handover probability versus traffic density for baseline model and proposed models respectively. It can be observed that the missing handover probabilities are less by around 70% in the proposed models. The results start to take reverse profile after the decision time of 15 time units.

## 5. CONCLUSION

In this work the probability models are developed using four states of the mobile node namely, cooperative, selfish, failed and malicious states. The mobile node can move from any one of the state to other states. What is important in this model is, one the cooperative state is useful for the handover and all the probabilities moving from and into this state is considered for calculating the wrong decision probability. It is observed that the wrong decision probabilities are predicted much accurately by at least 60% higher than the baseline model proposed in reference 6. It is also demonstrated that at certain values of the decision time the results start to take the reverse profiles for the factors used in this model. Based on this one can decide on the values of factors for the kind of probabilities that is of interest to the user. These models can be further extended to consider the signal strength, movement of the mobile node, access delays etc.

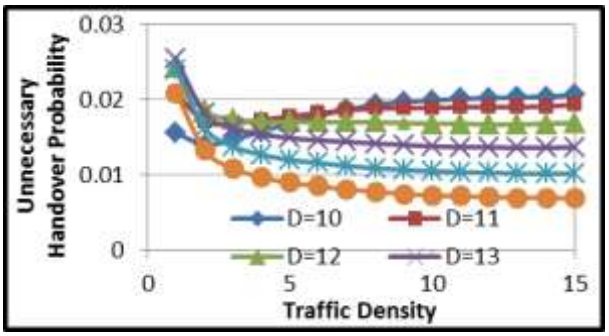



Figure 4: Unnecessary Handover Probability versus Traffic Density for Baseline Model

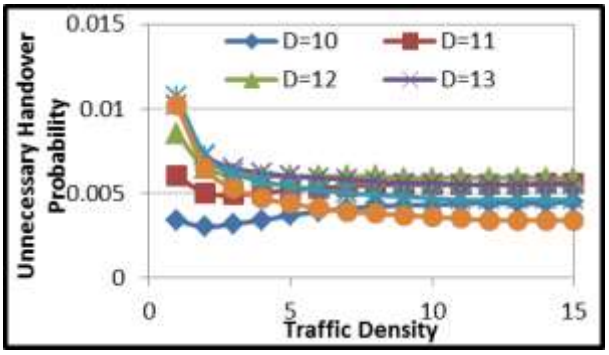



Figure 5: Unnecessary Handover Probability versus Traffic Density for Proposed Model

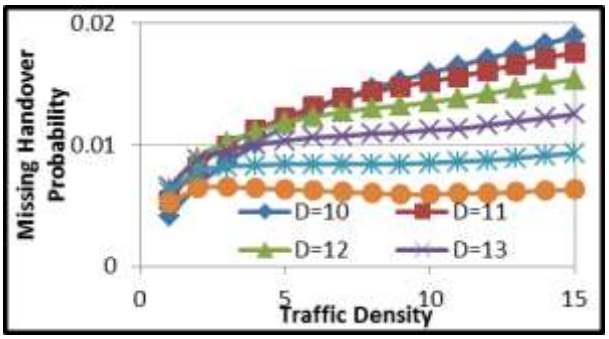



Figure 6: Missing Handover Probability versus Traffic Density for Baseline Model

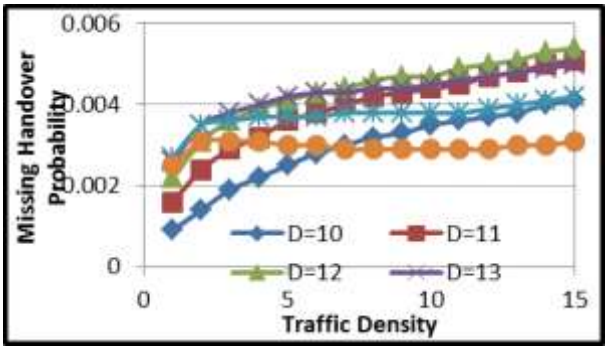



Figure 7: Missing Handover Probability versus Traffic Density for Proposed Model

## 6. REFERENCES

[1]. A. H. Zahran, B. Liang, and A. Saleh, "Signal threshold adaptation for vertical handoff in heterogeneous wireless networks," Mob. Netw. Appl., vol. 11, no. 4, pp. 625–640, 2006.

[2]. Chuanxiong Guo, Zihua Guo, Qian Zhang, and Wenwu Zhu, "A Seamless and Proactive End-to-End Mobility Solution for Roaming Across Heterogeneous Wireless Networks", IEEE Journal on Selected Areas in Communications,Vol. 22, No. 5, June 2004, pp.834-848.

[3]. W. Zhao, R. Tafazolli, and B. G. Evans, "Internetwork handover performance analysis in a gsm-satellite integrated mobile communication system," Selected Areas in Communications, IEEE Journal on, vol. 15, no. 8, pp. 1657–1671, 1997.

[4]. M. N. Halgamuge, H. Le Vu, K. Rarnamohanarao, and M. Zukerman, "Signal-based evaluation of handoff algorithms," Communications Letters, IEEE, vol. 9, no. 9, pp. 790–792, 2005.

[5]. H.Sanneck and G.Carle, "A Framework Model for Packet Loss Metrics Based on Loss Runlengths", Proceedings of the SPIE/ACM SIGMM Multimedia Computing and Networking Conference 2000.

[6]. C. Chi, X. Cai, R. Hao and F. Liu "Modeling and Analysis of Handover Algorithms" IEEE GLOBECOM 2007 proceedings.

[7]. Q. Zhang, C. Guo, Z. Guo, and W. Zhu, "Efficient Mobility Management for Vertical Handoff between WWAN and WLAN," IEEE Commun. Mag., November 2003.

[8]. K. Egevang and P. Francis, "The IP network address translator (NAT)," RFC1631.

[9]. Daniel Heyman and Mattlew Sobel. Stochastic Models in Operations Research. McGraw-Hill, 1982.

[10]. G. Pollini, "Trends in handover design," IEEE Commun. Mag., March, 1996.

[11]. K. Pahlavan, P. Krishnamurthy, A. Hatami, M. Ylianttila, J. Makela, R. Pichna, and J. Vallstrom, "Handoff in Hybrid Mobile Data Networks," IEEE Personal Communications, Apr. 2001.

[12]. M. Ylianttila, M. Pande, J. Makela, and P. Mahonen, "Optimization scheme for mobile users performing vertical handoffs between IEEE 802.11 and GPRS/EDGE networks," in Proc. IEEE Globecom'01.

[13]. Xiang-Yang Li, Peng-Jun Wan, Yu Wang, and Chih-Wei Yi. Fault Tolerant Deployment and Topology Control in Wireless Networks. In Proc. of ACM MobiHoc '03, pages 117–128, Jan. 2003.

[14]. EURESCOM Project P1013-FIT-MIP, "First steps towards UMTS: Mobile IP services, a European testbed", http://www.eurescom.de/public/projects/P1000-series/p1013/default.asp.

[15]. J. McNair, I. Akyildiz, and M. Bender, "An inter-system handoff technique for the IMT-2000 system," in Proc. IEEE infocom 2000.

[16]. C. Perkins, Editor, "IP mobility support for IPv4," RFC 3344, http://www.ietf.org/rfc/rfc3344.txt

[17]. C. Perkins, et al., "Mobility support in IPv6," IETF draft, http://www.ietf.org/internet-drafts/draft-ietfmobileip- ipv6-24.txt, work in progress, June, 2003.

[18]. Hao Yang, Haiyun Luo, Fan Ye, Songwu Lu, and Lixia Zhang. Security in Mobile Ad Hoc Networks: Challenges and Solutions. IEEE Wireless Communications, pages 38 – 47, Feb 2004.

[19]. A. Valko, "Cellular IP: A New Approach to Internet Host Mobility," Computer and Communication Review, vol. 29, no.1, pp. 50-65, January 1999.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 56 - 62, 2014

# Presenting an Excusable Model of Enterprise Architecture for Evaluation of Reliability using Colored Petri Nets

Mohammad Azizi
Department of Computer,
Science and Research branch,
Islamic Azad University,
Ardabil, Iran

Ali Harounabadi
Department of Computer
Engineering, Central Branch,
Islamic Azad University,
Tehran, Iran

Sayed Javad Mirabedini
Department of Computer
Engineering, Central Branch,
Islamic Azad University,
Tehran, Iran

**Abstract:** Upon increasing application of unified modeling language diagrams for description of enterprise architecture and importance of evaluating the non-functional requirements in the enterprise architecture, establishing an executable model of these diagrams is necessary. On the other side, unified modeling language diagrams have not the capability of evaluating the non-functional requirements of system directly. The simplicity and characteristics of these diagrams must be reserved and the capability of evaluating these requirements to be incorporated thereto. To achieve this goal, an executable model of these diagrams to be Created.

In this paper, it is assumed that architecture of a system has been described by two use case and collaboration diagrams of unified modeling language. The role of these diagrams in evaluation of reliability and annotations related thereto has been examined and extended by a algorithm to an executable model means colored Petri net. In this study, the procedure of establishing an executable model that can analyze the reliability of artifacts in C4ISR framework. According to the results obtained from simulation of this model and their analysis, we could identify the problems in planning phase and improve out artifacts in order to avoid the extreme time and economic costs of implementation.

**Keywords**: Enterprise Architecture, C4ISR Framework, Petri Net, Non-functional requirements, Executable Model

## 1. INTRODUCTION

In the current century, the enterprises are complex and integrated systems which have been constituted of processes, enterprise units, personnel, information and support technologies as well as dependences and communications between different elements. In order to achieve and retain the enterprises' performance, recognizing, engineering and managing these social, technical and infrastructural factors is very crucial. This necessity has resulted in establishment of enterprise architecture.

The modern enterprises are complex and integrated systems which have been constituted of processes, enterprise units, personnel, information and support technologies as well as dependences and communications between different elements. In order to achieve and retain the enterprises' performance, recognizing, engineering and managing these social, technical and infrastructural factors is very crucial.

Recently, a strategic approach has been created for enterprise architecture titled enterprise systemic architecture that assumes the enterprise as an integrated system including different viewpoints in an integrated framework.

Today, various definitions have been presented for enterprise architecture. The respective definition in this paper is as follows:

The enterprise architecture is a macro attitude to enterprise missions and duties, working processes, information existences, communication networks, hierarchy and order of works performance in an enterprise,that is implemented with the purpose of establishing integrated and efficient information systems [3].

Whereas enterprise architecture is a complex process that includes all sections of an enterprise and a lot of people with varied skills are engaged therein, it is obvious that controlling and leading such an extensive and great process may not be possible without following the predetermined and unified model and structure [5].

One of successful methods in such extensive process is using enterprise architecture frameworks. Therefore, the framework is deemed as the key concept in enterprise architecture. Enterprise architecture for effective achievement to its goals is extremely dependent to the framework concept [2]. Frequent frameworks have been presented for enterprise architecture, out of them C4ISR architectural framework due to appropriateness is assumed as architectural promoter. Zachman framework explains the procedure of their classification for architecting and C4ISR provides a process for its establishment. Therefore, C4ISR completes Zachman.

The executable architectural models are extracted from operational components of system, tasks, messages, events, scheduling and messages transmission mechanism. These models provide the capability of analyzing the interaction between components, quality of tasks allocation to components, tasks control and system's temporal and local characteristics, to the architecture.

One of the most important non-functional requirements that affects the software development is reliability. Because, the reliability problems may made considerable changes at each stage of software life cycle particularly at the primary stages of extension. Different works have been performed in relation to non-functional requirements particularly reliability in primary levels of software development and architecture. But, application of each one of these methods is limited in complex and pragmatic applied instances and none of methods evaluates these requirements completely.

The difference between method presented in this paper to the other works is that the architect evaluates the architectural level of software without interference of analyzer and annotates the information required for reliability to the architectural descriptions.

There are different executable models for evaluation of non-functional traits the most important ones include queue nets, Petri nets, process algebra, random processes etc. Queue is appropriate for using in random analysis, but not appropriate for analysis of qualitative traits. The difference between proposed method and performed activities is that use case and collaboration diagrams were used for evaluation of enterprise architecture performance and the obtained executable model is formed based on petri nets. For evaluation of performance by petri nets, against queue nets, an executable model is required and upon changing the value and type of tokens available in these nets, inserting the phrases to arcs for computing the respective requirement or inserting the additional place or transition to the net, may evaluate other non-functional requirements. Thus a general framework is provided by these networks for evaluation of enterprise architecture in relation to different non-functional requirements.

## 2 . BACKGROUND

In this section, a general description of architectural framework and unified modeling language and colored Petri nets are presented.

### 2.1 Evaluation framework of architectural artifacts

In general, the framework is a mean for classification of objects and whereas our subject is related to enterprise, the respective objects are descriptions of enterprise aspects and contexts. These descriptions (models) is usable by every person who needs them providing to have access permission.

C4ISR framework is formed based on work performance process, therefore the framework accuracy and delicacy majorly focuses on execution procedure of processes available in enterprise. This framework has been constituted of three main components as follows:

Three standard viewpoints that indicate different viewpoints to architecture (figure 1) are as follows:

Operational viewpoint, duties and activities: this viewpoint includes text and graphic artifacts that defines and describes the operational elements, assigned activities and duties and required processes or information between these nodes for performing a specific mission. For describing the flow of information, an information process is required to specify the information types, above information exchange frequency, tasks and duties that are protected and supported by this process as well as information exchanging method (only until clearing the manner of interaction between elements) [3].

Systems architecture viewpoint: This viewpoint presents the appropriate degree of interactivity (operational internodes) to a collection of required systems capabilities, identifies the current systems that are used for protection of operational requirements and facilitates the comparison of current implemented or planned systems with the required capabilities.

Technical viewpoint: This viewpoint provides the systems implementation guide and recommends the procedure of systems. These guides are used in affairs such as extraction of engineering specifications, implementation and execution of main blocks and establishment of production lines [9].

The architectural viewpoint is a technique including a collection of different standards, agreements, rules, conditions and states that is applied as a series of profiles to all systemic services, user interfaces and their communications. Each one of these profiles are related to one of system architectural viewpoints and in connection with specific operational viewpoint.

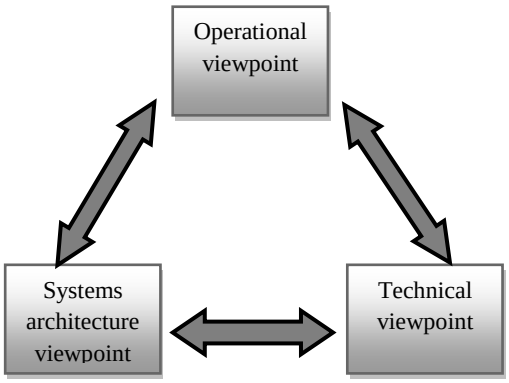



Figure 1 . Different viewpoints of C4ISR framework

### 2.2 Unified modeling language

Providing the model from a system causes to analyze the system from respective aspect. Various methods have been raised so far for developing the information systemsthat may be in general classified in two categories including structured and objectivistic. Lack of a unified modeling language in structured methods reduces the legibility and efficiency of these methods. This problem in the objectivistic methods was obviated by raising UML (unified modeling language) language. UML language as a powerful language is able to support all objectivistic concepts so that today in addition to objectivistic methodologies, this language is also used in objectivistic database. Uncertainty in information systems is unavoidable due to the normal nature of requirements. In order to develop the systems, objectivistic methodologies explain their respective concepts by UML language. The software may be described in different levels of abstraction and by different attitudes according to different diagrams of unified modeling language.

During software development process, quality of interaction between elements for performing the tasks is described by one of sequence or collaboration diagrams. The difference between these two diagrams is that the collaboration diagram focuses on participants, their role and relationship between them, whilst the sequence diagram shows the temporary execution of communication models. This diagram includes a group of roles that are played by the samples, moreover the required relations in this diagram are created in a specific context, in other word, this diagram shows a group of samples collaborated with each other together with their relations. In addition, this diagram may indicate an interaction that in this case, a group of messages are defined to determine the interaction between samples playing the roles in a cooperation for reaching to an appropriate result [7].

### 2.3 Colored Petri nets

These nets provide a graphic and clear exhibition together with a mathematical approach and may show the communication models, controller models and information process. These nets provide a framework for analysis, validation and evaluation of performance. The Petri nets are formed based on the graph and informally is a two-component directed graph consisted of place and transition elements. These nets are formed based on status not based on event and this explicit modeling of status makes any case possible. The Petri models provide the models of structural and behavioral aspects of an extensive event system. As well as, provide a framework for analysis, validation and evaluation of performance and reliability.

# 3. RELATED WORKS

Within the recent decades, efforts have been taken for presenting methods, models and frameworks of enterprise architecture evaluation including as follows:

A method for transforming UML diagrams to object stochastic activity networks (OSAN) has been analyzed. OSAN model due to supporting the objectivism and elements available in this model such as colored place etc. caused this model to be a capable tool for evaluation of architecture particularly evaluation of architecture performance [11].

In this method, the enterprise architecture program process is evaluated. The presented evaluation method evaluates the extraction process of enterprise architecture program's essential components in details that includes the current status architecture, appropriate status architecture and transmission strategy. In this method, the federal enterprise architecture framework has been used as an index and criterion for evaluation of enterprise architecture. So, upon evaluating the essential components of framework, the enterprise architecture evaluation has been achieved. In fact, the process evaluates the definition of each one essential components of enterprise architecture program separately and in details. One of important advantages of presented method may refer to precise evaluation manner and possibility of enterprise architecture program at each stage of its definition process [2].

A method has been raised for establishment of a formal model of UML diagrams. In this research, a methodology is presented for establishment of a formal model of system for analysis and UML behavioral modeling. In fact, UML diagrams indicate the sequence of an object's states during its life and UML collaboration diagrams have been transformed to object Petri networks. Then, using the presented formal model, the likelihood of UML behavioral characteristics has been analyzed to discover and detect the simultaneity-based behaviors [6].

In addition, a method has been presented for evaluating the accuracy of behavior and validation of UML sequence diagram. In this method, the messages origin and destination has been used in sequence diagram, and sequence diagrams have been mapped in promela language. Later, SPIN tools have been used for simulation [4].

Louis & Wagenhalder presented a method and evaluated the software architectures based on colored petri nets. The quality of architectural design of a software system has a great effect on achieving the nonfunctional requirements of a system. In this study, a technique has been presented for analysis and explanation of behavioral aspects of software architectures based on colored petri nets (CPNs) and a technique for evaluation of their quality by colored petri nets. In addition, models have been presented for evaluation of security, performance and reliability of software architecture and their integration by descriptions of colored petri nets. Then, their quality has been evaluated and exhibited on CPN analysis tools by means of simulation of colored petri nets. In the following research work, C4ISR architectural artifacts which have been produced by objectivistic approach have been transformed to colored petri nets for an executable model for evaluation [10].

# 4. THE PROPOSED METHOD

In the proposed method, among the common architectural frameworks, C4ISR framework has been considered. Whereas architectural framework C4ISR uses UML modeling language for modeling and architectural artifacts may be produced by UML modeling language. In this method, UML has been used for documentation of architecture. On the other side, colored petri nets that despite of simplicity have strong mathematical support as well as its supporting strong tools such as CPNTOOLS has been used for establishment of executable model of architecture.

To establish the executable model, firstly UML diagrams must be transformed to colored petri nets. Whereas in this paper, out of UML diagrams, collaboration diagram is used, the products are produced considering the requirement benefitting from C4ISR framework and UML diagrams have been transformed to colored petri nets by means of presented algorithm, and the system has been evaluated by annotations related to reliability. According to the results of simulating this model and their analysis we could identify the problems in planning phase and improve our artifacts to avoid the expensive time and economical costs of implementation.

## 4.1 Effect of collaboration diagrams on evaluation of reliability

The enterprise architecture specifying the structural and behavioral describes the system in high levels of abstraction. During the software development process, quality of interaction between for performing the duties is described by one of sequence or collaboration diagrams. Collaboration diagram similar to sequence diagram exhibits the messages transmitted among components based on their occurring time. This diagram shows the functionality of a use case. Moreover, it focuses on interaction between samples and is drawn based on messages transmitting time. To model the system load in collaboration diagram, specific elements and structures are used for reliability.

## 4.2 Algorithm for transforming UML collaboration diagrams to colored petri nets

This diagram shows the sequence of messages transmitted among elements based on their occurring time. Transformation is centralized on messages and their transmission procedure, as well as on the message transmitter and receiver object, therefore implementation of this method is simpler and its understanding is easier. The purpose at this stage is transmitting the message transmitter and receiver components to petri net. The collaboration diagram in reliability evaluation indicates that how a customer of a specific kind moves among service centers. For each collection of scenarios that attributed to a customer, a colored

may be considered for the token in petri net. In continue each one of message transmitter and receiver components and messages between them are transformed to petri nets. In this state, each one of message receiver and transmitter components is transformed to place-transition-place [4].

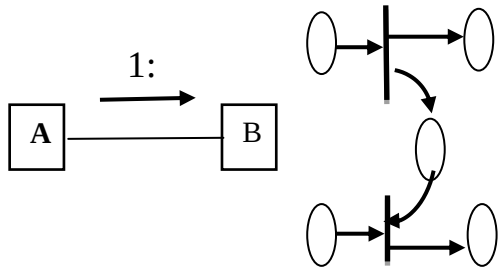



Figure 2. Asynchronous message and its equivalent colored petri net

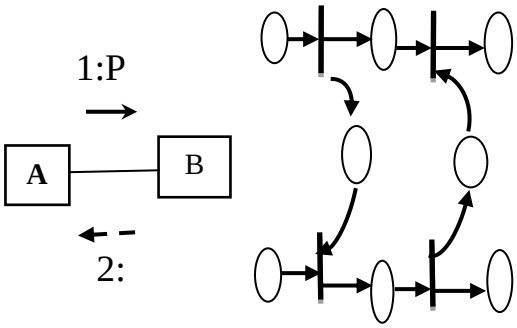



Figure 3. Synchronous message and its equivalent colored petri net

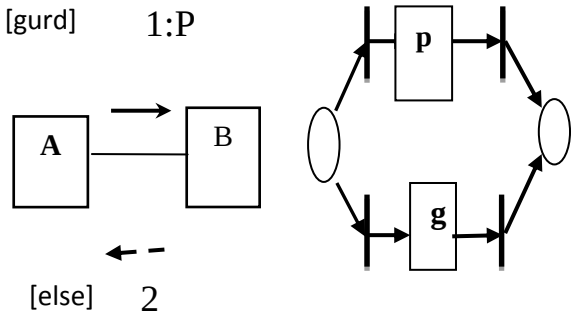



Figure 4. Selection structure and its equivalent colored petri net

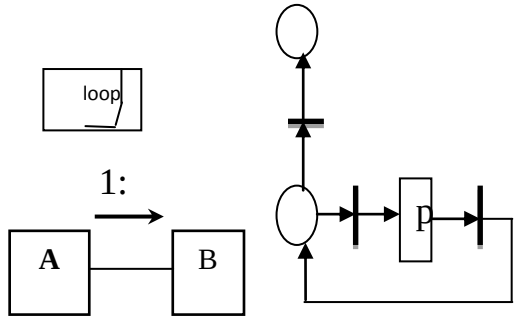



Figure 5. Ring structure and its equivalent colored petri net

## 4.3 Annotation related to reliability

In this paper, a collection of profiles has been introduced for evaluation of reliability. These profiles are inserted as stereotypes and tags to unified modeling language describing the architecture. The profiles provide the quantitative data in unified modeling language and quantitative analysis is provided by these data.To model the system load in collaboration diagram, specific elements and structures are used for reliability. In the first mode, it is assumed that the time spent for messages transmission is not very significant. A term is attributed to each message available in the diagram that explains the message transmission probability. Moreover, several terms that term has been attributed thereto may be transmitted from one point and implemented parallel. If the messages are placed in a repeating structure, may be sent for several times. This diagram is used for evaluation of performance, working load and delay of each message, and for evaluation of reliability, the periods of each component in busy mode and each scenario's failure probability are used.

## 4.4 Evaluation of reliability in system

In this paper, a method is presented for evaluation of reliability at architecture stage that reliability data are inserted as tag and stereotype to unified modeling language diagrams. The method is formed based on using three types of data including user profile, unified failure probability of each one of components and connectors and their productivity. REcomponent stereotype is annotated together with Recomfailprob and REpb. Moreover, the interaction between components is annotated with REconnector stereotype and REconnfailprob tag. Reliability prediction algorithm produces beta distribution from components failure rate and requires inserting confidence interval by the user for failure rate of each component. In this paper, it is assumed that components failure is independent from other components and not transferred to other components. After failure of each component and within two components was calculated in each transition, the mean failure probability in each transition is calculated according to equation (1) as below:

$$\theta_s = 1 - \sum_{j=1}^{k} pj \left[ \prod_{i=1}^{n} [(1] - \theta_i)^{bpij} \prod_{(j,j)} (1 - \psi l.1, j)^{linteract(l,i,j)} \right] \quad (1)$$

Ultimately, all transitions are calculated by a failure probability and service time. To calculate the reliability of the whole system, sum total of failures and service time in all transitions is calculated.

## 4.5 Creating the executable model

To create executable model in this paper, colored Petri nets and *CPN* Tools is used. In fact, an executable model of architecture is assumed as formal description of architecture through which we may evaluate the final behavior before implementation of architecture and get aware of problems and inefficiency and implement the architecture with better confidence and also avoid its extra costs even its failure. Furthermore, it calculates and analyzes evaluation of each one of respective meters of performance such as service time, failure rate in each component and between competent.

# 5. CASE STUDY

In this section, an applied example of E-business system is introduced and the algorithm presented in this paper is applied thereon. The respective reliability meters are examined by means of proposed method and simulation of executable model in CPN Tools.

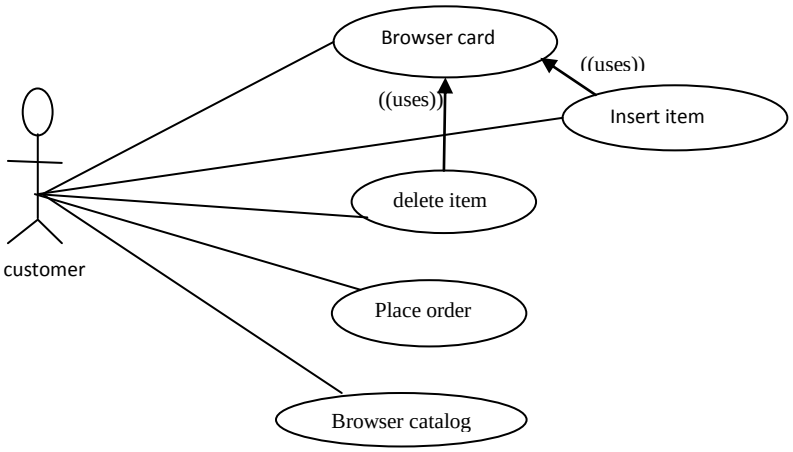



Figure 6. E-business use case diagram

To transform the UML diagrams to petri nets, out of UML diagrams, collaboration diagram has been used. Figure (7) shows the use case collaboration diagram of figure (6). In this diagram, the system reliability annotations have been performed for computation of failure of each component and between components.

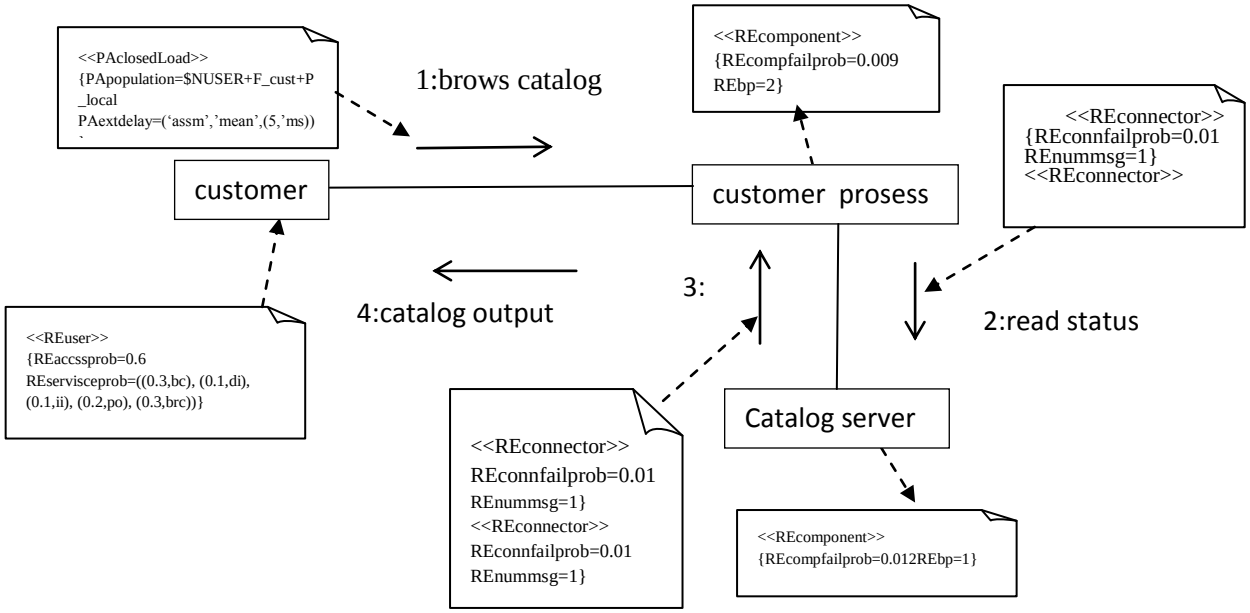



Figure 7. Collaboration diagram of catalog selection transition

To transform the collaboration diagram to colored petri net, the presented algorithm is used. Figure (8) shows the colored petri net related to catalogue selection.

To calculate the failure rate between two components, considering the performed annotations, one failure coefficient has been attributed to each one of components. Calculation of failure rate between components and reliability in system is simulated by applying the said equation for respective meters in CPN Tools and summary of results indicates the respective meters.

**Table 1. Failure rate and reliability of executable model in catalogue selection transition**

| Parameter | components | Failure rate | reliability |
|---|---|---|---|
| $\psi 1$ | Customer prosess | 0.0184 | 0.9817 |
| $\Theta 1$ | Catalog server | 0.0163 | 0.9930 |

According to equation (1), to compute the mean failure probability of system in catalogue selection transition and system's reliability, the following results are obtained:

**Table 2. Average of failure rate and reliability in transition of catalog selection**

| | |
|---|---|
| **Average of failure rate** | 0.01735 |
| **Average of reliability** | 0.98735 |

According to the model simulation in CPN Tools, evaluation of system requirements and features is available. Tables (1) and (2) provide the summary of colored petri net of catalogue selection. In continue, the proposed method is evaluated by other methods and results thereof are observed in table (3).

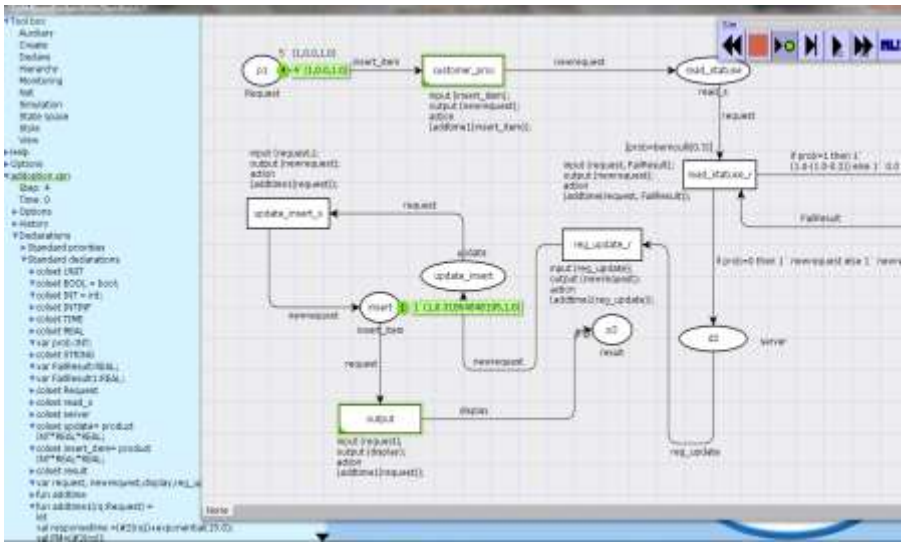


Figure 8.colored Petri net of catalog selection

**Table 3. evaluation of other methods with the proposed method**

| Evaluation criteria | Levis model | OSAN model | Archimate model | proposed model |
|---|---|---|---|---|
| Applicability | ✓ | ✓ | X | ✓ |
| Abilityof performance evaluation | ✓ | ✓ | ✓ | ✓ |
| Evaluation of reliability | X | X | ✓ | ✓ |
| Evaluation of other non- functional requirements | X | ✓ | X | X |
| Support of object-oriented | ✓ | ✓ | X | ✓ |

## 6. CONCLUSION

This paper firstly focused on evaluation of reliability in enterprise architecture level regardless of characteristics of hardware bedand then on an executable model based on petri nets of architectural descriptions. What is raised in this paper includes presenting a new method for evaluation of nonfunctional requirements of enterprise architecture. For this purpose, we attempted to present an executable model of architectural artifacts (UML diagrams) by means of colored petri nets so that can simulate the designed architecture and find out its weaknesses and strengths before execution of architecture and take the required measure for improvement of designed architectural artifacts. In order to establish an executable model for evaluation of reliability in system, UML collaboration diagram profiles means messages, messages transmission and receiving events and their origin and destination were used and the system requirements were explained as stereotype. Then, upon execution of model in CPN Tools, we simulated the architecture and according to the obtained results, examined if the established model meets the system features and requirements or not, so we could analyze the system reliability.

## 7. REFRENCES

[1] Esma,Y., Aubry, A. , Hervé ,P. 2008. Formal measures for semantic interoperability assessment in cooperative enterprise information systems ,Computers in Industry, Vol. 63, no. 5, pp. 443-457.

[2] Javadpour, R. 2006, Presentation of an Executable Model for Evaluation of Enterprise Architecture by means of Colored Petri Nets, M.Sc. Thesis, ShahidBeheshti University,

[3] Kamandi, A., Abdollahi, A. ,Movaghar ,A. 2006. Transformation of UML Models into Analyzable OSAN Models, Journal of Computational Information Systems,Vol.159 , no. 3 , pp. 3-22.

[4] Khayami, R, 2006.Qualitative characteristics of enterprise architecture. Procedia Computer Science, 3,1277–1282

[5] L. Xiao-hua, W. Gang, X. Lin, D. Maosheng, 2005. Reliability Analysis of Digital Protection's Sofware Based on Architecture, IEEE/PES Transmission and Distribution Conference & Exhibition: Asia and Pacific Dalian, China, 2005.

[6] Rezaei, R. 2006, Presentation of a Method for Evaluation of Enterprise Architecture, M.Sc. Thesis, Islamic Azad University, Science & Research Branch

[7] Rezaei, R., & Shams, F. 2008. A methodology to create data architecture in Zachman framework. WorldApplied Science Journal, 3(2), 343-349

[8] Saldhana , J. , Shatz , S. 1998. UML diagrams to object petri net models: an approach for modeling and analysis. Department of Software Engineering and Computer Science Blekinge Institute of Technology ,Vol. 13 , no. 2 , pp. 102-184.

[9] Saldhana, J. , Shatz , S. 2000. UML diagrams to object petri net models : an approachfor modeling and analysis, In proc Of international conference on softweareengineering and knowledge engineering ,Chicago , pp.103-110.

[10] Shin, M. , Levis, A. , Wagenhals, L. 2007. Transformation of UML-based System Model to Design/CPNmodel for Validating System Behavior In proc Of Compositional Verification of UML Models, Workshop on Compositional Verification of UML'03 conference ,USA ,pp.126-145.

[11] Wagenhals, L., Haider , S. ,Levis, A. 2003. Synthesizing Executable Models ofObject Oriented Architectures, Journal of Systems Engineering ,Vol. 6 , No. 4, pp52-94.

[12] Wagenhals,L., Shin, L. , Levis, A. 2006. Creating executable models of influence nets with coloured Petri nets, International Journal of STTT ,Spring , Vol. 87, No .4, pp15-103.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 63 - 67, 2014

# Job Scheduling on the Grid Environment using Max-Min Firefly Algorithm

Sajjad Asadzadeh Chalack
Department of Computer Science and Research Branch
Islamic Azad University
Ardabil, Iran

Seyed Naser Razavi
Computer Engineering Department, Faculty of Electrical and Computer Engineering, University of Tabriz, Iran

Ali Harounabadi
Member of Science Board of Computer Group in Azad Islamic university of Tehran Center, Iran

**Abstract**: Grid computing indeed is the next generation of distributed systems and its goals is creating a powerful virtual, great, and autonomous computer that is created using countless Heterogeneous resource with the purpose of sharing resources. Scheduling is one of the main steps to exploit the capabilities of emerging computing systems such as the grid. Scheduling of the jobs in computational grids due to Heterogeneous resources is known as an NP-Complete problem. Grid resources belong to different management domains and each applies different management policies. Since the nature of the grid is Heterogeneous and dynamic, techniques used in traditional systems cannot be applied to grid scheduling, therefore new methods must be found. This paper proposes a new algorithm which combines the firefly algorithm with the Max-Min algorithm for scheduling of jobs on the grid. The firefly algorithm is a new technique based on the swarm behavior that is inspired by social behavior of fireflies in nature. Fireflies move in the search space of problem to find the optimal or near-optimal solutions. Minimization of the makespan and flowtime of completing jobs simultaneously are the goals of this paper. Experiments and simulation results show that the proposed method has a better efficiency than other compared algorithms.

**Keywords**: Scheduling, Grid computing, Firefly algorithm, Max-Min algorithm, Makespan, Flowtime.

## 1. INTRODUCTION

Grid computing enables the sharing of a wide integrating of distributed resources including supercomputers, data storage systems, data resources, and as well as special tools that are available to certain organizations and provides the ability to solve complex problems in science, engineering and in commerce. The main motivation of the grid computing was created when the available resources in a range of management was not available to resolve a scientific problem which requires huge computations or data. There is an ability on grid computing that Influences human life surprisingly such as the effect of electricity grids, and be the next revolution after the Internet and World Wide Web revolutions. Grid is composed of a set of virtual machines that each has its varied resources and services and provides access to resources based on its special policy. Hence, grid's resources and services are very different and distributed in different geographical areas. Grid resources are recorded within one or more of information service. Users submit their requests to the resource broker and grid resources broker discoveries appropriate resources to apply this request by searching in grid resources, and then schedules them on the discovered resources. Computational grid is a software and hardware infrastructure that provides reliable, stable, comprehensive and cheap access to other local resources [4]. Computational grid is a shared environment that is implemented by establishment of lasting and standard service. These services support the creation and sharing of distributed resources. Nowadays increasing grid efficiency is a problem. To increase the efficiency of the grid, a properly and efficiency scheduling is needed. Unfortunately, dynamic nature of grid resources as well as different demands of users caused the complexity of grid scheduling problem. Dynamic of resource efficiency is due to Heterogeneous, autonomy and being shared of grid resources [7]. The goal of grid scheduling problem is the optimal assignment of jobs to resources.

Research has shown that heuristic optimization methods inspired from nature have more impact and efficient than other methods. Most of these methods try to minimize the maximum execution time of jobs. Swarm Intelligence is a kind of artificial intelligence methods based on swarm behavior. Many swarm Intelligence algorithms are proposed for optimization such as: Ant colony, Particle swarm and firefly algorithms. Among them, firefly algorithm is the best Heuristics method because of features such as high speed convergence, being insensitive to the initial values, flexibility and having the high error tolerability. On the other hand it has the local search, falls into the trap of local optimality and acts weakly in global search. It is proven that this algorithm efficiency is improved by combining other methods.

## 2. LITERATURE

Scheduling approach of Min-Min algorithm is a heuristic method that has a relatively reasonable efficiency and starts with a group of unallocated jobs that consists of two stages. In the first stage, a set of jobs are computed by minimum time. In the second stage, job is selected with the minimum completion time and be allocated to resources. Then allocated job is removed from unallocated jobs and this process is repeated for other unallocated jobs [5][2]. Scheduling approach of Max-Min algorithm is similar to Min-Min method and consists of two stages. In the first stage, a set of jobs are computed by minimum time. In the second stage, job is allocated to resources with the maximum completion. In most cases, efficiency and Max-Min load balancing is better than Min-Min in grid resources [5][3]. PSO is parallel search algorithms based on population that starts with a set of random answers (Particles), then PSO continues searching in problem space to find optimal answer by updating particles positions. Each particle is specified as multi-dimensional (depending on the type of problem). An important problem in

the use of particle swarm optimization algorithm for solving optimization problems is how to create a mapping between problem and particles vector. The problem dimension in this algorithm is the number of jobs considered. So length of each particle and the velocity vector is considered as the number of jobs. Each particle as $x_{id}$ has a position vector, a velocity vector and its own fitness value[6]. During each algorithm iteration, values of positions and velocity are changed by the flowing equations.

$$v_{id}(t+1)= wv_{id}(t)+c_1r_1(pBest_{id}- x_{id}(t))+ c_2r_2(gBest_d-x_{id}(t))$$

$$x_{id}(t+1)=x_{id}(t)+v_{id}(t+1))$$

In the above equation, w is the inertia weight factor, $pBest_{id}$ is the best previous position of particle, gBestd is the best previous position of all particles, $v_{id}$ is the velocity of ith particle at iteration t, $x_{id}$ is the position of particle ith at iteration t, $r_1$, $r_2$ are random numbers, and $c_1$, $c_2$ are constants [5].

## 3. SCHEDULING JOBS PROBLEM ON GRID

Scheduling problem of independent jobs, include N jobs and M machines. Each jobs should be processed somehow by each of M machines, to minimized the total length of schedule at last. In the proposed algorithm the quality of service parameters, makespan and flowtime of jobs are considered respectively. Each job can be run only on one source and does not stop until the end of the run. In the proposed algorithm the ETC matrix model is used that is explained in [3]. Since the scheduling algorithm proposed is the static, it is assumed that expected execution time for each job j on each resource i, has already been determined and is located within the matrix ETC [i, j].

Completion_Time [i,j] is equal time that job j be completed on resource i and is computed as follows.

(1)Completion_Time[i,j]= ETC[i,j]

Makespan: maximum completion_Time [i,j], that is computed as follows.

(2)Makespan= Max (Completion_Time[i,j]) $1\leq j\leq N$ , $1\leq i\leq M$

Flowtime: sum of the completion time of jobs [i, j] over all resources, that is computed by follow equation.

(3) $\text{flowtime} = \sum_{i=1}^{m} \sum \text{completion\_Time}[i,j]$

The goal of scheduling in the proposed algorithm is to submit each of the jobs to each of the resources to minimize makespan and flowtime of the jobs at last.

## 4. THE PROPOSED SCHEDULING ALGORITHM

In the proposed scheduling algorithm, hybridization of firefly with Max-Min algorithm in the scheduling problem solving of independent jobs in the computational grid is used. Before presenting the algorithm, it is necessary to examine what parameters are needed to solve the scheduling problems using the firefly algorithm[8].

### 4.1 Fireflies representation

Search procedure in the firefly algorithm is each firefly compares with others, if firefly light is less than compared firefly, moves towards firefly with more lights (problem of finding maximum point), this act causes particles focuses around a particle which has more lights, and in the next iteration of the algorithm, if a particle with more lights exist, particles move towards it again. Search stages must be iterated in more numbers. If the population size is increased in the wide problem space such as grid system, time complexity of algorithm rises and algorithm efficiency falls strongly. In such algorithms, the solution of this problem in generating initial population. This means that we should apply solutions to reduce initial population size as far as possible. For this goal we should generate more qualitative initial population until the number of comparisons reduces. So we can reduce the number of iterations of algorithms and achieve to optimal or near-optimal solution in the less time. We have used the Max-Min algorithm to generate part of the initial population. One of the features of this algorithm is that maintains load balancing and establishes it at the same first stage. One of the most important points in the use of firefly algorithm in scheduling problem solving of independent jobs is that how to convert a scheduling problem to a solution, or indeed how can create a mapping between solutions and fireflies in firefly algorithm. In scheduling firefly algorithm, each firefly is a solution for allocation of tasks, so the length of each firefly vector is N, which N is the total number of input tasks. Each element inside the firefly vector is a random number between 1 to M (M is the total number of resources).

| | T1 | T2 | T3 | T4 | T5 |
|---|---|---|---|---|---|
| **Firefly1** | R2 | R5 | R3 | R2 | R1 |
| **Firefly2** | R1 | R4 | R5 | R3 | R2 |
| **Firefly3** | R1 | R3 | R5 | R4 | R2 |

Figure 1. Typical Fireflies

### 4.2 Generation of the initial population of fireflies

In the proposed method, a part of the initial population individuals is generated by Max-Min algorithm and some randomly. With this method the generated population has a good qualification and the other part of the population such that a random number between 1 to M indicating the number of resource is generated until the job specified is executed on it. Randomness helps to maintain population variety and the selection chance to be given to individuals of the population. The length of each of fireflies would be equivalent to the number of jobs. The size of fireflies indicates the number of candidate solutions or the value of searching in the problem space.

### 4.2.1 The method of generating primary population by Max-Min algorithm

Max-Min scheduling approach is a heuristic method. At first a set of jobs with minimum completion time is computed and the job with maximum completion time is selected and allocated to the resources.

For example, matrix ETC in table 2 is indicated for 6 jobs over 4 resources.

**Table 1. Matrix ETC for 6 jobs over 4 resources**

| Task\Resource | $R_0$ | $R_1$ | $R_2$ | $R_3$ |
|---|---|---|---|---|
| $T_0$ | 200 | 250 | 220 | 300 |
| $T_1$ | 150 | 170 | 190 | 160 |
| $T_2$ | 300 | 320 | 180 | 360 |

| | | | | |
|---|---|---|---|---|
| **$T_3$** | 400 | 380 | 350 | 310 |
| **$T_4$** | 100 | 120 | 140 | 160 |
| **$T_5$** | 220 | 250 | 280 | 200 |

According to the Max-Min method, a set of jobs with minimum completion time of each matrix row is computed and job with maximum completion time is selected and allocated to resources. The result of this job is as follows.

[200, 150, 180, 310, 100, 220]

[310]

Namely the task of $T_3$ is allocated to $R_3$ resources. This job is done for other unallocated tasks with regard to executed time of $R_3$ resources. Finally the diagram of tasks allocation to resources is in the form of figure (2).

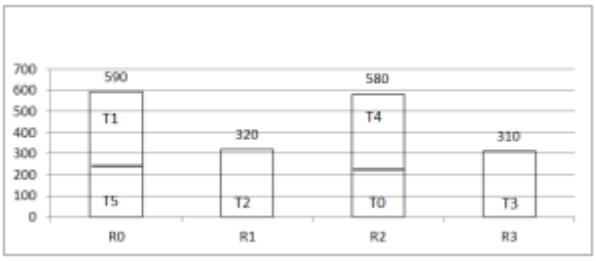



Figure 2. Tasks allocation by Max-Min algorithm

## 4.3 Evaluation of the fireflies

The next stage is measurement of fireflies light that depends on the considered problem. For this goal we have used the multi objective fitness function that includes two parameters of QOS (flowtime, makespan) for evaluation of fireflies.

### 4.3.1 The first fitness function

One of the parameters of evaluating the goodness of schedules in this problem is measurement of makespan that can be computed using the following equation.

(4) Min( Max(Completion_Time[i,j]))

### 4.3.2 The second fitness function

Sum of the completion time of all jobs should be minimum flowtime that can be computed by the following equation.

(5) $\text{Min}(\sum_{i=1}^{m} \sum \text{completion}_{\text{time}i,[i,j]})$

## 4.4 Distance

Distance between any two firefly i and j such as $x_i$ and $x_j$ can be defined by Cartesian distance $R_{ij}$ using the below equation respectively, such that $x_{i,k}$ is the k'th section of the spatial coordinates $x_i$ of firefly i, and d is the number of dimensions of the problem.

(6) $r_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^{d}(x_{i,k} - x_{j,k})^2}$

## 4.5 The attractiveness

Computation of the attractiveness function for a firefly is shown in the following equation, where r is the distance between each pair of fireflies, $\beta_0$ is the first attractiveness (r=0), and $\gamma$ is the absorption coefficient that controls light intensity.

$$\beta_0 = \beta_0 e^{-\gamma r_{i,j}^2} \quad with \quad m \geq 1$$

(7)

## 4.6 The movement

The firefly I is attracted by firefly j, that is brighter, according to the equation (7), where $x_i$ is the current position or solution of a firefly, $\beta_0 e^{\gamma r_{i,j}^2 (x_i - x_j)}$ is the rate absorption of the firefly by adjacent fireflies. *α(rand-1/2)* is the random movement rate of a firefly, α coefficient is a random parameter with the problem interest, α=[0-1]whereas rand of the obtained numbers is determined from the uniform distribution in the space.

(8) $x_i(t+1) = x_i(t) + \beta_0 e^{-\gamma r_{ij}^2}(x_i - x_j) + \alpha(rand - 1/2),$

$$x_i(t+1) = x_i(t) + \alpha(rand - 1/2)$$

## 4.7 Termination conditions

To finish of swarm Intelligence algorithms such as firefly, it must be mentioned the termination conditions. This algorithm will be terminated after reaching maximum iteration.

# 5. PSEUDO CODE OF THE PROPOSED ALGORITHM

Max-Min Firefly Algorithm

*1. Begin*

*2. Finding some part of the solutions by Max-Min*

*3. Initialize population by the result of Step 2*

*4. Objective function f(x), x = (x1,..., xd)T*

*5. Generate initial population of fireflies xi (i = 1, 2,..., n)*

*6. Light intensity Ii at xi is determined by f(xi)*

*7. Define light absorption coefficient γ*

*8. while (t <MaxGeneration)*

*a. for i = 1 : n all n fireflies*

*b. for j = 1 : n all n fireflies (inner loop)*

*c. if ($I_i$ < $I_j$), Move firefly I towards j; end if*

*c.1. Vary attractiveness with distance r via exp[-γr]*

*c.2. Evaluate new solutions and update light intensity*

*end for j*

*end for i*

*9. Rank the fireflies and find the current global best g**

*10. End while*

# 6. SIMULATION

In this section, efficiency of the presented algorithm in the previous section that was carrying out scheduling jobs in a computational grid will be evaluated. This algorithm tries to carry out scheduling act of a number of independent jobs in a grid media. This jobs are belong to an application that user is delivered it to grid for run. The user accompany with program specifies the quality of considered service, namely time optimization strategy for system. By selecting time optimization strategy, user can ask from grid system to run its applied program in the least possible time. It can be shown by

simulation that how is the efficiency of time optimization algorithm in compared together. All of the experiments is done on a system with dual core processor of 2.40 MHz, 3 GB memory and windows 7. Simulation is done by Matlab R2010a and all of the algorithms are simulated in this environment. The proposed algorithm with several other scheduling algorithms is examined according to conditions of table 2 and for suitable consider of jobs length is considered equal in all models experiment. The parameter of the range of jobs length indicates the range of uniform distribution of jobs length. The numbers of jobs parameter indicates that for obtaining runtime of program using the available algorithms, 400 repeats is done and then the mean values are selected for consideration. Figure (3), shows that in the two proposed methods, makespan is minimized than other compared algorithms. Figure (4), shows that in the two proposed methods, flowtime is minimized than other compared algorithms.

**Table 2. Primary values of scheduling algorithms parameters**

| Algorithm | Parameter | Value |
|---|---|---|
| **Firefly** | Population size | 40 |
| | light absorption coefficient($\gamma$) | 1 |
| | the randomization parameter($\alpha$) | 0.2 |
| | maximum attractiveness value($\beta_0$) | 2.0 |
| **Particle swarm optimization** | Population size | 40 |
| | Self-consciousness study factor $C_1$ | 1.49 |
| | Swarm consciousness study factor $C_2$ | 1.49 |
| | Inertia factor | 0.9 |

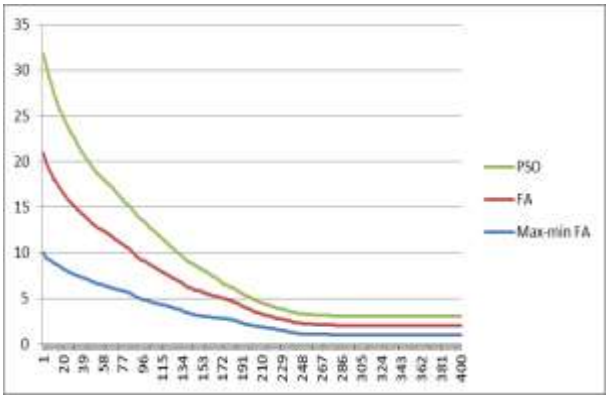



Figure 3. Diagram of makespan

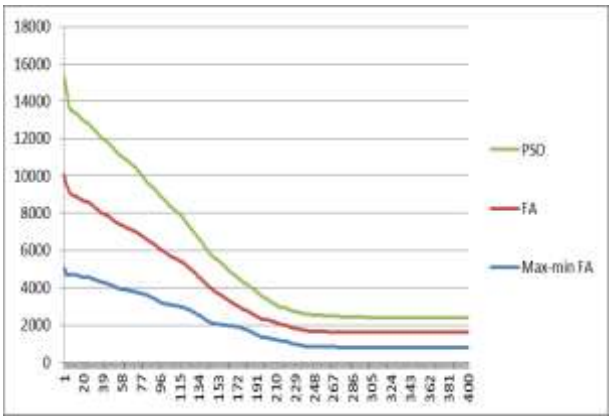



Figure 4. Diagram of flowtime

## 7. CONCLUSION

The computational grids provide reliable and cheap available to other computational resources. These resources are as Heterogeneous and distributed and are used shared. On the other hand, resources in grid are belonged to various organizations that have specific management policy and used for different users at different times. Hence, owners and users of resources have different aims, strategies and supply and demand. In this complicated media management it cannot be used traditional methods for resources management that try to optimize the efficiency rate at the system level. In this paper, proposed method was presented for scheduling jobs in computational grid. In the proposed method combination of the firefly, Max-Min was used. The most concentration was over two factors, first was generation of primary population done using Max-Min algorithm and cause to be better and thus the convergence algorithm speed rises and the optimum solution is reached sooner. The second factor was evaluation of the solutions problem. For this purpose we used multi purposes function. Simultaneously two parameters makespan and flowtime evaluate service quality and minimum sum of the three mentioned parameters. In the proposed method, we have improved the mentioned parameters of service quality such as time of jobs implementation. Mentioned parameters are simulated carefully. The results show the superiority of the proposed method than the compared methods.

## 8. REFERENCES

[1] Braun, T.D., et al. 2001 A Comparison of Eleven Static Heuristics for Mapping a Class of Independent Tasks onto Heterogeneous Distributed Computing Systems. Journal of Parallel and Distributed computing.

[2] Casavant, T.L., Kuhl, J.G. 1988 A taxonomy of scheduling in general-purpose distributed computing systems. IEEE Transactions on Software Engineering.

[3] Chauhan, S.S. and R. Joshi. 2010 A weighted mean time min-min max-min selective scheduling strategy for independent tasks on grid. Advance Computing Conference (IACC) IEEE 2nd International Patiala.

[4] Foster, I., Kesselman, C., Nick, J. and Tuecke, S. 2002 The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration, Computer.

[5] He, X.S., Sun, X.H. and G. Von L. 2003 QoS guided min-min heuristic for grid task scheduling. Journal of Computer Science and Technology.

[6] Hesam, I., Behrouz, T.L., Ajith, A., Vaclav, S. 2010 A DISCRETE PARTICLE SWARM OPTIMIZATION APPROACH FOR GRID JOB SCHEDULING. International Journal of Innovative Computing, Information and Control.

[7] Sarkar, V. 1989 Determining average program execution times and their variance. in: Proceedings of the ACM SIGPLAN Conference on Programming Language Design and Implementation.

[8] Yang, X. 2010 Nature-Inspired Metaheuristic Algoritm, University of Cambridge: Luniver Press.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 68 - 72, 2014

# A Survey of Job Scheduling Algorithms Whit Hierarchical Structure to Load Balancing In Grid Computing Environments

Maryam Masoudi Khorsand
Departement of
Computer Engineering
Islamic Azad University
Sciences and Research Branch
Ardabil, Iran

Shahram Jamali
Department of
Computer Engineering
University of
Mohaghegh Ardabili
Ardabil, Iran

Morteza Analoui
Department of
Computer Engineering
University of
Science and Technology
Tehran, Iran

**Abstract**: Due to the advances in human civilization, problems in science and engineering are becoming more complicated than ever before. To solve these complicated problems, grid computing becomes a popular tool. a grid environment collects, integrates, and uses heterogeneous or homogeneous resources scattered around the globe by a high-speed network. Scheduling problems are at the heart of any Grid-like computational system. a good scheduling algorithm can assign jobs to resources efficiently and can balance the system load. in this paper we survey three algorithms for grid scheduling and compare benefit and disadvantages of their based on makespan**.**

**Keywords**: Grid Computing; Load Balancing; Hierarchical; Scheduling; Makespan.

## 1. INTRODUCTION

Computational grids as a new approach to solve large-scale problems in science, engineering and business have been most focused on large scale resource sharing, cooperation of several organizations and their use in new applications. Thus Grid computing systems involve a collection of programs and resources that are distributed among the machines of the grid. [1]. The goal of grid computing is to manage the system such that set the jobs to be done in shortest time. one of the aspects that improve the efficiency and decrease the time to performance, is the jobs Scheduling. the concept of scheduling jobs this is that to be specified jobs when and in which machine be performed. in grid environment, scheduling and effective management of resources is vital and important. appropriate scheduling causes the one of the best service to users and applications, and from the other side from all the grid resources, and occur optimal and maximum utilization. the other hand fair load divided into between computers in network is from goals and benefits in grid computing. dividing the load on the grid means that the grid management system for any assigned application, select the best source for run an program assigned to the that computer to runing. therefore, by jobs effective scheduling using Appropriate distribution of the applications among processor resources, the total load will be divided among all computers.[2]

In this paper be reviewed scheduling algorithms with hierarchical structure that overall goal in all of them is increase the efficiency and reduce the makespan for grid computing.

## 2. BACKGROUND

### 2.1 Grid Computing

The grid is a network of computers and supercomputers is capable of much higher speed data processing made it possible. In other words, the execution time of jobs that would take hours to cut a few seconds. Grid computing is a new technology that the goal is to share resources for the job to come. In computing grid, job scheduling is a very important problem. A good scheduling algorithm can assign jobs to resources efficiently and can balance the system load. Grid computing has tried everything so far in the field of high performance computing, including cluster computing, peer-to-peer computing, which is done brought together.

### 2.2 Resource Discovery In Computational Grid

In grid environment, resource is a reusable entity that serves to fulfill a job or request is. Resource can be a car or any other service network (a combination of machines, networks, and software) is. The grid does not have any limitations in terms of geographic area coverage and the resources available in a grid of geographically distributed, heterogeneous, and may also be placed in different locations have the resources belong to different organizations. shared resources in a Grid environment are May be a type of desktop systems, clusters, large data sets, storage devices, etc.

When a remote user to a remote resource to run or access a data the resource is required. The mechanism that the grid provides the structure must be able to appropriate resource for the user to discover. Thus one of the most basic jobs in a Grid is support of resource discovery mechanism. The purpose of discovery of resources, identify a list of available resources to assigning jobs. When a user requests a resource in the Grid environment to run a specific application, a method of discover the resource must be able to discover resources to user requests with low cost and time for the user so the user can quickly to the resource uses. Grid resource discovery algorithms in different ways to use it. Based on resource discovery are three models: centralized, distributed and

hierarchical. Went on to is a summary of the hierarchical model, we evaluate the algorithms, this paper benefited of this model.

## 2.3 Model of Hierarchical Resource Discovery

In this model, there is a central resource management system and a number of local resource management system and a central resource management system with local resource management systems to discover of resources interact does. One way Interact from the central management of to local schedulers, as shown in Figure 1.

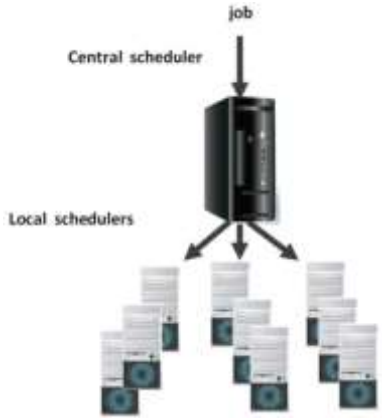



Figure1. Hierarchical scheduling model

As the centralized model, this model has problem of extensibility and fault-tolerance in that is low. But compared with the centralized model, an advantage of the hierarchical model is the central resource management and local resource management systems may have different policies.

## 2.4 Strategies of Load Balancing Algorithms

For the proper distribution of users requests, load balancing is very important in the grid environment. The load balancer receives users request and sends them to suitable resources. according to some specific distribution logic. Generally speaking, load balancing can be classified as static, dynamic, and adaptive. In static load balancing, the decision regarding the allocating resources to requests is predefined. In dynamic balancing, the request allocation decisions are decided at run time based on the current state of the system. In adaptive balancing, the allocation parameters and polices can vary depending on the run time information about the system, such as the current state of the system, previous decisions, etc [3,4].

## 3. REVIEWS OF HIERARCHICAL SCHEDULING ALGORITHMS

In the literature, many scheduling algorithms have been proposed. Most of them can be applied to the grid environment with suitable modifications. In general, they can be separated into two types: batch mode and on-line mode. In this section, we will introduce some scheduling algorithms in on-line mode.

### 3.1 On-line Mode Heuristic Scheduling Algorithm

Jobs are scheduled when they arrive. Since the Grid environment is a heterogeneous system and the speed of each processor varies quickly, the on-line mode heuristic scheduling algorithms are more appropriate for the Grid environment.

### 3.1.1 Most Fit Job Scheduling Algorithm (MFTF)

MFTF algorithm tries to make jobs compatible with resources for job transmitter and assigns the resource to transmitter based on fitness function. [5]

Fitness function is calculated as follows:

$$\text{fitness}(i,j) = \frac{10000}{1+|W_i/S_i - E_i|} \quad (1)$$

$W_i$ is load of ith job $S_j$ is CPU speed in jth node. $E_i$ is expected time of ith job. Wi/Sj is the performance time used in the given node. Wi/Sj – Ei is the performance time of predicted jobs. Ei is identified by machines or transmitters:

$$E_i = A + n \times S \quad (2)$$

If A is mean responding time for 100 done jobs. N is nonnegative cardinal number and S is standard deviation of responding time for 100 done jobs. When performance time is measured near $E_i$, it means that this node is more suitable for job to be assigned. Since some processors may have much load and others remain free, this algorithm improves the efficiency of distributed system through load coordination and process capability of the system to smooth the periods with load traffic in the nodes. This is done by transmitting some loads from the nodes with heavy load to other processors in order to process.

Although many problems related to scheduling are solved with this algorithm, it can occur in the real environments with wrong scheduling because this algorithm doesn't consider the efficiency of resources.

### 3.1.2 Dynamic Load Balancing Algorithm (DLBA)

In [6] the authors propose an scheduling algorithm that performs an intra-cluster and inter-cluster load balancing. This algorithm in load index and other common parameters of each node in dynamic scheduling, the jobs. Intra-cluster load balancing is performed depending on the cluster manager cluster manager decides whether to begin the local balancing based on the current workload of the cluster of its child nodes. Inter-cluster load balancing is performed when some cluster manager unsuccessful to balance their workload. The local balancing unsuccessful may be due to a saturation of the cluster. In this position the cluster jobs excessive loaded,

according to selection strategies, has will be transferred to another cluster which is underloaded.

to justify the cluster is overloaded or not, they were introduced threshold called a threshold balance as Ψ. If the load of cluster more than Ψ, load balancing will be executed. This algorithm is feasible and improves the performance of the system.

However, the value of balanced threshold is fixed and set by its cluster, but the balanced threshold may not be suitable for the dynamic environment in the Grid system.

### 3.1.3 The Hierarchical Load Balancing Algorithm (HLBA)

This algorithm, adaptive load balancing between clusters. Since assume jobs are computing-intensive jobs, and consider the computing power of each resource as the standard for the selecting resources.

Therefore, the system can assign the job to the appropriate resource according to the updated information that arrived from information server. The Information Server discovers resource nodes registered with the system, and records the information of the resource, such as CPU speed, idle CPU percentage, memory utilization, average load of each cluster, etc.

The Portal provides an interface for users to submit jobs. The job scheduler accepts the job from the portal and uses the HLBA to choose the proper cluster and compare its load with the system.

Then, it selects the resource with the integrates computing power in the cluster to execute the submitted job. After the job is finished, the result and the new status of the resource will be sent back to the Information Service for another scheduling[7].The system framework is shown in Figure 2.

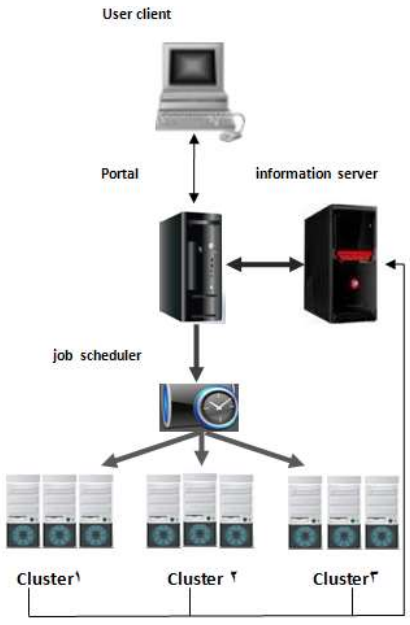



Figure2. system structure for HLBA

When the scheduler receives a job submitted by a user, it will transfer a request to the Information Service in order to obtain the necessary information such as the idle CPU percentage of each resource, average load of each cluster and average load of the system. Then the scheduler chooses a cluster which has the fastest average computing power (ACP). The average computing power of the cluster is defined as:

$$ACP = \frac{\sum_{k=1}^{n} CPU\_SPEED_k \times (1 - CPU_k)}{n} \quad (3)$$

where CPU-Speed$_k$ is the CPU MIPS of resource k in cluster i; CPU$_k$ is the current CPU utilization of the resource k in the cluster i, expressed as a percentage, and n is the number of resources in cluster i.

After the scheduler selects the cluster which has the fastest ACP, it will compare the average load of the chosen cluster with the average load of the system. The average load of the cluster is defined by the average load of each resource in cluster i. Load$_{k,i}$ represents the load of each resource k in the cluster i.

In HLBA, consider three load attributes, CPU utilization of the resource (CPU$_k$), the memory utilization of the resource (MU$_k$) and the utilization of network (NU$_k$). So the Load$_{k,i}$ may become:

$$Load_{k,i} = \sqrt{\alpha_1 CPU_k^2 + \alpha_2 NU_k^2 + \alpha_3 MU_k^2} \quad (4)$$

where a1 is the weight of the load attribute CPU$_k$, a2 is the weight of the load attribute NU$_k$, and a3 is the weight of load attribute MU$_k$. The average load of each cluster i is defined as

$$ALC_i = \frac{1}{n} \sum_{k=1}^{n} Load_{k,i} \quad (5)$$

The average load of the system is defined as:

$$AL = \frac{1}{m} \sum_{i=1}^{m} ALC_i \quad (6)$$

the average load of each cluster i, ALCi , to be less than the balance threshold of the system. Hence, a threshold called balance threshold, denoted as Ψ and defined as below:

$$\Psi = AL + \sigma \quad (7)$$

where σ is the standard deviation of the load of the system and defined as below:

$$\sigma = \sqrt{\frac{1}{m} \sum_{i=1}^{N} (x_i - \bar{x})^2}, \quad \text{for all i} \quad (8)$$

where x is equals AL and xi is equals ALC, and N is the number of clusters in the system.

When a job is to be assigned to a cluster with the highest ACP$_i$, the load of the selected cluster will be checked first to see if it is already overloaded. If the cluster's average load is more than Ψ, the cluster is marked as overloaded and the job will seek the cluster with the next highest ACP. If a cluster is

assigned a job, local update will perform. Namely recalculate $ACP_i$, $Load_{k,i}$, $ALC_i$, AL, and Ψ immediately.
It will also affect the global update. But the average computing power and average load of other clusters are not affected. Therefore, the computation cost is less than a global update described below. The overall flowchart of the system is shown in Figure 3.

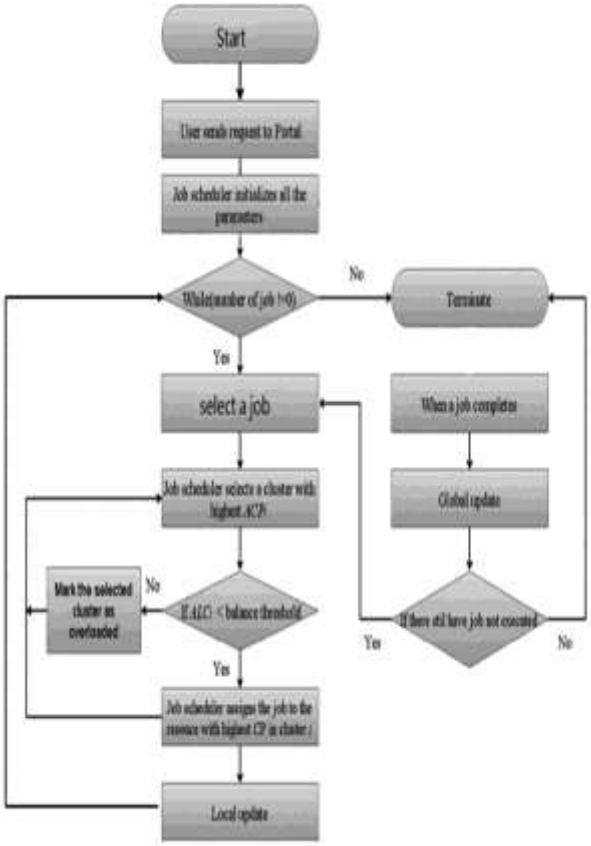



Figure3. Flowchart of HLBA

When a job is completed, the job had to release all resources and a global update will recalculate all Parameters again. Overloaded clusters can become available again after a global update.

### 3.1.4 Improvement of Hierarchy Load Balancing Algorithm (IHLBA)

This algorithm is an improved algorithm from HLBA. In this algorithm use the same parameters for scheduling but with different orders in HLBA. In HLBA, first choose the cluster with highest average computing power than compare its average load with Ψ. Considering Figure 3 there is a loop while selecting the cluster. This step may be the loss of time and impact on the makespan. when scheduler receives a job and obtains necessary information from the job, will sort clusters by their average loads If the average load of cluster exceeds Ψ, means that the cluster is overloaded. In IHLBA sort the clusters which are underloaded and choose the cluster with the maximum ACP within those clusters. After choosing the proper cluster, then select the resource with the best computing power in this cluster and assign the job. Local update and global update are also done in IHLBA [7]. If jobs have dependences and/or communications, their precedence relationship graph must first be drawn. Then a linearization is performed on this precedence graph. Jobs are then scheduled according to this linear order. If two jobs need to be executed at the same, they can be scheduled together as one job. The overall flowchart of the system is shown in Figure 4.

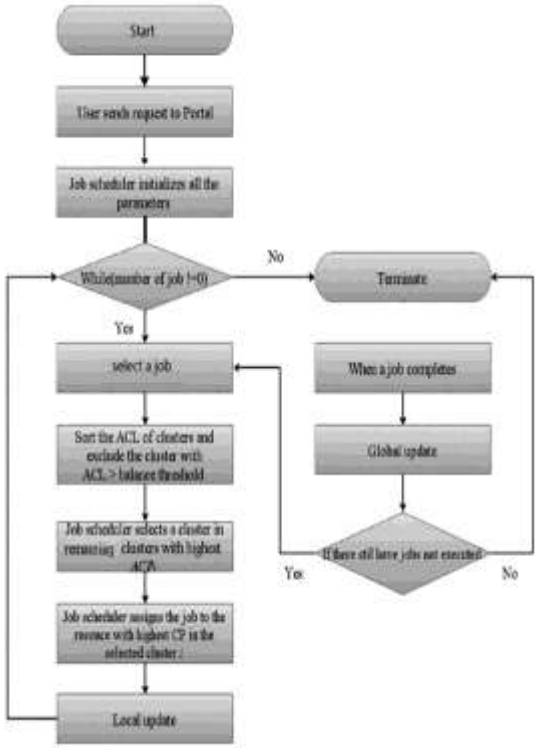



Figure4. Flowchart of IHLBA

assume that the cluster is static in this paper. Namely, the resources in a cluster are static. A dynamic cluster can also be handled by this algorithm since the dynamicity will be reflected in the cluster parameters. Another question is how the range of average computing power of clusters might affect the performance of this algorithm.
also, when a job arrives, the scheduler selects a cluster which is ALC Less than balance threshold and assigns the job to the resource with the maximum computing power in the cluster.

At the beginning of scheduling, the job scheduler will assign jobs to the resources in this cluster because it has the more average computing power than others. As time goes by, the load of the cluster becomes heavy and the ALC of the cluster exceeds the balance threshold. Therefore, in IHLBA algorithm may be removed when the cluster is overloaded, however is a cluster whit high throughput. Thus, this clusters because has a little overload is not allowed to participate in competition clusters for resource allocation.

## 4. THE COMPARISON OF MENTIONED SCHEDULING ALGORITHM

In Table 1 Scheduling algorithms listed are compared for load balancing and makespan. The benefit and the disadvantages of for each is checked.

Table1. comparison of scheduling algorithms

| parameters | advantages | disadvantages |
|---|---|---|
| MFTF Algorithm | It is compatible. It has high resistance in dynamic environments. | It doesn't consider the efficiency of resources. |
| DLBA Algorithm | This algorithm causing to load balancing. | Balance threshold may not appropriate for the characteristics of the dynamic grid systems. |
| HLBA Algorithm | This algorithm causing to load balancing and is appropriate for grid dynamic environments | In the time computing, it is not optimum. |
| IHLBA Algorithm | Cause Load balancing is and the ratio of the HLBA use optimum of time. | Due to few Overload may be to ignore the high processing resources |

## 5. CONCLUSION AND FUTURE STUDIES

In this paper, four scheduling algorithm based on load coordination whit hierarchical structure in the grid systems were investigated. although there have been a lot of studies concerning job scheduling in grid systems, a new challenge still can be interesting and many research projects can be done. The present study was an attempt to focus on current scheduling algorithms and the mentioned algorithms were compared considering a variety of aspect and advantages and disadvantages of each one were explained. In the future, these algorithms can be tested on heterogeneous processors.

## 6. REFERENCES

[1] Arora, S.K, Das, R and Biswas A. 2002. Decentralized Scheduling and Load Balancing Algorithm for Heterogeneous Grid Environments, in Proc. Of International Conference on Parallel Processing Workshops (ICPPW'02), pp. 499 – 505

[2] Braun, R., Siegel H., Beck N, Boloni .L and Maheswaran M. 2001. A Comparison of Eleven Static Heuristics for Mapping a class of Independent Jobs onto Heterogeneous Distributed Computing Systems, Journal of Parallel and Distributed Computing, pp. 810-837.

[3] Casavant, T.L. and Kuhl J.G. 1988. A taxonomy of scheduling in general purpose distributed computing system, IEEE Transaction on Software Engineering 14 (2) 141–154.

[4] Mehta, H, Kanungo P and Chandwani, M. 2008. Performance enhancement of scheduling algorithms in web server clusters using improved dynamic load balancing policies, in: 2nd National Conference, INDIACom-2008 Computing For Nation Development, New Delhi, Feb, pp. 651–656.

[5] Wang S.D., Hsu I.T. and Huang. Z.Y. 2005. Dynamic scheduling methods for computational grid environments, International Conference on Parallel and Distributed Systems pp 22–28.

[6] Suri P.K. and Manpreet S. 2010. An efficient decentralized load balancing algorithm for grid, in: IEEE 2nd International Advance Computing Conference, IACC, pp. 10–13.

[7] Lee, Y.H., Leu, S and Chang, R,S. 2011. Improving job scheduling algorithms in a grid environment, Future Generation Computer Systems 27, pp.991-998.

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 73 - 78, 2014

# Modeling and Evaluation of Performance and Reliability of Component-based Software Systems using Formal Models

Fakhraddin Farjaminejad
Department of Computer, Science and Research branch, Islamic Azad University, Ardabil, Iran

Ali Harounabadi
Department of Computer Engineering, Central Branch, Islamic Azad University, Tehran, Iran

Sayed Javad Mirabedini
Department of Computer Engineering, Central Branch, Islamic Azad University, Tehran, Iran

**Abstract**: Validation of software systems is very useful at the primary stages of their development cycle. Evaluation of functional requirements is supported by clear and appropriate approaches, but there is no similar strategy for evaluation of non-functional requirements (such as performance and reliability). Whereas establishing the non-functional requirements have significant effect on success of software systems, therefore considerable necessities are needed for evaluation of non-functional requirements. Also, if the software performance has been specified based on performance models, may be evaluated at the primary stages of software development cycle. Therefore, modeling and evaluation of non-functional requirements in software architecture level, that are designed at the primary stages of software systems development cycle and prior to implementation, will be very effective.
We propose an approach for evaluate the performance and reliability of software systems, based on formal models (hierarchical timed colored petri nets) in software architecture level. In this approach, the software architecture is described by UML use case, activity and component diagrams, then UML model is transformed to an executable model based on hierarchical timed colored petri nets (HTCPN) by a proposed algorithm. Consequently, upon execution of an executive model and analysis of its results, non-functional requirements including performance (such as response time) and reliability may be evaluated in software architecture level.

**Keywords**: software architecture; colored petri nets; object-oriented design; non-functional requirements; unified modeling language; performance modeling

## 1. INTRODUCTION

Within the recent decades, the software complexities have been increased day to day and demands for more powerful and high quality software have been increased. Therefore, software development based on principles and methodologies that in addition to reduction of costs, meet all expected features of shareholders (functional and non-functional requirements) seems to be necessary. Establishing non-functional requirements in software engineering was raised recently whilst they have considerable effect on success of software systems. Software Architecture (SA) is established at the first stages of design and has a significant effect on access to nonfunctional requirements of software system. Therefore, establishment of an executive model of SA and evaluation of nonfunctional requirements thereby is a cheap solution for prevention of time and cost waste for achieving the qualitative goals for development of software systems.
Unified modeling language (UML) is a semiformal and standard language for easy description of software, but performance and reliability of SA may not be evaluated thereby. Therefore, to evaluate the performance and reliability, pragmatic model (UML) must be transformed to formal model (HTCPN).
HTCPN is very suitable for displaying the behavior of systems with concurrent and interactive components and in addition to having a virtual structure and behavior have the capability of graphic display, hence modeling by them is easy. Moreover, these nets provide a framework for analysis, validation and evaluation of nonfunctional requirements such as performance and reliability in complex systems [10].
In this paper, we have proposed an approach, therein SA is described by UML Use Case, Activity and Component diagrams. Then, the required information related to non-functional requirements are annotated to these diagrams as stereotypes and tags. In continue, an algorithm is offered which has been assumed from UML diagrams and establishes the executive model based on HTCPN. Ultimately, upon execution of this model and analysis of its results, nonfunctional requirements including performance (such as response time) and reliability may be evaluated in SA level.
The next sections of this paper have been organized as follows: In section 2, a general description of performance modeling in UML and hierarchical timed colored petri net models has been presented. In section 3, some works related to paper subjects have been reviewed and in section 4, procedure together with details are described. In section 5, a case study is investigated for evaluation of suggested method and ultimately in section 6, a general conclusion of suggested method is explained.

## 2. BACKGROUND

In this section, a general description of performance modeling in UML and timed colored petri net models is presented.

### 2.1 Performance Modeling in UML

The UML may describe the behavioral and structural aspects of SA. But in order to evaluate the SA, features of SA are required that UML has not the capability of their exhibition. Accordingly, a strategy has been presented by OMG including profiles consisted of stereotypes, tags and limitations that provide the capability of exhibiting these features for UML. These profiles and their complete details have been explained in [11].

### 2.2 Timed Colored Petri Nets

Colored petri nets are used for formal description of activities flow in the complex systems and provide the requirements of concurrency and parallelism exhibition. Classic petri nets are not suitable for modeling the systems with large space or a complex temporary behavior. In these cases, we must use a developed petri net model having color and time. This model is the base of

a framework that is used for solving the problems related to design and control in complex systems.
In these networks, the concept of time is introduced by global element called global time. The values selected by this time explains the model time. This model may be an integral number that indicates the discrete time or maybe a true number explaining the continuous time. This value of time that is pertained to each token is referred to as stamp time that indicates the first time of model therein token may be used. As a result, these nets will be appropriate for evaluation of qualitative requirements (response time etc.) and reliability in SA. Specifications of colored petri nets and its different types have been explained in [7].

## 3. RELATED WORKS

In the past various techniques have been presented for evaluation of nonfunctional requirements of performance and reliability in component-based software systems. In general, these techniques may be divided in various groups. Some of these groups such as models based on state and path may be referred that commonly are used for evaluation of reliability. It is notable that in plenty of these methods, evaluation of performance and reliability nonfunctional requirements have not been considered at the primary stages of software systems development cycle and prior to implementation stage. For instance, in place-based models [5, 9], control graph is used to describe the software architecture, so that control graphs are commonly extracted from programs code source. Therefore, applying these techniques will be possible after implementation stage. There are some other methods that provides the prerequisites for evaluation of performance or reliability nonfunctional requirements at the design stage. It is notable that in most of these methods, no integrated executive model has been used for evaluation of performance and reliability of nonfunctional requirements and their analysis. In continue, some of these methods are raised.
Zhu and Wang [13] have introduced a platform for evaluation of software systems performance by UML and hierarchical timed colored petri net. In this method, the software system is modeled by UML Use Case, Collaboration and Deployment diagrams and then these models are transformed to a hierarchical timed colored petri net model. Consequently, an evaluation of software system performance is provided.
Fukuzawa and Saeki [4] presented a method therein software architecture is described by UML Component diagram. Then, the above algorithm has been transformed to colored petri net by an algorithm and ultimately the performance is evaluated, so that the own component and its connector are transformed to a colored petri net but its interface is transformed to a place of colored petri net.
Balsamo and Marzolla [2] presented a method therein software architecture is described by UML Use Case, Activity and Deployment diagrams, then operational profiles related to performance are annotated therein. Ultimately, to evaluate the performance, UML diagrams are transformed to an executive model based on Queuing Networks.
Balsam et al [1, 3] presented a method therein software architecture is described by UML Use Case, Collaboration and State diagrams, in this method, to evaluate the performance of software architecture, an executable model based on generalized stochastic petri nets is established. The main objective in this method is transforming the State diagram and each one of objects of Collaboration diagram to Petri Net that a stochastic petri net is established upon combination thereof and indicates the whole system.

In general, whereas the mentioned methods may provide better basic techniques for evaluation of performance or reliability of software system, but a few ones may be used at the primary stages of software systems development cycle and prior to implementation. In addition, in these methods, there is no unified model for evaluation of performance and reliability simultaneously. Therefore, the techniques will be appropriate and important that use an executive model for evaluation of several nonfunctional requirements (such as performance, reliability).
Therefore the main objectives of technique presented in this paper are as follows:

- Development of a method based on probability for evaluation of reliability in SA Level and prior to implementation stage;
- Capability in studying the performance and reliability of components and connectors so that the system architect is enabled to use an implementer (collection of activities) superseding the components, in case of non-complying with appropriate performance and reliability features;
- Establishing a unified model for evaluation of performance and reliability of nonfunctional requirements simultaneously.

## 4. THE PROPOSED METHOD

The main approach in this paper is establishing the HTCPN-based execution model of UML diagrams and evaluation of performance (such as response time) and reliability of nonfunctional requirements of software architecture.
For this purpose, firstly the SA is described by UML, later operational profiles related to performance and reliability features are annotated therein. In continue, an algorithm is offered for transformation of UML model to HTCPN model and ultimately the said nonfunctional requirements are evaluated by suggested techniques at the SA level.

### 4.1 Description of Software Architecture by UML Diagrams

Different methods has been presented for description of SA by UML. In this paper, to describe the SA, a UML-based method is used, therefore UML Use Case, Activity and Component diagrams are applied for description of SA structure and behavior. In continue, the effect of diagrams and annotations related to performance and reliability therein is explained.

#### *4.1.1 The Role of Use Case Diagram and Annotation of Performance Specification Therein*

Use case diagram describes the functional requirements of system and interaction between system and environment [12]. In this paper, this diagram is used for exhibition of functional requirements and working load applied to the system in SA description. Annotations related to performance in this diagram are related to actors that requesting service from system.
The actors indicating a sequence of unlimited requests out of system are annotated by “PAopenLoad” stereotype and actors indicating a fixed population of requests from system are annotated by “PAclosedLoad” stereotype.

"PAclosedLoad" stereotype has a tag called PAoccurrence that indicates the interarrival time between two subsequent requests. "PAclosedLoad" stereotype has two tags named PApopulation and PAextDelay that respectively indicates "the number of requests" and "the time spent by each completed request before the next interaction with the system". An annotated use case diagram is exhibited in figure 1.

### *4.1.2 The Role of Activity Diagram and Annotation of Performance Specifications Therein*

Activity diagrams are used for description of further details of contents of each use case. This diagram specifies the sequence, order and conditions of operation execution. The operation sequence from beginning to the end is referred to as an activity performed by the system [12]. In this paper, this diagram is used for exhibition of activities flow inside and out of components. Showing the boundaries of each component in activity diagram is not possible but it may be shown by Swimline of activity diagram.

Annotations related to performance and reliability in this diagram are related to actions. In this diagram, each action is annotated by "PAstep" stereotype that indicates the service demand from an active source of system.

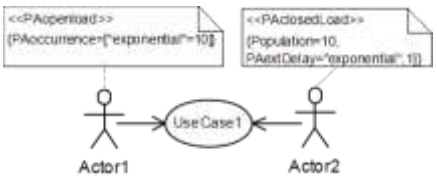
"PAstep" stereotype has two tags named PAdemand and failprob that respectively indicates "service demand" and "failure probability in actions". It is notable that failprob tag has not been defined based on OMG standard for "PAstep" stereotype, but here "PAstep" stereotype tag has been placed for easy to show that. Transitions of activity diagrams are annotated by PAprob probability that indicates the probability of applying a specified transition and is signified when we have several output transitions from an action; in this mode, sum total of transitions probability outputted from same action must be equal to 1. In figure 1, an annotated activity diagram has been shown.

### *4.1.3 The Role of Component Diagram and annotation of Performance Specifications Therein*

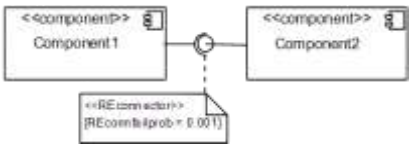
The component diagram indicates the logical structure of a software system. In addition, each component may use the services provided by other components. Services provided by each component is accessible by its interfaces [12]. In this paper, this diagram is used for describing the logical structure of software system, relationship between components and show the interfaces of each component.

Annotations related to reliability in this diagram are related to the connectors. In this diagram, each connector is annotated by "REconnector" stereotype that indicates the reliability specifications in connectors.

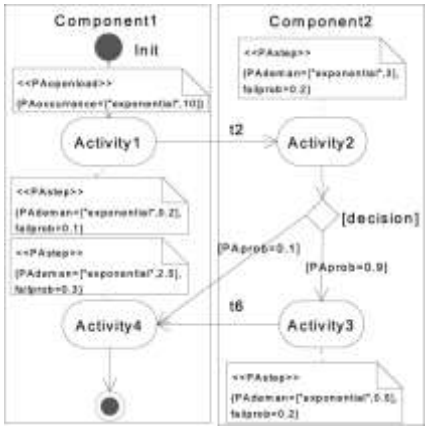
"REconnector" stereotype has a tag named REconnfailprob that indicates the probability failure in connector. An annotated component diagram is shown in figure 1.




(a) Annotated UML Use Case Diagram




(b) Annotated UML Component Diagram




(c) Annotated UML Activity Diagram

**Figure 1. Annotated UML Diagrams**

## 4.2 Suggested Algorithm for Transformation of Annotated UML Model to HTCPN Model

Our suggested algorithm has been assumed for transforming annotated UML model to HTCPN-based executive model according to a three stage design including as below:

***First step: Transformation of component diagram to CPN model***

At the first stage, the component diagram is transformed to a cpn model. For this purpose, each component is transformed to a sub cpn, each interface to a place and each connector to two transitions, so that one of these transitions is used for transmitting the request and the other for receiving. In table 1, transformation maps are shown.

According to table 1, cpn model related to components consists of two places and one transition. The places are applied as component interfaces, and transition as component implementer.

***Second stage: Determination of hierarchical structure***

At this stage, the activities to be performed in the components are specified; in other word, at this stage, each component is implemented. The activity diagram of each component is transformed as per procedure provided at third stage.

***Third stage: Transformation of activity diagram to CPN model***

At this stage, the activity diagram is transformed to a CPN model. Procedure of transforming activity diagram to CPN model will be in accordance with method presented in [8]. Therefore, each action and transition in activity diagram is transformed to one transition and place respectively in cpn. Procedure of transforming branching, joint and fork nodes has been provided in table 2.

Ultimately, considering the stages of suggested algorithm, the final HTCPN model established from UML model as per figure 1 will be as HTCPN model shown in figure 2. It is notable that the different demand classes of a component are related to different colors in HTCPN model.

**Table 1. The general notation mapping in UML componet diagram to cpn model**

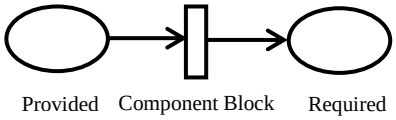
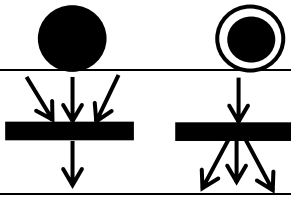
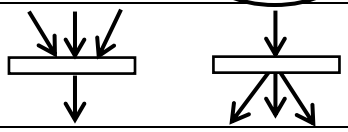
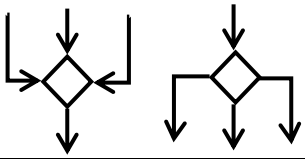
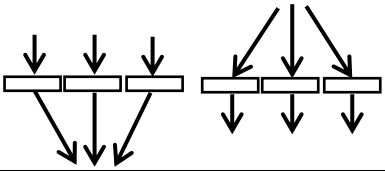
| Item Name | Component Diagram notation | Association CPNs notation |
|---|---|---|
| Component |  |  |
| Connector |  |  |

**Table 2. The general notation mapping in UML activity diagram to cpn model**

| Item Name | Activity Diagram Notation | Association CPNs Notation |
|---|---|---|
| Action node / Transition |  |  |
| Initial / Final node |  |  |
| Fork / Join node |  |  |
| Branching node |  |  |

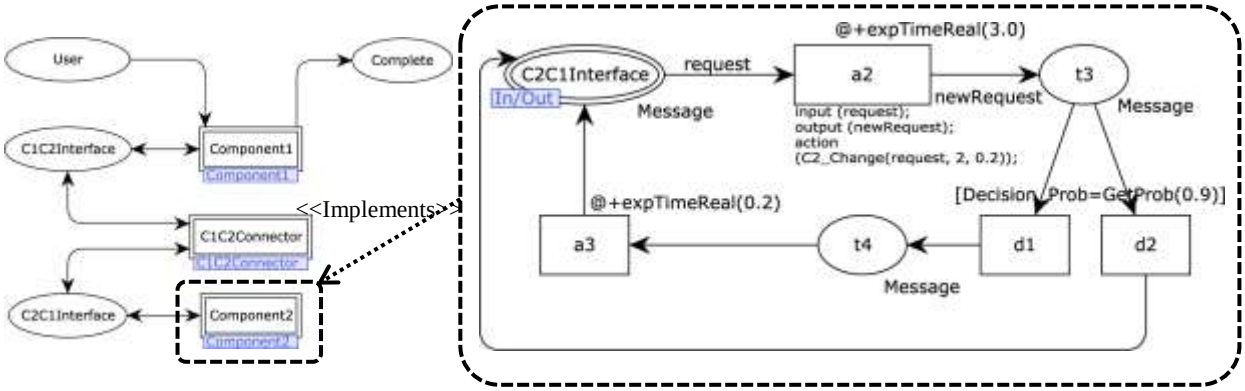



**Figure 2. Generated HTCPN model from UML model**

## 4.3 Evaluation of reliability in software architecture Level

In this study, to evaluate the reliability, extension of method presented in [4] is used so that therein the failure probability has not been considered in the internal activities of each component. Accordingly, to evaluate the reliability of SA, failure probability in components connectors and actions performed in the components are used. Tokens of colored petri nets carry a f value as reliability. Figure 3 shows the manner of calculating reliability for reaching to points that failure probability may occur therein (e.g. components connectors and actions).

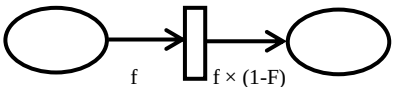



**Figure 3. Reliability**

According to figure 3, variable F indicates the failure probability and its values in UML diagrams are specified by REconnfail and failprob tags that respectively indicates failure probability in connectors and actions. Finally, the total reliability is equal to f value in final place of HTCPN model.

## 5. CASE STUDY

In this section, we use an internet sale system for evaluation of suggested method. In this system, that a part of which is examined, we consider the candidate scenario of products information display. It is notable that for drawing UML diagrams, Enterprise Architecture (Version 9) and for establishment of HTCPN models, CPNTools have been applied. Figures 4, 5 and 6 respectively show UML use case, component and activity diagrams of internet sale system.

According to figure 6 that represents the activity diagram for products information display scenario, five components are involved therein. In this scenario, firstly the user requests for display of products information in Client section, from system. Then, the products information display requests sent to Webserver component via Client component. In continue, as per activity diagram of figure 6, products information display operation is continued and consequently the products information is displayed for the user in Client section.

Here, to evaluate the nonfunctional requirements (such as performance and reliability) for product information display scenario in SA level, UML model shown in figures 4,5 and 6 must be transformed to HTCPN model. Therefore, considering our suggested algorithm in this paper, the final HTCPN model will be as per figure 7.

For evaluation of performance (such as response time) and reliability, it is assumed that 20 requests are inserted to the system for execution of products information display scenario. Therefore, upon execution of HTCPN model that indicates the system behavior in servicing for the products information display requests by the users, we can achieve a series of valuable results related to evaluation of nonfunctional requirements in SA level. Table 3 represents the mean response time and reliability of internet sale system for providing services to 20 requests of products information display scenario.

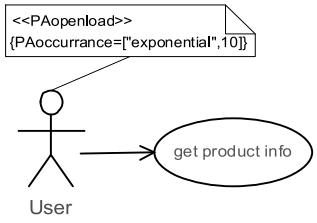



**Figure 4. Annotated Use Case Diagram for Internet Sale System**

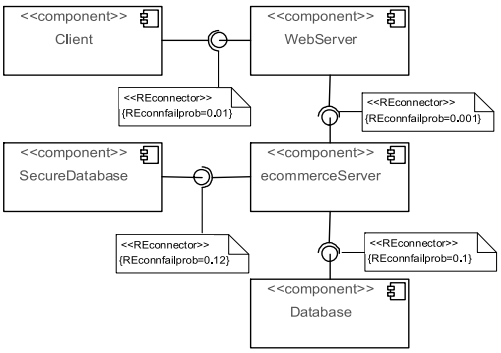



**Figure 5. Annotated Component Diagram for Internet Sale System**

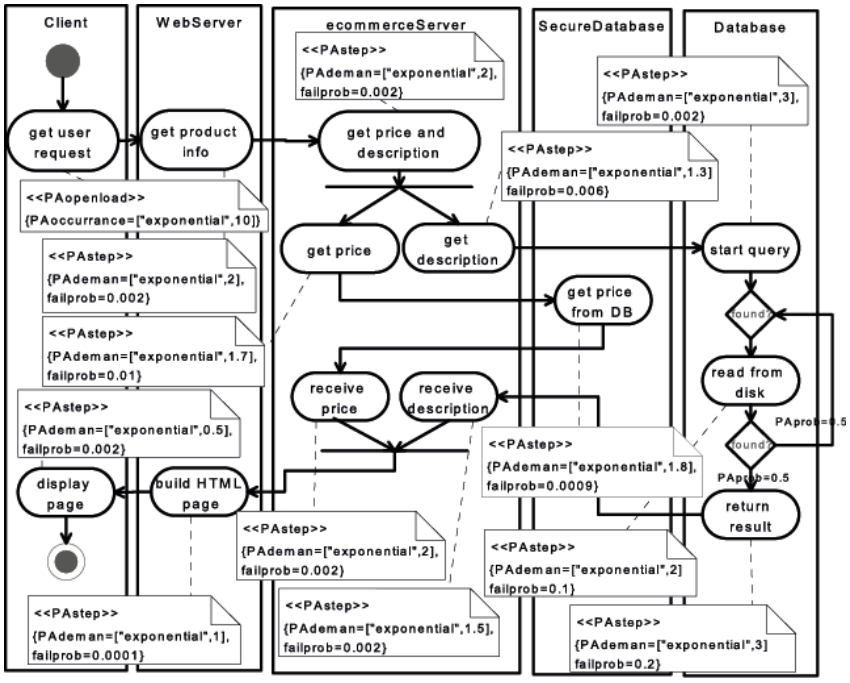



**Figure 6. Annotated Activity Diagram for products information display**

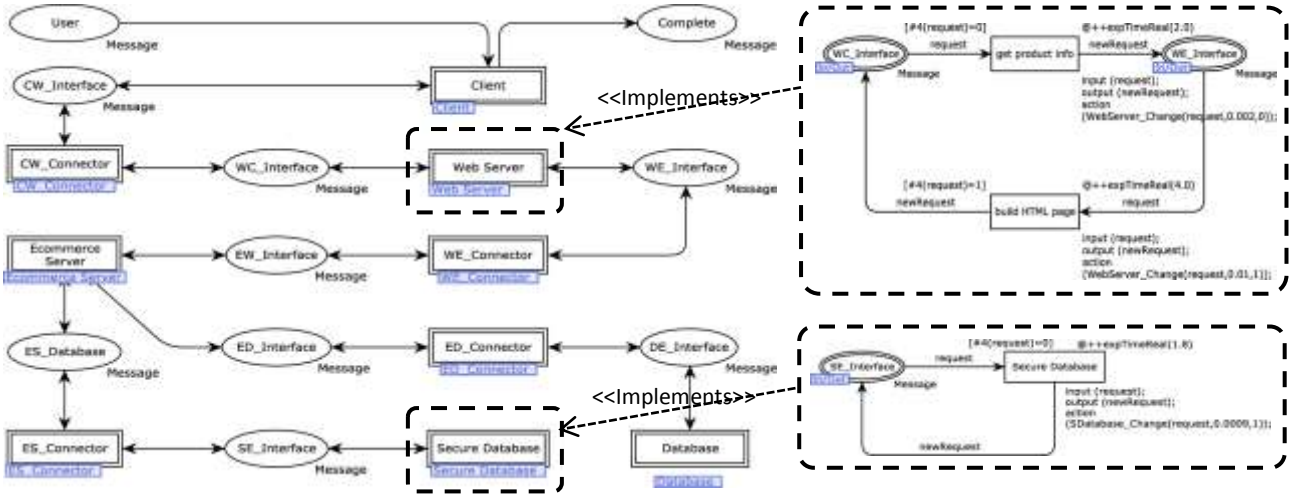



**Figure 7. The full HTCPN model**

**Table 3. he mean response time and reliability**

| Number of Requests | Responste Time (ms) | Reliability |
|---|---|---|
| 20 | 21.058 | 0.956 |

## 6. CONCLUSION

In this paper, we have presented a strategy for evaluation of performance and reliability of nonfunctional requirements in software architecture modeled by UML diagrams. So, the software system may be validated for meeting or not meeting the nonfunctional requirements of case at the primary stages of software systems development cycle. The general analysis framework in this method is formed based on formal models (HTCPN) that accordingly is free of ambiguity. Whereas in this method, UML diagrams are used for description of SA, therefore description of SA by means of achievements of analysis and design stages will be very reasonable and low-cost. On the other side, a transformation algorithm has been presented for establishment of a HTCPN-based executive model from UML model for description of SA, hence the gap between architect and analyzer is removed and this process is performed easily.
In this strategy, further researches are also possible. There are a lot of tools for working with UML models and UML models may be transformed to HTCPN-based executive model automatically. In addition, other nonfunctional requirements may be evaluated by means of other architectural specifications.

## 7. REFERENCES

[1] Balsamo, S., Marco, A. D., Inverardi, P. and Simeoni, M. 2004. Model-Based Performance Prediction in Software Development: A Survey. IEEE Transaction On Software Engineering, Vol. 30, NO. 5.

[2] Balsamo, S. and Marzolla, M. 2005. Performance Evaluation of UML Software Architectures with Multiclass Queueing Network Models. ACM Workshop on Software and Performance (*WOSP*).

[3] Balsamo, S. and Simeoni, M. 2001. Deriving Performance Models from Software Architecture Specifications. In Proceedings of the 15th European Simulation Multi Conference (ESM2001) on Computer Simulation.

[4] Fukuzawa, K. and Saeki, M. 2002. Evaluating Software Architectures by Coloured Petri Nets. in SEKE02 14th International Conference on Software Engineering and Knowledge Engineering, ACM, Ischia, Italy.

[5] Ghokale, S., Lyu, M. and Trivedi, K. 1998. Reliability simulation of component based software systems. In Proceedings of the 9th International Symposium on Software Reliability Engineering (ISSRE'98).

[6] Gyarmati, E. and Strakendal, P. 2002. Software Performance Prediction-Using SPE. Master Thesis Software Engineering, Department of Software Engineering and Computer Science Blekinge Institute of Technology, Sweden.

[7] Jensen, K. and Kristensen, L. 2009. Coloured Petri nets: modeling and validation of concurrent systems. Springer-Verlag.

[8] Lai, Chien-Yuan., Shih, Dong-Her., Chiang, Hsiu-Sen. and Chen, Ching-Chiang. 2010. Transformation of UML activity diagrams into analyzable systems and software blueprints construction. WSEAS Transactions on Information Science and Applications, vol. 7, no. 3.

[9] Littlewood, B. 1979. Software reliability model for modular program structure. IEEE Transactions on Software Engineering, Vol. 28, NO. 3.

[10] Murata, T. 1989. Petri Nets: Properties, Analysis, and Applications. In Proceedings of the IEEE, Vol. 77, NO. 4.

[11] Object Management Group (OMG). 2002. UML Profile for Reliability, Schedulability, Performance and Time Specification.

[12] Object Management Group (OMG). 2005. Unified Modeling Language (UML). Version 2.0,

[13] Zhu, L. and Wang, W. 2012. UML Diagrams to Hierarchical Colored Petri Nets: An Automatic Software Performance Tool. Science Direct International Workshop on Information and Electronics Engineering (*IWIEE*).

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 79 - 83, 2014

# An AMBA Advanced High performance Bus Tracer with Real time Compression for Support of SoC

Sridhar B
Vardhaman College of Engineering
Hyderabad, India

K.Harikrinshna
Vardhaman College of Engineering
Hyderabad, India

**Abstract**: AMBA (Advanced Microcontroller based Bus Architecture) consists of AHB (Advanced High performance Bus), APB (Advanced Peripheral Bus), ASB (Advanced System Bus) and AXI (Advanced extensible Bus). This Paper proposes the debugging and analyses of system on chip (SoC) at various test conditions by verifying the functional aspects of the on-chip bus. Here we select an Advanced High performance bus (AHB), since the AHB bus signals are hard to examine as they are extremely embedded in the system on chip and there are no sufficient I/O pins to access these signals. Hence we embed a bus tracer in SoC to capture and compress the bus signals with different compression mechanisms. The tracer is successfully verified in FPGA SPARTAN 3E (XC3S500E). Tools used in this manuscript Modelsim for simulation and XILINX ISE II for RTL Synthesis.

**Keywords**: AMBA AHB, SoC debugging, On-chip bus, FPGA

## 1. INTRODUCTION

The proposed multi resolution AHB on chip Bus tracer is named as SYS-HMRBT. The bus tracer adopts three trace compression mechanisms to attain the high trace compression ratio. Multi resolution tracing is supported by capturing traces at different timing and signal abstraction levels. The dynamic mode change is the add on feature to allow user to switch the resolution for different portions of the trace to match specific debugging/analysis need.

## 2. RELATED WORK

Since the traced data is very huge that limits the trace memory and there are some hardware methods to compress the traces they are lossy trace compression and lossless trace compression. Maximum compression ratio is achieved by lossy trace compression technique but the accuracy reduced. Anis and Nicolici [5] technique is suitable for repeatable and deterministic systems. But the complex System on chip with different IPs are neither repeatable nor deterministic therefore the appropriate method for on-chip bus tracers is lossless trace compression.

## 3. EXISTING SYSTEMS

The AMBA navigator[3] that is capable of tracing all the AHB bus signals but not providing any compression support and AMBA AHB trace macro cell(HTM) [2] developed by ARM provides the AHB bus trace with compression techniques and the Data value trace is not supported by HTM, AMBA Navigator, HTM both are having restricted abstraction ability in timing dimension only.

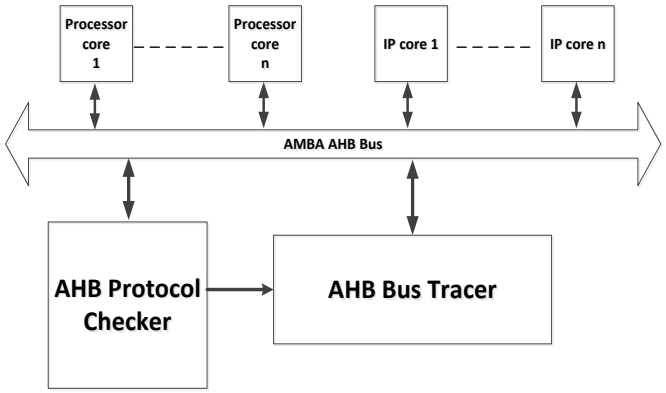



**Fig 1: AMBA AHB Bus Tracer Block Diagram**

## 4. PROPOSED SYSTEM

The SYS-HMRBT shown in figure 1 above mainly consists of four modules they are Event Generation, Abstraction, Compression and Packing as seen in figure 2. Firstly the Event generation module decides the starting &stopping of the trace and its trace mode it consists of configurable event registers which specify the triggering events on bus and it also have corresponding matching circuit to compare the bus activities with the other events in the register. Generally, this module can accepts events from external modules like AHB bus Protocol checker(HPChecker)[6].The format of event generation contains four parameters they are trigger conditions, trace mode, trace direction and trace depth among these, the trigger condition can be of any address value which can be the combination of address value, data value and control signal values. There is a mark field for each value to

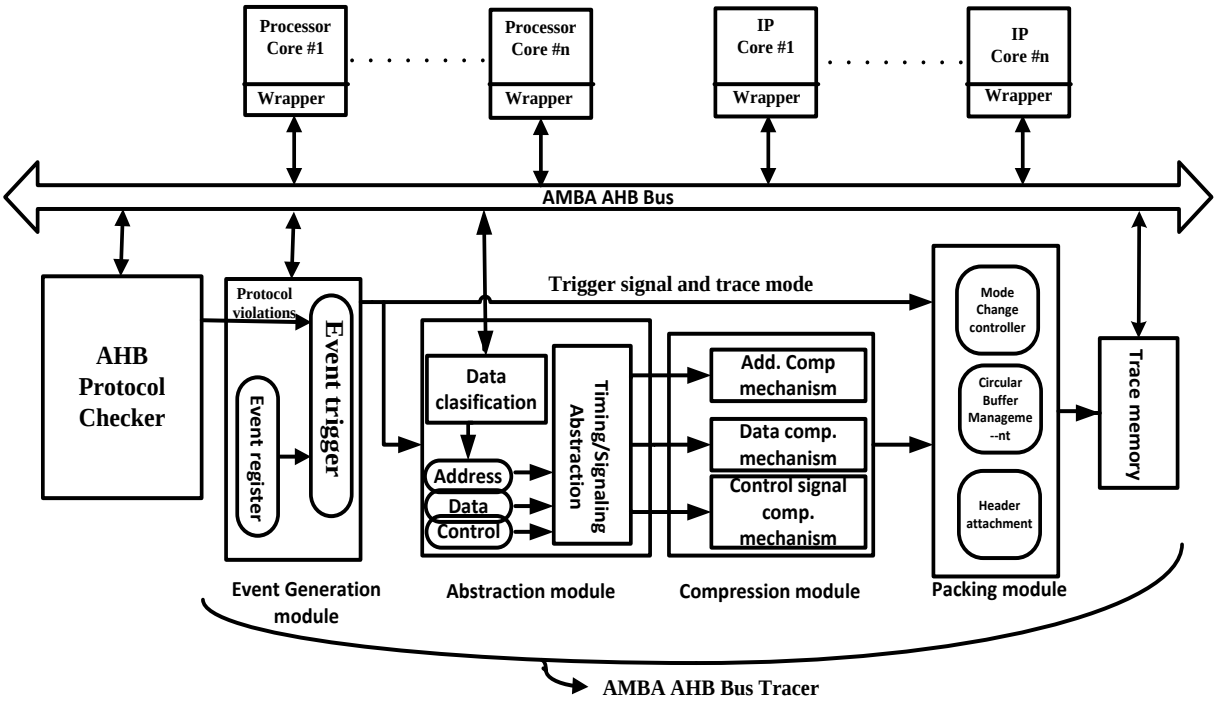



**Fig 2: AHB Bus Tracer Block Diagram**

enable partial match. For every triggering condition, we can assign a desire trace mode, which allows to be dynamically switched between events.

The next module after the Event generation module is Abstraction module and the main function of abstraction module is, it monitors the AMBA AHB bus, filters the signals. The abstraction is in two dimensions timing, signal. At the timing dimension cycle level, transaction level two levels of abstractions are there, we can define three levels of abstraction for signal dimension they are bus state, full signal and master operation. The AHB bus signals classified into four categories they are program address, protocol control signals, access control signals and data address/value. All the signals are captured by the full signal level. Bus state level will captures all the signals and encodes the protocol control signals. The bus master transfer activity captured at the master state level. Integrating the abstraction levels in both dimensions we can get five modes they are mode full signal cycle level (FC), mode full signal transaction level (FT), mode bus state cycle level (BC), mode bus state transaction level (BT) and mode master state transaction level(MT). We can dynamically change the trace mode in real-time to analyze the bus trace and also achieves the dynamic mode change feature.

The compression module is to compress the signals to reduce trace size. Here we achieves the compression by using three effective compression mechanisms as shown in fig.2, they are address compression mechanism, data compression mechanism and control signal compression mechanism.

## 4.1 Address Compression Mechanism

The program addresses can be compressed in three phases by using three compression techniques they are branch/target filtering technique, dictionary based compression technique [4] and slicing technique [3].

### *4.1.1 Branch/target filtering*

Mostly the program address is sequential so the address of the first instruction (target) and last instruction (branch) are recorded and the size is further reduced by dictionary based compression.

### *4.1.2 Dictionary Based Compression:*

In this approach a dictionary stores the frequently appeared target-branch address; we use a comparator to compare data with the previous data in the dictionary if it found that existed then simply stores the index value otherwise it stores data in the dictionary.

### *4.1.3 Slicing:*

This is the last phase in the address compression if there is any missed address in the previous phase can be compressed by slicing approach. In the slicing approach the address are divided in the form of slices with equal size, in this approach we use a register and a slice comparator the register is used to store the previous address and the slice comparator compares the present data with the previous data. If the data differs with the previous data in one slice then only that slice is recorded remaining are ignored for example x = 0000 8066 is the previous data and y = 0000 8020 is the present, after slicing only 20 is recorded for y

## 4.2 Data Compression Mechanism

The data address/value mostly random and irregular. We propose difference method based on subtraction, in this method the present data subtracted from the previous one and also removes the leading zeros and then stored, if the difference is greater than 65535 then it records present 32-bit data value

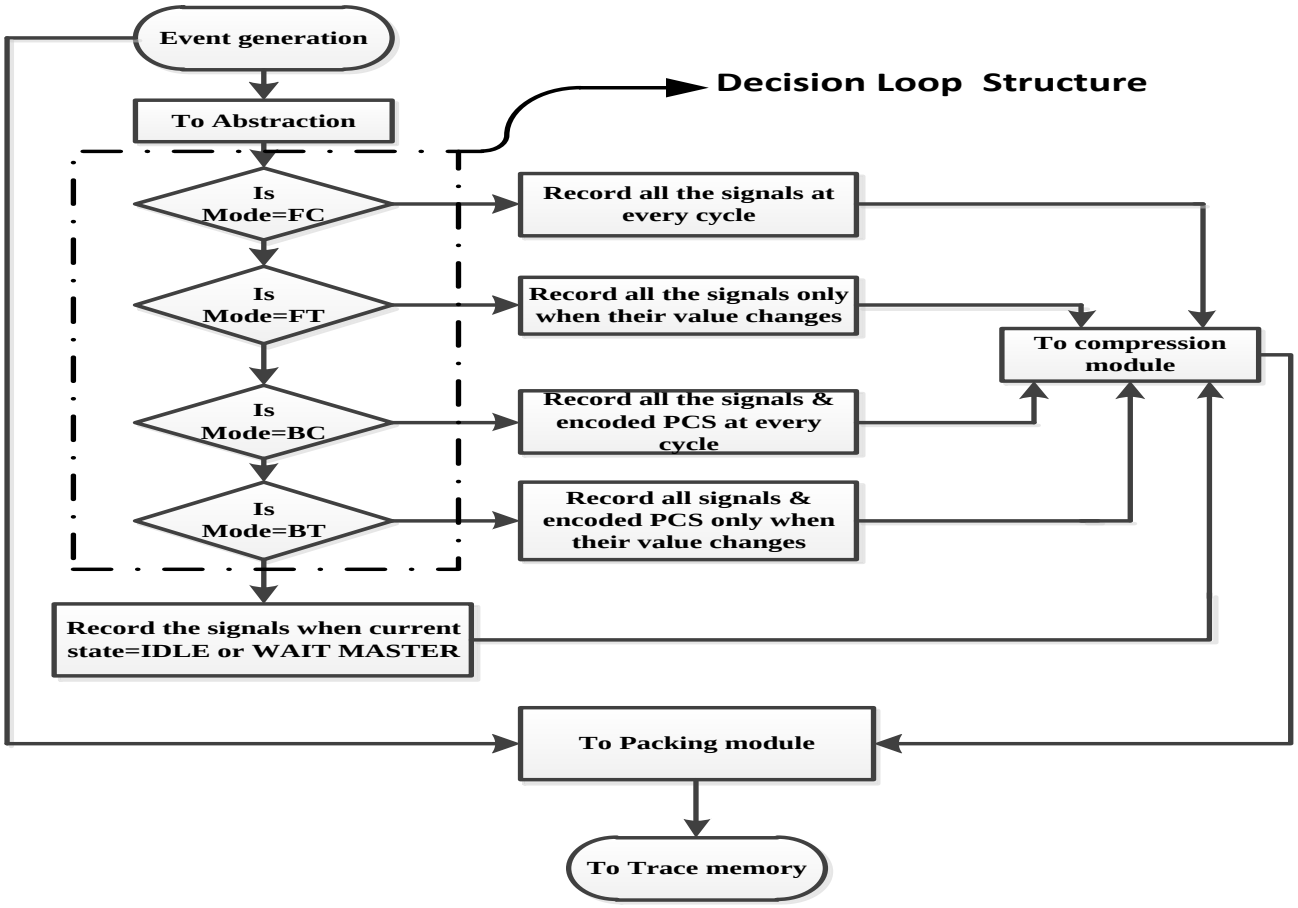



**Fig 3: Tracer Implementation Flow**

### 4.3 Control Signal Compression Mechanism

The AMBA AHB Control signals are classified in to two groups one is access control signals and another is protocol control signals.in these some of the signals are frequently repeat and some signals doesn't occur or occur rarely so that we chooses the Dictionary Based Compression technique.

## 5. FLOW CHART

As we illustrated in the flow diagram in figure 3 the functionalities of the four modules in the proposed system now the implementation flow of the tracer, in this the AHB protocol checker will gives protocol which describes the errors caused by protocol violation to the event trigger and also it has event registers with these the Event trigger triggers the starting &stopping of the trace and its trace mode and the abstraction module has five modes in which if the designer selects Mode FC then the tracer records all the signals at every cycle, if it is Mode FT then the tracer records all the signals only when their value changes, if the selected mode is BC then it records all the signals and encodes the protocol control signals at every cycle and the abstraction module sends this information to the compression module to reduce the traced data size. The compressed data finally packed with proper headersthen it is written to trace memory by packing module.

## 6. ALGORITHM DESIGN

The algorithm according to the flow chart as illustrated above is as follows

a. Start
b. Initialize the event generation module
c. Scale down to abstraction
d. Enter into the decision loop structure
e. Record the signals when the current state is IDLE or WAIT MASTER
f. If the condition in decision loop structure is true then
    i. Record all signals at every cycle
    ii. Record all signals only when their value changes
    iii. Record all the signals and encode Protocol Control Signal at every cycle
    iv. Record all the signals and encode Protocol Control Signal only when their value changes.
g. Send all the above true values into the compression modules.
h. Perform the compression
i. Obtain traced and compressed data
j. Send the data to the packing module

## 7. RESULTS OBTAINED

After performing the compression the traced and compressed data is sent to the Packing module here the header attachment is done by providing the mode change information and circular buffer management and writes it to the trace memory.

The output results obtained are as illustrated in figures 4 to 8

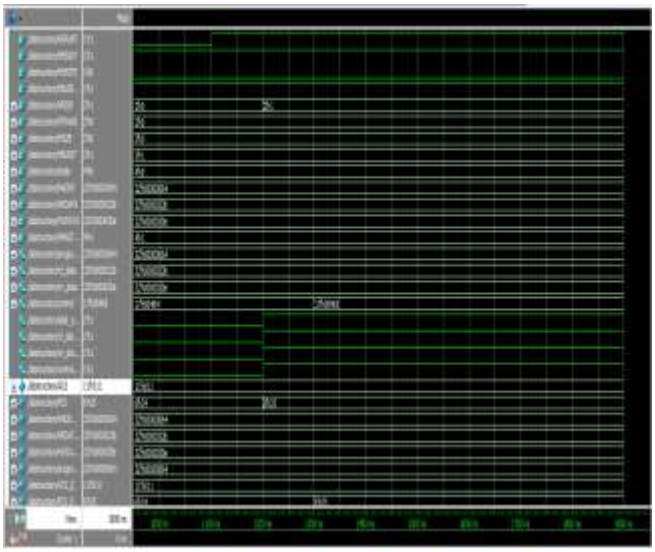


**Fig 4. Simulation result of abstraction module**

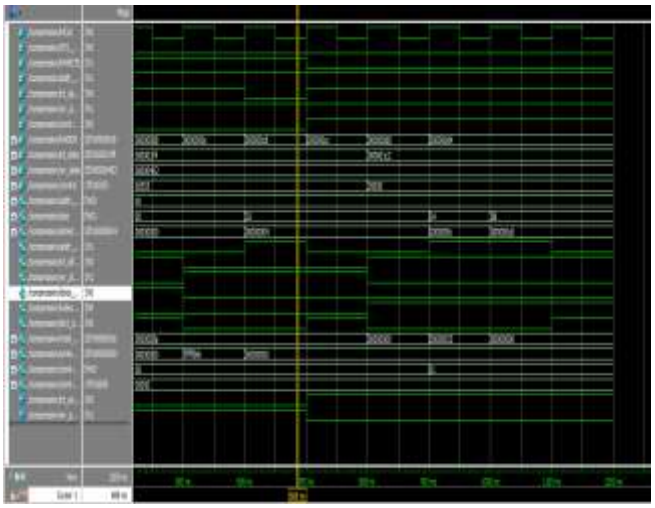


**Fig 5. Simulation result of compression module**

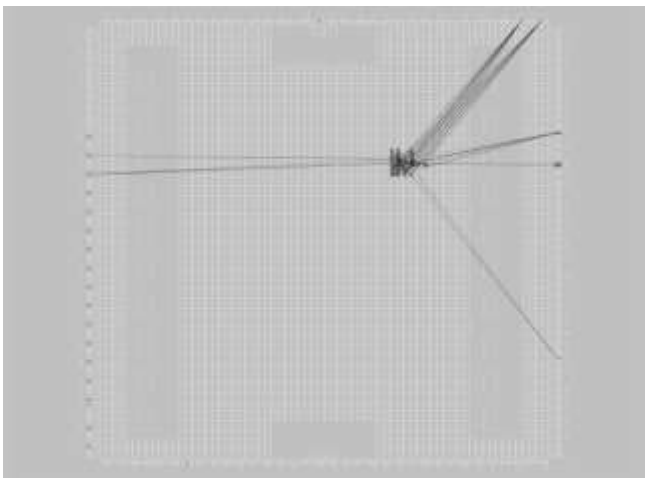


**Fig 6. Routing diagram of AHB Tracer**

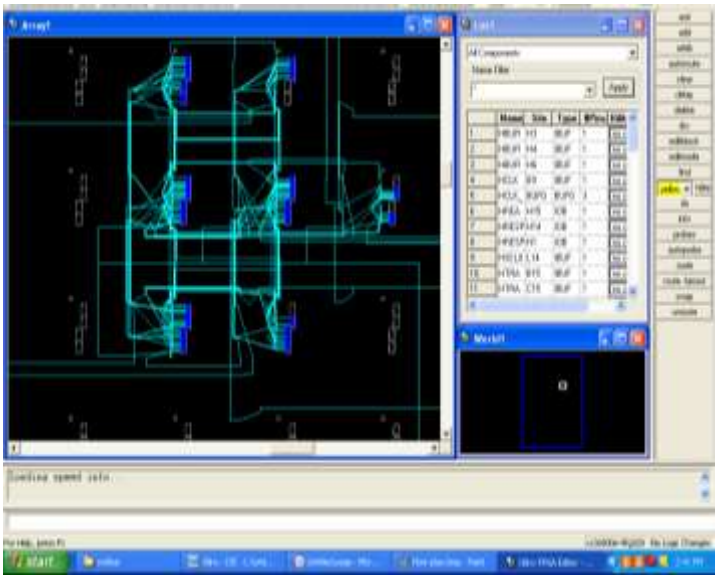


**Fig 7. Complete Placement and Routing of AHB Tracer**

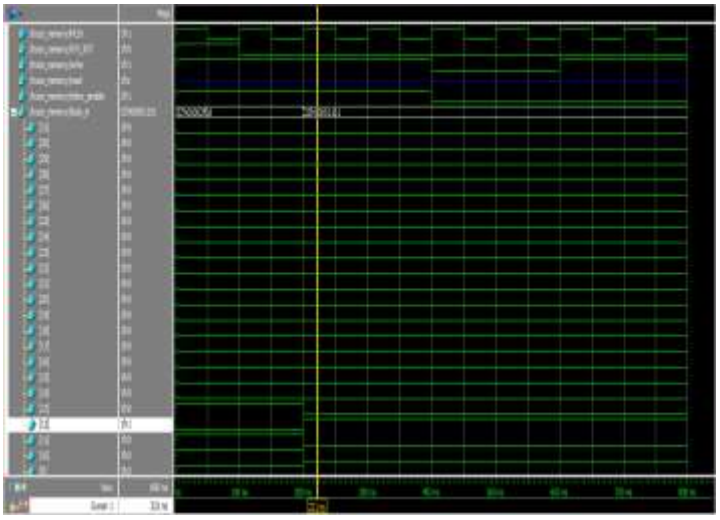


**Fig 8. Simulation result of AHB Tracer**

## 8. CONCLUSION

We have presented the system on chip (SoC) debugging and analyze its behaviour at several test conditions by verifying the functional aspects of the on-chip bus. To capture and compress the bus tracer we embed a bus tracer in System on Chip (SoC). The tracer is successfully verified on FPGA, The FPGA here we selected is XC3S500E SPARTAN 3E and also we obtained the complete placement and routing of our tracer and also the simulation results are obtained that are matching with theoretical calculations. This will shows our tracer achieves a good compression ratio ranging. In addition, the proposed arbiter selects the best possible arbitration schemes based on the priority-level notifications and the desired transfer length from the masters to allow the arbitration to lead to the maximum performance. Experimental results show that, although the area of the proposed arbitration scheme is larger than

those of other arbitration schemes, our arbiter improves the throughput compared with other schemes. We therefore expect that it would be better to apply our arbitration scheme to an application-specific system because it is easy to tune the arbitration scheme according to the features of the target system. For future work, we feel that the configurations of the arbitration scheme with the maximum throughput need to be found automatically during run time.

## 9. REFERENCES

[1] ARM Ltd., San Jose, CA, "AMBA Specification (REV 2.0) ARM IHI0011A"

[2] ARM Ltd., San Jose, CA, "AMBA AHB trace macrocell (HTM) technical reference manual"

[3] First Silicon Solutions (FS2) Inc., Sunnyvale, CA, "AMBA navigator spec sheet"

[4] Heikkinen J and Takala J, "Programmability in Dictionary based compression" *IEEE symposium on* System on chip 2006.

[5] E. Anis and N. Nicolici, "Low cost debug architecture using lossy compression for silicon debug," in *Proc. IEEE Des., Autom. Test Eur.Conf.*,Apr. 16–20, 2007

[6] Y.-T. Lin, C.C.Wang and I.-J. Huang, "AMBA AHB bus protocol checker with efficient debugging mechanism," in *Proc. IEEE Int. Symp.Circuits Syst.*, Seattle, WA, May 18–21, 2008, pp. 928–931.

[7] Y.T. Lin, W.C. Shiue, and I.J. Huang, "A multi-resolution AHB bus tracer for read-time compression of forward/backward traces in a circular buffer,"

[8] A. B. T. Hopkins and K. D. Mcdonald-Maier, "Debug support strategy for systems-on-chips with multiple processorcores,"

[9] J. Gaisler, E. Catovic, M. Isomaki, K. Glembo, and S. Habinc, "GRLIB IP core user's manual, gaisler research,"

International Journal of Computer Applications Technology and Research
Volume 3– Issue 1, 84 - 87, 2014

# Mobile Device Oriented Image Scaling for Reducing Memory Consumption in storing in Android

Rupali Sachin Vairagde
Computer Engineering Department
Pune Institute of Computer Technology
Pune, India

Rekha A. Kulkarni
Computer Engineering Department
Pune Institute of Computer Technology
Pune, India

**Abstract**: In Android operating system, efficient memory consumption is an important feature for better performance. It is very important to efficiently use and manage the internal and external memory space present inside the mobile operating system. Various techniques has been used and implemented to reduce the memory usage in android. One of the technique for better utilization of memory is image compression using scaling technique in android. As per the previous work, image compression techniques are widely used in the field medical science for transparency of reports and effective treatment practices. Other studies didn`t show good results and facing problems such as data loses, poor visibility, large image size and incompatibility issues with the devices and the size of application developed for image compression is too large. Here in this proposed system, we are developing an android mobile application where device oriented image scaling and user oriented image scaling for different screen sizes can be done to reduce the memory consumption of android mobile devices without data loss and visibility.

**Keywords**: Image Scaling; Android OS; Application; memory consumption; mobile device; storage management;

## 1. INTRODUCTION

Nowadays we are facing the increasing use of images in many parts of our life. 3D vision systems, satellites, cameras, medical equipments etc. All of these equipments use or produce image for different purposes. For example the images, for competitive examination or profile pictures we need to upload the compressed image[1]. We have to save image on mobile device or when transmit them, then because of the limitation in disk space and channel bandwidth we almost always need image compression for decreasing the size of data which must be save or transmit[1]. There are several methods for image compression based on the criteria and conditions.

Some of these criteria are compression ratio, compression quality, compression time. Image compression is minimizing the size in bytes of a graphics file without degrading the quality of the image. The reduction in file size allows more images to be stored in a given amount of disk or memory space. It also reduces the time required for images to be sent over the Internet or downloaded from android [13]. The JPEG method is more often used for photographs, while the PNG method is designed to work well in online viewing applications like web browsers so it is fully streamable with a progressive display option. The PNG (Portable Network Graphics) file format was created as the free, open-source successor to GIF. Android SDK multiple drawable directories exist for different screen resolutions. There are low, medium, and high DPI specific directories, drawable-ldpi, drawable-mdpi, drawable-hdpi respectively. This allows you to create images at different DPI to enhance the appearance of your application.

## 2. RELATED WORK

Image compression is very vast area where new concepts and methods are used day-by-day. Android JPEG2000 is a image compression engine or tool which helps in transplanting a image on smart phone based on Android to enhance efficiency of image compression compared with traditional compression. They use the NDK for the compression. A Quantization approach is proposed for the image compression. This method is useful for enhancement of compression quality when each kind of neural network is used to compress the image.

In Android operating system image compression techniques are widely used in the field medical science for transparency of reports and effective treatment practices. Other studies didn`t show good results and facing problems such as data loses, poor visibility, large image size and incompatibility issues with the devices and the size of application developed for image compression is too large.

## 3. IMAGE COMPRESSION TECHNIQUE

Image compression addresses the problem of reducing the amount of data required to represent a digital image. It is a process intended to yield a compact representation of an image, thereby reducing the image storage/transmission requirements [15].

### 3.1 Advantages of Image Compression

1. It provides a potential cost savings associated with sending less data over mobile phone network.
2. It not only reduces storage requirements but also overall execution time.
3. It also reduces the probability of transmission errors since fewer bits are transferred.
4. It also provides a level of security against illicit monitoring.
5. It is important to efficiently use and manage the internal and external memory space present inside the mobile operating system.

The image compression techniques are broadly classified into two categories:

1. Lossless Technique
2. Lossy Technique

In lossless compression techniques, the original image can be perfectly recovered from the compressed (encoded) image.
In Lossy compression techniques, it provides much higher compression ratios than lossless schemes. lossy method can produce a much smaller compressed file than any known lossless method, while still meeting the requirements of the application.
Lossy methods are most often used for Compressing sound, images or videos. The compression ratio (that is, the size of the compressed file compared to that of the uncompressed file). Lossily compressed still images are often compressed to 1/10th their original size. In this project we used the lossy compression method it will provide the good compression ratio. Our objective is to reduce the size of file and save the more memory & reduced the consumption in storing in android. Now days we are using lots of application it will consume lots of space & in browsing if we used compressed image it take less time & less bandwidth to transfer the data. In Android it store the images in DCIM folder and create thumbnails for images it will take the more space to store files that's why rather than lossless method the lossy method is more beneficial in our project.

# 4. PROPOSED SYSTEM ARCHITECTURE

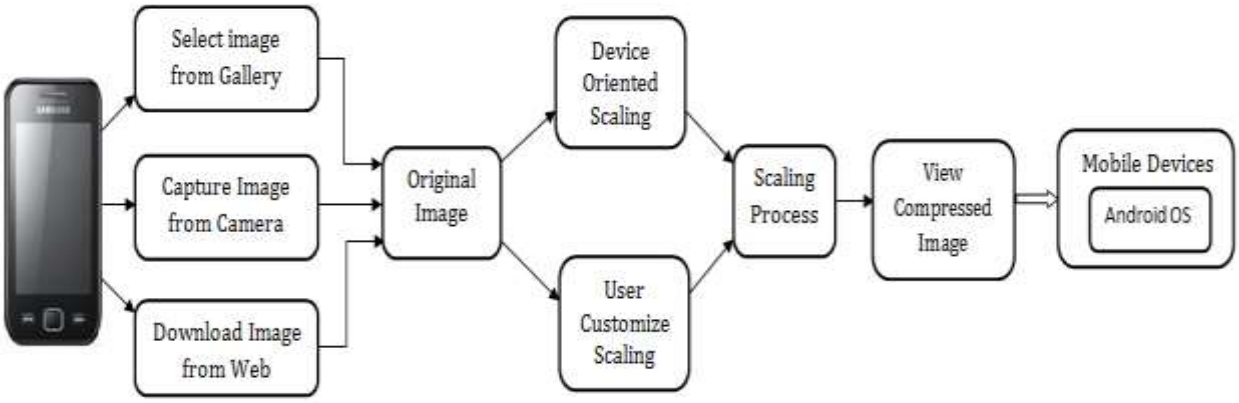



Fig 1. Proposed System Architecture

In proposed System architecture for Device Oriented Image Scaling for reducing memory consumption in storing in Android. Image Compression in Android for Mobile System it will run android application on any mobile device where the android operating system is present then Input of Images for application it will take Images from Gallery of Mobile Phones, Image capture from Camera and Images downloaded from WAP or Internet.
Either select single image or multiple Images for Image compression or scaling, for Single Image it will show the file size of the Image i.e original image stored in mobile device and the resolution of the image and the extension or type of the image for ex. jpeg or .png. Our objective is to reduce the file size of the image and save more memory it will reduced the memory consumption in storing in android for different screen devices. So, that after selection of image user will decide two ways for reducing the size of image with the process of scaling in android.

## 4.1 Device Oriented Scaling

In Device oriented scaling it depends upon the different screen sizes and densities of the individual devices. The system performs scaling and resizing to make your application work on different screens, it fit the screen on their devices.

## 4.2 User Customize scaling

In User Customize scaling it depends upon the user what is the scaling factor to compress the image. We can scale the image to reduce the size of images as per the requirement of the user. But, it will reduce the quality of the image. We need to decide the scaling factor it will not reduced the quality of the image.
The image compression tool for android mobile devices it finally result it into the compressed image and it displayed all the details for example size of the original image and size of the compressed image, Compression ratio(CR). Android by default it will store the type .jpeg it will already compressed but in our proposed work we reduced the size of .jpeg image & .png image for multiple screen devices. Finally the resulted compressed image displayed on the mobile screen[8].

# 5. IMAGE SCALING IN ANDROID

When you may want to use an image but it is too big to use it causing a memory problem for the devices. The way around the problem is to resize the image and scaling of Image is needed. In computer graphics, image scaling is the process of resizing a digital image.[10] Scaling is a non-trivial process that involves a trade-off between efficiency, smoothness and sharpness[16].

In android for images scaling it will consider three factors height of the image, width of the image, resolution of image it measured in a DPI. DPI is simply the "Dots Per Inch" in your image[11]. That simply means it's the measure of the resolution of your image based on the number of pixels or printer dots per inch. In android it considers the definition for the image properties is [6]:
**Resolution** = number of pixels available in the display, scale-independent pixel = **sp**

**Density** = how many pixels appear within a constant area of the display, dots per inch = **dpi**

**Size** = amount of physical space available for displaying an interface, screen's diagonal, **inch**

**Density-independent pixel** = virtual pixel that is independent of the screen density, **dp**

### 5.1 Density Classes in Android:

| Class | Name | Density | Factor |
|---|---|---|---|
| Ldpi | Low Density | 120 dpi | sp = 3/4 * dp |
| Mdpi | Medium Density | 160 dpi | sp = dp |
| Hdpi | High Density | 240 dpi | sp = 1.5 x dp |
| Xhdpi | Extra High Density | 320 dpi | sp = 2 x dp |
| xxhdpi | Extra Extra High Density | 480 dpi | sp = 3 x dp |
| xxxhdpi | Extra Extra Extra High Density | 640 dpi | sp = 4 x dp |

Fig. 2 Density Classes in Android

In our proposed work we adjust the DPI for multiple Screen and it will not loss the quality of image because we resize the image. resizing an image changes the size of its pixels, not its number of pixels[12]. We're not increasing or decreasing the number of pixels. The final output for the proposed system looks like this in a mobile device of android:


Fig. 3. Result of Image Compression Tool

## 6. MATHEMATICAL MODEL

**Let S be the System such that,**

**S = {Input, DD, NDD, SI, Sc, CR, Mt, Mavi, Macc, Succ, Failure, Output}**

Input = {P1, P2, P3} = {Number of Images from Gallery, Camera & WAP}

P1= {Image Capture from Camera type is . jpeg or .png}

P2= {Image is Downloaded from web in Mobile Devices type is .jpeg or .png}

P3= {Image is already present in a Gallery in jpeg or png}

DD = Deterministic Data
= CI ϵ AI

Where, CI = Required Images requested by client to compress

AI = Available Images in Mobile Device

NDD = Non-Deterministic Data
= CI ϵ AI

= Required Images resources are not available.

SI = {Size, Resolution, DPI} = {Single Image Properties from P1, P2, P3}

Sc = {De_Or, Ur_Cu} = {Device oriented Scaling, User Customized Scaling}

CR = Compression Ratio
= Original Image / Compressed Image

Mt = {1……n} = Total Memory

Mavi = {1…….n} = Available Memory

Mocc = {1……m} = Occupied Memory

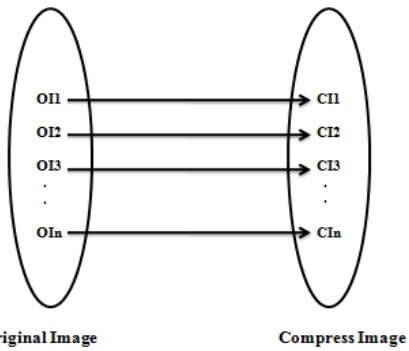
$$Mavi = \sum_{i=1}^{n} Mt - \sum_{i=1}^{m} Mocc$$

Output: Image is compressed and Save the more Memory of Android Devices.

Success: Reduced the file size of Image & reduced consumption of memory.

Failure: The File size of the image is not reduced & more consumption of memory.

### 6.1 Functional Mapping of Original Image and Compressed Image



## 7. CONCLUSION

The increasing use of images in many parts of our life for example in multimedia technology and for online environment we used digital image & that required image we need compress. Image compressions for decreasing the size of data which must be save or transmit. It is important to efficiently use and manage the internal and external memory space present inside the mobile operating system. In this project we used the lossy compression method it will provide the good compression ratio. Our objective is to reduce the size of image file and save the more memory & reduced the memory consumption in storing in android. Now days we are using lots of application for mobile devices it will consume lots of memory space, if we compressed image it take less time , less space & less bandwidth to transfer the data for that in proposed work the android mobile application is developed.

## 8. ACKNOWLEDGMENTS

I would like to express my gratitude to all those who helped me during the writing of this paper. A special thanks to my guide Prof. Rekha A. Kulkani for her valuable suggestions & with the reference material of great value. Finally I wish to extend my thanks to my family & all my friends for support.

## 9. REFERENCES

[1] Hui-yang wang·, xiang-dong, Study of jpeg2000 on Android , School of Tnfonnation and Communication Engineering, Beijing University of Posts and Telecommunications, proceedings of IEEE, 2012

[2] Seyed Mehdi Moghadas, Hossein Pourghasem , New Image Compression Algorithm using Proposed Quantization Approach, 2011 International Conference on Pattern Analysis and Intelligent Robotics 28-29 June 2011, Putrajaya, Malaysia.

[3] Robert D. Dony, Simon Haykin, Neural Network Approaches to Image Compression, Proceedings of the IEEE, vol. 83, no. 2, february 1995.

[4] Lalit Kumar , Kanchan Sharma , Web Based Novel Technique for Watermarking Colour Images on Android Mobile Phones , International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 7, July 2013 ISSN: 2277 128X .

[5] Miss. Jagtap R.S.,Mrs. Sadalage J.A, Image Processing as Android Application, IJSRT , MAY - JUNE 2013

[6] DeepalinKayande and urmila shrawankar , Performance Analysis for Improved RAM utilization for Android Application, Software Engineering (CONSEG), 2012 CSI Sixth International Conference.

[7] http://stefan222devel.blogspot.in/2012/10/android-screen-densities-sizes.html

[8] http://en.wikipedia.org/wiki/Image_file_formats

[9] http://developer.sonymobile.com/2011/0627/how-to-scale-images-for-your-ndroid-application/

[10] http://developer.android.com/guide/practices/screens_support.html

[11] http://www.serif.com/appresources/HPX4/Tutorials/engb/photoplus_tutorials/photo_image_size.htm

[12] http://www.ccbirding.com/dcsig/howto/dpi/dpi.pdf

[13] http://www.webdesignerdepot.com/2010/02/the-myth-of-dpi/

[14] http://argillander.wordpress.com/2011/11/24/scale-image-into-imageview-then-resize-imageview-to-match-the-image/#comment-52

[15] http://www.higherpass.com/Android/Tutorials/Working-With-Images-In-Android/3/

[16] http://searchciomidmarket.techtarget.com/definition/image-compression

[17] http://en.wikipedia.org/wiki/Image_scaling