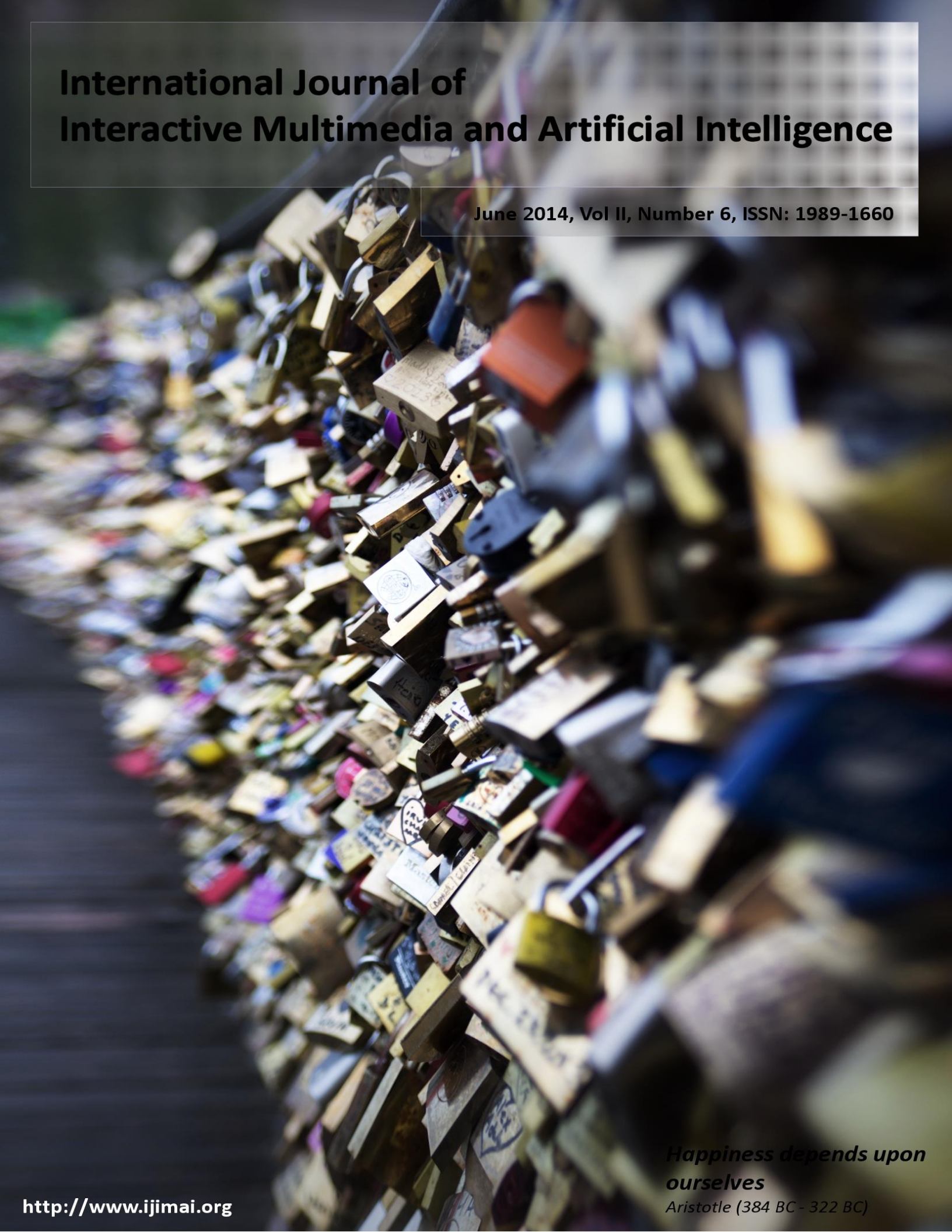


International Journal of Interactive Multimedia and Artificial Intelligence

June 2014, Vol II, Number 6, ISSN: 1989-1660

Happiness depends upon ourselves

Aristotle (384 BC - 322 BC)

http://www.ijimai.org

INTERNATIONAL JOURNAL OF ARTIFICIAL INTELLIGENCE AND INTERACTIVE MULTIMEDIA
ISSN: 1989-1660 - VOL. II, NUMBER 6

Editor's Note

The International Journal of Interactive Multimedia and Artificial Intelligence provides an interdisciplinary forum in which scientists and professionals can share their research results and report new advances on Artificial Intelligence and Interactive Multimedia techniques.

The research works presented in this issue are based on various topics of interest, among which are included: 3D image reconstruction, Persian texts, usability evaluation methods, user experience, oriented matroids, flexible job-shop scheduling, business and social behavior, mobile computing and mobile devices, intelligent tutoring systems and geography optimization.

Pacheco et al. present a novel hybrid methodology, composed by 10 phases that combine active and passive methods, using images and a laser in order to supplement the missing information and obtain better results in the 3D object reconstruction. The proposed methodology proved its efficiency in two complex topological complex objects [1].

Noferesti, S. and Shamsfard M., write about Persian Texts and the Ezafe construction. They introduce a framework for combining genetic algorithms with rule-based models. This framework was used for recognizing the position of Ezafe constructions in Persian written texts. At the first stage, the rule-based model was applied to tag some tokens of an input sentence. Then, in the second stage, the search capabilities of the genetic algorithm were used to assign the Ezafe tag to untagged tokens using the previously captured training information. The proposed framework was evaluated on Peykareh corpus and it achieved 95.26 percent accuracy. Test results show that this proposed approach outperformed other approaches for recognizing the position of Ezafe constructions [2].

Schön, E.M. et al. describe a procedure to analyze and optimize scientific Internet information services that can be accomplished with relatively low effort. It consists of a combination of methods that already have been successfully applied to practice: Human beings, usability inspections, Online Questionnaire, Kano model and Web Analytics [3].

José Antonio Valero Medina and Ivan Lizarazo Salcedo describe the implementation of a prototype REST service for triangulation of point sets collected by mobile GPS receivers. The first objective of their work is to test functionalities of an application, which exploits mobile devices' capabilities to get data associated with their spatial location. A triangulation of a set of points provides a mechanism through which it is possible to produce an accurate representation of spatial data. Such triangulation may be used for representing surfaces by Triangulated Irregular Networks (TINs), and for decomposing complex two-dimensional spatial objects into simpler geometries. The second objective is to promote the use of oriented matroids for finding alternative solutions to spatial data processing and analysis tasks. This study focused on the particular case of the calculation of triangulations based on oriented matroids. The prototype described used a wrapper to integrate and expose several tools previously implemented in C++ [4].

Celia Gutiérrez proposes a work based on an algorithm where each objective (resource allocation, start-time assignment) is solved by a genetic algorithm (GA) that optimizes a particular fitness function, and enhances the results by the execution of a set of heuristics that evaluate and repair each scheduling constraint on each operation. The flexible Job-shop Scheduling Problem (fJSP) considers the execution of jobs by a set of candidate resources while satisfying time and technological constraints. The aim of this work is to analyze the impact of some algorithmic features of the overlap constraint heuristics, in order to achieve the objectives at a highest degree. To demonstrate the efficiency of this approach, experimentation has been performed and compared with similar cases, tuning the GA parameters correctly [5].

Zouhair et al. present a work in the field of Intelligent Tutoring System (ITS), in fact, there is still the problem of knowing how to ensure an individualized and continuous learners follow-up during learning process, indeed among the numerous methods proposed, very few systems concentrate on a real time learners follow-up. This research develops the design and implementation of a Multi-Agents System Based on Dynamic Case Based Reasoning which can initiate learning and provide an individualized follow-up of learner. This approach involves the use of Dynamic Case Based Reasoning to retrieve the past experiences that are similar to the learner's traces (traces in progress), and the use of Multi-Agents System. Through monitoring, comparing and analyzing learner traces, the system keeps a constant intelligent watch on the platform, and therefore, it detects the difficulties hindering progress and avoids possible dropping out. The system can support any learning subject. To help and guide the learner, the system is equipped with combined virtual and human tutor [6].

Sati, M. et al. outline a solution about a fault-tolerant mobile computing model based on scalable replica. The most frequent challenge faced by mobile user is stay connected with online data, while disconnected or poorly connected store the replica of critical data. Nomadic users require replication to store copies of critical data on their mobile machines. Existing replication services do not provide all classes of mobile users with the capabilities they require, which includes: the ability for direct synchronization between any two replicas, support for large numbers of replicas, and detailed control over what files reside on their local (mobile) replica. Existing peer-to-peer solutions would enable direct communication, but suffers from dramatic scaling problems in the number of replicas, limiting the number of overall users and impacting performance. Roam is a replication system designed to satisfy the requirements of the mobile user. Roam is based on the Ward Model, replication architecture for mobile environments. Using the Ward Model and new distributed algorithms, Roam provides a scalable replication solution for the mobile user.

DOI: 10.9781/ijimai.2014.260

They describe the motivation, design, and implementation of Roam and report its performance. Replication is extremely important in mobile environments because nomadic users require local copies of important data [7].

Bhaskar, V.S. et al., describe the work to elaborate swarm intelligence for business intelligence decision making and the business rules management improvement. Their paper introduces the decision making model which is based on the application of Artificial Neural Networks (ANNs) and Particle Swarm Optimization (PSO) algorithm. Essentially the business spatial data illustrate the group behaviors. The swarm optimization, which is highly influenced by the behavior of creature, performs in group. The Spatial data is defined as data that is represented by 2D or 3D images. SQL Server supports only 2D images till now. As they know that location is an essential part of any organizational data as well as business data: enterprises maintain customer address lists, own property, ship goods from and to warehouses, manage transport flows among their workforce, and perform many other activities. By means to say a lot of spatial data is used and processed by enterprises, organizations and other bodies in order to make the things more visible and self-descriptive. From the experiments, they found that PSO is can facilitate the intelligence in social and business behaviour [8].

Dixit, P. et al. write about the Internet security. This topic deals with the information secure and the integrity of the data. Sending messages over the Internet secretly is one of the major tasks as it is widely used for passing the message. In order to achieve security there must be some mechanism to protect the data against unauthorized access. A lossless data hiding scheme is proposed in this paper which has a higher embedding capacity than other schemes. Unlike other schemes that are used for embedding fixed amount of data, the proposed data hiding method is block based approach and it uses a variable data embedding in different blocks which reduces the chances of distortion and increases the hiding capacity of the image. When the data is recovered the original image can be restored without any distortion. The experimental results indicate that the proposed solution can significantly support the data hiding problem. We achieved good Peak signal-to-noise ratio (PSNR) while hiding large amount of data into smoother regions [9]

Dr. Enrique Herrera Viedma.

REFERENCES

[1] A. Pacheco et al.; " Reconstruction of High Resolution 3D Objects from Incomplete Images and 3D Information"; International Journal of Interactive Multimedia and Artificial Intelligence vol. 2 (6). (2014); pp 7-16; DOI: 10.9781/ijimai.2014.261

[2] S. Noferesti, M. Shamsfard; “A Hybrid Algorithm for Recognizing the Position of Ezafe Constructions in Persian Texts"; International Journal of Interactive Multimedia and Artificial Intelligence vol. 2 (6). (2014); pp 17-25; DOI: 10.9781/ijimai.2014.262

[3] E.M. Schön et al.; "Usability Evaluation Methods for Special Interest Internet Information Services"; International Journal of Interactive Multimedia and Artificial Intelligence vol. 2 (6). (2014); pp 26-32; DOI: 10.9781/ijimai.2014.263

[4] J.A. Valero and I. Lizarazo; "A REST Service for Triangulation of Point Sets Using Oriented Matroids"; International Journal of Interactive Multimedia and Artificial Intelligence vol. 2 (6). (2014); pp 33-40; DOI: 10.9781/ijimai.2016.264

[5] C. Gutierrez.; "Overlap Algorithms in Flexible Job-shop Scheduling"; International Journal of Interactive Multimedia and Artificial Intelligence vol. 2 (6). (2014); pp 41-47; DOI: 10.9781/ijimai.2014.265

[6] A. Zohuair et al.; "Our System IDCBR-MAS: from the Modelisation by AUML to the Implementation under JADE Platform"; International Journal of Interactive Multimedia and Artificial Intelligence vol. 2 (6). (2014); pp 48-57; DOI: 10.9781/ijimai.2014.266

[7] M. Sati et al.; "A Fault-Tolerant Mobile Computing Model Based On Scalable Replica"; International Journal of Interactive Multimedia and Artificial Intelligence vol. 2 (6). (2014); pp 58-68; DOI: 10.9781/ijimai.2014.267

[8] V. Bhaskar et al..; "Business and Social Behaviour Intelligence Analysis Using PSO"; International Journal of Interactive Multimedia and Artificial Intelligence vol. 2 (6). (2014); pp 68-76; DOI: 10.9781/ijimai.2014.268

[9] P. Dixit.et al..; "Robust Lossless Semi Fragile Information Protection in Images"; International Journal of Interactive Multimedia and Artificial Intelligence vol. 2 (6). (2014); pp 68-76; DOI: 10.9781/ijimai.2014.269

TABLE OF CONTENTS

OPEN ACCESS JOURNAL

ISSN: 1989-1660

Reconstruction of High Resolution 3D Objects from Incomplete Images and 3D Information

Alexander Pacheco[1], Holman Bolívar[1], Jordán Pascual Espada[2], Rubén González Crespo[3]

[1]*Universidad Católica de Colombia, Bogotá, Colombia*
[2]*Universidad de Oviedo, Asturias, Spain*
[3]*Universidad Internacional de La Rioja - UNIR, Madrid, Spain*

***Abstract* — To this day, digital object reconstruction is a quite complex area that requires many techniques and novel approaches, in which high-resolution 3D objects present one of the biggest challenges. There are mainly two different methods that can be used to reconstruct high resolution objects and images: passive methods and active methods. This methods depend on the type of information available as input for modeling 3D objects. The passive methods use information contained in the images and the active methods make use of controlled light sources, such as lasers. The reconstruction of 3D objects is quite complex and there is no unique solution- The use of specific methodologies for the reconstruction of certain objects it's also very common, such as human faces, molecular structures, etc. This paper proposes a novel hybrid methodology, composed by 10 phases that combine active and passive methods, using images and a laser in order to supplement the missing information and obtain better results in the 3D object reconstruction. Finally, the proposed methodology proved its efficiency in two complex topological complex objects**

***Keywords* — Visual hull, octree, ICP (iterative closest point), marching cubes, iso-surface, iso-level.**

I. Introduction

THE object and image reconstruction is a very useful technique that is commonly applied in many areas, such as human faces recognition [1], quality production control [2], aerial images and maps [3], etc. The high-resolution 3D objects are one of the biggest challenges in the field of objects and image reconstruction methods. Overall, the reconstruction of high resolution 3D objects is an expensive and tedious process, since a user usually has to spend so much time on manual work [4]. The objects reconstruction technology has a very important role in archeology and ancient history; these disciplines often require simple, robust and cheap methods for scanning objects, so these can be studied. The poor state of some discovered antiques implies having complex topologies [5].

The 3-D object reconstruction is a procedure for recovering 3-D geometry and color information from an object [6]. Majorly, there are two kinds of methods to make digital reconstructions of 3D objects: the passive methods [7] and the active methods [8]. The passive methods make use of a set of digital images of the object and the active methods make use of data obtained by laser, which is used to scan the object. Each method has its own advantages and disadvantages, each object reconstruction poses its own problems and appropriate solution must be sought, depending on the requirements.

The active methods are quite expensive in comparison to the passive methods, as they make use of controlled light sources, such as a laser scanner. The object data obtained with the active methods is usually very accurate and the reconstruction algorithms are generally computationally faster than those used in passive methods. Generally, these computational processes are responsible for surface constructions and data registration. In some cases, the configuration of these methods is not easy because the laser scanners require proper calibrations to obtain high quality data. Another aspect of these methods is that they are often not able to capture the actual texture of the physical object; it would need to be combined with the use of cameras. The brightness of objects can also be a problem when trying to scan the object with the laser, the light reflected from the object does not allow laser triangulation and therefore data occlusion is generated.

Passive methods can be used with less technical resources and different types of digital cameras. Recovering the 3D object's geometry by using passive methods involves solving problems related to the camera's calibration, occlusion, correspondence and data fusion. Nowadays, passive methods are widely used in reconstruction and recognition systems, and in many cases they can reconstruct objects with a high level of quality. Generally, these systems involve high computational costs [9] , as they obtain many data sets from digital images and these must be processed through several algorithms. As in active methods, passive methods can also have problems obtaining data shape of bright objects because the bright parts of the object generate texture degradation. Passive methods are very appropriate when the goal is to obtain a real object's reconstruction and is necessary to obtain its texture.

DOI: 10.9781/ijimai.2014.261

There are several methodologies based on passive or active methods; these methodologies often pursue different objectives, such as: improving the quality of the object reconstructions, improving the processing speed, reducing the computational cost of the process, reconstructions of objects in real time, etc. As noted above, both passive and active methods have advantages and disadvantages. This research aims to achieve a specific methodology in order to get high resolution reconstructions of topologically complex objects with great detail and with a relatively low computational cost, or at least lower than the cost involved in passive methods. The proposed method must be able to obtain the object texture in order to achieve a more realistic reconstruction. For the rebuilding of an object of great quality we will combine passive and active methods, the objective is to obtain the best data for the reconstruction of the object. The combination of two methods presents challenges, especially if we take into consideration the objectives of the proposed methodology. The need of applying two different methodologies increases the computational costs, which is a problem to be solved. The combination of active and passive methods in a unique methodology raises the number of challenges, is quite possible that the computational cost due to the application of the two methods is very high. This is a problem that the proposed methodology has to solve, combining the appropriate tasks and algorithms for data processing and reusing the object information and tool calibration.

The proposed method is composed by 10 phases that combine the use of high resolution images and laser scanner in order to supplement the missing information, the object's information obtained is analyzed with image reconstruction algorithms to obtain the best result for the 3D high resolution reconstruction of the object. In order to evaluate the effectiveness and quality of the proposed methodology we have applied it to the reconstruction of topologically complex, detailed objects. We intend to take advantage of the two types of reconstruction methods in order to achieve successful results. Although there is a proposed methodology which is very similar to the one being developed [10], the proposed method uses the NextEngine laser scanner and digital cameras with 13 Mpixels.

This paper is divided into the following sections: Section 2 describes related work. Section 3 describes the proposed hybrid methodology for the reconstruction of 3D objects. In Section 4 we use the proposed methodology to reconstruct two topologically complex 3D objects. Then, in Section 5 we present the results of the application of the proposed methodology. Finally, Section 6 presents conclusions and future work.

II. Related Work

Within the area of reconstruction of high resolution objects from data obtained from images (passive methods [7]) a wide range of solutions have been proposed, depending on the physical property we want to retrieve, such as Space Sculpting Algorithm, Set Level Algorithm, Graph Cut Algorithm, Visual Hull Algorithm [11], Stereo Fitting Algorithm [12, 13, 14, 15]. Several proposals have been combining some of these classic algorithms to improve some aspects of the image reconstruction systems, the main goal of this has been to improve the quality of the reconstruction and the computational costs [8]. Although they are widely used, most of these algorithms do not obtain perfect reconstructions of 3D objects [16, 17]. However, in some cases several solutions achieved high quality reconstructions of particular types of 3D objects, such as human faces [1], industrial parts, simple topological objects, etc. Either way, there is no general solution based on passive methods that is capable of making perfect reconstructions of any type of 3D object, especially in the case of topologically complex objects. Another very common drawback of these solutions is that, in order to slightly increase the quality of the object reconstruction, a great increase of the computation costs becomes inevitable [15]. Additionally, these methods which use digital cameras generate the need to control parameters such as light, color or movement.

There are several types of algorithms that are used commonly in passive methods for the reconstruction and recognition of objects. Some of the most relevant algorithms are: space sculpting, set level, graph cut, Visual hull and stereo fitting.

The space sculpting algorithm consists in initially creating a volume surrounding the surface of interest in order to be able to make iterations on it and reconstruct the shape of the object. In most cases, the refinement is performed under the photo-consistency conditions and starting from an initial volume represented by voxels. Therefore it's an algorithm that commonly uses the color information method and takes volumetric orientation. An example of such solution can be found in [1,18]

The set level algorithm represents the surface as an initial set, which varies in time according to the volumetric density function defined. These algorithms are numerically stable and handle topology changes automatically. Several authors have worked on this type of solution, the most representative being Jean Pons, R.Keriven, and O. Faugeras [19], M. Habbecke and L.Kobbelt in [13].

The graph cut algorithm consists in making a representation of the object's volume in a graph structure with weight on the edges. The problem is finding the most efficient way to cut in a graph with the lowest cost. The weight indicates the discrepancies between images, in other words, it integrates the function of photo-consistency. Several authors have proposed solutions on this methodology, Kolmogorov and R. Zabih in their work of scene reconstruction via graph cuts [20], G. Vogiatzis P. Torr and R. Cipolla in their work of multi-view stereo via volumetric graph-cut [11], A. Hornung and L. Kobbelt in their work of hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding [21].

Visual hull algorithm is defined as the intersection of all the possible cones containing the object. It was proposed by Laurentini [22] and several other authors which made use of this algorithm as an initial step to make refinements on it. [23, 24] Visual hull is the maximal object shape that is consistent with the silhouettes of the object. Supposedly, it can be obtained by the intersection of back-projected visual cones of all the silhouettes [24].

The stereo fitting algorithm is designed to find the correspondence of objects between two images. With the use of this information we can find the fundamental matrix that relates the image space with the object space, and then reconstruct the object through triangulation. The following authors have worked with this type of solution: Y. Furukawa and J. Ponce [13]; and Derek Bradley, T. Boubekeur, and W. Heidrich in [25]. Depending on the physical property we want to analyze and characterize in the object, the restrictions and properties described in a method may be applied. There are three types of methods shown in figure 1: the pattern through shading, pattern through shape and the use of the color information.

Besides, the type of algorithm can be applied to other important factors that determine the quality of the reconstruction of the object. Pattern through the shading requires light calibration in order to assume the surface is Lambertian and make use of the properties this assumption implies. Pattern through the shape requires space calibration to allow the use of segmentation and use geometry properties to obtain a representation of the object's pattern. The color information allows photo-consistency verifications between images. The orientation of the algorithm depends on how it is generating the structure of the object. It may be by volumetry (voxel), or through representation by means of the construction of surfaces.

Within the area of reconstruction of high resolution objects from data obtained by scanners (active methods [8]), reconstruction algorithms are based on parametric surface orientation or implicit functions. [26] is an example of algorithms based on parametric surfaces (Delaunay triangulation) and [27] shows an example of algorithms based on implicit functions (signed functions followed by iso-surface extraction [28]).

Active methods have proven their effectiveness and quality in many solutions, but due to the technical cost involved by its use they are not suitable for all scenarios, not always have the resources or can be used to that extent. There are several solutions using hybrid methods, based on a combination of passive and active methods. [29] Combines the active and passive vision and determines viewpoints in a similar way to human intuition, but in the words of the authors the recovered surface is not very accurate, so the solution is not very effective for reconstructing topologically complex objects.

The proposed solution aims to define a novel methodology to reconstruct objects with high quality, including texture mapping. To obtain the best possible quality data, this methodology merges data from active and passive methods. This methodology is based on the solutions that use volumetric techniques to merge laser scan data and digital images for the reconstruction of the surface. This approach will reduce the characteristic reconstruction errors of active and passive methods by merging data, which can be useful when reconstructing topologically complex objects with better quality. To achieve this, the proposed methodology has to solve the problem of computational cost derived from the combination of two methods, this can be done by combining the appropriate tasks and algorithms for data processing and reusing the object's information. Also, if the information is not reused properly, the calibrations in the cameras and the scanner can also mean a great loss of efficiency. For this purposes, the proposal introduces a new approach to reduce the complexity presented by the scanner's laser calibrations, it uses the box enclosing the 3D object convex hull to register the models. Finally, the geometric 3D object is calibrated through the high resolution shapes, obtained from photography, to make texture mapping as realistic as possible.

III. Hybrid Methodology for the Reconstruction of 3D Objects

As mentioned above, the proposed methodology uses both the data obtained from digital images and the data acquired via laser scanner to retrieve the geometric information of the object. This methodology is divided in two stages. The first one would be to use the digital image information alone to generate a reconstruction of the initial model, and the second one is to register both models: the one obtained in the first stage and the object reconstructed by the laser scanner, in order to make a refinement and find the final volume that best represents the 3D object we want to rebuild. The first phase, reconstruction from images, uses space sculpting algorithm, pattern methods through shape and volumetric orientation. Figure 1 shows in detail the methodology and each of its steps. The first stage consists of phases 1 to 7, and the second stage comprises the phases of 8 to 10.

We will describe the tasks involved in each phase of the proposed methodology below.

A. STAGE 1

1) Phase 1: Data Acquisition

The objective of this phase is to collect information concerning the object, mainly its shape, but also the calibration pattern for each object view. There are some physical factors that can affect this data collection, one of them is the brightness of the object. Bright objects, like shiny metal objects are difficult to reconstruct with either the active or the passive method. In the active method, the light reflected from the object does not allow laser triangulation and, therefore, data occlusion is generated; in the passive method, the bright parts of the object generate texture degradation.

In this first stage we must minimize the effect of light in

order to obtain acceptable images to be processed in the following phases. We have developed a photo-montage with the following elements: Light box (own production), professional lights, a semi-professional digital camera, a revolving board and a computer. Figure 2 shows the montage developed for this phase of data acquisition.

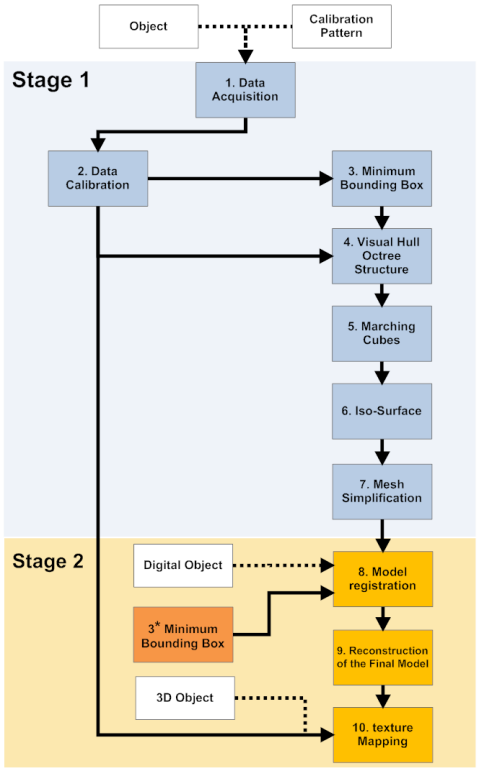


Fig 1. The flow of phases of the methodology for 3D objects reconstruction.

The revolving board allows us to obtain up to 36 views of the object. The lights are calibrated with the light box and the digital camera to minimize the reflection of light on the object as much as possible. The pictures have been taken from the computer in order to obtain more stability and consistency between images.

The background has been arranged as uniformly as possible (white color). However, shadows and dark lines that may affect the process of thresholding of the images can still appear

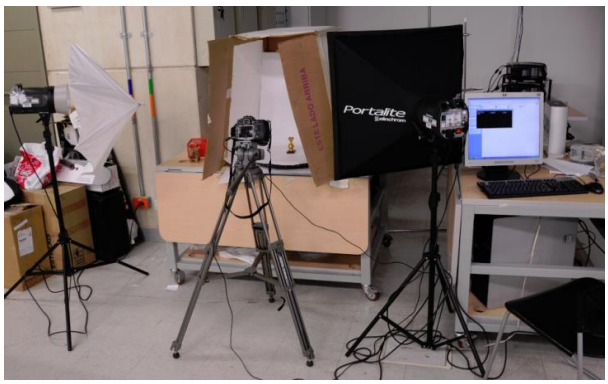


Fig. 2. Photo-montage for data acquisition.

As noted, this phase has the object and the calibration pattern as inputs. The object is photographed in different views and afterwards a thresholding pre-processing is performed to obtain the shape of the object. In practice, many image-editing software can be used to obtain an accurate shape of the object. The calibration pattern corresponds to a table with black and white squares, the result of the pictures taken in each view of the object.

As shown, this first phase has two types of data sets as output. The first data set corresponds to images containing shapes of the object and the second data set corresponds to images with the calibration pattern for the object in each view.

2) Phase 2: Data Calibration

In order to calibrate data we can use a specific open source computer vision software or commercial software. In this case we used Matlab Toolbox [30] developed by Jean-Yves Bouquet. Matlab toolbox allows us to find both the intrinsic and extrinsic parameters of the camera quickly and accurately. In addition, intermediate results can be seen graphically and projections can be created on the calibration to minimize the errors. The output of this phase is the projection matrix M3x4 for each view of the object. The projection matrix is:

$$(1)\quad M_{3X4} = K[R|T]$$

Where K is the matrix, containing the intrinsic parameters of the camera, R is the rotation matrix and T is the translation vector of the object in relation to the camera.

3) Phase 3: Minimum Box Surround

In order to speed up the computational processing of data we must know the minimum work space. From **[18]**, "*the 2D projections of N boxes define a 3D convex hull formed by 4N planes*". The minimum 3D box that encloses the 3D convex hull can be calculated if posed as an optimization problem where the objective function becomes one of the 6 variables that define the 3D box we want to recover under the constraints of each 2D plane of the box. The restrictions imposed by each 2D plane of the box are:

$$(2)\quad a_i x + b_i y + c_i z + w \le 0, \quad i = \{1,...,4n\}$$

We can find the six variables that define the 3D box by using the simplex method. In practice, it is necessary to calculate the 2D convex hull of the object in each of the images to find the box enclosing the 2D object, and then project the sides of the box (lines) to find the equation of the 3D plane.

4) Phase 4: Visual Hull

There are several ways to construct the visual hull of the 3D object depending on several factors: the algorithm, the method and approach used, etc. In this method we will use volume-

oriented algorithms. In order to do so we have decided to represent the object with octree structure.

The structure of the octree is described in **[31]**. The advantage of using this structure is that it uses little memory as it is described by a string of '1's, '0's, '('s that represent the position of each cube (inside, outside and above respectively). Moreover, the order of the chain is due to the pre-order path in the tree. In order to generate the octree we first take a cube, we project it in all images and we check the iso-level function's value to determine the position of the cube. If the position is 'inside', the cube is drawn, if it is 'outside' it is not drawn, and if it is 'above' it must be subdivided in eight cubes.

This process continues recursively projecting and determining positions to smaller and smaller cubes until we reach the degree of resolution that is wanted. In this case, the digital images have a resolution of 2008x3040 pixels; the recommended depth level of the tree will be 8 or 9.

To define the iso-level function the images that match the shapes must be encoded so the background has a value (e.g. -0.5) and the figure has a value with opposite sign (e.g. 0.5). This way, the boundary of the given object will have an intermediate value, 0. Using these values, the iso-level function would be:

$$(3)\quad f_i(c) = MinV_i(P_i c), \quad i = \{1,...,n\}$$

Where $V_i(P_i c)$ is the value taken by the projection of the 3D point in the 2D image, depending on whether it is inside or outside the silhouette or not.

In order to determine the status of the cube we must compare the iso-level values from the eight points of the cube. If all of them were negative, it would mean that the cube is outside the valid volume range for the model. If all of them were positive, the cube would be within the valid range, but if we'd find both positive and negative values, we wouldn't be able to validate it with certainty, and we'd must therefore continue subdividing.

In practice, we should project points that are not only at the vertex but on the edges of the cube, in order to avoid sampling problems and to be able to determine the position of the cube correctly.

5) Phase 5: Marching cubes.

Once we have obtained the volumetric representation of the object, phases 5 *Marching cubes*, 6 *Iso-surface* and 7 *Simplification* are applied to obtain the mesh that best represents the surface of the object. The marching cubes algorithm takes the iso-level values containing the vertexes of the cube and determines the points cutting the iso-surface to generate the corresponding mesh configuration. As they are cubes, there are 256 meshing possibilities that can be reduced to 15. Once the mesh in each cube has been created, they are merged in order to create the surface that represents the object.

6) Phase 6: Iso-surface & Phase 7: Simplification

In the Phase 6, the iso-surface for the proposed algorithm corresponds with values equal to zero. Then we proceed to apply the phase 7: Simplification. Once the mesh has been obtained, the next step is to go through an optimization process by applying mesh simplification algorithms. These are algorithms of decimation and our intention is to maintain the shape of the surface for a smaller quantity of sampled data.

B. STAGE 2

After the first stage, based on a passive method, the second stage began. At this stage the methodology focuses on data obtained using laser scanner. In order to reduce the computational cost during this stage we use data obtained from the preceding stage.

1) Phase 7: Minimum bounding box

This phase consists in the calculation of the scanned object's workspace. As the object is already a 3D object and is also aligned with the coordinate system so it doesn't present inclination along any axis, the minimum containing box is determined by finding the six variables defining the cube through simple minimum and maximum functions.

2) Phase 8: Model Registration

At this stage we aim to register the two geometric models obtained through the two methods we have used so far: passive and active. As the object that has been reconstructed by passive methods is being calibrated, we want to transport the model obtained by the laser scanner to the coordinate system of the model obtained from the digital image data.

The common way to register models is to calibrate the laser, using methods like the one described in **[32]**, and then apply ICP algorithms (iterative closest point) and find the rotation and translation matrix that relates both models. However, due to the way the acquisition of data has been performed; this paper suggests that both models are brought to register through the minimum box enclosing the object.

The registration process from the bounding boxes can be a manual process in practice, but significantly reduces the complexity of the problem, bearing in mind that the models have been reconstructed by different means and have considerable variations.

3) Phase 9: Reconstruction of the final model

Once we have brought the two objects to register we proceed to generate the final surface of the object detailed and free of gaps.

The idea of this algorithm is to take the visual hull represented by the octree and redefine the states of the cube depending on the intersection of the line defined by the laser spot and the source point of the camera. This algorithm makes sculpting in 3D space on the volume defined by the octree guided by 3D points obtained using the laser scanner. In order to generate the model correct, several cases

must be considered, such as the one described in **[10].**

Each octree cube which is 'inside' or 'on' changes to 'off' if the ray defined by the 3D point and the origin of the camera intersects the surface. Finally, the marching cubes algorithm is applied in order to reconstruct the final, gap-free volume of the object.

4) Phase 10: Texture mapping

To offer realistic 3D digital objects, these should include textures. The textures we used were collected from digital images. As the data is calibrated, texture mapping can be done with any algorithm that may be developed for the purpose. In this instance we are proposing a mapping based on particles such as in **[33,34]**. After phase 10 the reconstruction process of the 3D object was successfully completed. As an output from the application of the methodology, a high definition 3D reconstruction of the object is obtained. This reconstruction has a high degree of quality and texture mapping.

IV. Application of the Proposed Metodology in a Real Scenario

To validate the objectives set at the beginning of this research and illustrate the use of the proposed methodology we used it to reconstruct topologically complex 3D objects. The objects involved in the reconstruction process are two replicas of archaeological objects from the Gold Museum in Bogota, Colombia. We choose these objects because of its morphological properties, great detail and brightness management (Phase 1 & 2). Figures 3a and 3b shows the images of the objects we want to rebuild.

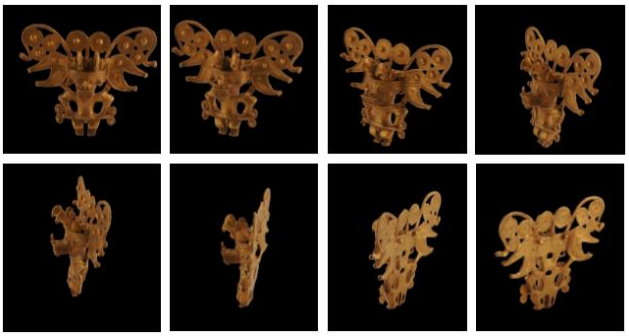


Fig. 3a. Replica from the Gold Museum, Objet 1.

Fig. 3b. Replica from the Gold Museum, Objet 2.

Figure 4 shows the graphic results of the minimum bounding box for the objects (Phase 3). The figure's vertexes are the source point of the cameras and the lines represent the projection of the 2D box surrounding the object. The minimum 3D box is defined in the interior space of intersection of the projections.

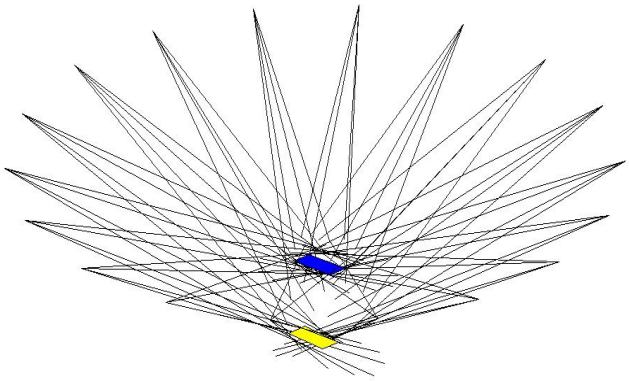


Fig. 4. Minimum Bounding Box (lower and top cover).

For the visual hull's phase we define the structure of the octree given in **[31]**. The right part of Figure 5 shows the simulation results for the octree represented on the left side of the figure.

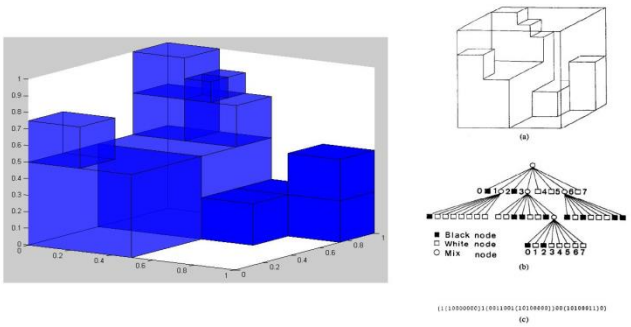


Fig. 5. Representation of the octree.

The results of the octree for the replica from the gold museum are shown in Figure 6. Figure 7 shows the results of phases: 5 Machine Cubes, 6 Iso-surface and 7 Mesh Simplification.

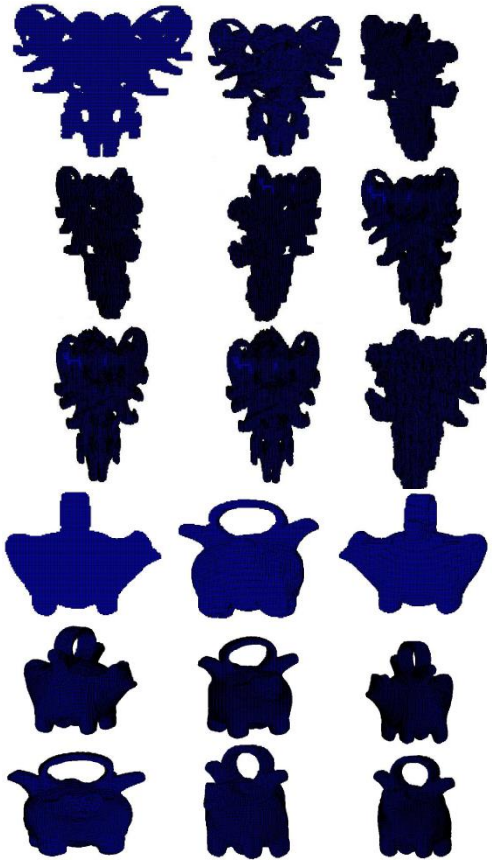


Fig. 6. Octree of objects reconstructed from different views.

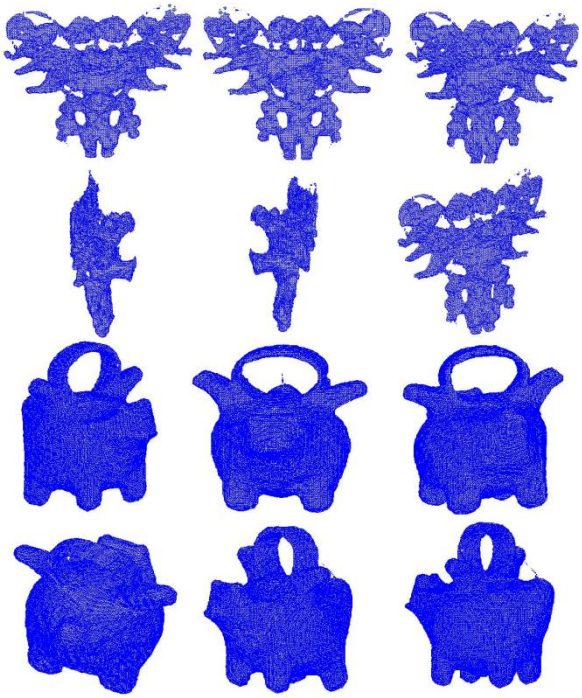


Fig. 7. Results of phases: 5 Marching Cubes, 6 Iso-surface and 7 Mesh Simplification.

For the second stage we used the NextEngine 3D desktop scanner. The data obtained by the laser scanner is shown in Figure 8. As shown, the objects have some deficiencies in the reconstruction.

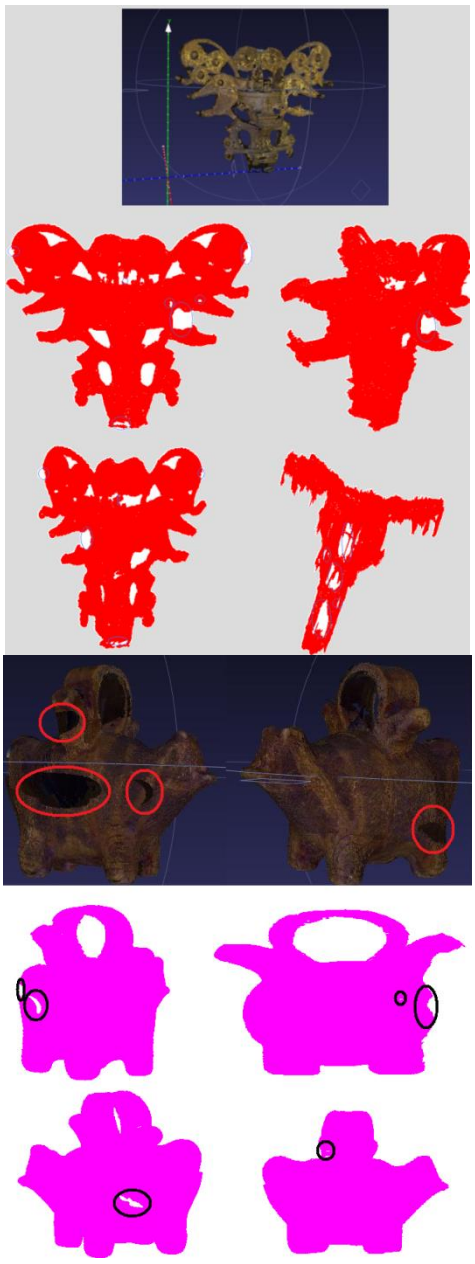


Fig. 8. Data obtained by the laser scanner. Highlighting shows where the failures of reconstructed objects are.

Then we proceed to implement Phase 8 of the methodology: Model Registration, bringing the two models into register, with the help of the minimum encasing boxes that contain the object, as shown in figure 9. In practice, the boxes do not have the same dimensions and the objects must be brought into register manually. However, they are close enough to apply registration algorithms such as ICP (iterative closest point) so that the object is eventually registered.

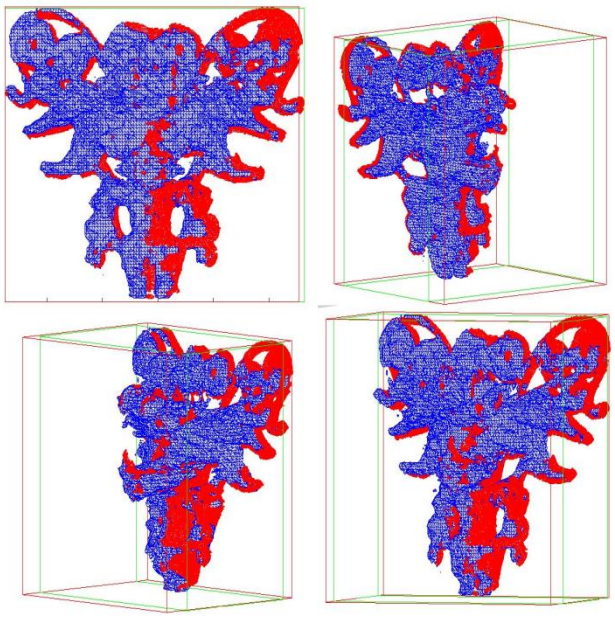


Fig. 9a. Model Registration reconstructed with different techniques. Reconstructed object from images (blue), object reconstructed from laser scanner (red).

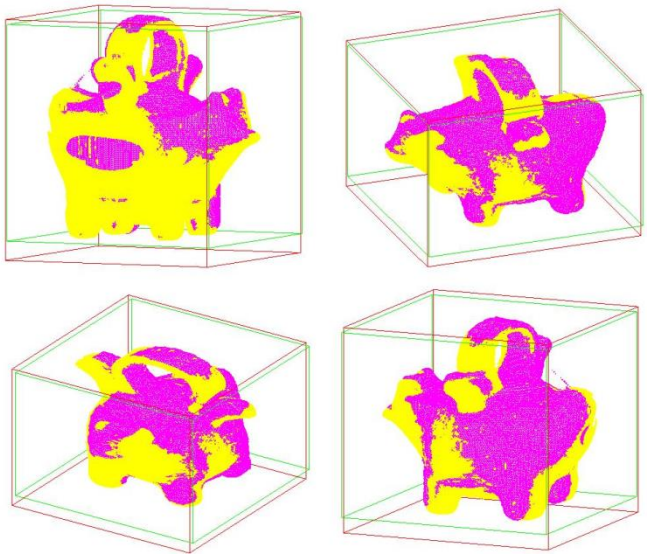


Fig. 9b. Model Registration reconstructed with different techniques. Reconstructed object from images (purple), object reconstructed from laser scanner (yellow).

Finally, in the Phase 10: Texture Mapping we must obtain the merged geometric model and make a photo-realistic texture mapping. An approximation of the final geometry of the 3D objects is shown in Figure 10. After applying the proposed methodology we have successfully managed to build two morphologically complex 3D objects, obtaining a high quality reconstruction of the object, with a detail level significantly higher than those obtained by using passive methods.

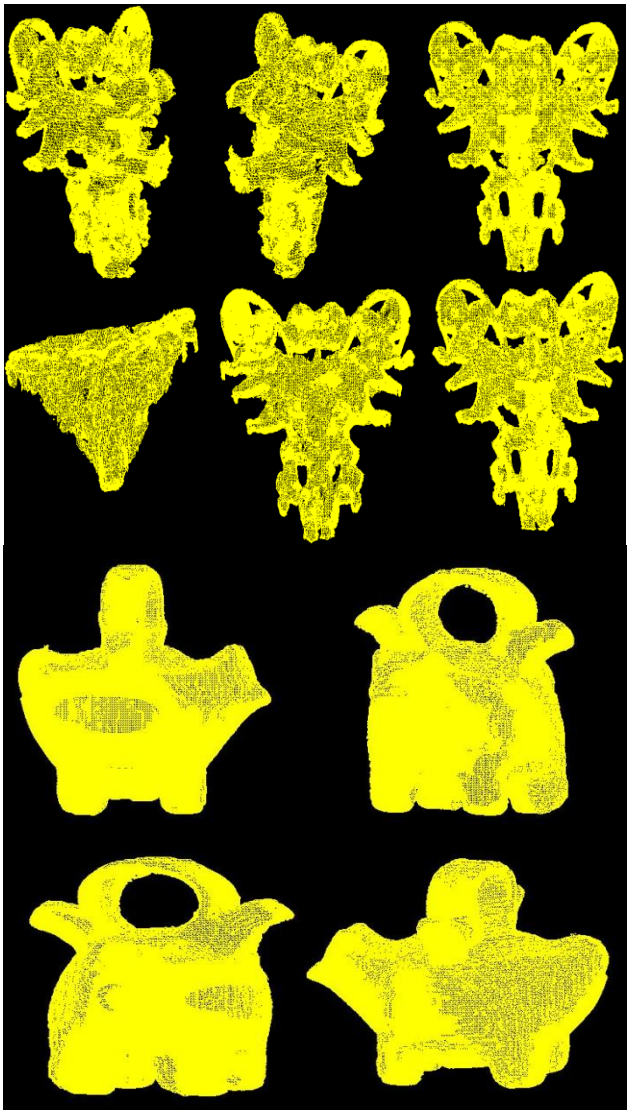


Fig. 10. Different views of the reconstructed 3D objects using images and laser scanner information.

V. Proposal Evaluation

The proposed methodology has been previously used in the reconstruction of two morphologically complex objects. This evaluation examines the relevant aspects related to the objectives set at the beginning of this research.

One of the objectives of the methodology was to obtain quality reconstructions from objects that are difficult to reconstruct, like topologically complex objects with shiny surfaces. To verify that the quality of the reconstructions has increased using the proposed methodology we will compare it with a passive method. A parameter that is often used to validate the quality of an object's reconstruction is the number of errors in the topology generated during the reconstruction. As an object is topologically more complex needs a larger number of voxels to be reconstructed, when more voxels are used the potential for reconstruction errors increases. The error relationship between the two methods is shown in Figure 11, observing results it is possible to detect a function with an associated trend to the occurrence of errors depending on to

the number of voxels of the object to be reconstructed.

The combination of two methods would increase the computational cost. The other major objective of this research was to achieve an adequate computational cost, which did not exceed the computational cost of a passive methodology. To achieve this requirement the proposed methodology reuses the object data captured and the calibrations calculated at different phases of the process. The trend of error occurrence in the calibration as a function of the number of voxels is potential with a coefficient of determination of 0.94. Second, calculate the relationship between the time spent on the reconstruction process of the 3D object and the object's voxels. In the graph of Figure 11b the comparison of the proposed methodology with a passive one is shown

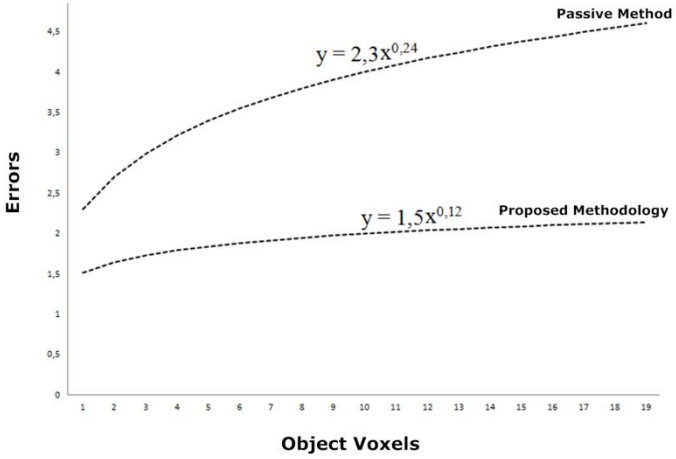


Fig. 11a. Graphical representation of the relationship between the errors and the voxels of the object in the proposed methodology and passive methodology.

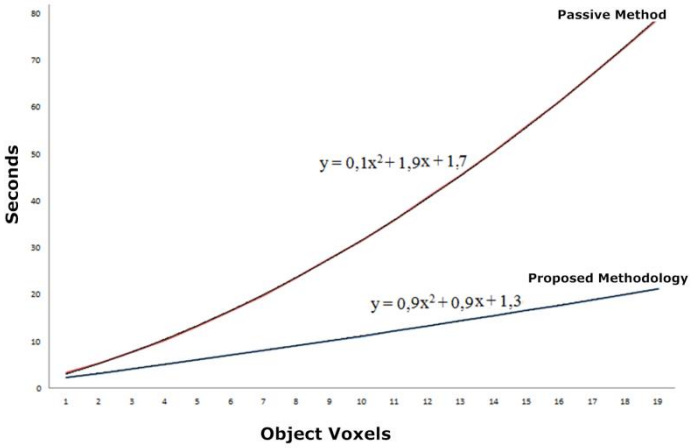


Fig. 11b. Relationship between processing time and the number of voxels of the reconstructed object.

The tendency of the processing time as a function of the number of voxels is polynomial and has a coefficient of determination of 0.99. As a result of this analysis it can be concluded that the proposed methodology improves in several aspects respect passive methods for reconstructing 3D objects, being able to reduce significantly the number of errors and also reducing the processing time used. The improvement of these features is complemented by a more optimal use of the technical resources that enable active methodologies, such as laser scanners.

VI. Conclusions and Future Work

In this paper we propose a novel hybrid methodology for reconstruction of high resolution 3D objects. These objects are also topologically complex and can have bright parts. The methodology is composed by 10 phases that combine active and passive methods, using data obtained from digital images and a laser scanner in order to supplement the missing information and obtain better results in the 3D object reconstruction. The proposed methodology takes advantage of the passive and active methods of object reconstruction, using the level of detail obtained with the laser scanner and the topology recovered by the digital images method. As it has been demonstrated in the Section 5 with the combination method, the proposed methodology has improved the quality of the reconstruction by significantly reducing the number of errors depending on the object's voxels. The proposed methodology solved the problems derived from the need to handle large volumes of object data from multiple methods and the need for different tool calibrations. In the different phases of the proposed methodology, several specific techniques have been applied to reuse object data, calibrations and optimize the computation cost.

Also, we observed that small scanner laser errors generate large errors in the reconstruction of the object. For this purposes the methodology introduces a new approach to reduce the complexity presented by the scanner's laser calibrations, it uses the box enclosing the 3D object convex hull to register the models. The use of digital images in this methodology allows us to obtain the texture-mapping of the object, making it well suited to achieve realistic object reconstructions. The geometric 3D object is calibrated with the high resolution shapes, obtained from photography, to make texture mapping as realistic as possible.

There are many directions for further research. The methodology can be modified at some phase to guide their tasks and algorithms towards achieving better quality and efficiency in the reconstruction of a specific type of topological complex objects, such as industrial material. A modification of the methodology it's being evaluated in order to optimize it for use it in real-time objects recognition under unfavorable circumstances for active recognition.

References

[1] Leo MJ, Manimegalai D, editors. 3D modeling of human faces- A survey. Trendz in Information Sciences and Computing (TISC), 2011 3rd International Conference on; 2011 8-9 Dec. 2011.

[2] Supak M, editor. Mechatronics in a household - sound recording. MECHATRONIKA, 2010 13th International Symposium; 2010 2-4 June 2010.

[3] Suveg I, Vosselman G. Reconstruction of 3D building models from aerial images and maps. ISPRS Journal of Photogrammetry and Remote Sensing. 2004;58(3–4):202-224.

[4] Park J-S. Interactive 3D reconstruction from multiple images: A primitive-based approach. Pattern Recognition Letters. 2005;26(16):2558-2571.
[5] Ducke B, Score D, Reeves J. Multiview 3D reconstruction of the archaeological site at Weymouth from image series. Computers & Graphics. 2011;35(2):375-382.
[6] Grum M, Bors AG Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on; 12-15 Oct. 2008; p. 209-212
[7] Lowe DG. Distinctive Image Features from Scale-Invariant Keypoints. Int J Comput Vision. 2004;60(2):91-110.
[8] Fischer A, Manor A, Barhak Y, editors. Adaptive parameterization for reconstruction of 3D freeform objects from laser-scanned data. Computer Graphics and Applications, 1999 Proceedings Seventh Pacific Conference on; 1999
[9] Modi RA, Goyal SR, editors. An efficient algorithm for 3-D object reconstruction from 2-D images. Image Information Processing (ICIIP), 2011 International Conference on; 2011 3-5 Nov. 2011.
[10] Yemez, Y. and Wetherilt, CJ. 2007. A volumetric fusion technique for surface reconstruction from silhouettes and range data. *Comput. Vis. Image Underst.* 105, 1 (Jan. 2007), 30-41.
[11] Chen G, Shen B, editors. A Visual Hull Algorithm of 3D Reconstruction Based on Interframe Coherence. Virtual Reality and Visualization (ICVRV), 2011 International Conference on; 2011 4-5 Nov. 2011.
[12] Vogiatzis, G., Torr, P. and Cipolla, R. Multi-view stereo via volumetric graph-cuts. CVPR 2005, pp. 391-398.
[13] Furukawa, Y. and Ponce, J. High-fidelity image-based modeling. Technical Report 2006-02, UIUC, 2006.
[14] Habbecke, M. and Kobbelt, L. A surface-growing approach to multi-view stereo reconstruction. CVPR 2007.
[15] Seitz, S., Curless, B., Diebel, J., Scharstein, D. and Szeliski, R. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms. CVPR 2006, vol. 1, pages 519-526.
[16] Huei-Yung L, Jing-Ren W, editors. 3D reconstruction by combining shape from silhouette with stereo. Pattern Recognition, 2008 ICPR 2008 19th International Conference on; 2008 8-11 Dec. 2008
[17] Jang W-S, Ho Y-S. 3-D object reconstruction from multiple 2-D images. Journal 3D Research. 2010;1(2):1-5.
[18] Hernandez, C. and Schmitt, F. Silhouette and stereo fusion for 3D object modeling. CVIU, 96(3):367-392, 2004.
[19] Pons, JP. Keriven, R. and Faugeras, O. Modelling dynamic scenes by registering multi-view image sequences. CVPR 2005, vol. II, pp. 822-827.
[20] Kolmogorov, V. and Zabih, R. Multi-camera scene reconstruction via graph cuts. ECCV 2002, vol. III, pp. 82-96.
[21] Hornung A. and Kobbelt, L. Hierarchical volumetric multi-view stereo reconstruction of manifold surfaces based on dual graph embedding. CVPR 2006.
[22] Laurentini, A. The visual hull concept for silhouette based image understanding, IEEE. Trans. on PAMI 16 (2) (1994) 150.162.
[23] Goesele, M., Curless, B. and Seitz, S. Multi-view stereo revisited. CVPR 2006.
[24] Chari V, Agrawal A, Taguchi Y, Ramalingam S, editors. Convex bricks: A new primitive for visual hull modeling and reconstruction. Robotics and Automation (ICRA), 2012 IEEE International Conference on; 2012 14-18 May 2012.
[25] Bradley, D., Boubekeur, T. and Heidrich, W. Accurate multi-view reconstruction using robust binocular stereo and surface meshing. CVPR 2008.
[26] Boissonnat, J. 1984. Geometric structures for three-dimensional shape representation. *ACM Trans. Graph.* 3, 4 (Oct. 1984), 266-286.
[27] Hoppe, H., DeRose, T., Duchamp, T., McDonald, J. and W. Stuetzle. Surface reconstruction from unorganized points. In *Computer Graphics (SIGGRAPH '92 Proceedings)*, volume 26, pages 71–78, July 1992.
[28] Na L, Bao-de F, editors. An efficient ISO-surface extraction method supported by topology correct. Image and Signal Processing (CISP), 2010 3rd International Congress on; 2010 16-18 Oct. 2010.
[29] Wei F, Bingwei H, editors. Automatic view planning for 3D reconstruction and occlusion handling based on the integration of active and passive vision. Industrial Electronics (ISIE), 2012 IEEE International Symposium on; 2012 28-31 May 2012.
[30] Bouquet, J. *Camera Calibration Toolbox for Matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/.*,2008
[31] Noborio, H., Fukuda, S., Arimoto, S., "Construction of the octree approximating a three-dimensional object by using multiple views," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.10, no.6, pp.769-782, Nov 1988.
[32] Zhang, Q., Pless, R., "Extrinsic calibration of a camera and laser range finder (improves camera calibration)," *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on* , vol.3, no., pp. 2301-2306 vol.3, 28 Sept.-2 Oct. 2004.
[33] Schmitt, F., Yemez, Y., "3D color object reconstruction from 2D image sequences," *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on* , vol.3, no., pp.65-69 vol.3, 1999
[34] Bolivar-Baron H, Pachecho A, Gonzalez-Crespo R. Semantics of immersive web through its architectural structure and graphic primitives. International Journal of Artificial Intelligence and Interactive Multimedia. 2010; Vol. 1, 47-52. http://dx.doi.org/10.9781/ijimai.2010.138.

Alexander Pacheco was born in August of 1983 in Bogota, Colombia. I studied at the University of Andes system engineering and electronic engineering. I have a master degree in computer science from Andes University. At this moment, I work in Wincor Nixdorf as a self-service software specialist. I worked at University Catolica de Colombia as a professor in computer science area. My research areas are computer vision, machine learning, image processing and artificial intelligence.

Holman Diego Bolívar B. is a Lecturer in the Engineering Faculty of the Catholic University of Colombia. Doctoral Student. from the Pontifical University of Salamanca in Computer Science. His research interests include Immersive Learning, Intelligent Agents, Computer Graphics and Complexity Analysis

Jordan Pascual Espada, **phD**. has developed research work at the University of Oviedo and is a guest professor in the same university. He also collaborates with companies like ElasticBox. He previously worked as a developer for mobile devices in Treelogic. Throughout his career he has participated in several national research projects, published several books and over 10 scientific articles in international journals. His areas of expertise include mobile platforms and the Web. In his work as a trainer has taught numerous courses on different technologies in the private sector.

Rubén González Crespo, phD., is the deputy director of the School of Engineering at the Universidad Internacional de La Rioja. Professor of Project Management and Engineering of Web sites. He is honorary professor and guest of various institutions such as the University of Oviedo and University Francisco José de Caldas. Previously, he worked as Manager and Director of Graduate Chair in the School of Engineering and Architecture at the Pontifical University of Salamanca for over 10 years. He has participated in numerous projects I + D + I such as SEACW, GMOSS, eInkPlusPlusy among others. He advises a number of public and private, national and international institutions. His research and scientific production focuses on accessibility, web engineering, mobile technologies and project management. He has published more than 80 works in indexed research journals, books, book chapters and conferences.

A Hybrid Algorithm for Recognizing the Position of Ezafe Constructions in Persian Texts

Samira Noferesti, Mehrnoush Shamsfard,

Faculty of Electrical and Computer Engineering,

Shahid Beheshti University, Iran.

***Abstract* — In the Persian language, an Ezafe construction is a linking element which joins the head of a phrase to its modifiers. The Ezafe in its simplest form is pronounced as –e, but generally not indicated in writing. Determining the position of an Ezafe is advantageous for disambiguating the boundary of the syntactic phrases which is a fundamental task in most natural language processing applications. This paper introduces a framework for combining genetic algorithms with rule-based models that brings the advantages of both approaches and overcomes their problems. This framework was used for recognizing the position of Ezafe constructions in Persian written texts. At the first stage, the rule-based model was applied to tag some tokens of an input sentence. Then, in the second stage, the search capabilities of the genetic algorithm were used to assign the Ezafe tag to untagged tokens using the previously captured training information. The proposed framework was evaluated on Peykareh corpus and it achieved 95.26 percent accuracy. Test results show that this proposed approach outperformed other approaches for recognizing the position of Ezafe constructions.**

***Keywords* — Ezafe construction, genetic algorithm, genitive construction, rule-based model.**

I. INTRODUCTION

THE "Ezafe"[1] is a Persian language grammatical construct which links two words together. Ezafe means "addition" and is an unstressed vowel –e– which marks genitive cases. The constructs linked by the Ezafe particle are known as "Ezafe constructions". Some common uses of the Persian Ezafe are [1]:

— a noun before an adjective:
e.g. توپ قرمز[2] (tu:p-Ezafe Germez) "red ball"

— a noun before a possessor:
e.g. کتاب علی (ketã:b-Ezafe Ali:) "Ali's book"

— some prepositions before nouns:
e.g. زیر میز (zi:r-Ezafe mi:z) "under the table"

The Ezafe in its simplest form is pronounced as –e, but generally not indicated in writing. In some cases the Ezafe has an explicit sign in writing. For example, with nouns ending in ا (ã:) or و (ou), the Ezafe appears as an ی (j) at the end; with nouns ending in a silent ه (h) (short e followed by a mute h), the Ezafe may appear as a superscript ء (hamze) or a ی (j).

The Ezafe is also found in Urdo [2], Kurdish [3] and Turkish [4]. The Persian Ezafe has been discussed extensively [5]-[8]. This construction raises several issues in syntax and morphology. There are three issues on the function of the Ezafe in the literature: (1) the Ezafe is a case marker [9], (2) the Ezafe is inserted at PF to identify constituenthood [10], and (3) the Ezafe is a phrasal affix [11].

Determining the position of an Ezafe construct may facilitate text processing activities in natural language processing (NLP) applications, such as segmenting a phrase or detecting the head word of a phrase [12]. Moreover, recognizing words which need an Ezafe is advantageous for tokenization [13], morphological analysis, and syntax parsing [14], and it is essential for speech synthesis [15].

Some NLP tasks in Persian, such as machine translation [16], construction of morphological lexicons [14], and grammar construction [17], have benefited from the availability of an Ezafe construction. However, they have determined the position of the Ezafe manually, exploited cases in which the Ezafe is visually represented, or extracted some insertion rules which are not general; therefore, they could not determine the Ezafe tags for all tokens in a text.

The Persian Ezafe has been discussed extensively in theory [5], [18], but there are few works on the automatic detection of this construction in Persian texts. The most completely reported works on this subject are one work based on probabilistic context free grammar (PCFG) [19] and another based on classification and regression tree (CART) [20]. The former uses a bank including trees of noun groups in Persian for training PCFG. Then, a bottom-up parser extracts the most probable noun groups of the input. Finally, using lexical analysis, the system determines which words need an Ezafe. The disadvantage of this method is that writing a PCFG requires a large amount of linguistic knowledge. In addition, it is not sensitive to lexical information. The latter uses morpho-syntactic features of words to train and construct binary classification trees to predict the presence or absence of an

[1] It is also known as Kasreh.

[2] For each Persian word or phrase we wrote its transliteration within parenthesis and its English meaning within double quotes. International Phonetic Alphabet (IPA) was used to represent Persian language pronunciations.

DOI: 10.9781/ijimai.2014.262

Ezafe between two adjacent words. In fact, there are two kinds of rules: rules which predict Ezafe words, and rules which predict non-Ezafe words. Although this method can predict the absence of an Ezafe with high accuracy, it is not sufficient in detecting words which need an Ezafe. In other words, the rules which predict the non-Ezafe words act more precisely.

The main contributions of this paper were (1) introducing a framework for combining genetic algorithms with rule-based models, and (2) using the proposed framework to develop an Ezafe tagger.

Combining genetic algorithms with rule-based models brings the advantages of both approaches and overcomes their problems. Genetic algorithms can detect general patterns in text, but sometimes they cannot handle exceptions and special cases. In such cases, rule-based models can provide significant improvements by defining rules for handling special cases and exceptions. In our proposed framework, for the rule-based model, linguistic rules were extended by analyzing errors of the genetic algorithm and defining new rules for handling these errors (named as correction rules). In contrast, a rule-based model needs a great deal of knowledge external to the corpus that only linguistic experts can generate. In fact, acquiring rules through interviews with experts is cumbersome and time-consuming. Furthermore, certain application domains are very complicated and may require a large number of rules. Therefore, the acquired rules may be incomplete or even partially correct. In order to overcome these problems, we can handle general patterns by genetic algorithms and only define correction rules to handle special cases, which means less time handling by expert humans. We can also define a set of general rules besides correction rules in order to reduce the run time of the genetic algorithm.

There is a remarkable amount of ongoing research on applying machine learning approaches to different tasks of NLP in the English language. Most machine learning approaches such as those methods based on hidden Markov models, use information extracted from a tagged corpus to assign a suitable tag to each word according to preceding tags. Since these approaches are purely statistical, as such they are most suitable for cases that have a corpus large enough to contain all possible combinations of n-grams. In contrast, evolutionary algorithms offer a more generalized method that can be applied to any statistical model. For example, they can be applied to perform tagging operations according to the Markov model (tag prediction for a current word based on preceding tags) or improve the Markov results by using more contextual information (for example, using tags of preceding words or those of following words). In other words, HMM or other models can be used as part of the fitness function in a genetic algorithm. Therefore, a genetic algorithm provides more flexibility than any of the other classical approaches such as HMM based methods. On the other hand, the effectiveness of using hybrid approaches has been demonstrated in different NLP tasks [21]. Thus a hybrid approach for determining the position of Ezafe construction was chosen for this study. Results of the tests in this study show that our proposed algorithm outperformed other algorithms for Persian Ezafe tagging as well as the classical MM based method.

The rest of the paper is organized as follows. Section 2 introduces the annotated corpus of Persian texts. Section 3 explains our proposed model. Experiment results are discussed in section 4. Finally, section 5 concludes the paper.

II. THE CORPUS

An annotated corpus of Persian text is needed in order to train and evaluate the Ezafe tagger. This corpus must be annotated with POS and Ezafe tags. For the current work, a subset of Persian POS tagged corpora known as Peykareh[3] [21] was used. This collection was gathered from daily news and common texts and contained about 2.6 million, manually tagged tokens. The main corpus was tagged with a rich set of POS tags consisting of 550 different tags from which 20 tags were selected for the system. Those that could be detected by the present Persian POS taggers were selected for use in the system applied to this study.

The tagged corpus was divided into three sets: (1) a training set including 423,721 tokens, (2) a held-out data set containing 1,010,375 tokens, and (3) a test set containing 39,850 tokens.

A big portion of the Peykareh corpus was set aside as a held-out dataset. The held-out dataset was used to find exceptions to general rules in the rule-based model. Since the exceptions occur only rarely, much more data was needed to determine the exceptions. Furthermore, to determine the classes of conjunctions and prepositions which never take an Ezafe or always require an Ezafe, the held-out data set was searched. Thus, a sufficiently large data set was needed to find as many words as possible.

III. THE PROPOSED ALGORITHM

This paper proposes a hybrid approach to determine the position of Ezafe constructions in Persian texts. The Ezafe tagger contained two phases. The first phase used the rule-based model to tag as many words as possible. Then the second phase ran the genetic algorithm to assign tags to the tokens which had not been assigned an Ezafe tag in the previous phase. Therefore, a faster genetic algorithm was achieved by producing more tagged tokens that had been generated from the rule-based model.

The Ezafe tagger assigned each word of an input sentence with one of two tags: *true* or *false*. Tag *true* for a word meant that it requires an Ezafe, and the tag *false* meant it does not require one.

A. The rule-based model

Initially, some general rules such as “verbs do not take an Ezafe” were defined using linguistic knowledge. Although the genetic algorithm could detect these tags correctly by training on annotated examples, we preferred to define such rules in

[3] This corpus also is known by its author’s name, Bijankhan.

order to reduce the run time of the algorithm. The more tokens detected by the rule-based model there were, the less chromosome length and lower number of generations were needed.

Next, the exceptions of each rule were explored on the held-out data, and some new rules were defined to handle exceptions. This process was repeated. In other words, if the generated rules had exceptions, new rules were defined. In some cases we could not find suitable rules to fix errors. In these cases, probabilistic rules were used.

The genetic algorithm tagged tokens according to the context; however, experiments showed that words which appeared in an infrequent context usually took an incorrect tag from the genetic algorithm. We tried to handle these cases through hand-crafted rules. Thus, the initial rule-based model and the genetic algorithm were run on the held-out data, and errors were analyzed to introduce rules that would fix them.

In this way, a set of 53 hand-crafted rules was developed. Then, the most suitable sequence of rules was determined in terms of avoiding bleeding and creeping; in fact, a tree was constructed. The first level of the tree contained some general rules. Level 2, consisted of some rules for handling exceptions of the first level and so it continued. However, each node in level *i* handled an exception of the rule of its parent node in level *i-1* (if that rule had exceptions).

At the first stage of the proposed algorithm, each rule was taken individually from the rule-set one at a time and the function was performed only if the rule was applicable to the input word.

Rules were categorized according to various dimensions:

— Deterministic vs. probabilistic rules: Deterministic rules are those which are always valid and correct; probabilistic rules may have exceptions. In other words, probabilistic rules are valid most of the time, but as they may have exceptions we apply them with a probability of less than 100%. This probability is extracted from the corpus.

— Negative vs. positive rules: Negative rules find and tag negative examples which are the structures which never take an Ezafe, while the positive rules determine structures which need an Ezafe.

— Syntactic, morphological and lexical rules: Syntactic rules use part-of-speech to tag words (either as the target word or a neighboring word) in a sequence to determine the Ezafe tag, while morphological rules consider internal and morphological structures of a word to do this task, and lexical rules consider real words.

In the rest of this section these categories are discussed in more detail some examples are given from each category.

1) Syntactic Rules

Some POS categories enforced a special tag on words or on neighboring words. The accusative case marker را (rã:) and verbs were among this set.

— Verbs

- In the Persian language the Ezafe is not used with verbs. The following rule dictated that verbs, which were shown in the corpus with the POS tag V, never take an Ezafe.
  If POS(X) =V Then EZ-Tag(X) =false
- If a verb appears as a stand-alone, the word before it does not take an Ezafe. We presented this by the following rule:
  If POS(X) =V Then EZ-Tag(X-1) =false
- If the verb is not a stand-alone and appears as an attachment (enclitic) to another word, then the previous word (before the combination) may take an Ezafe. This is the case for some of the enclitics representing the copula verb 'to be' such as ی (i:) "to be- single second person" and ًم (æm) "to be- single first person". These enclitics are ambiguous and, in addition to copula verbs, can be interpreted as an indefinite marker or as a single first-person possessive pronoun, respectively. For example, the word شاعری (ʃã:?eri:) may mean شاعر هستی (ʃã:?er hæsti:) "you are poet" or یک شاعر (jek ʃã:?er) "a poet". In the first case, even though the whole word was tagged as a verb in the corpus, it is actually a combination of a noun and a verb. Even though its verb part and its previous word do not take an Ezafe, the word before the noun part of it may take one. As another example, in the following sentence the word دولتم (dolætæm) "I'm government" is an abbreviation of دولت هستم (dolæt hæstæm) "I am government". In Peykareh corpus, this word was tagged as verb with POS tag V,AJCC. However, the previous word takes an Ezafe.
  من در استخدام دولتم.
  I am a government employee.
  Thus, we used POS tag V, ACJJ for this kind of verbs to prevent applying the previous rule for them.

— The accusative case marker را (rã:)

- The Persian language has an accusative case marker را (rã:) that follows the direct object, adverb or prepositional object. The following rule dictated that the accusative case marker, which was shown in the corpus with the POS tag POSTP, never takes an Ezafe.
  If POS(X) =POSTP Then EZ-Tag(X)=false
- The word before را (rã:) does not take an Ezafe too.
  If POS(X) =POSTP Then EZ-Tag(X-1)=false
- In some cases the accusative case marker is attached to the previous noun or pronoun. For example the word مرا[4] (mærã:) is an abbreviation of من را (mæn rã:), in which را (rã:) is an object marker and من (mæn) "me" is a pronoun. This rule was written as follows:
  If postfix(X) = accusative-case-marker EZ_Tag(X) =false

2) Morphological rules

The following rules are examples of morphological rules that determine structures that take an Ezafe.

—When a word ending in the plural suffix ها (hã:) needs the

[4] Sometimes it means 'my' and other times it means 'me'

Ezafe, the letter ی (j) must be attached to the end of the word in writing. Thus, if a plural word ends in ها (hã:), this word should not be followed by an Ezafe unless it is followed by a ی (j) clitic.

If postfix(X)= ها Then EZ-Tag(X)=false

Consider the following example.

نامه های علی را خواندم.

I read Ali's letters.

The word نامه ها (nãːme hãː) "letters" is the plural form of نامه (nãːme) "letter". When this word requires Ezafe, we add ی (j) at the end of it.

—If the last character of a word is اً (Tanvin)[5], then it does not take an Ezafe.

If LastChar(X) = اً Then EZ-Tag(X)=false

3) Lexical rules

Lexical rules consider real form of words as shown in the following examples:

—Prepositions

Reference [23] showed that the class of prepositions in the Persian language is not uniform with respect to the Ezafe. Some prepositions reject the Ezafe (These prepositions were called Class P1.), while others either permit or require it. We divided the latter group into two classes. The first class which always requires an Ezafe was called Class P2. The other class which permits an Ezafe but does not necessarily require one was called Class P3. Table I shows some examples of each class. We applied the following rules to handle prepositions:

If POS(X) =P and WORD(X)∈ClassP1 Then EZ-Tag(X)=false

If POS(X) =P and WORD(X)∈ClassP2 Then EZ-Tag(X)=true

TABLE I
EXAMPLES OF PREPOSITION CLASSES

Class name	Examples
Class P1	به (be) "to" از (æz) "from" با (bãː) "with" در (dær) " in, on"
Class P2	وسط (væsæt) "in the middle" دور (duːr) "around" بیرون (biːruːn) "outside" داخل (dãːxel) "inside"
Class P3	زیر (ziːr) "under" رو (ruː) "on" بالا (bãːlãː) "up" جلو (dʒoulou) "in front of"

— Conjunctions

Same as prepositions, we divided conjunctions into two classes. Some conjunctions never take an Ezafe (These conjunctions were called Class C1), while others always take an Ezafe (These conjunctions were called Class C2). In order to determine these classes we searched 300 files of Peykareh corpus which were selected as held-out data. Some examples of Class C1 and Class C2 are presented in Table II. The following rules applied to conjunctions:

If POS(X) =CONJ and WORD(X)∈ClassC1 Then EZ-Tag(X)=false

If POS(X) =CONJ and WORD(X)∈ClassC2 Then EZ-Tag(X)=true

TABLE II
EXAMPLES OF CONJUNCTION CLASSES

Class name	Examples
Class C1	و (væ) "and" زیرا (ziːrãː) "because" یا (jãː) "or" که (ke) "that" یعنی (jæniː) "means"
Class C2	علی رغم (ʔælãːræɣ me) "in spite of" باستثناء (beestesnãːʔe) "except" سوای (sævãːje) "except" برخلاف (bærxælãːfe) "in spite of"

—Adverbs

We also divided Persian adverbs into three classes. Class A1 contained adverbs which never take an Ezafe; class A2 included adverbs with an obligatory Ezafe; class A3 contained adverbs with an optional Ezafe. Examples of these classes are shown in Table III. The following rules applied to adverbs:

If POS(X) =ADV and WORD(X)∈ClassA1 Then EZ-Tag(X)=false

If POS(X) =ADV and WORD(X) ∈ClassA2 Then EZ-Tag(X)=true

TABLE III
EXAMPLES OF ADVERB CLASSES

Class name	Examples
Class A1	بویژه (beviːʒe) "specially" هیچگاه (hiːtʃgãːh) "never" شاید (ʃãːjæd) "maybe"
Class A2	مثل (mesle) "like" مانند (mãːnænde) "like" از قبیل (æz Gæbiːle) "such as"
Class A3	گذشته (gozæʃte) "past" سالیانه (sãːliːjãːneh) "annual"

4) Probabilistic rules

We also defined 5 probabilistic rules which were correct and valid in most cases but had some exceptions in a few cases. Defining each rule, the probability of that rule was calculated according to the corpus. The lowest probability among these rules was 0.95. Here, we discuss some of the probabilistic rules.

—Long vowels

There are three long vowels in Persian: ا (ã:), ی (i:) and و (u:). Generally, when a word ending in ا (ã:) or و (u:) needs an Ezafe, the letter ی (j) is added to the end of it. However, this rule has some exceptions.

In the case of ا (ã:), these exceptions happen when we replace أ (Alef Hamze) by ا (ã:) (single alef). Alef Hamze is a single Arabic character that represents the two-character

[5] This sign was taken from Arabic alphabet

combination of Alef plus Hamze and in Persian writing is sometimes replaced by the letter ا (ã:). Consider the following example:

آقای احمدی منشا فساد را فقر می داند.

"Mr. Ahamdi believes that the source of evil is poverty."

The word آقا (ã:Gã:) "Mr." takes a ی (j) at the end because it requires the Ezafe; however, the word منشا (mænʃæ) "source" also ends in ا (ã:) and needs the Ezafe, but it does not get ی (j). In fact, the last character of this word is أ (Alef Hamze) which is written the same as ا (ã:).

To compute the probability of the rule, the algorithm searched the held-out data set and computed the percentage of words ending with ا (ã:) and the Ezafe which had the letter ی (j) added. In other words, this probability was computed by the following formula:

$$p = \frac{\text{count(words ending with ی ا (ã:je) and } True \text{ Tag)}}{\text{count(words end ing with ی ا (ãje))}} \quad (1)$$

Thus, the following rule was defined with a 95% probability:

If LastChars(X) =ی ا Then with a 0.95 probability EZ-Tag(X) =true

In the same way, the following rules were defined:

If LastChars(X) = ا Then with a 0.9978 probability EZ-Tag(X) =false

If LastChars(X) =و ی Then with a 0.96 probability EZ-Tag(X) =true

—Tanvin

Tanvin is a sign which is derived from Arabic. The following rule says that with a probability of 96.24% the word preceding a word that has ً (Tanvin) as the final character does not take an Ezafe:

If LastChars(X) = ً Then with a 0.9624 probability EZ-Tag(X-1) =false

Frequency counts for the rule-categories are shown in Table IV.

TABLE IV
NUMBER OF EXTRACTED RULES IN EACH CATEGORY

Syntactic	Morphological	Lexical	Probabilistic
25	4	19	5

After running the rule-based model, some of the tokens remained untagged. Thus, a genetic based algorithm was used to tag the remaining words.

B. Genetic tagging algorithm

The proposed genetic algorithm receives a natural language sentence and assigns a corresponding tag according to previously computed training information from the annotated corpus. Formally, given a sequence of *n* words and corresponding POS tags, the aim is to find the most probable Ezafe tag sequence.

In our implementation, each gene can take values: *true* or *false*. Individuals of the first generation were produced randomly. After producing an individual, all tokens of a given sentence were assigned Ezafe tags (some of tokens get Ezafe tag by the rule-based model and others get Ezafe tag by the genetic algorithm).

An initial population was created randomly by assigning a random value to each untagged gene (some genes were assigned Ezafe tags from the rule-based model). These individuals were sorted according to fitness value of individuals from high to low.

Three genetic operations were used for producing the next generation.

—Selection: All individuals in the population are sorted according to fitness, so the first individual was the best fit in the generation. To perform crossover, the *i*th and *(i+1)*th individuals of the current generation were selected, where $i=1,2,...,[(p+1)/2]$ and *p* was the population size. The aim of selection was to choose the fitter individuals.

— Crossover: Selected two chromosomes, crossover exchanges portioned of a pair of chromosomes at a randomly chosen point called the crossover point.

— Mutation: Selected an untagged gene randomly and toggled its value, for example if its value was *true*, it was reset to *false* and vice versa.

1) Fitness Functions

To evaluate the quality of Ezafe tags generated for an individual, four functions were used; F1, F2, F3 and F4. These functions considered the context in which a word appeared. Context consisted of a current word, one tag to the left and another to the right and the previous and next word.

F1 considered the sequence of POS tags of a sentence. The probability of the sequence of POS tags of a sequence of *n* words was as follows:

$$F1 = \sum_{i=1}^{n} P(EZ - POS_i | POS_{i+1}) \quad (2)$$

Where, $P(EZ - POS_i | POS_{i+1})$ represents the probability that the current word with POS_i tag gets the Ezafe when the next word has the POS_{i+1} tag. The probability of assigning the Ezafe to a word given the next POS tag was computed as:

$$P(EZ - POS_i | POS_{i+1}) = \frac{count(EZ - POS_i, POS_{i+1})}{count(POS_i, POS_{i+1})} \quad (3)$$

Where, count(POS_i, POS_{i+1}) was the number of occurrences of the (POS_i, POS_{i+1}) sequence within the training corpus, and $count(EZ - POS_i, POS_{i+1})$ was the number of (POS_i, POS_{i+1}) occurrences when the first token has an Ezafe within the same corpus. In order to compute F1 function, the HMM model can be used with the Viterbi algorithm [24].

For computing F2 function, a data driven approach was applied to calculate the probability that a specific word has the Ezafe.

$$F2 = \sum_{i=1}^{n} P(EZ_i | word_i) \quad (4)$$

F2 was defined because some words in Persian are mostly assigned a special tag. For example, the word تـقـريـبـاً (tæGri:bæn) "approximately" never take an Ezafe.

F3 function was the probability that a token gets an Ezafe when it occurs before a specific word in the training corpus.

$$F3 = \sum_{i=1}^{n} P(EZ_i | word_i , word_{i+1}) \quad (5)$$

This function was defined to handle compound words such as اخـتـلاف نـظـر (extelã:f-Ezafe næzær) "difference in opinion".

In Persian, some words such as سـايـر (sã:jer) "other" get the Ezafe most of the time. Therefore, we defined the F4 function to consider these words. The F4 function was the probability that a specific word occurs after a word with the Ezafe in the training corpus.

$$F4 = \sum_{i=1}^{n} P(word_{i+1} | EZ_i) \quad (6)$$

The following fitness function was used to evaluate the genetic algorithm:

$$fitness - function = \sum_{i=1}^{4} w_i . F_i \quad (7)$$

Where $w_i s$ are constant parameters chosen from [0,1) and show relative importance of syntactic and lexical information. It was assumed that 0 is a legal value to show the effect of removing one or more functions from the formula. To adjust w_i parameters in the fitness function formula, variable structure learning automata were applied on chunked held-out data. For more information you can see [25]. Finally, the values of w_1, w_2, w_3 and w_4 were set to 0.8, 0.5, 0.1 and 0.1 respectively.

IV. Experiments

The proposed algorithm was implemented using java language and was run on a Pentium IV processor. First, the rule-based model was run followed by the genetic algorithm, and the best solution was selected. Approximately 78% of the tokens were tagged with the rule-based model, because about 80% of the tokens selected as test data did not require an Ezafe, and most of them were tagged by the rule-based system.

To evaluate the performance of our proposed algorithm, three measures were taken: accuracy (the percentage of correctly tagged tokens), precision (the percentage of predicted tags that were correct) and recall (the percentage of predictable tags that were found).

Since performance was related to both precision and recall, the *F*-measure was given as the final evaluation.

$$F - measure = \frac{2.precision.recall}{precision + recall} \quad (8)$$

A. Tuning Parameters of the Genetic Algorithm

The efficiency of a genetic algorithm greatly depends on how its parameters are tuned. To adjust the genetic parameters, a subset of 34,832 tokens from held-out data set was selected. Then, the proposed algorithm was run on this set.

Beginning with a baseline configuration, such as Dejong's setting [26] with 1000 generations, 50 chromosomes in each generation and 0.6 for crossover probability, the algorithm was run for different mutation probabilities (P_m) from 0.01 to 0.3. Fig. 1 shows that the best results were obtained using the mutation probability 0.05.

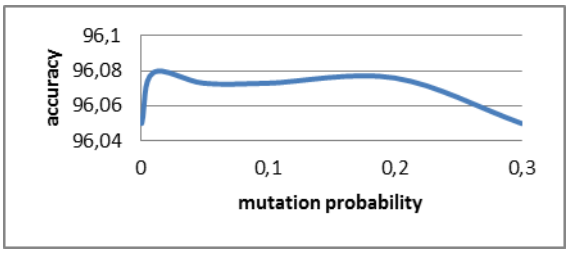


Fig. 1. Average fitness values of executing of the GA using different mutation probabilities

In the same way, crossover probability was set to 0.6. In Fig. 2 the results of running the genetic algorithm using mutation probability 0.05, crossover probability 0.6, population size 50 and different number of generations are shown.

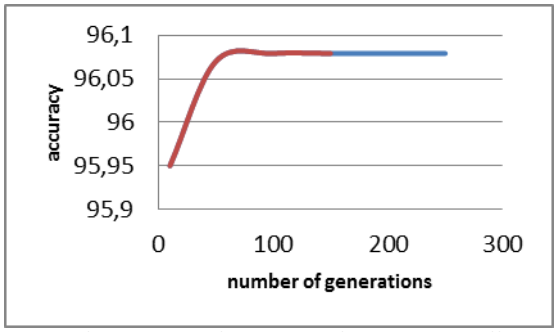


Fig. 2. Average fitness values of executions of the GA using different number of generations

Table V shows the optimum values of genetic algorithm parameters.

TABLE V
SELECTED VALUES FOR GENETIC ALGORITHM PARAMETERS

P_m	P_c	population size	generations
0.05	0.6	50	150

B. Effectiveness of the Proposed Algorithm

The experiment applied 423,721 annotated tokens as the training set and 39,850 tokens as the test set. Parameter settings shown in Table V were used for the genetic algorithm.

Table VI compares our approach with a baseline method and other available methods based on PCFG [19] and CART [20]. We also implemented the binary Markov model with Viterbi decoding (a typical algorithm widely used for stochastic tagging). As can be seen, our proposed algorithm

outperformed these algorithms in terms of F-measure. The baseline assumed all words have an Ezafe, resulting in 100% recall but very low precision (15.79%). We could define another baseline where no word has the Ezafe tag. In this case, we would achieve 84.21% precision, but the recall would be 0.

TABLE VI
EVALUATING OUR PROPOSED ALGORITHM IN TERMS OF F-MEASURE

	recall	precision	F-measure
Our proposed algorithm	88.81	87.85	88.33
Baseline	100	15.79	27.27
PCFG method [19]	86.74	87.54	87.14
CART method [20]	88.85	84.13	86.43
Viterbi method	95.51	78.63	86.25

Since we had no access to the corpus that was used for training in the CART method [20] or a description of the exact features used, we could not regenerate the exact results. For this reason, we used two approaches for comparison. In the first approach, we compared results of our proposed method with the best results reported in [20], and in the second one, we implemented the CART method using the same features as our proposed method and tested it on our test corpus. Since the performance of the second approach was much lower than what was reported in [20], we only presented the results of the first approach in Table VI.

In the above-mentioned experiments, correct POS tags were used, because results from the proposed algorithm were compared to those from other available Ezafe taggers. Since these taggers had used correct tags, we also used the correct tags to enable the comparison. By using a Tnt tagger, the proposed algorithm achieved a 95.08% accuracy, while with correct tags it achieved an accuracy of 95.49%. This indicates that the tagging error decreased the Ezafe detection accuracy by only about 0.41%. The reason for this is that both the rule-based model and the genetic algorithm consider other features besides POS tags, and these features can, to some extent, cover the errors of the POS tagger.

Considering accuracy as the percentage of correctly assigned tags, we evaluated the performance of the proposed algorithm from two different aspects: (1) the overall accuracy by taking all tokens in the test corpus into account, and (2) the accuracy for words with an Ezafe and without an Ezafe, respectively. Table VII shows that the overall accuracy of the proposed algorithm was around 95.26%. Additionally, the accuracy for detecting words without an Ezafe was significantly higher than that for words with an Ezafe (96.89% versus 88.81%).

TABLE VII
EVALUATING OUR PROPOSED ALGORITHM IN TERMS OF ACCURACY

Number of correctly tagged tokens			
	with Ezafe	without Ezafe	Total
Corpus	8054	31796	39850
Our proposed Algorithm	7153	30807	37960
Accuracy	88.81	96.89	95.26

Table VIII compares overall accuracy from the combination of the rule-based model and the genetic algorithm. Approximately 78% of tokens were tagged by the rule-based model with 99.21% accuracy. In fact, from tokens in the test set, 30,972 tokens were tagged by the rule-based model and among them 30,728 tokens were assigned correct tags. In contrast, the genetic algorithm assigned correct tags to 7,232 tokens from 8,878 tokens and achieved 81.46% accuracy.

TABLE VIII
COMPARING THE ACCURACY OF THE RULE-BASED MODEL VERSUS GENETIC ALGORITHM

	Number of tagged tokens	Number of correctly tagged tokens	Accuracy
Rule-based model	30972	30728	99.21
Genetic algorithm	8878	7232	81.46

Table IX compares the accuracy of the rule-based model versus the genetic algorithm. In RBM1, we ran the rule-based model and assigned the false tag to tokens which did not get the Ezafe tag after applying the rules. In contrast, the untagged tokens got true tags in RBM2. We also ran the genetic algorithm alone (without the rule-based model). Results show that the combination of the rule-based model and the genetic algorithm outperformed both individual algorithms. As might be expected, the main problem of the RBM models was missing rules, which caused some tokens remained untagged, and the main problem of the genetic algorithm was special cases that could not be handled by general patterns.

TABLE IX
COMPARING THE ACCURACY OF THE RULE-BASED MODEL VERSUS GENETIC ALGORITHM

	Accuracy
RBM1	85.29
RBM2	91.21
GA	89.21
Combination of rule-based and GA	95.26

Since the ratio of words with an Ezafe to words without an Ezafe was low, the Kappa coefficient was used to evaluate the proposed algorithm. This measure was first suggested for linguistic classification tasks [27] and has since been used to avoid dependency of the score on the proportion of non-breaks in the text. The Kappa coefficient (K) was calculated as:

$$K = \frac{Pr(A) - Pr(E)}{1 - Pr(E)} \quad (9)$$

Where, Pr(A) was accuracy, and Pr(E) was the ratio of words without an Ezafe to total words. Table X shows how to evaluate an algorithm in terms of Kappa value. Using (9) the Kappa coefficient became 0.77. According to Table X, our proposed algorithm is assessed as good.

TABLE X
DECISION MAKING BY USING KAPPA [19]

Kappa values	Strength of agreement
K<0.2	bad
0.2<K≤0.4	average
0.4<K≤0.6	relatively good

0.6<K≤0.8	good
0.8<K≤1	very good

Table XI shows that our proposed algorithm outperformed previously reported algorithms in terms of Kappa value.

TABLE XI
COMPARING THE PERFORMANCE OF THE PROPOSED ALGORITHM WITH OTHER METHODS

Kappa value		
Our proposed algorithm	PCFG method	CART method
0.77	0.74	0.72

In the final experiment, we assessed the impact of training corpus size on the performance of the proposed algorithm. The corpus size was reduced slightly until it reached 32% of the initial training corpus size. The results are presented in Fig. 3. As can be seen, the proposed algorithm's accuracy did not show a significant drop when reducing the training corpus size from 100% to 60%.

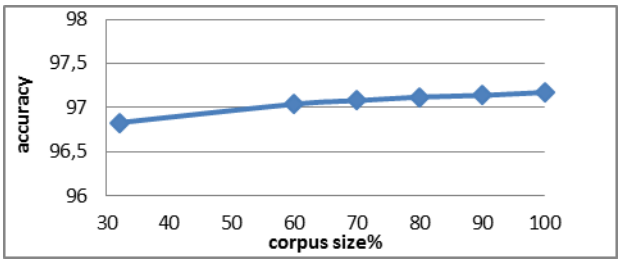


Fig. 3. The impact of training corpus size on performance

V. CONCLUSION AND FUTURE WORK

This paper proposes a framework for recognizing the position of Ezafe constructions in Persian written texts that combines genetic algorithms with rule-based models. Genetic algorithms provide a search strategy to learn general Ezafe patterns in text optimizing a measure of probability that is effective globally. However, the rule-based model handles special cases and exceptions to general patterns. Results of the tests reported in this study show that the proposed algorithm outperformed other algorithms for Persian text Ezafe tagging and classical HMM based methods.

Although this paper presents an algorithm for Persian Ezafe tagging, the principles can be applied to other NLP tasks such as POS tagging or chunking in any language. A genetic algorithm can be used for any language to find common statistical patterns for tagging. Obviously, there may be exceptions to these patterns, so some rules are defined to handle exceptions in the rule-based model that serve to improve performance of the genetic algorithm. In fact, combining a genetic algorithm with the rule-based model improves performance of the tagging process.

In addition, we showed that the accuracy of the proposed algorithm does not depend highly on the training corpus size. This feature is advantageous for practical applications, because annotating training corpora for text analysis purposes is an extremely demanding task.

In future work, linguistic rules may be extended by analyzing errors of test data. It is also observed that input of the Ezafe-tag set has a major influence on accuracy. Errors in the training data have caused some problems, and these can be reduced by correcting the training data.

In addition, it is intended that new attributes be added to the fitness function of the genetic algorithm. One advantage of the genetic algorithm compared to other classical approaches such as HMM based methods is that new attributes can be added to the system and this facilitates examination of the effect of different attributes on tagging without altering the system's basic structure. Thus, tests will be done on new attributes applied to the fitness function of the genetic algorithm and to evaluate effects on tagging accuracy.

It was also observed that high accuracy is extremely influenced by input tag set. A richer tag set with POS information produces more accurate results. For example, we can consider additional information with a POS noun, such as time, location, and so on. In addition, in [5] there is a class of lexical words called eventive adjectives, and they cannot co-occur with an Ezafe in contrast with other lexical words. Consider the following examples. Predicative adjectives may only appear in Light Verb Constructions (a) and not in Ezafe Constructions (b).

(a) علی کتاب را فراموش کرد "Ali forgot the book."

(b) * فراموش کتاب توسط علی "forgetting the book by Ali"

We are going to enrich the tagset with more POS tags such as eventive adjectives to define more accurate rules.

REFERENCES

[1] A. Kahnemuyipour, "Persian Ezafe construction: case, agreement or something else," in *Proceedings of the 2nd Workshop on Persian Language and Computer, Tehran University, Tehran, Iran*, 2006.

[2] T. Bögel, M. Butt, and S. Sulger, "Urdu ezafe and the morphology-syntax interface," *Proceedings of LFG08*, 2008.

[3] A. Holmberg and D. Odden, "The Izafe and NP structure in Hawrami," Durham Working Papers in Linguistics, 2004.

[4] G. Van Schaaik, *The noun in Turkish: Its argument structure and the compounding straitjacket*: Otto Harrassowitz Verlag, 2002.

[5] G. Karimi-Doostan, "Lexical categories in Persian," *Lingua*, vol. 121, pp. 207-220, 2011.

[6] A. K. Ahranjani, "The head noun in possessive construction in English and Persian languages," *International Journal of Academic Research*, vol. 2, 2010.

[7] A. K. Ahranjani and R. Tohidian, "Ezafe construction in complex noun phrases in Persian medieval poems," *International Journal of Academic Research*, vol. 3, 2011.

[8] A. Moinzadeh, "The Ezafe phrase in Persian: How complements are added to Ns and As," *Journal of Social Sciences & Humanities of Shiraz University*, vol. 23, pp. 45-57, 2006.

[9] R. Larson and H. Yamakido, "Ezafe and the deep position of nominal modifiers," in *Barcelona Workshop on Adjectives and Adverbs, Barcelona*, 2005.

[10] J. Ghomeshi, "Non-projecting nouns and the ezafe: construction in Persian," *Natural Language & Linguistic Theory*, vol. 15, pp. 729-788, 1997.

[11] P. Samvelian, *A (phrasal) affix analysis of the Persian Ezafe*: Cambridge Univ Press, 2007.
[12] P. Samvelian, "The Ezafe as a head-marking inflectional affix: Evidence from Persian and Kurmanji Kurdish," *Aspects of Iranian Linguistics: Papers in Honor of Mohammad Reza Bateni*, pp. 339-361, 2007.
[13] M. Ghayoomi and S. Momtazi, "Challenges in developing Persian corpora from online resources," in *Asian Language Processing, 2009. IALP'09. International Conference on*, 2009, pp. 108-113.
[14] B. Sagot and G. Walther, "A morphological lexicon for the Persian language," in *Proceedings of the 7th Language Resources and Evaluation Conference (LREC'10)*, 2010.
[15] M. Sheikhan, M. Tebyani, and M. Lotfizad, "Continuous speech recognition and syntactic processing in Iranian Farsi language," *International Journal of Speech Technology*, vol. 1, pp. 135-141, 1997.
[16] J. W. Amtrup, H. M. Rad, K. Megerdoomian, and R. Zajac, *Persian-English machine translation: An overview of the Shiraz project*: Citeseer, 2000.
[17] S. Muller and M. Ghayoomi, "PerGram: A TRALE implementation of an HPSG fragment of Persian," in *Computer Science and Information Technology (IMCSIT), Proceedings of the 2010 International Multiconference on*, 2010, pp. 461-467.
[18] M. Bijankhan, J. Sheykhzadegan, M. Bahrani, and M. Ghayoomi, "Lessons from building a Persian written corpus: Peykare," *Language resources and evaluation*, vol. 45, pp. 143-164, 2011.
[19] S. Isapour, M. Homayounpour, and M. Bijabkhan, "The Prediction of Ezafe Construction in Persian by Using Probabilistic Context Free grammar," in *In Proceedings of 13th Annual Conference of Computer Society of Iran*, Kish Island, 2008.
[20] A. Koochari, B. Qasemzade, M. Kasaeiyan, and M. Namnabat, "Ezafe Prediction in Phrases of Farsi Using CART," in *Proceedings of the I International Conference on Multidisciplinary Information Sciences and Technologies*, 2006, pp. 329-332.
[21] R. Dehkharghani and M. Shamsfard, "Mapping Persian Words to WordNet Synsets," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 1, 2009.
[22] M. Bijankhan, "The role of the corpus in writing a grammar: An introduction to a software," *Iranian Journal of Linguistics*, vol. 19, 2004.
[23] V. Samiian, "The Ezafe construction: some implications for the theory of X-bar syntax," *Persian Studies in North America*, pp. 17-41, 1994.
[24] G. D. Forney Jr, "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, pp. 268-278, 1973.
[25] S. Noferesti and M. Rajaei, "A Hybrid Algorithm Based on Ant Colony System and Learning Automata for Solving Steiner Tree Problem," *International Journal of Applied Mathematics and Statistics*, vol. 22, pp. 79-88, 2011.
[26] K. A. De Jong and W. M. Spears, "An analysis of the interacting roles of population size and crossover in genetic algorithms," in *Parallel problem solving from nature*, ed: Springer, 1991, pp. 38-47.
[27] J. Carletta, "Assessing agreement on classification tasks: the kappa statistic," *Computational linguistics*, vol. 22, pp. 249-254, 1996.

Samira Noferesti is a Ph.D student at Shahid Beheshti University. Her interests include artificial intelligence, natural language processing and opinion mining.

Dr. Mehrnoush Shamsfard has received her BS and MS both on computer software engineering from Sharif University of Technology, Tehran, Iran. She received her PhD in Computer Engineering- Artificial Intelligence from AmirKabir University of Technology in 2003. Dr. Shamsfard has been an assistant professor at Shahid Beheshti University from 2004. She is the head of NLP research Laboratory of Electrical and Computer Engineering faculty. Her main fields of interest are natural language processing, ontology engineering, text mining and semantic web.

Usability Evaluation Methods for Special Interest Internet Information Services

Eva-Maria Schön[1], Jens Hellmers[2], Jörg Thomaschewski[3]

[1]*SEVEN PRINCIPLES, Germany*
[2]*University of Bremen, Germany*
[3]*University of Applied Sciences Emden/Leer, Germany*

***Abstract* — The internet provides a wide range of scientific information for different areas of research, used by the related scientific communities. Often the design or architecture of these web pages does not correspond to the mental model of their users. As a result the wanted information is difficult to find. Methods established by Usability Engineering and User Experience can help to increase the appeal of scientific internet information services by analyzing the users' requirements. This paper describes a procedure to analyze and optimize scientific internet information services that can be accomplished with relatively low effort. It consists of a combination of methods that already have been successfully applied to practice: Personas, usability inspections, Online Questionnaire, Kano model and Web Analytics.**

***Keywords* — Scientific Portal, Usability, User Experience, Questionnaire, UEQ, Kano model**

I. Introduction

NOWADAYS, people expect scientific content to be provided by research institutions online via the internet. Therefore, scientific internet portals for a wide range of research areas were developed and can be easily accessed.

For an institution that provides such a service it is essential that the offered information is useful for the user as it can help to increase the provider's reputation. This includes not only the information itself, but also the way the information is presented to the user. A successful use of an internet information service can be monitored e.g. by web traffic analysis: an increasing number of users, a high number of recurring users or direct accesses to the web pages implicate a satisfying information service.

On the other hand, usually the bulk scientific information within the internet is provided by research groups at universities or institutes. Very often the operation of such a service is not part of the main work and is carried out additionally to the daily research routine. Therefore, it is in the interest of the providers to run a successful internet information service with as low as possible requirements for time and effort.

In this paper we would like to outline a procedure for analyzing and optimizing scientific internet information services using common methods of *Usability Engineering* and *User Experience*.

Usability Engineering within the scope of the World Wide Web traditionally is used for internet services applying to a broad variety of users, like news (paper) portals or shop systems. Corresponding methods are given by Scholtz [1] and Hornbæk [2]. Such broadly ranged internet presences usually are accessed by a large number of users which are already connected content wise to the service. Therefore, standard procedures like questionnaires [3] or A/B-Tests [4] can be applied for achieving fast and valid results.

For business software it is different. Here, usually so-called business users who frequently use such systems are at hand for evaluation. As a result, usability methods based on direct user participation like interviews as well as usability tests can be used.

In contrast to the above outlined internet presences and business software, a scientific information service is accessed by a small spread target audience. In this paper, we would like to introduce the term 'compact target audience' to describe this group of users. It is characterized by:

- Limitation: the number of (potential) users is small,
- Internationality: users access from all over the world,
- Homogeneity: the users share a comparable (academic) background,
- Focus: the information users look for is very specific.

This 'compact target audience' brings some limitation to the reasonably applicable methods of Usability Engineering and User Experience. Mainly, due to the relatively low number of (potential) users a reliable statistical analysis is difficult to achieve.

II. Proposed Process Model

Because of the limitations by a 'compact target audience' we would like to propose a combination of methods especially for scientific internet information services. Despite being a mixture of several different procedures the overall amount of work is still relatively low as the collected data is not that extensive and therefore easy to handle.

DOI: 10.9781/ijimai.2014.263

We recommend the following methods: Personas, usability inspections, User Experience Questionnaire, product usage related questions, product related questions, Kano Model and Web Analytics. The correlations of the particular methods are outlined in Fig. 1.

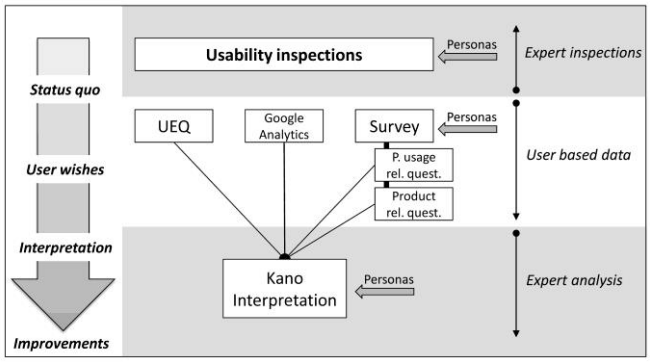


Fig. 1 Overview of the process model

A corresponding prototypical approach has been undertaken for the internet information portal regarding electromagnetic and light scattering *ScattPort* [5]. This procedure considers an already operational internet service which is currently accessed by the corresponding scientific community. That is why access data exists that can be used for the analysis.
For the conception of a new scientific service we would like to refer to Beyer et al. [6], Winter et al. [7] and the process of *Human Centered Design* described in DIN EN ISO 9241-210 [8] for more detailed information.

III. Evaluation of Special Interest Internet Information Services

A. Personas

To ensure the acceptance of a scientific information service it is essential to know the potential users. Knowledge of their motivation and expectations helps to constructively adapt the web pages in regard to information architecture, interaction design and content editing. As in this case there is a 'compact target audience' a convenient approach would be the use of Personas.

The method Personas originates from the User Centered Design [9]. A Persona concretely describes the profile of a potential user. This profile consists of demographic data like name, age and origin. Also a photograph of the user is presented. Information on the user's motivation to access an offer and usage patterns are part of the Persona, too. This includes short notes about personal quirks and characteristics. The data is given in short list-form.

Fig. 2 shows an example for the general structure of a Persona. A specific example for a Persona can be found e.g. at Winter et al. [7].

A Persona does not describe a specific, existing person. Instead it represents a whole, concrete group that is part of the target audience. Therefore, several Personas are created to cover the whole range of the users the target audience consists of. A big advantage of this method is that it can help to prevent providers and developers of a product – in this case the scientific information service – from projecting their own assumptions onto the target audience [10]. Instead an empathic understanding for the real users and their motivation is established.

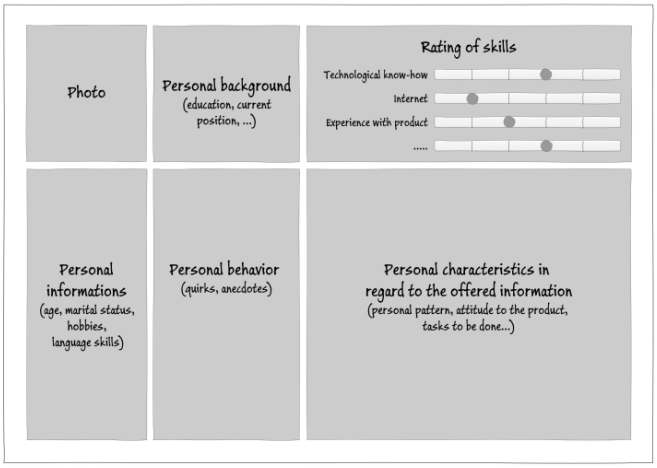


Fig. 2. General structure of a Persona

There are several ways to develop Personas. One way is to collect data on the target audience by interviews or surveys. This data is then abstracted into Personas [9]. Another way is to develop so-called *Ad-Hoc Personas* [11]. These are based on observations and experiences of (potential) users by the providers and developers themselves. It has been mentioned that this joint development of Personas within a project team increases the willingness to use Personas on a permanent base [12]. Such Ad-Hoc Personas were e.g. developed in the frame of a project to investigate the scientific internet information service regarding the topic of electromagnetic and light scattering *ScattPort* [13]. Here, we would like to exemplarily list the six different types of users who were identified and for which Personas were created:

- The established professor with many years of experience,
- The dynamic young professor who just started a career,
- The PostDoc scientist,
- The Ph.D. student,
- The established scientist working outside an university or institute,
- The engineer working in the industry.

Such an approach can be easily transferred to any kind of scientific information service by identifying the relevant user groups and developing corresponding Personas, altogether covering the target audience.

When the target audience is known, the specified context of use should be analyzed [14]. This comprises the used environment and the equipment. For example, a scientific information service could especially provide data for the access during conferences (environment) via mobile devices (equipment).

B. Usability inspections

Usability inspections (see Fig 1.) should be used to analyze the web content to check the information architecture and to identify usability problems. The corresponding methods *Cognitive Walkthrough* and *Heuristic Evaluation* are described by Scholtz [1]. In the frame of the Cognitive Walkthrough an usability expert defines user-orientated intentions based on Personas. The resulting scenarios are then simulated using the actual user interface. This process is reviewed by an expert who especially examines whether a user would be able to execute the outlined operation and if this action would lead to the correct result or not. On the other hand for the Heuristic Evaluation an interface is reviewed using defined and approved rules for Usability [15], [16], [17]. Again, Personas are applied for this process.

C. User Experience Questionnaire UEQ

The User Experience Questionnaire (UEQ) [3] is an established questionnaire for the quantitative evaluation of the User Experience [8]. The main idea of the questionnaire is to collect the immediate and spontaneous response by a user regarding a product (in this case the product is an internet service on the whole, which includes structure and content).

The best way for this is to survey users directly and personally, e.g. during a conference. In cases when this is not possible the UEQ can be set up online, e.g. by prominently placing a corresponding link on the starting page of the information service.

The advantage of the UEQ in comparison to other surveys (e.g. IsoMetrics [18], SUMI [19]) is that usability aspects (efficiency and effectiveness) are expanded by *Hedonic Quality* [20] (attractiveness, stimulation, novelty). For this reason additionally to Usability, User Experience can also be evaluated.

A different survey for the evaluation of User Experience was developed by Hassenzahl [21].

The UEQ covers six dimensions: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. It is based on the semantic differential of 26 bipolar pairs of adjectives (Fig. 3).

The items are scaled from -3 to +3. Thus, -3 represents the most negative answer, 0 a neutral answer, and +3 the most positive answer.

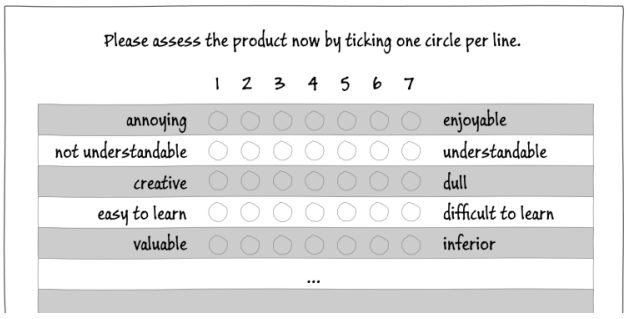


Fig. 3. Example of the UEQ

Note: these are the values actually used for the numerical analysis of the impressions given by the users. The survey presented to the users itself (on paper or online) shows a slightly different scale. Here, the values range from 0 to 7. This is done to avoid a subliminal influence on the users during the questioning caused by negative numbers (see Fig.3).

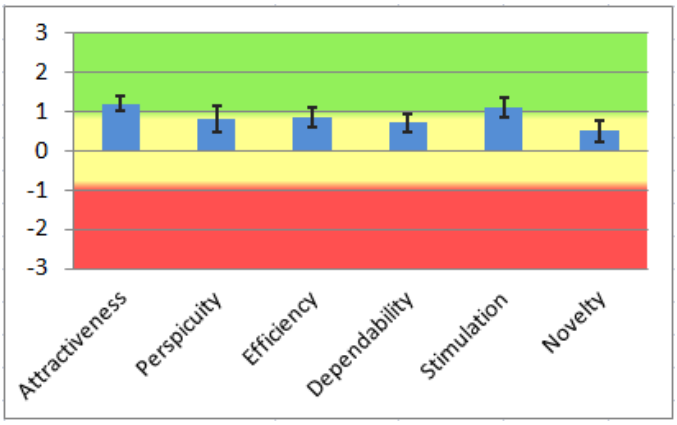


Fig. 4. Example of an UEQ result.

For the analysis the following aspects should be considered. Scale values above +1 indicate a positive impression of the users concerning this scale, values below -1 a negative impression. Due to well-known answer effects, like the avoidance of extremes, observed scales means are in general in the range of -2 to +2. More extreme values are rarely observed, so a value near +2 represents a very positive near optimal impression by the participants. Fig. 4 shows an example for an overall result including error bars. The graphic is automatically generated by the data analysis sheet (Excel) that can be downloaded together with the questionnaire.

Rauschenberger et al. [22] give specific recommendations for the practical work with the UEQ and its interpretation. More information and the UEQ itself are available online [23].

D. Product usage related questions

The UEQ presents pre-defined questions and answers that can be universally applied to a wide range of products and is an easy and fast method to analyze User Experience. But as a result of its universality it cannot cover usage and specific product attributes. Therefore, a survey (see Fig. 1) should contain corresponding fields allowing users to describe their own usage of the product. It is important to keep the number of such fields low as too many fields will have a negative influence on the users' motivation, which could lead in the end to non-usable data. In this sense the length of a questionnaire plays an important role as it has an impact on the response behavior [24]. The response rate to a short questionnaire is larger than the response to long questionnaires as they tend to lead to a higher drop-out rate [25]. To avoid common errors in the development of questionnaires appropriate guidelines should be used [26].

Exemplarily, we would like to suggest questions like the following:

- How often do you use <*the internet information service*>?
- On which devices do you use <*the internet information service*>?
- How high is your motivation to add content yourself?

This should be added by questions regarding demographic data like age, position and gender. The results of these questions can be used to verify or optimize the Personas.

E. Product related questions

So far a user cannot state comments on observed flaws or suggest possible improvements. Here, product related questions (see Fig. 1) can help to collect individual impressions to identify more user requirements. Exemplarily, we would like to suggest questions like the following:

- How do you like the general presentation of the internet service?
- How do you like the presentation of <*specific product feature*>?
- How would you rate <*a possible specific product feature for the future #1*>?
- How would you rate <*a possible specific product feature for the future #2* >?
- *etc.*

We suggest the usage of a Likert scale [27] followed by a text field for every question to get qualitative and quantitative data. Additionally one text field can be given to write down personal opinions on how to improve the information service.

F. Kano Model

The Kano model (see [28], [29], [30]) sorts requirements into different categories. Each category has specific influence on a user's satisfaction. This allows identifying the relevance of product attributes. This is outlined in Fig. 5.

Table 1 describes this scheme in more detail. There are three main categories: *basic requirement*, *performance requirement* and *enthusiasm requirement* [29]. The *basic requirement* is expected by a user. Without it the product misses its purpose. A user will not feel any satisfaction from it, but will be definitely dissatisfied if it is missing. Contrary to that, a *performance requirement* has a direct connection to the user's satisfaction. The *enthusiasm requirement* provides a sort of 'extra'. It will not be missed, but can help to improve the satisfaction. Additionally to the main categories there are *indifferent factor*, *reverse factor*, and *questionable factor*.

For constructing a Kano questionnaire a pair of questions for each product feature is formulated: a functional question ("If the product contains the feature, how do you feel?") and a corresponding dysfunctional question ("If the product doesn't contain the feature, how do you feel?"). Because of these combinations of seemingly similar questions the questionnaire might get exhausting or even boring for the user [26]. Producing an answer for these kind of questions is no longer a motivating challenge, leading to the effect that the attention of the user is not focused on the questions. This is especially disadvantageous as the target audience is already small in numbers. As a result the usable data set could be decreased significantly.

TABLE I
KANO MODEL – PROPERTIES AND IMPACT (SEE ALSO [5], [29]).

Category	Impact	Description
basic requirement	Must-be	user expects it, will not be satisfied if fulfilled, but dissatisfied if not fulfilled
performance requirement	One-dimensional	the better it is fulfilled the more satisfied a user will be
enthusiasm requirement	Attractive	user does not expect it, will not be dissatisfied if not fulfilled but satisfied if fulfilled
indifferent factor	Indifferent	user has no use for it, is neutral
reverse factor	Reverse	user expects the opposite
questionable factor	Questionable	user misunderstood question or the question was phrased incorrectly

Because of this, in our proposed procedure the Kano model is not applied 'step-by-step' to its full extent, instead its basic-core with categories as outlined in Table 1 is used to interpret the answers (context see Fig. 1).

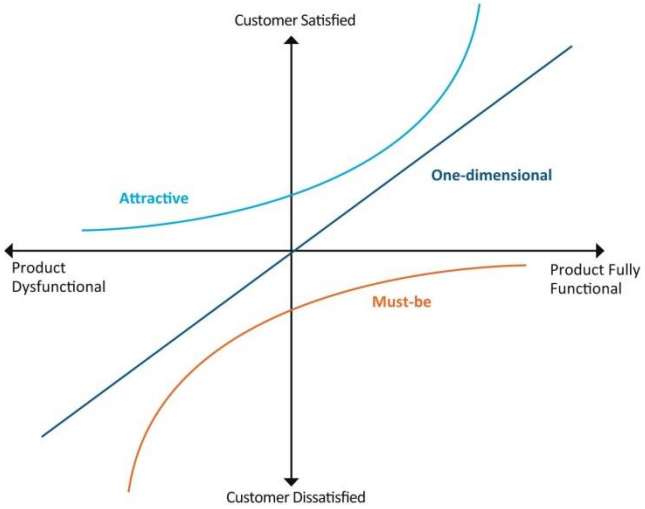


Fig. 5. Kano's model of customer satisfaction (see [30])

By this, a questionnaire can be kept compact. Nevertheless, it enables not only to analyze whether a user requirement is fulfilled or not, but also how satisfied a user is. To give an example how to use the Kano categories for the interpretation of the product related questions we like to refer to the article about the evaluation of the *ScattPort* internet information portal [5].

To address the problem of possible self referential assumptions Personas were applied to the interpretation, helping to focus on the users' mental models and needs. This procedure proved itself very helpful for the evaluation of the statements given by the users.

For example for the question "How do you like the presentation of the program list" 24 Likert scale answers and

12 usable text field comments were given. The vast majority expressed satisfaction with the current presentation [5].

The users' comments stressed the program list as the central feature of the information service *ScattPort* [31]. This was interpreted as a *Must-be* feature and thereby a *basic requirement* (see Table 1).

The other product related questions were evaluated accordingly. This procedure shows how the users rank the current implementation, general availability or lack of certain product features.

G. Web Analytics

For already existing special interest information services the analysis of the access data can provide useful information about its perception by the users [32]. Web Analytics tools help to collect and interpret data about user behaviour, which then can also be used to verify conclusions gathered by the methods described above. Suitable tools for Web Analytics are *Google Analytics*, *Piwik* or *Open Web Analytics (OWA)*.

It is possible to compare statements (from the UEQ or Kano analysis) regarding the appeal of certain sections with the actual number of accesses to those parts of the service. For example, popular and highly accessed webpages can be considered as a central part of the offered information and therefore indicate a *Must-be* feature (compare Table 1). Such data can be cross-checked with statements collected during a Kano-based evaluation respectively interpretation. Also the kind of access contains information. A direct request for a page (e.g. by a bookmark) indicates that this page fulfils a user's needs. The next step then should be to analyze the rate of recurring accesses and users. Returning users indicate the successful fulfilment of their needs. Additionally the usability of the information service can be estimated: a poorly designed information structure leads to users' frustration and would prevent them from returning. Redirections from result pages of a search engine on the other hand can indicate first time visits. The corresponding rate in combination with the average time a user stays on the page can also be used to determine the appeal of areas of the internet service: immediate leaving of a page accessed via web-search indicates that a user was expecting a different kind of information. Subsequent accesses of other pages (starting from the firstly viewed page after a web-search) are more difficult to interpret: it can either mean that a user got curious and checks out the other pages and their information, or the user simply got lost. Here the time a user stays on subsequent pages can help with the interpretation as interested users should stay longer on these pages, reading them thoroughly, than frustrated ones. Now, indications for satisfaction or frustration can also be derived for example from the UEQ and thereby compared to the web analytic statistics.

While the described methods for usability inspections help to expose flaws in the presentation of the information service, web analytics can help to set priorities for their adjustment. Highly accessed areas should be fixed first.

An In-Page-Analysis shows how users interact with a webpage, e.g. which links they follow and on which areas on the screen they focus. This allows determining whether the content is presented in a way users prefer or not. As a result the landing page of a service can be adjusted and improved by rearranging the content. Placing interesting – means: popular – topics in front of the user the first time he or she accesses the service will lead to a higher satisfaction and thereby to a better chance for recurring visits.

In general, Unique Visitors, Visits/Sessions and Page Views are considered the most important metrics for Web Analytics [33]. Additionally Entry Page, Landing Page, Visit Duration, Click-through und Page Views per Visit can be useful [34].

A more detailed example for the interpretation of data provided by Web Analytics in relation to statements collected by UEQ and Kano analysis can be found at Hellmers et al. [5].

IV. SUMMARY

This article describes a procedure for the evaluation of scientific special interest internet services using methods of Usability Engineering and User Experience.

It especially takes into account that the user base is a 'compact target audience': there is a limited number of (potential) users who are looking for very specific information and who share specific user requirements. This is a disadvantage for the conventional evaluation methods used nowadays as they rely on a sort of 'critical data mass' for a successful statistical analysis.

Now, the importance of small-scale internet information services for compact audiences like a specific scientific community should not be underestimated. By increasing the awareness-level they can help to boost the scientific progress, to improve the reputation of the providing institution, and to gain funding for scientific projects. Therefore, a successfully designed special interest internet service is as advantageous for the provider as it is for its users.

The approach proposed here consists of a combination of methods that already has been successfully applied in practice: Personas, usability inspections, User Experience Questionnaire (UEQ), product usage related questions, product related questions, Kano-based interpretation and Web Analytics.

The idea is to check whether the corresponding statements and results collected by each individual method support each other or not. Matching data can be considered reliable and therefore used to analyze the users' needs and their impression of the usability of an already existing service. This helps to improve the information service further. Inconsistent or even contradictory data hint basic flaws of a service; the corresponding concept should be checked.

By this it is possible to gather information about the users' perception of an internet service in a relative easy way and with low effort, even in cases where the dataset is low.

Overall, it helps to develop and improve a scientific internet information service in regard to Usability and User Experience.

- Personas help understanding the target audience: its composition and also the resulting user requirements,

- Usability inspections help identifying usability problems,
- The User Experience Questionnaire (UEQ) gives information on the users perception of an existing internet service,
- Product usage related questions expand the UEQ to gather information about aspects of the usage and specific attributes of the internet service,
- Product related questions allow users to describe their impressions of an existing service, to comment specific requirements, and to suggest ways how to improve the offer,
- A Kano analysis indicates the functional satisfaction by reviewing certain attributes and sorting them into specific categories,
- Web Analytics can be used to analyze the users' behavior and to check data collected by the other methods for plausibility.

The approach outlined not only suits scientific internet information services. It also should be applicable in other cases with a 'compact target audience'.

REFERENCES

[1] J. Scholtz. (2004). Usability Evaluation [Online]. Available: www.itl.nist.gov/iad/IADpapers/2004/Usability%20Evaluation_rev1.pdf

[2] K. Hornbæk, "Current practice in measuring usability: Challenges to usability studies and research" in International journal of human-computer studies, 64 (2), 2006, pp. 79–102.

[3] B. Laugwitz, T. Held, and M. Schrepp, "Construction and Evaluation of a User Experience Questionnaire" in Lecture Notes in Computer Science (5298), 2008, pp. 63–76.

[4] J. Nielsen. (2005). Putting A/B Testing in Its Place [Online]. Available: www.nngroup.com/articles/putting-ab-testing-in-its-place/

[5] J. Hellmers, J. Thomaschewski, E.-M. Holt, and T. Wriedt, "Usability evaluation methods for a scientific internet information portal" in Journal of Universal Computer Science (JUCS) 18 (10), 2012, pp.1308–1322.

[6] H. Beyer, *User-centered agile methods.* San Francisco, Calif.: Morgan & Claypool, 2010

[7] D. Winter, E.-M. Holt, and J. Thomaschewski, "Persona driven agile development. Build up a vision with personas, sketches and persona driven user stories", in *2012 Proc. Information Systems and Technologies (CISTI)*. Vol. 2, pp. 307-310.

[8] DIN EN ISO 9241-210:2011-01, Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems, Beuth, Berlin, 2011

[9] A. Cooper, *The inmates are running the asylum*. Indianapolis, Ind.: Sams, 1999

[10] A. Cooper, R. Reimann, and D. Cronin, *About face 3: The Essentials of Interaction Design.* Indianapolis, Ind.: Wiley, 2007

[11] D. Norman. (2004). Ad-Hoc Personas & Empathetic Focus [Online]. Available: jnd.org/dn.mss/personas_empath.html

[12] Å. Blomquist and M. Arvola, "Personas in action: ethnography in an interaction design team" in *Proc. of the second Nordic conference on Human-computer interaction*, 2002, pp. 197–200.

[13] J. Hellmers, E.-M. Holt, and J. Thomaschewski, "Personas für das wissenschaftliche Internetportal ScattPort zum Thema Lichtstreuung", in *Proc. 2. Symposium Wissenschaftliche Portale und Informations-netzwerke*, Institut für Werkstofftechnik, Bremen, 2011

[14] DIN EN ISO 9241-11:1998, Ergonomic requirements for office work with visual display terminals (VDTs) – Part 11: Guidance on usability, Beuth, Berlin, 1999

[15] R. L. Mack and J. Nielsen, *Usability inspection methods*. Wiley & Sons, 1994

[16] B. Shneiderman and C. Plaisant, *Designing the user interface. Strategies for effective human-computer interaction*. 5. ed., internat. ed. Upper Saddle River, NJ: Addison-Wesley/Pearson, 2010

[17] DIN EN ISO 9241-110: 2006, Ergonomics of human-system interaction - Part 110: Dialogue principles, Beuth, Berlin, 2008

[18] G. Gediga, K. Hamborg, and I. Düntsch, "The IsoMetrics usability inventory: an operationalization of ISO 9241-10 supporting summative and formative evaluation of software systems", Behaviour & Information Technology, 18 (3), 1999, pp. 151–164.

[19] J. Kirakowski and M. Corbett, "SUMI: The software usability measurement inventory" in British journal of educational technology, 24 (3), 1993, pp. 210–212.

[20] M. Hassenzahl, "The Effect of Perceived Hedonic Quality on Product Appealingness" in International Journal of Human-Computer Interaction 13 (4), 2001, pp. 481–499.

[21] M. Hassenzahl, M. Burmester, and F. Koller, "AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität" in *Proc. Mensch & Computer 2003*, pp. 187–196.

[22] M. Rauschenberger, M. Schrepp, M. Perez-Cota, S. Olschner, and J. Thomaschewski, "Efficient Measurement of the User Experience of Interactive Products. How to use the User Experience Questionnaire (UEQ). Example: Spanish Language Version" in IJIMAI: International Journal of Interactive Multimedia and Artificial Intelligence, Bd. 2 (Special issue "Artificial Intelligence and Social Application", Vol 2. Nr. 1), 2013, pp. 39-45.

[23] User Experience Questionnaire [Online]. Available: www.ueq-online.org

[24] M. Bosnjak, B. Batinic, "Understanding the Willingness to Participate in Online-Surveys – The Case of e-mail Questionnaires" in Online social sciences, 2002, pp. 111–116.

[25] T. Tuten, D. Urban, M. Bosnjak, "Internet Surveys and Data Quality: A Review" in Online social sciences, 2002, 7–27.

[26] L. Gräf, "Assessing Internet Questionnaires: The Online Pretest Lab" in Online social sciences, 2002, 73–93.

[27] R. Likert, "A Technique for the Measurement of Attitudes" in Archives of Psychology No. 140, 1932

[28] E. Sauerwein, F. Bailom, K. Matzler, and H.H. Hinterhuber, "The Kano Model: How to delight your customers" in Preprints Volume I of the IX. International Working Seminar on Production Economics, Innsbruck/Igls/Austria, February 19-23 1996, 1996, pp. 313–327.

[29] E. Sauerwein, "Experiences with the reliability and Validity of the Kano-Method: Comparison to Alternate Forms of Classification of Product Requirements" in Transactions of the 11th Symposium on QFD, 1999, pp. 416–429.

[30] C. Berger, R. Blauth, D. Boger, C. Bolster, G. Burchill, W. DuMouchel, F. Pouliot, R. Richter, A. Rubinoff, D. Shen, M. Timko and D. Walden: "Kano's Methods for Understanding Customer-defined Quality"; David Walden (Ed): Center for Quality Management Journal, Vol. 2, Nr. 4 (Fall 1993) (Vol. 2, Nr. 4 (Fall 1993)), pp. 3–36.

[31] T. Wriedt and J. Hellmers: "New Scattering Information Portal for the light-scattering community"; Journal of Quantitative Spectroscopy & Radiative Transfer 109, 2008, pp. 1536–1542

[32] L. Hasan, A. Morris, and S. Probets, "Using Google Analytics to Evaluate the Usability of E-Commerce Sites", Lecture Notes in Computer Science, 2009, pp. 697–706.

[33] J. Burby, A. Brown and WAA Standards Committee, "Web Analytics "Big Three" Definitions Version 1.0", 2006

[34] J. Burby, A. Brown and WAA Standards Committee, "Web Analytics Definitions Version 4.0", 2007

Eva-Maria Schön received her Bachelor of Engineering in Media technology in 2010. Now she is studying the Master of Science in Computer Science and Media Applications at the University of Applied Science Emden/Leer. She works as a Consultant at SEVEN PRINCIPLES in Germany. She focuses on agile product management for mobile and web-based internet portals. As a product owner Eva has great passion for outstanding User Experience. Since 2004 Eva is a specialist in target-group-oriented design. Her research interests are: interaction design, agile software development, Usability and User Experience.

Dr. Jens Hellmers received a diploma in Physics from the University of Oldenburg, Germany in 1998 and a Doctor Degree in Engineering from the University of Bremen, Germany in 2008. Currently, he is Researcher at the Process Engineering department of the University of Bremen. In several projects he investigated the potentials and feasibility of internet information services for a scientific community from technical points of view, but also in regard of Usability and User Experience. Dr. Hellmers is co-creator of the ScattPort internet information portal on the topic of light scattering.

Prof. Dr. Jörg Thomaschewski. He became full professor at the University of Applied Sciences Emden/Leer, Germany from September 2000. He received the Doctor degree in Physics from the University of Oldenburg, Germany, in 1996. His research interests are in the fields "Internet Applications" focusing on human computer interaction, e-learning and software engineering. He is author of various online modules, e.g. "Human Computer Communication", which is used in the Virtual University (Online) at six university sites. He has wide experience in usability training, analysis and consulting.

A REST Service for Triangulation of Point Sets Using Oriented Matroids

José Antonio Valero Medina and Ivan Lizarazo Salcedo

Francisco José de Caldas District University, Colombia

***Abstract* — This paper describes the implementation of a prototype REST service for triangulation of point sets collected by mobile GPS receivers. The first objective of this paper is to test functionalities of an application, which exploits mobile devices' capabilities to get data associated with their spatial location. A triangulation of a set of points provides a mechanism through which it is possible to produce an accurate representation of spatial data. Such triangulation may be used for representing surfaces by Triangulated Irregular Networks (TINs), and for decomposing complex two-dimensional spatial objects into simpler geometries. The second objective of this paper is to promote the use of oriented matroids for finding alternative solutions to spatial data processing and analysis tasks. This study focused on the particular case of the calculation of triangulations based on oriented matroids. The prototype described in this paper used a wrapper to integrate and expose several tools previously implemented in C++.**

***Keywords* — Mobile devices, oriented matroids, spatial data representation, services, triangulation.**

I. Introduction

IN the context of spatial data representation, triangulation of a set of points is essential for many applications. A typical example is the generation of triangulated irregular networks (TINs) to represent ground topography. Another example is the representation of spatial objects using *simplicial complexes* which allows breaking complex objects into simpler underlying geometries. In the case of complex objects in the two-dimensional Euclidean space, simple geometries such as triangles provide a close approximation of spatial objects.

Whatever be the purpose of triangulation, a computational solution is necessary to perform an optimal calculation. Traditional algorithms are based on strategies that rely on geometric representation both of the point set and the Euclidean space embedding it. That is the case of greedy algorithms ([1]) which build triangles using an iterative process based on three criteria: (i) selecting the minimum distance between pairs of points, (ii) not revisiting diagonals that have already been visited, and (iii) testing intersection of new diagonals with those previously obtained. This algorithm has $O(n^3)$ time complexity.

A proposal for simple polygons [2] breaks down the input polygon into monotonous chains to subsequently perform triangulation of each sub-polygon, thus obtaining $O(n)$ time complexity. Another proposal initially finds the Voronoi diagram [3] associated with the set of points, and then calculates the Delaunay triangulation. All of the above cited proposals use traditional geometric representations leading to algorithms suffering from a significant computational complexity. However, in recent decades, there has been an interest for using alternative mechanisms to improve computational efficiency for spatial data processing. [4] has studied the particular case of triangulation of a set of points based on oriented matroids, who implemented several tools using the C++ language.

This article describes the implementation of a prototype REST service for triangulation of a set of points obtained by global positioning system (GPS) receivers. This application targets mobile devices such as tablets and cell phones, using Android OS. The prototype wraps components generated by [4] with the purpose of testing oriented matroids as a means to solve problems in which spatial location matters.

The next section of this paper introduces basic concepts on matroids and oriented matroids as well as the proposal of [4] for triangulation of a point set using only purely combinatorial oriented matroids. The third section describes REST services based on entities that are stored in a persistent repository. This section also makes a description of the application development architecture for Android environment, based on the Eclipse IDE.

The fourth section discusses the architecture of the system and its scheme of access and operation. The fifth section makes a presentation of an overview of the used tools. The sixth section presents a prototype test and the last section formulates conclusions and final remarks.

II. From the Geometric to the Combinatorial

This section introduces basic concepts of oriented matroids. For a more comprehensive treatment, interested readers are referred to [5].

A. Matroid

The concept of matroid was introduced by [6] in the article

DOI: 10.9781/ijimai.2014.264

entitled "On the abstract properties of linear dependence". Whitney describes the approach with respect to a given matrix and its columns, in such a way that any subset of these columns forms a matrix with a particular range. Considering the columns as abstract elements, a matroid with range given by the number of linearly independent columns is formed.

According to [**¡Error! No se encuentra el origen de la referencia.**] from a matrix over any field, it is possible to define a matroid. In particular, the interest from a computational point of view is on finite fields, such as the Galois fields $GF(p^k)$, which give exactly p^k elements when p is a prime number. With $k = 1$ the $GF(p)$ field can be seen as the set $\{0,1,2,\ldots,p-1\}$ with operations of addition and multiplication module *p*. The obtained matroid is called a *vector matroid*.

The concept of graphic matroid is introduced in [7] based on the set of edges of a graph and the set of subsets of arcs of the same graph that does not contain all the arcs in any cycle, with the peculiarity that non isomorphic graphs may have isomorphic matroids.

The formulation of affine matroid [8] is based on the concept of affine independence against affine sets (in $\mathbb{R}^2$ they are the empty set, the points themselves, the straight lines and the plane itself); thus, given a set of points in the plane and a set of affine independent subsets of them, there is a matroid called *affine matroid*.

B. Oriented Matroids

According to [8], geometry of matroids is based on what affine or linear sets provide but it lacks a structure of convexity. There is neither notion of duality or intermediation between the points of a straight line, nor existence of hyperplanes separating the space into two half-spaces. According to [8], vector spaces do not have enough structure to support a theory of convexity; therefore, it is not sound to expect matroids to do it. Authors state the need of providing an additional framing (i.e. an orientation) to the matroids from which the convexity may arise. In a nutshell, that is the theory of oriented matroids.

In [9], it is contended that the oriented matroid of a finite set of points draws information from the relative position and orientation from the configuration, which can be provided by a list of signs that encode the orientation of all its bases. Although in the passage from a specific point's configuration to its oriented matroid, metric information is lost, many of the structural properties have their counterpart in the combinatorial level of oriented matroid. That is to say that oriented matroids describe the structure of incidence between the elements of the matroid and their respective hyperplanes, as well as they encode the position of the elements relative to the hyperplanes; i.e. which items fall on the positive side, which ones on the negative side and which ones inside the hyperplane [10].

C. Computational calculation of triangulations

In [4], it is introduced a procedure to perform triangulation from a set of points based only on the combinatorial structures of the associated oriented matroid. Authors state that triangulations are a basic means for decomposing complex objects into simpler ones. According to [4], for a configuration $\mathcal{A}$ with n $d-dimentional$ points, a T subset of subsets of $\mathcal{A}$, each one made of $d+1$ elements, is a triangulation if and only if

$$\bigcup_{\sigma \in T} conv\ \sigma = conv\ \mathcal{A} \quad (1)$$

and

$$conv\ \sigma \cap conv\ \sigma' = conv(\sigma \cap \sigma')\ \ \forall \sigma, \sigma' \in T \quad (2)$$

The contribution of [4] is that, instead of using expensive linear programming with exact arithmetic to ensure the two aforementioned conditions, authors use purely combinatorial checks based on the oriented matroid $\mathcal{A}$. To check whether the first condition is met the set of all circuits of $\mathcal{A}$ $is\ used$ and it is verified that for every pair of subsets of $\sigma, \sigma' \in T$ there exists a circuit $(Z^+, Z^-) \in \mathcal{A}, Z^+ \subseteq \sigma\ y\ Z^- \subseteq \sigma'$.

On the other hand, to check purely combinatorial fulfillment of the second condition, the set of all co-circuits of $\mathcal{A}$ is used. There is a co-circuit for each affine hyperplane $(d-1)-dimensional$ extending by subsets of $\mathcal{A}$, which includes all points of $\mathcal{A}$ that are on the positive side ($\mathcal{C}^+$) and all the points of $\mathcal{A}$ that are on the downside ($\mathcal{C}^-$) of the hyperplane. For final checking of the second condition, it is verified that, for each $\sigma \in T$, each cocircuit of $\mathcal{A}$ contained by itself, whose sets $\mathcal{C}^+$ and $\mathcal{C}^-$ are both not empty, must have a $\sigma' \in T$ that also contains itself.

In accordance with [4], as long as circuits of $\mathcal{A}$ determine in a purely combinatorial fashion their co-circuits and vice versa, and since the matroid of $\mathcal{A}$ is defined by anyone of such co-circuits, the number of possible triangulations of $\mathcal{A}$ set depends only on its oriented matroid. The passage from the geometry to the combinatorics associated with $\mathcal{A}$ is established through the concept of its chirotopo [11] [12].

$$\chi: \begin{cases} \binom{\mathcal{A}}{d+1} \to \{+,-,0\} \\ (i_1, i_2, \ldots, i_{d+1}) \mapsto sign\left(\det\left(a_{i_1}, a_{i_2}, \ldots, a_{i_{d+1}}\right)\right) \end{cases} \quad (3)$$

A chirotopo assigns its orientation to each ordered base of $\mathcal{A}$. The circuits of $\mathcal{A}$ can be calculated based on all subsets of $d+2$ elements of $\mathcal{A}$ and its cocircuits, using all its subsets of d elements and the calculation of the associated chirotopos. In [4] proposes several algorithmic solutions for triangulation, ranging from obtaining one of the several possible triangulations to obtaining all possible number of triangulations, by using only combinatorial structures associated with the oriented matroid of the points set. These algorithms were implemented in the package TOPCOM (Triangulations of Point Configurations and Oriented

Matroids) looking for a minimum time complexity and a maximum efficiency [4].

III. TECHNOLOGICAL DEVELOPMENT ENVIRONMENT

This section discusses web development concepts useful for using oriented matroids as alternative means for representation and implementation of computational solutions involving spatial location of data.

The main reason for adopting a Web implementation was to provide high availability for the triangulation functionalities proposed by [4]. In a similar way, the choice of the mobile devices environment with Android support leaned on the idea of having a source of sets of points that would be highly available. As it is well known, using a mobile device such as a GPS enabled tablet or a cell phone is something quite common these days.

A. *REST webservices*

Possibilities for providing online Web capabilities are very different, including solutions based on APIs such as Sockets, XML-RPC or RMI-based components. This article used Web services REST - REpresentational State Transfer [13], as a platform for implementation that provides transparent access to resources possibly including persistent repository storage.

REST was originally introduced as an architectural style for building systems hypermedia distributed on a large scale. The architectural style REST rests on four principles:

--Identification of resources via URIs which outline a set of resources identifying which elements customers interact with.

--Uniform interface which allows user to manipulate the resources using a predefined set of operations: *create* (PUT), *read* (GET), *update* (POST) and *delete* (DELETE).

--Self-descriptive messages, through which resources are uncoupled from their representation to allow content's access in a variety of formats (XML, JSON, etc.).

--Interactions with state through hyperlinks are enabled causing that resource's interactions be stateless, i.e. request messages are self-contained.

The REST services build on well-established W3C protocols (HTTP, XML, etc); thus, their creation should be fairly simple. Effort to building REST services clients is low since testing can be done using any Web browser.

On the other hand, since messages are contained in the URI, a major constraint arises when the associated data set is large enough to exceed the maximum size to be considered a well formed URI.

REST constrains the interface of a resource to its generic uniform interface with predefined operations, and there is very little to choose in terms of available operations. Therefore, a lot of effort should be put on defining which resources need to be exposed. Furthermore, it is necessary to assess whether the four operations are applicable to each resource exposed. In addition, it is also necessary to establish what is the application's semantics associated to each individual resource.

Since REST services are built directly over the HTTP transport protocol, nothing has to be decided about the communication protocol to use. On the other hand, REST has no preset format to adjust the data with. As this issue can be negotiated, it is possible to use different formats such as XML, JSON, or even the SOAP itself.

B. *Development environments for Android*

Android is undoubtedly one of the most popular platforms for mobile devices in the world such that the manufacturers of this devices use it in a very high proportion. Android is an operating system based on Linux, an open, free and cross-platform operating system, so there are a very nice set of tools for application development. Android provides all the necessary interfaces for developing applications that require access to the basic functionality of mobile devices (i.e. network interface, GPS, etc.) in the Java programming language.

All of the functionalities required in the development of an application for Android are available in its Software Development Kit (SDK). As alternative it is possible to use libraries released by Google as part of its Google Play Services.

The primary IDE environment is the open source tool Eclipse, which also has a plugin, which provides an Android image, with which it is possible to perform application development for Android on any operating system using this emulator for testing [14].

Each Android application uses an own process identified by its ID, and it is the only process accessing user files. The devices have a unique focus, the main application, which is the application visible on the screen, but they can run several applications in the background, each one with its own stack of tasks. The pile of tasks is the sequence of execution of processes in Android. They consist of activities that are stacking as they are invoked, and can only be terminated when the tasks above are completed.

IV. WEB SERVICE FOR TRIANGULATION OF SETS OF POINTS

As mentioned above, the prototype service of point set triangulation was planned in order to be a highly available service, so it was arranged as a Web service REST with the access mode shown in **Fig. 1.**

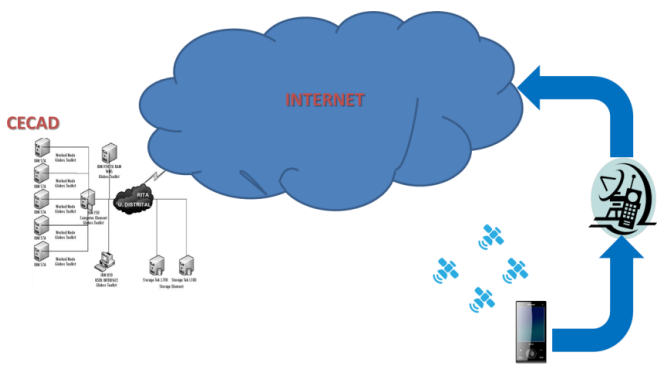


Fig. 1 Access mode of triangulation service based on oriented matroids.

The REST web service was deployed on GlassFish in the Center for High Performance Computing (CECAD) at *Francisco José de Caldas District University*. The service can be accessed from any mobile device connected to the internet via a provider of cellular mobile data, or discovered through a wireless network. The mobile device should have a good coverage of GPS satellites to guarantee a minimum level of accuracy and availability of geographical coordinates.

Fig. 2 shows the components architecture diagram of the system following [15]. For a set of points obtained from the mobile device, it was created a REST server with two REST services from Entity Classes based on the repository defined on the service provided by Java Persistence. *ConjuntodepuntosFacadeREST* is a REST service responsible for generating a group ID of points per session triggered by a connected mobile device with the *read* operation (GET) enabled. This operation returns a list of points formed by a single point with identifier and coordinates dummy but with the new group ID. The reading based on identifier operation (GET {id}) for this REST service was also adapted to allow any connected client to obtain the associated set of points saved to the repository within a given time interval.

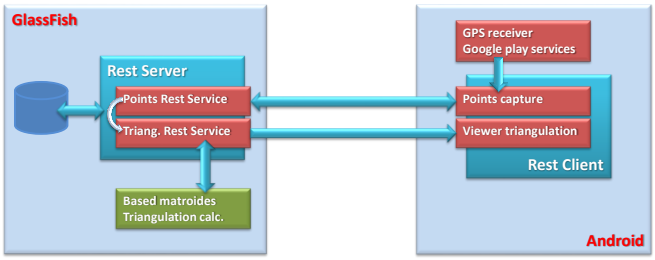


Fig. 2 Triangulation service based on oriented matroids architecture diagram.

TriangulacionesFacadeREST is a REST service created to get the possible triangulations from the current set of points. For this REST service, it was adapted the reading operation based on identifier (GET {id}). Thus, a connected client can request all possible triangulations at any time. For the calculation of the triangles, the REST service was supplemented with a thread responsible for carrying out the wrapper of the *points2triang* component of the software implemented by [4].

The legacy *points2triang* component receives the set of points by its standard input and delivers the calculated triangulations (if it is possible to find any) by its standard output. In order to enable the communication between this component and *TriangulacionesFacadeREST* service, the standard input and output of the component was respectively redirected from and to two auxiliary files. The former is created by the REST service with the coordinate transformation (from the geographic system coordinates to Transversal Mercator) of the points set that the mobile client provide. The latter is generated by the legacy component and used by feed backing the *TriangulacionesFacadeREST*.

Finally, *TriangulacionesFacadeREST* returns back to the mobile client a set of edges for the different possible triangulations associated with the set of points in a session with a client.

The client was implemented based on asynchronous tasks in order to reduce the complexity of the main thread and prevent application crashes.

As mentioned above, the client was developed for the Android system using the API provided by Google Play Services to get coordinates obtained from the receiving device's GPS.

V. TOOLS

This section describes tools used for implementation of the prototype.

A. Building of the web service

The Netbeans IDE was used for the construction of the web service (WS-Rest), following a six-step procedure:

--Creation of tables *ConjuntoDePuntos* and *Triangulaciones*.

--Creation of a new project and addition of connection drivers to the database in Derby.

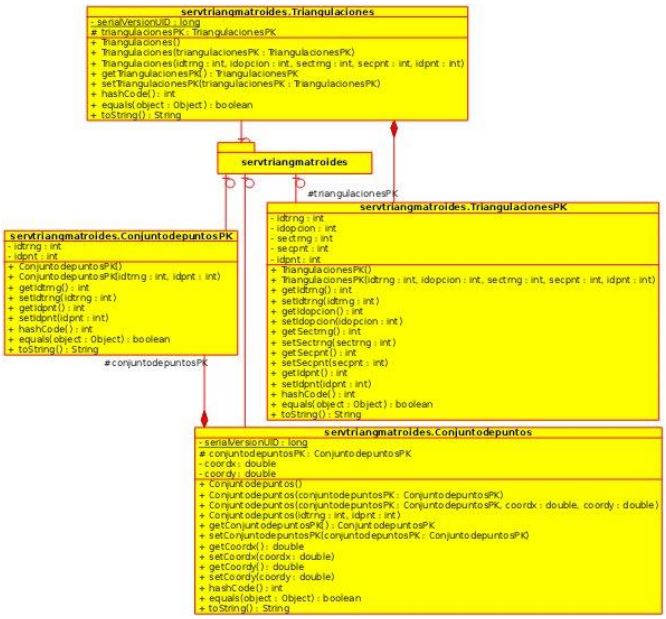


Fig. 3 Entity classes class diagram.

--Creation of a resource (persistent entity) for each table in the database (**Fig. 3**).

--Creation of *ConjuntodepuntosFacadeREST* and *TriangulacionesFacadeREST* REST Web services for each Entity Class created in the previous step (**Fig. 4**).

--Creation of methods that implement the REST operations that will be used in each REST service created. The *ConjuntodepuntosFacadeREST* class is adapted using the *find* (GET {id}) and *findAll* (GET) methods. The *TriangulacionesFacadeREST* class is adapted using the *find* (GET {id}) method.

--Creation of the *Trng* class for the Wrapper of the legacy component responsible for carrying out the triangulation based on the oriented matroid associated to the set of points.

The REST services were available at a public URL allocated by the CECAD of the District University.

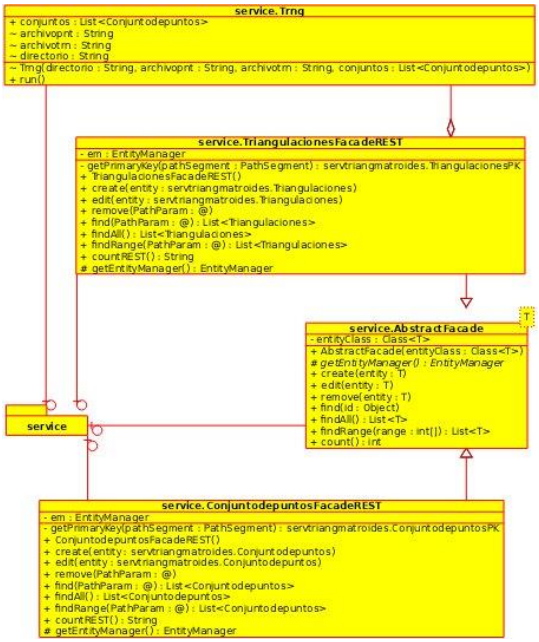


Fig. 4 Services class diagram.

B. Building of the Android (WS-Res) client

For the construction of the Android client providing access to the web service (WS-Rest), following tools were used: Eclipse IDE, the Android Development Tools (ADT) plugin for Eclipse and the Android SDK. Steps for the installation of the mentioned tools were as follows:

--Installing the Eclipse IDE, downloading it from the URL *http://www.eclipse.org/downloads/*. The installation is doing by simple unzip of the downloaded file and run the eclipse program from the eclipse folder.

--Installing the SDK Android available from the URL *http://developer.android.com/sdk/index.html*. After downloading the Android SDK, the installer was run, specifying a path to the JDK. Then, it was needed, through the SDK Manager, to install the "*Android SDK Platform-tools*", "*Android 4.3 (18 API)*" and "*Android 2.2 (8 API)*" platform, and extra "*Android Support Library*" package.

--It was also necessary to create an environment for Android emulation, using the AVD (*Android Virtual Device*) manager of the Android SDK. This emulated environment is useful for test development for Android without using an actual mobile device.

--To complete the tools installation it was necessary to install the *Plugin Android Development Tools* (ADT) for Eclipse using the option "*Install New Software of Eclipse*", providing the URL *https://dl-ssl.google.com/android/eclipse/*, and selecting the two *Developer Tools* and *NDK Plugins* packages.

--Once the software tools were installed, the Android application project in Eclipse and the layout of the GUI client were created.

To support all of the required functionality, it was necessary to make the main activity (*MainActivity*) to extend the *FragmentActivity* class and to implement *ConnectionCallbacks* and *OnConnectionFailedListener* interfaces of the *GooglePlayServicesClient* package, as well as the *LocationListener* interface of the package *com.google.android.gms.location* (**Fig. 5**).

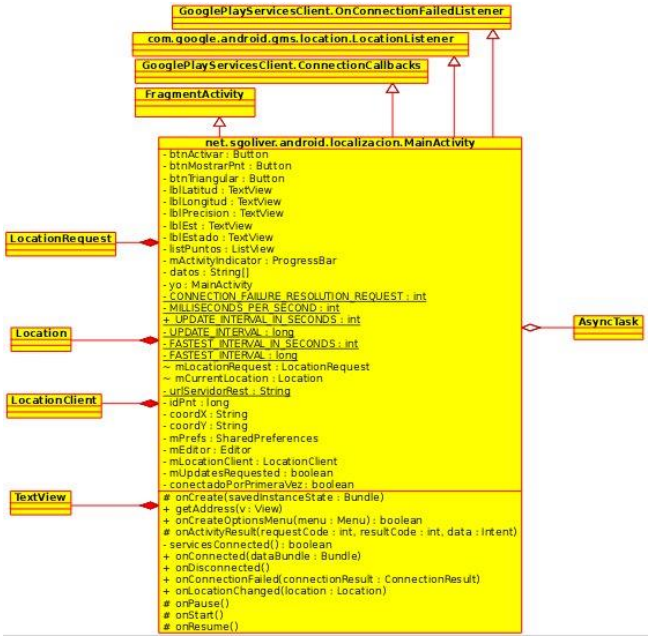


Fig. 5 Client class diagram.

VI. PROTOTYPE TEST

For interaction with the points set triangulation REST service, four asynchronous tasks were defined.

-- A task to obtain the ID of the points set associated with the session initiated by the client; this task starts from the *onCreate* method of the main activity class (**Fig. 6**).

Fig. 6 Triangulation ID assigned on client interface starting.

--When an *onLocationChanged* event is thrown by the Google Play Services software component, a popup message is put on the user interface whit the geographic coordinates provided by the GPS device and the client component put them in the Posición Actual controls zone ().

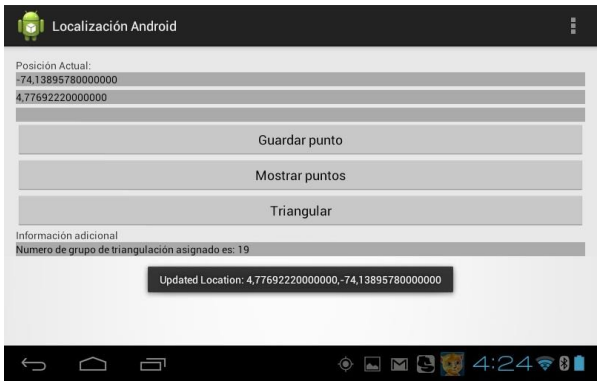


Fig. 7 Detection of Updated Location events on the user interface.

--A second task for the insertion of the point selected by the user (through the action of the "*Guardar punto*" button on the user interface) provided by the GPS receiver and displayed in the user interface in the field "*Posición actual*" (**Fig. 8**).

Fig. 8 Storage by the service using the user interface.

--A third task for obtaining the set of points stored so far from the REST service (through the action of the "*Muestre puntos*" button of the user interface) (**Fig. 9**).

Fig. 9 Stored Points set provided by the service.

--Finally, a task to obtain the edges of the possible triangulations reachable from the saved set of points. Both the list of points and the edges of the possible triangulations are presented in the *ListView* that is shown at the bottom of the user interface.

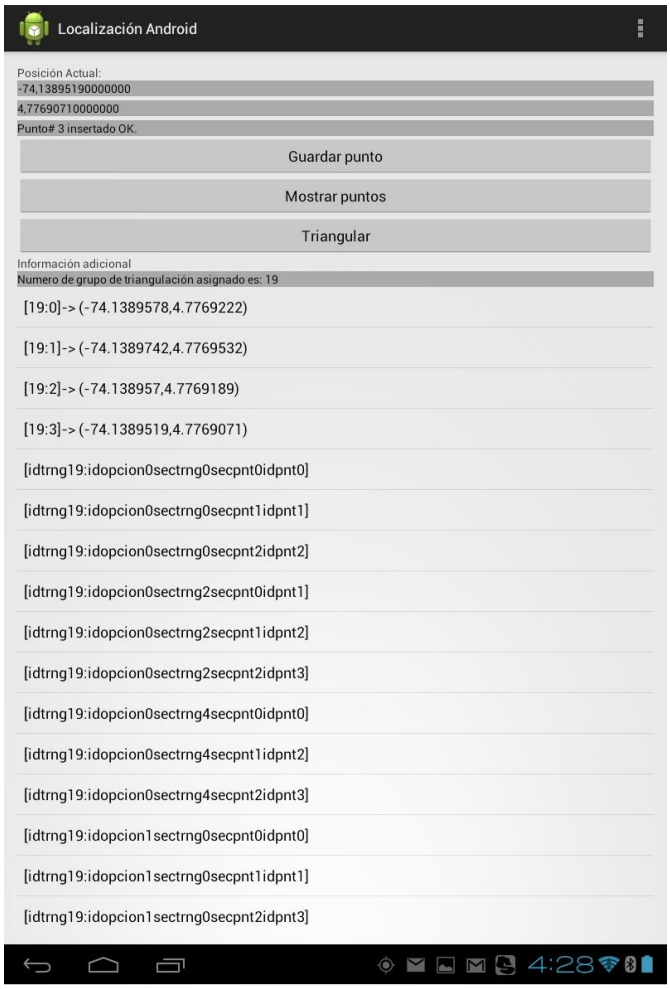


Fig. 10 Stored points set obtained from the service and its possible triangulations.

Fig. 10 shows the user interface with four points provided by the mobile device (a Toshiba Excite 10 inches tablet with a wireless link) and their possible triangulations. For each possible triangulation (identified by *idopcion*), the triangle is shown (identified by *sectrn*). For each triangle the point sequence (indicated by *secpnt*) and the point's id (identified by *idpnt*) is shown. The test used four very near points only a few meters apart and in this case, the particular point's arrangement produced the only two possible triangulations in agreement with the based matroid combinatorial algorithm used.

Coordinate transformation from the original World Geodetic System (WGS-84) spatial reference system into the Transverse Mercator projection was conducted using the minimum latitude and longitude of the set of points as the *latitude of origin* and *central meridian* parameters. This spatial

coordinate's projection is needed as GPS receivers provide geographic coordinates.

The test was also conducted using a LG Optimus L7 Cell smartphone. The following explanation focuses on illustrating the REST's server side of this test. **TABLE I** shows the 12 points provided by the smartphone. These points were located at a farther distance than the points in the previous test.

TABLE I
POINT SET LIST ON SERVICE SIDE

Geographic coordinates JSON format
World Geodetic System (WGS-84)


```
[{"conjuntodepuntosPK":{"idtrng":23,"idpnt":0},
    "coordx":-74.1465482,"coordy":4.7710368},
 {"conjuntodepuntosPK":{"idtrng":23,"idpnt":1},
    "coordx":-74.1429136,"coordy":4.7686211},
 {"conjuntodepuntosPK":{"idtrng":23,"idpnt":2},
    "coordx":-74.141702,"coordy":4.7678158},
 {"conjuntodepuntosPK":{"idtrng":23,"idpnt":3},
    "coordx":-74.13895,"coordy":4.7768594},
 {"conjuntodepuntosPK":{"idtrng":23,"idpnt":4},
    "coordx":-74.1328556,"coordy":4.7716231},
 {"conjuntodepuntosPK":{"idtrng":23,"idpnt":5},
    "coordx":-74.1354637,"coordy":4.7729735},
 {"conjuntodepuntosPK":{"idtrng":23,"idpnt":6},
    "coordx":-74.1429136,"coordy":4.7686211},
 {"conjuntodepuntosPK":{"idtrng":23,"idpnt":7},
    "coordx":-74.1366037,"coordy":4.7763712},
 {"conjuntodepuntosPK":{"idtrng":23,"idpnt":8},
    "coordx":-74.1365644,"coordy":4.7764313},
 {"conjuntodepuntosPK":{"idtrng":23,"idpnt":9},
    "coordx":-74.1465482,"coordy":4.7710368},
 {"conjuntodepuntosPK":{"idtrng":23,"idpnt":10},
    "coordx":-74.140837,"coordy":4.7762281},
 {"conjuntodepuntosPK": {"idtrng":23,"idpnt":11},
    "coordx":-74.1409072,"coordy":4.7760984}]
```


The point primary key *conjuntodepuntosPK* is composed of the triangulation id *idtrng* and the point id *idpnt*. *coordx* and *coordy* are the point´s meridian and latitude respectively.

When the client mobile device asks to performing the triangulation, the original coordinates are projected as shown in **TABLE II**, using for this test the values *4.7678158* and *-74.1465482* (the minimums) as the *latitude of origin* and *central meridian* parameters respectively.

TABLE II
PROJECTED POINT SET LIST

Id	X	Y
0	3561	0
1	890	4032
2	0	5376
3	10000	8429
4	4210	15190
5	5703	12296
6	890	4032
7	9460	11031
8	9527	11075
9	3561	0
10	9302	6335
11	9159	6257

The projected coordinates were scaled by a 10 factor.

TABLE III shows the legacy component result returned to the triangulation service REST.

TABLE III
THE LEGACY COMPONENT RESULT


```
[T[0]:=[0->12,3:{{0,1,2},{0,2,3},{2,3,4},{3,4,7},{4,7,8},{3,7,8},{0,3,10}}];]
[T[1]:=[1->12,3:{{0,1,2},{0,2,3},{2,3,4},{3,4,7},{4,7,8},{3,7,8},{3,10,11},
    {0,10,11},{0,3,11}}];]
[T[2]:=[2->12,3:{{0,1,2},{0,2,3},{4,7,8},{3,7,8},{0,3,10},{2,4,7},{2,3,7}}];]
[T[3]:=[3->12,3:{{0,2,3},{2,3,4},{3,4,7},{4,7,8},{3,7,8},{0,3,10},{0,2,6}}];]
[T[4]:=[4->12,3:{{2,3,4},{3,4,7},{4,7,8},{3,7,8},{0,3,10},{1,2,3},{0,1,3}}];]
[T[5]:=[5->12,3:{{0,1,2},{3,4,7},{4,7,8},{3,7,8},{0,3,10},{0,3,4},{0,2,4}}];]
[T[6]:=[6->12,3:{{0,1,2},{0,2,3},{2,3,4},{0,3,10},{3,4,8}}];]
[T[7]:=[7->12,3:{{2,3,4},{3,4,7},{4,7,8},{3,7,8},{1,2,9},{2,3,9},{3,9,10}}];]
[T[8]:=[8->12,3:{{0,1,2},{0,2,3},{3,4,7},{4,7,8},{3,7,8},{0,3,10},{3,4,5},{2,4,5},
    {2,3,5}}];]
[T[9]:=[9->12,3:{{0,1,2},{2,3,4},{3,4,7},{4,7,8},{3,7,8},{2,3,10},{0,2,10}}];]

…
[T[514]:=[514->12,3:{{0,1,2},{0,2,10},{2,4,8},{3,8,10},{2,8,10}}];]
[T[515]:=[515->12,3:{{0,1,2},{2,4,5},{2,3,10},{0,2,10},{2,3,8},{4,5,8},{2,5,8}}];]
[T[516]:=[516->12,3:{{4,7,8},{3,7,8},{2,4,7},{3,9,10},{1,3,9},{1,3,7},{1,2,7}}];]
[T[517]:=[517->12,3:{{3,4,7},{4,7,8},{3,7,8},{3,10,11},{0,10,11},{1,2,3},{3,4,5},
    {2,4,5},{2,3,5},{0,1,11},{1,3,11}}];]
[T[518]:=[518->12,3:{{3,4,7},{4,7,8},{3,7,8},{3,10,11},{1,2,3},{3,4,5},{2,4,5},
    {2,3,5},{1,3,9},{9,10,11},{3,9,11}}];]
[T[519]:=[519->12,3:{{3,4,7},{4,7,8},{3,7,8},{3,10,11},{0,10,11},{0,3,11},
    {0,1,3},{3,4,5},{2,4,5},{1,3,5},{1,2,5}}];]
[T[520]:=[520->12,3:{{3,4,7},{4,7,8},{3,7,8},{3,10,11},{0,10,11},{0,3,11},
    {0,2,6},{3,4,5},{2,4,5},{0,3,5},{0,2,5}}];]
[T[521]:=[521->12,3:{{0,1,2},{3,7,8},{0,3,10},{0,2,4},{3,5,7},{0,3,5},{0,4,5},
    {4,5,8},{5,7,8}}];]
```


The point primary key *conjuntodepuntosPK* is composed of the triangulation id *idtrng* and the point id *idpnt*. *coordx* and *coordy* are the point´s meridian and latitude respectively.

In this case, the particular point's arrangement produced 521 triangulations (all the possible ones). TABLE III shows only the result's head and tail. Each possible triangulation takes no more than the point id's as the process uses only the point set combinatorial structure.

Finally, the four last triangles of the last possible

triangulation of the response to the client mobile device are shown in **TABLE IV** using the JSON format.

TABLE IV
TRIANGULATIONS LIST ON SERVICE SIDE

…

{"triangulacionesPK":{"idtrng":23,"idopcion":521,"sectrng":10,"secpnt":0,"idpnt":0}},
{"triangulacionesPK":{"idtrng":23,"idopcion":521,"sectrng":10,"secpnt":1,"idpnt":3}},
{"triangulacionesPK":{"idtrng":23,"idopcion":521,"sectrng":10,"secpnt":2,"idpnt":5}},
{"triangulacionesPK":{"idtrng":23,"idopcion":521,"sectrng":12,"secpnt":0,"idpnt":0}},
{"triangulacionesPK":{"idtrng":23,"idopcion":521,"sectrng":12,"secpnt":1,"idpnt":4}},
{"triangulacionesPK":{"idtrng":23,"idopcion":521,"sectrng":12,"secpnt":2,"idpnt":5}},
{"triangulacionesPK":{"idtrng":23,"idopcion":521,"sectrng":14,"secpnt":0,"idpnt":4}},
{"triangulacionesPK":{"idtrng":23,"idopcion":521,"sectrng":14,"secpnt":1,"idpnt":5}},
{"triangulacionesPK":{"idtrng":23,"idopcion":521,"sectrng":14,"secpnt":2,"idpnt":8}},
{"triangulacionesPK":{"idtrng":23,"idopcion":521,"sectrng":16,"secpnt":0,"idpnt":5}},
{"triangulacionesPK":{"idtrng":23,"idopcion":521,"sectrng":16,"secpnt":1,"idpnt":7}},
{"triangulacionesPK":{"idtrng":23,"idopcion":521,"sectrng":16,"secpnt":2,"idpnt":8}}]

Each possible triangulation (identified by *idopcion*) shows the triangle (identified by *sectrn*), the point sequence (indicated by *secpnt*), and the point's id (identified by *idpnt*).

VII. CONCLUSION

The implementation of the prototype presented here allowed the reuse of a legacy tool implemented in C++ for triangulation of point configurations in two-dimensional spaces (for a discussion of performance details see [4]). The inherited tool was integrated in a Rest Web service using the point settings obtained from an Android mobile device.

Android client's interaction for coordinates' collection and use of the triangulation service was greatly simplified by the adoption of REST services. However, adaptation of the semantics of the operations GET and PUT, did not turn out as "natural" as it might be expected.

As definition of the services REST, for both the sets of points and the possible triangulations, was based on the Entity Class associated with the persistence provided by Java Persistence Service, the interchange format output by the triangulation component was not as self-descriptive as the one supplied by the service.

The projected coordinates were scaled by a ten factor in order to have integer values (as required by *points2triang*, the legacy component) but having a decimeter precision. The introduction of this scale factor did not affect in any way the outcome because the procedure used by the legacy component takes in account only the combinatorial structure of the associated oriented matroid.

The REST service asynchronous interaction facilitated the client implementation faults' tolerance, mainly the GPS receiver disconnection one, being necessary only to define the state variables with REST service session as static due that Google Play Services destroys and creates newly its objects after each disconnection.

No statistical validation was done because that was beyond the scope of the work, but this could be the subject of future work.

The matroid-based triangulation service can be reached at the IP address 200.69.103.29 on the http port 22095. The Android Client component (UbicacionGeografica.apk) and the other resources (i.e., IDE projects for Netbeans and eclipse-kepler for the services and client respectively) are available at IDE projects temporary site.

ACKNOWLEDGMENTS

We would like to thank the Francisco José de Caldas District University's Center for High Performance (CECAD) for their prompt and ongoing support, and to providing the computational resources required for the implementation of the prototype and the allocation of a public IP address to access the REST services.

REFERENCES

[1] Worboys, M. and Duckham, M. (2004) GIS: A Computing Perspective, 2Nd Edition, CRC Press, Inc., Boca Raton, FL, USA.

[2] Garey, M. R., Johnson, D. S., Preparata, F. P., and Tarjan, R. E. (1978) Information Processing Letters 7(4), 175 – 179.

[3] deBerg, M., Cheong, O., vanKreveld, M., and Overmars, M. (2008) Computational Geometry, Springer Berlin Heidelberg, third edition.

[4] Pfeifle, J. and Rambau, J. (2003) Computing triangulations using oriented matroids In Michael Joswig and Nobuki Takayama, (ed.), Algebra, Geometry and Software Systems, pp. 49–75 Springer Berlin Heidelberg.

[5] Valero, J. A. (2013) UD y la Geomática 0(7).

[6] Whitney, H. (1935) American Journal of Mathematics 57, 509–533. Oxley, J. (2007).

[8] Stell, J. and Webster, J. (2007) Computers, Environment and Urban Systems 31(4), 379 – 392 Topology and Spatial Databases.

[9] Richter-Gebert, J. and Ziegler, G. M. (2004) In Jacob E. Goodman and Joseph O'Rourke, (ed.), Handbook of Discrete and Computational Geometry, Boca Raton, FL, USA: CRC Press, Inc.. pp. 129–151.

[10] Anderson, L. and Delucchi, E. (2012) Discrete & Computational Geometry 48(4), 807–846.

[11] De Loera, J. A., Rambau, J., and Santos, F. (2010) Triangulations: Structures for Algorithms and Applications, Springer Publishing Company, Incorporated, 1st edition.

[12] Tsukamoto, Y. (2013) Discrete & Computational Geometry 49(2), 287–295.

[13] Pautasso, C., Zimmermann, O., and Leymann, F. (2008) In Proceedings of the 17th International Conference on World Wide Web WWW '08 New York, NY, USA: ACM. pp. 805–814.

[14] Manuel Báez, Álvaro Borrego, J. C., Cruz, L., González, M., Hernández, F., Palomero, D., deLleraand Daniel Sanz, J. R., Saucedo, M., Torralbo, P., and ÁlvaroZapata (2010) Introducción a Android, E.M.E. Editorial ©.

[15] Kovachev, D. and Klamma, Framework for Computation Offloading in Mobile Cloud Computing, R. 12/2012 2012 International Jorunal of Interactive Multimedia and Artificial Intelligence 1(7), 6–15. DOI: 10.9781/ijimai.2012.171

José Antonio Valero Medina is professor at the Facultad de Ingeniería at Francisco José de Caldas District University in Bogotá Colombia. He holds Msc degree in Teleinformática from Francisco José de Caldas District University. Actually he is a student of the Engineer PhD program there. His main areas of research are distributed systems and Geoinformatics.

Ivan Lizarazo Salcedo is professor at the Engineer Faculty at Francisco José de Caldas District University in Bogotá Colombia. He holds Master in Geographic Information Science and PhD In Geography degrees from University Of London. His main areas of research are Geographic-Object Based Image Analysis and Integration of agent-based simulation and GIS.

Overlap Algorithms in Flexible Job-shop Scheduling

Celia Gutiérrez

Universidad Complutense de Madrid, Spain

***Abstract* — The flexible Job-shop Scheduling Problem (fJSP) considers the execution of jobs by a set of candidate resources while satisfying time and technological constraints. This work, that follows the hierarchical architecture, is based on an algorithm where each objective (resource allocation, start-time assignment) is solved by a genetic algorithm (GA) that optimizes a particular fitness function, and enhances the results by the execution of a set of heuristics that evaluate and repair each scheduling constraint on each operation. The aim of this work is to analyze the impact of some algorithmic features of the overlap constraint heuristics, in order to achieve the objectives at a highest degree. To demonstrate the efficiency of this approach, experimentation has been performed and compared with similar cases, tuning the GA parameters correctly.**

***Keywords*— Algorithm, Flexible Job-Shop Scheduling, GA parameters, Local improvement, Overlap heuristics.**

I. Introduction

A Job-shop Scheduling Problem (JSP) is based on the concept of jobs, which are composed of operations that must be processed by the resources of different type in a sequential order. Each operation has a completion time. One machine can only process one job at a time and an operation cannot be pre-empted. The objective is to minimize the total makespan (the time to complete all jobs). The simplification of this problem is enunciated like this: there are *n* jobs to be scheduled on *m* machines in a general job-shop problem, *G*, minimizing the total completion operation time, C_{max}, $n/m/G/C_{max}$.

Flexible Job-shop Scheduling Problem is a generalization of the JSP, where the resource is selected among a set of suitable ones, giving place to two subproblems: routing and allocation of operations. The first one produces the start-time of the operations, and the second one the assignment of operations on resources.

Both JSP and fJSP have been solved by the use of metaheuristic algorithms, like GAs.The application of a GA on the simple basis as in [1] has poor performance because no domain knowledge is inserted, leading to non-feasible results. One way to insert knowledge into the algorithm is by hybridizing the GA with heuristics that provide local search. This paper follows the last approach, and goes beyond a deep analysis of GAs. It fact, it is an extension of [2], that explains how to achieve optimal results in the hybridization of GA with local search techniques to solve fJSP. This work provides a further analysis of the overlap constraint operators. In this way, the previous work provides a macroperspective view of the whole solution, and the present work is a microperspective view. It is structured in this way: section 2 covers the problem background; section 3 introduces the complete algorithm and the codification of information regarding the resources and fitness functions; section 4 shows the algorithms of a heuristic operator variants; section 5 shows the results of the experimentation phase; section 6 contains the comparison with similar approaches; and section 7 has the conclusions and future work.

II. Problem Background

Hybrid approaches that mix GA and heuristics are a well-known solution that has proven to be efficient, as heuristics provide domain knowledge that the simple GA cannot [3]. This focus can be applied in two ways: embedding the heuristics into the GA loop (integrated approach), or outside it (hierarchical approach), [3].

Literature shows examples of hybrid GA with intelligent genetic operators than produce optimal schedules. This is the case of [4], that describe an effective hybridation of both techniques, applying improved crossover and mutation operators when there are non-feasible schedules.[5] describes a hybrid GA solution by the use of two vector chromosome and bottleneck shifting procedure. The representation is made by two vectors: one for the machine assignment and the another one for the operation sequence. [6] solve the same problem by the use of an artificial immune algorithm. It uses several strategies for generating an initial population and selecting the best individuals. It also has operators that reorganize the operations (by a mutation). [7] adopt the hybrid GA by the use of the approach by localization to initialize the GA, and improving it by reordering jobs and machines, and by searching for a global minimum [4] have improved operators constraint and mutator operators that consider constraint violations.

The second way to include the heuristics has also been widely implemented, though the existing algorithms vary in the order of application, heuristic methods, goal of the application,

DOI: 10.9781/ijimai.2014.265

and even domain. [8] follows this paradigm by means of a local search by the definition of the neighborhood.

This work follows the second approach. Having proven the efficiency of the mentioned algorithms, the objective of this research is to provide the designer with relevant issues that improve the algorithm performance when using local improvements within a hybrid GA under a hierarchical architecture. This is also considered a multi-objective fJSP, because the solution achieves three goals:

- To minimize the makespan of the operations.
- To minimize the maximal machine overload, i.e., the maximum working time spent at any machine.
- To satisfy the maximum number of constraints.

There are also recent approaches to solve the problem of JSP, like [9], where they solve the problem of scheduling independent tasks in a grid computing system. They use a new evaluation (distributed) algorithm inspired by the effect of leaders in social groups, the group leaders' optimization algorithm (GLOA). In contrast, the present work analyzes some design features of the hybrid algorithm, preferably the overlap constraint repairer.

III. HIERARCHICAL DESIGN FEATURES

This work constitutes the extended version of the previous work, providing deeper details of the heuristics design and argumentation for the parameters tuning. So, whereas [2] and [10] provide a solution to a general fJSP, the current work provides design and execution details in order to achieve the goals of the algorithm.

This research has been analysed following a hierarchical approach that decomposes the resource and the start-time assignment in two different problems solved by different and independent GA, like in [5]. Previous to both GA running, there is a module that calculates the limits for the start-time for each operation, and after both GA running the module of the heuristics solve the unfulfilled constraints. The adaptation of the algorithm to JSP claims a simpler architecture, where the resource GA module does not appear. Other variations concerning the heuristics are also discussed in the section 4.

A. Codification of the Resource GA Chromosome

The chromosome and fitness function for both GA are described in the previously cited works. There are subtle differences in the morphology of both chromosomes: while the solution for time GA is directly codified into the chromosome, the chromosome for resource GA stores as many genes as operations, which must be decoded to get the resource number. For example, for the set of 4 orders, 3 products per order (maximum), 1 product instance per product (maximum), 5 operations per instance (maximum), and 4available resources in the job-shop, the gene value must cover 4 x 3 x 1 x 5 x 4 values, so the range is [0-239]. To decode a gene value, successive divisions must be applied using this algorithm that involves equation (1) to equation (8):

$$\textit{Resource number} = \textit{gene MOD number of resources} \quad (1)$$

$$\textit{cant} = \textit{gene / number of resources} \quad (2)$$

$$\textit{product instance identification} = \textit{gene MOD number of product instances} \quad (3)$$

$$\textit{cant} = \textit{cant / number of product instances} \quad (4)$$

$$\textit{operation number} = \textit{cant MOD number of operations} \quad (5)$$

$$\textit{cant} = \textit{cant / number of operations} \quad (6)$$

$$\textit{product identification} = \textit{cant MOD number of products} \quad (7)$$

$$\textit{order number} = \textit{cant / number of products} \quad (8)$$

For a gene value of 69, the decoding process gives the following values for the parameters:

- resource number = 1
- product instance identification = 0
- operation number = 2
- product identification= 0
- order number = 1

B. GA fitness functions

There is one fitness function for each GA. Both functions incorporate penalizations that depend on the domain they are evaluating. For both GAs, the objective is to minimize the values obtained by the fitness functions.The following subsections contain their codification:

1) Fitness function for Resource GA

This function evaluates the sums of deviations between the assignment of operations to certain resource and the ideal assignment. In other words, this fitness function penalizes non-balanced assignments of operations among the resources of the same type. The ideal assignment is the number of operations assigned to the resources of the same type, divided into the number of resources of that type, as equation (9) shows:

$$\textit{Fitness} = f \times \sum_{i=0} |O_{i,t} - (O_t / R_t)| \quad (9)$$

where:

f is a the penalty factor (For simplicity, f=1),
i represents each resource in the job-shop,
$O_{i,t}$ is the number of operations assigned to the i resource, that belongs to the t type of resource,
Ot is the number of operations assigned to the resources of t type,
R_t is the number of resources of t type.

2) Fitness function for Time GA

This function sums up the starting times of all operations, with a penalization when an operation violates a constraint, as in equation (10):

$$Fitness = \sum_{i=0} t_i + p_i \quad (10)$$

where:

i represents each operation in the job-shop,
t_i is the starting time of the *i* operation,
p_i is the sum of quantities derived from penalizations for order and overlap violated constraints, in the way equations (11) and (12) show:

-if an order constraint is violated, the fitness must be severely penalized, so that this chromosome does not to pass to the next generation:

$$p_i = p_i + 100000000 \quad (11)$$

-if overlap constraint is violated, the fitness is penalized proportionally to the amount of the overlap. :

$$p_i = p_i + |t_{f,j} - t_i| \quad (12)$$

where $t_{f,j}$ is the finishing time of the *j* overlapped operation.

Notice that range constraint is not contemplated in the penalization equation because the time GA assigns the start-times within the range limits. Therefore the solutions provided by the time GA are always valid according to this constraint.

C. Heuristic algorithm

A relevant design issue is the organization of constraints in the heuristic stage. In a Constraint Satisfaction Problem (CSP) like this, a dilemma appears on the order of repairment of the constraints, claiming a further analysis. As the repairment of a constraint can modify the degree of satisfaction of the remaining constraints, the evaluation of the constraint of each operation must be followed by each repairment, so its start-time is updated. The algorithm below shows the workflow of the heuristic stage. It ends when it reaches a maximum number of iterations (MAX_IT). This parameter is tuned depending on the size of the orders, as explained in subsection 5.2.

```
Step 1: Point to 1st operation
Step 2: Get operation data
Step 3: Point to 1st constraint
Step 4: Heuristic evaluator
Step 5: Heuristic repairer
Step 6: If no more constraints
              then go to step 8
        otherwise go to step 7
Step 7: Point to next constraint
Step 8: If more operations
              then go to step 9
        otherwise go to step 10
Step 9: Point next operation
Step 10: Termination condition.
        If iterations = MAX_IT
              then exit
        otherwise go to step 1
```

IV. VARIANTS FOR THE OVERLAP CONSTRAINT

As mentioned before, each constraint has one module to evaluate, and another one to repair. Whereas Range and Order heuristics are simple and described in [2], Overlap heuristics requires a deeper design: the evaluator is more complex than the other ones, and the repairer presents different variants.

Previously to running this repairer, a conflict appears about which of the overlapped operations has the priority to get repaired, which is not necessarily the operation appointed by the main algorithm. This is solved by the designation of the *critical operation*. The overlap repairer goal is to find an interval where the operation can be shifted while respecting the range constraint, so the critical operation must have the narrowest margin for start-time assignment (i.e. it is the most restrictive), as equation (13) says:
i is critical over *j* if:

$$|tmax_i - tmin_i| < |tmax_j - tmin_j| \quad (13)$$

i, *j* are the overlapped operations
$tmax_i$ is the start-time upper limit for *i* operation
$tmin_i$ is the start-time lower limit for *i* operation

Each overlap repairer solves one overlap of a pair of operations, so if an overlap has more than two operations like equation (14) says, it will be solved in *k+1* iterations of the repairer. At each iteration, there will be a different designation for the critical operation.

$$k + 2, k > 0 | \quad (24)$$

Apart from these variables, there are others that participate in subsequent algorithms:

- O is the current operation of the algorithm defined in section 3. It is the operation that is being evaluated/repaired at each iteration of the main program.
- J is the operation that is being compared to the O at each evaluator/repairer iteration.
- C is the critical operation in an overlap.
- t_i is the start time of i operation.
- I is the current interval of the R. An interval is considered when there is a period of time when R is not assigned to any operation, so it remains not active.
- R_i is the resource assigned to i operation.
- T_R is the type of R resource.
- S is the resource currently appointed to.
- L is the list of operations that overlap with O.
- L_I is the list of I.
- L_R is the list of resources of the same type as R_o

The structure for the evaluator and the repairer variants are described in the following subsections.

A. Overlap Evaluator

The following algorithm includes the steps to evaluate if the current operation overlaps other one(s) on the same resource:

```
Step 1: Store (O, L)
Step 2: Point J at the 1st operation
assigned to Ro
Step 3: Stop condition:
          if no more operations for Ro
          then stop
          oterwise go to step 4.
Step 4: If J not = O, and J overlaps O
          then store (J, L)
Step 5: Point J at the next operation in
Ro
Step 6: Go to step 3.
```

Operations are overlapped if an operation begins before the other one has finished. The information that results from this stage is a list of operations that overlaps the current one. This list is the input of the overlap repairer stage.

B. Overlap Repairer

The overlap repairer includes several stages (i.e. Interval Search, OperationExchange, Resource Mutation), which are successively executed if the previous one has not been successful, as [2] show.

Other design issues come out when handling constraints that interfere with others. In this case, there are two possibilities:

1. To consider a blind repairment, so that the constraint is repaired without considering the other ones. Such is the case of the order and range repairers.
2. To consider an intelligent repairment, so that the constraint is repaired taking the other ones into consideration. Overlap repairer follows this approach. There are several ways to incorporate these considerations, producing two variants for overlap repairer: the first one (pure variant) considers the range constraint for its amendments; the second one (hybrid variant) considers both the range and the order constraints. The mentioned stages can be designed in both ways:

1) Algorithms for Pure Variants.

a) Algorithm for Interval Search

```
Step 1: Find LI for Ro
Step 2: Find C among two overlapped in L
Step 3: Position I at the beginning of
LI
Step 4: Stop condition:
          if no more intervals in LI
          then go to step 8.
Step 5: If I suitable for C
          Then tc = max (tminC, tminI)
          Exit
Step 6: Position I at next interval of
LI
Step 7: Go to step 4.
Step 8: Exit.
```

An interval is suitable if it matches the assignment conditions for the critical operation, in terms of operation duration and start-time range limits.

b) Algorithm for OperationExchange.

```
Step 1: Find C among two overlapped in L
Step 2: Position J in previous operation
in Rc
Step 3: Stop condition:
if no more previous operations,
then exit.
Step 4: If J suitable for C
          then exchange (tj, tC)
          exit.
Step 5: Position J in the next previous
operation in Rc
Step 6: Go to Step 3.
```

A current operation is suitable if its start-time fulfills the range constraint of the critical one.

2) Algorithms for Hybrid Variants.

a) Algorithm for Interval Search.

It remains the same as the PureVariant, except the suitability condition is step 5. In *this case, a*n interval is suitable if it matches the assignment conditions for the critical operation, in terms of operation duration and start time range bounds, and not belonging to the same job (to assure it fulfills the order constraint).

b) Algorithm for OperationExchange .

It *remains the same as the PureVariant, except the suitability condition is step 4. In this case,* an operation is suitable if it does not belong to the same job (to assure it fulfills the order constraint), and its start-times fulfills the range constraint of the critical.

c) Algorithm for Mutation Operator.

This operator assigns the operation to another resource of the same type, while preserving the start-time. This amendment does not interfere with the other type of constraints, but it can produce overlaps in the new resource.

```
Step 1: Find C among two overlapped in L
Step 2: Position S in 1st resource in
the job-shop
Step 3: Stop condition:
          if no more resources
         then go to step 7.
Step 4: If S not = RC and Ts = TRC
          then store (S, LR)
Step 5: Position S in next resource in
the job-shop
Step 6: Go to step 3.
Step 7: Random assignment of RC among
the candidates in LR.
```

V. Experimentation Results

Tests have been performed for the complete algorithm, putting special emphasis on the variants of the overlap repairers. The machine has been a Sun Sparc workstation running Solaris operating system. There has been a preliminary stage, to configure the GA, and a main stage, to validate the complete algorithm.

A. Tuning the GA Parameters

Beside the algorithmic issues, the success of the algorithm lies on several factors, like the correct tuning of the GA parameters. Several works have inserted in the code the way to tune them dynamically like the fuzzy logic controller (FLC), which methods are described in [11]. The key of success of applying FLC to GA is a well-formed fuzzy sets and rules [12]. In this work there has been previous experimentation to analyse the best values for the GA, by testing the different GA isolatedly. The most successful configuration for the parameter set population size/number of generations/mutation rate/selection type is 50/60/0.01/tournament for the resource GA and 8/10/0.01/elite for the time GA.

B. Configuration of the Hybrid GA

Testbeds have been configured varying the number of orders from 1 to 4, number of jobs from 1 to 3, number of products from1 to 4, number of product instances from 1 to 2, number of operations from 1 to 4, and operation processing times from 24 to 100, 5 resources belonging to 4 types, with the total number of executions per testbed of 25. The number of iterations for the heuristics stage has varied with the number of orders: for one order only 100 have been needed, while for four orders more than 200. Results collect the average of the executions.

Heuristic optimization algorithms can be evaluated in two ways [13]: by measuring the solution quality and measuring the solution time. In this case we have measured the solution quality by two criteria:

Considering this problem as a CSP, the solution quality must measure the constraint satisfaction rate. In this work, we consider the mean error (ME) parameter, as the percentage of constraints not satisfied. Figure 1 shows the results for the pure and hybrid variants of interval and exchange operators, distributed horizontally by the number of orders and vertically by the ME. This figure reflects that for few operations the pure repairer is better, but when the number of operations increases, the hybrid one is better. In this case, the ME is higher than 0, due to the technological limitations, i.e. more operations for the same number of resources produces more operations with unfulfilled constraints, and therefore reduces the number of fulfilled constraints. The reason for this improvement using the hybrid repairer is that the design of that heuristics has been made in such a way that the improvement in the overlap does not worsen any other constraint, in contrast with the pure repairer. The disadvantage of that is that fewer amendments can be applied with this variant, because it is more restrictive.

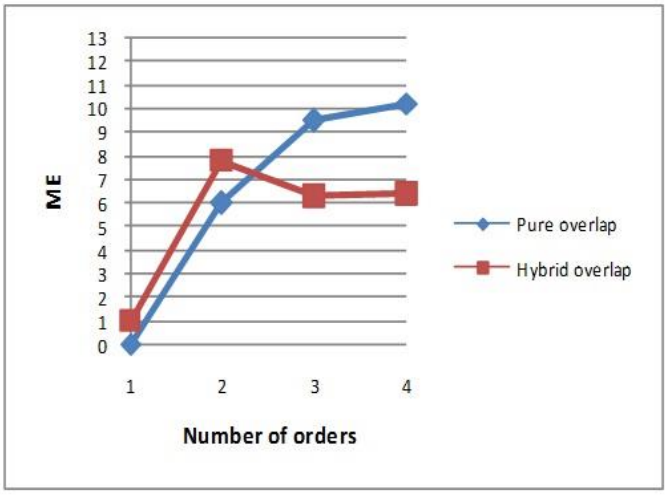


Fig. 1. ME of the two variants of overlap repairers

Considering it as a fJSP, the quality measurement is the time GA fitness. Table I shows the results for the time GA, as it is related to the constraints. PRf and HRf columns contain the Pure Repairer fitness and the Hybrid Repairer fitness respectively. Def(HRf, PRf) provides information about the percentage difference of both fitness values as equation 15 shows:

$$Def(HRf, PRf) = 100 * \frac{(HRf - PRf)}{PRf} \quad (15)$$

Table I
Fitness values for the overlap variant

Number of orders	PRf	HRf	Def(HRf, PRf)
1	300	316	5.33%
2	352	379	7.67%
3	380	397	4.47%
4	411	419	1.95%

There is a relationship between the values for ME in Figure 1 and the fitness values shown in Table I. The fitness function is penalized when the range and overlap constraints are not fulfilled. The fewer the number of orders, the lower (and better) fitness results. Results are also better for the pure variant than the hybrid one. The reason is that the former reorder the overlapped operations trying to fulfill the range constraint, and the latter must also makes sure that the reorganization also fulfills the order constraint. This complexity means that the search interval does not always find the earliest interval suitable, and even does not find and interval, delaying more operations of the jobs than in the pure variant.

Besides that, the evolution of Def(HRf, PRf) is to decrease when the number of orders increases. This also shows that the fitness values in both repairers tend to be very similar for high number of orders. Therefore, it is recommendable to use the Hybrid Repairer in these cases, because they will provide similar fitness values than the Pure Repairer but with lower ME values.

VI. Comparison with Alternative Solutions

To test the efficiency of our algorithm, Table II collects the comparison with respect the makespan using [8] benchmark. It contains the best results of a set of executions. It consists of ten problems mk1-mk10, with the number of jobs are in the range 10-20, the number of machines are in the range 6-15, number of operations are in the range 5-15. Other configuration information is: n x m, that refers to the number of jobs per number of machines; (LB, UB) with the optimum makespan if known [14]; otherwise, it reports the best lower and upper bound known; Flex. with the average number of equivalent machines per operation. This work compares the mentioned fJSP experiments of hGA from [5], AIA [6] and GA [7], and TWS for the best results achieved among the different rules in [8]. The information presented in Table 2 has been partially obtained from [2].

The proposed algorithm of GAH has achieved lower results of makespan for some fJSP instances and similar results of makespan for the remaining fJSP instances. These results combined with the ME results in section 5, demonstrate that the algorithm shows excellent quality solution as a fJSP and a CSP.

TABLE II
Comparison with best known makespan for ten fJSP instances

Problem	n x m	Flex.	(LB, UB)	hGA	AIA	GA	TWS	GAH
Mk01	10 x 6	2.09	(36, 42)	40	40	40	42	40
Mk02	10 × 6	4.10	(24, 32)	26	26	26	32	26
Mk03	15 × 8	3.01	(204, 211)	204	204	204	211	204
Mk04	15 x 8	1.91	(48, 81)	60	60	60	81	60
Mk05	15 × 4	1.71	(168, 186)	172	173	173	186	172
Mk06	10 x 15	3.27	(33, 86)	58	63	63	86	57
Mk07	20 × 5	2.83	(133, 157)	139	140	139	157	139
Mk08	20 x 10	1.43	523	523	523	523	523	523
Mk09	20 × 10	2.53	(299, 369)	307	312	311	369	308
Mk10	20 × 15	2.98	(165, 296)	197	214	212	296	196

VII. Conclusions and Future Work

This work has described the algorithms of a complex heuristic, like the overlap evaluator and repairers, in a hybrid GA applied to fJSP, a multi-objective problem. The most relevant issue concerns the use of two variants for the repairer: one that does not take into consideration the other constraints (pure), and the other one that incorporates them (hybrid). When adopting this approach, designers may consider what the experimentation has revealed: pure variant is better for fJSP with few operations, producing better ME results; in contrast, it is recommendable the use of the hybrid variant when the number of operations increases. It also shows that it maintains the level of quality of other algorithms, in terms of makespan. Finally, it is also recommendable an appropiate tuning of GA parameters.

The future work opens a high number of possibilities. Concerning the inclusion of intelligent operators, we are working in the design of hybrid variants for the range and precedence repairers. In the same way, we are making another variant of the ResourceMutation substage, which assures that the new resource assignment does not cause the overlap of other operations. Finally, new constraints adapted to concrete JSP and fJSP are to be incorporated and experimented. Re-design of the model is done using the FactoryMethod design patron, where a family of constraints can be chosen depending on the application that is used. The collection of classes in [2], will be transformed in the collection shown in Figure 2. The fJSP class is the superclass which the concrete application inherits from: in the described work, this application is GAH, which uses the order, range, and overlap concrete constraints. When using OtherApplication, it will use OtherConstraints (containing the measurer or evaluator), which has the corresponding OtherConstraint_unfulfilled subclass (containing the repairers for that constraint). The construction of the repairer will also contemplate the inclusion of pure and hybrid variants. The choice on which one to use will depend on the number of operations handled by the fJSP. The results of the mentioned modifications will be compared with the current version, to see how they affect to the ME and the *makespan*.

VIII. Acknowledgment

C. Gutiérrez thanks Jose Maria Lazaro for his continuous help on the applications of fJSP. She also wants to thank Juan Jose Merelo for his guidance in the customization of his GAGS. This research work has been funded by the Spanish Ministry for Economy and Competitiveness through the project “SOCIAL AMBIENT ASSISTING LIVING - METHODS (SociAAL)” (TIN2011-28335-C02-01).

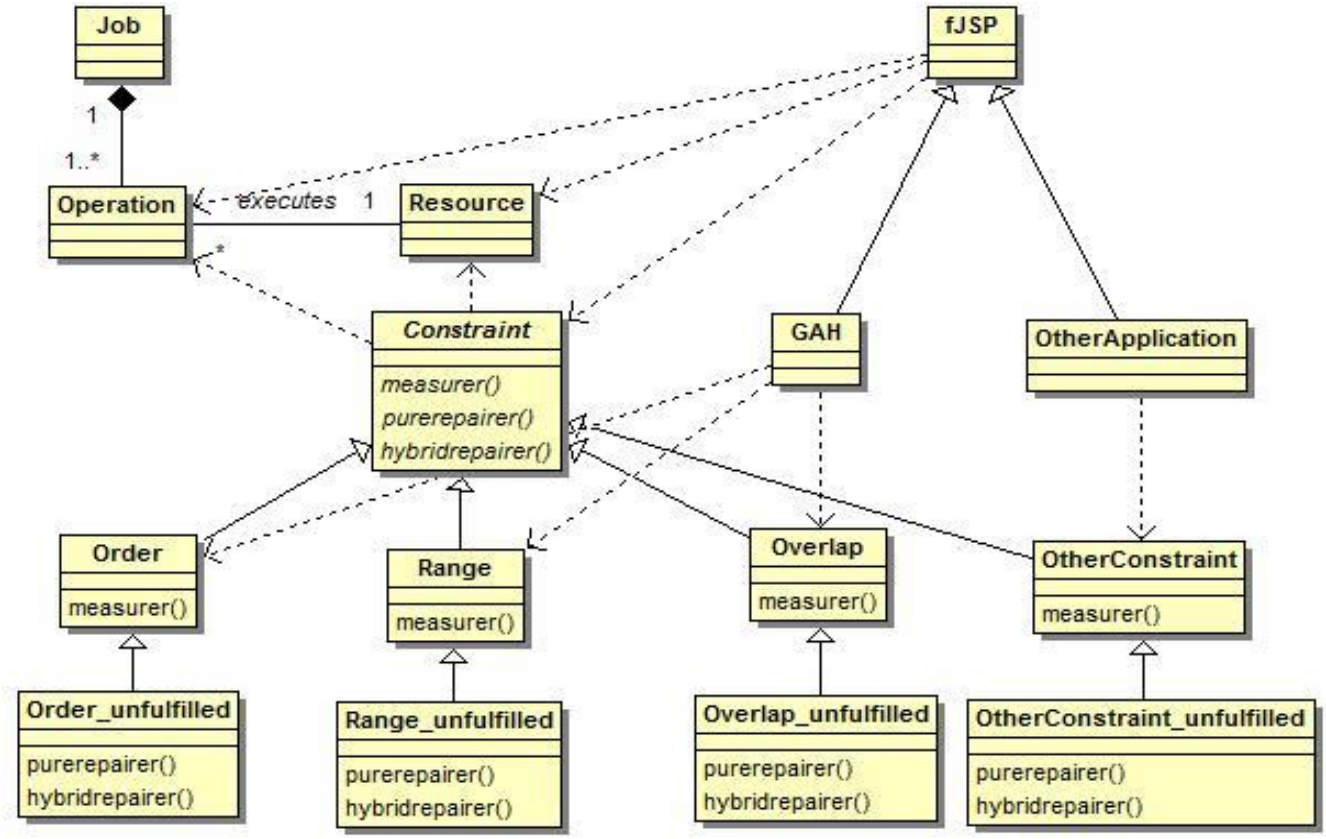


Fig. 2. Re-design of the classes for adaptation to other problems.

REFERENCES

[1] J.-J. Lin, "A GA-based Multi-Objective Decision Making for Optimal Vehicle Transportation," Journal of Information Science and Engineering, vol. 24, pp. 237-260, 2008.

[2] C. Gutierrez and I. García-Magariño, "Modular design of a hybrid genetic algorithm for a flexible job–shop," Knowledge-Based Systems, vol. 24, pp. 102–112, 2011.

[3] Y. Yun, M. Gen, M., and S. Seo, "Various hybrid methods based on genetic algorithm with fuzzy logic controller," Journal of Intelligent Manufacturing, vol. 14, pp. 401-419, 2003.

[4] J. Dorn and M. Girsch, "Genetic operators based on constraint repair," in ECAI'94 Workshop on Applied Genetic and other Evolutionary Algorithms.

[5] J. Gao, M. Gen, and L. Sun, "A hybrid of genetic algorithm and bottleneck shifting for flexible job shop scheduling problem," in 8th Annual Conference on Genetic and Evolutionary Computation, 2006, pp. 1157-1164.

[6] A. Bagheri, M. Zandieh, I. Mahdavi, and M. Yazdani, "An artificial immune algorithm for the flexible job-shop scheduling problem," Future Generation Computer Systems, vol. 26, pp. 533-541, 2010.

[7] F. Pezella, G. Morganti, and G. Ciaschetti, "A genetic algorithm for the Flexible Job-shop Scheduling Problem," Computers & Operations Research, vol. 35, pp. 3202-3212, 2008.

[8] P. Brandimarte, "Routing and scheduling in a flexible job shop by tabu search," Annals of Operations Research, vol. 41, pp. 157-183, 1993.

[9] Z. Pooranian, M. Shojafar, J. H. Abawajy, and M. Singhal, "GLOA: A New Job Scheduling Algorithm for Grid Computing," International Jorunal of Interactive Multimedia and Artificial Intelligence, vol. 2, no. 1, pp. 59-64, 2013.

[10] C. Gutierrez, "Heuristics in a General Scheduling Problem," in Proc. Conference on Evolutionary Computation, vol. 1, 2004, pp. 660-666.

[11] M. Gen and R. Cheng, Genetic Algorithms and Engineering Optimization. New Jersey: John Wiley and Sons, 2000.

[12] F. Cheong and R. Lai, "Constraining the optimization of a fuzzy logic controller using an enhanced genetic algorithm," IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, vol. 30, pp. 31-46, 2000.

[13] R. Rardin and R. Uzsoy, "Experimental Evaluation on Heuristic Optimization Algorithms: A Tutorial," Journal of Heuristics, vol. 7, pp. 261-304, 2001.

[14] M. Mastrolilli and L.-M. Gambardella, "Effective neighbourhood functions for the flexible job shop problem," Journal of Scheduling, vol. 3, pp. 3-20, 2000.

C. Gutierrez was born in Bilbao, Spain, in 1969. She received the B.Eng. and Ph.D. degrees in Computer Science from the University of Deusto (Spain) in 1992, and from the University of the Basque Country (Spain) in 2000, respectively. She has worked for Labein technological research center, in Bilbao (Spain). She has also worked for the Basque Government and Indra, as a software Engineer. After doing teaching and research at some private universities, such as Instituto Tecnológico de Estudios Superiores de Monterrey (México D.F.), she joined Universidad Complutense de Madrid (Spain) where her current position is Assistant Professor . She is a member of the Grasia research group, on agent technologies. She has published in relevant journals, like Knowledge-based Systems. She has also directs financially supported projects. Her current interests focus on Data Mining, Software Quality, Multiagent Systems and Accessibility.

Our System IDCBR-MAS: from the Modelisation by AUML to the Implementation under JADE Platform

Abdelhamid Zouhair[1], El Mokhtar En-Naimi[1], Benaissa Amami[1], Hadhoum Boukachour[2], Patrick Person[2], Cyrille Bertelle[2]

[1]*LIST Lab, The University of Abdelmalek Essaâdi, FST of Tangier, Morocco*
[2]*LITIS Lab, The University of Le Havre, France*

***Abstract* — This paper presents our work in the field of Intelligent Tutoring System (ITS), in fact there is still the problem of knowing how to ensure an individualized and continuous learners follow-up during learning process, indeed among the numerous methods proposed, very few systems concentrate on a real time learners follow-up. Our work in this field develops the design and implementation of a Multi-Agents System Based on Dynamic Case Based Reasoning which can initiate learning and provide an individualized follow-up of learner. This approach involves 1) the use of Dynamic Case Based Reasoning to retrieve the past experiences that are similar to the learner's traces (traces in progress), and 2) the use of Multi-Agents System. Our Work focuses on the use of the learner traces. When interacting with the platform, every learner leaves his/her traces on the machine. The traces are stored in database, this operation enriches collective past experiences. The traces left by the learner during the learning session evolve dynamically over time; the case-based reasoning must take into account this evolution in an incremental way. In other words, we do not consider each evolution of the traces as a new target, so the use of classical cycle Case Based reasoning in this case is insufficient and inadequate. In order to solve this problem, we propose a dynamic retrieving method based on a complementary similarity measure, named Inverse Longest Common Sub-Sequence (ILCSS). Through monitoring, comparing and analyzing these traces, the system keeps a constant intelligent watch on the platform, and therefore it detects the difficulties hindering progress, and it avoids possible dropping out. The system can support any learning subject. To help and guide the learner, the system is equipped with combined virtual and human tutors.**

***Keywords* — Intelligent Tutoring Systems (ITS), Multi-Agents System (MAS), Incremental Dynamic Case Based Reasoning (IDCBR), Similarity Measure, Traces.**

I. INTRODUCTION

E-LEARNING is a computer system which offers learners another means of learning. Indeed it allows learner to break free from the constraints of time and place of training. They are due to the learners' availability. In addition, the instructor is not physically present and training usually happens asynchronously. However, most E-learning platforms allow the transfer of knowledge in digital format, without integrating the latest teaching approach in the field of education (e. g. constructivism [23], ...). Consequently, in most cases distance learning systems degenerate into tools for downloading courses in different formats (pdf, word ...). These platforms also cause significant overload and cognitive disorientation for learners. Today, it is therefore necessary to design and implement a computer system (i. e. intelligent tutor) able to initiate the learning and provide an individualized monitoring of the learner, who thus becomes the pilot of training. The system will also respond to the learner's specific needs.

Solving these problems involves first, to understand the behaviour of the learner, or group of learners, who use platform to identify the causes of problems or difficulties which a learner can encounter. This can be accomplished while leaning on the traces of interactions of the learner with the platform, which include history, chronology of interactions and productions left by the learner during his/her learning process. This will allow us the reconstruction of perception elements of the activity performed by the learner.

We consider an Intelligent Tutoring System (ITS), that is able to represent, follow and analyze the evolution of a learning situation through the exploitation and the treatment of the traces left by the learner during his/her learning on the platform. This system is based, firstly on the traces to feed the system and secondly on the reconciliation between the course of the learner (traces in progress) and past courses (or past traces). The past traces are stored in a database. Our system is able to represent, follow and analyze the evolution of a learning situation through the exploitation and the treatment of the traces left by the learner during his/her learning on the platform. The analysis of the course must be executed continuously and in real time which leads us to choose a Multi-Agents architecture allowing the implementation of a dynamic case-based reasoning. Recently, several research works have been focused on the dynamic case based reasoning in order to push the limits of case based reasoning system static, reactive and responsive to users. All these works are based on the

DOI: 10.9781/ijimai.2014.266

observation that the current tools are limited in capabilities, and are not able of evolving to fit the non-anticipated or emerging needs. Indeed the reuse of past experiences causes several problems, such as:

- Modeling: formalization of experience acquired (cases), indeed a few CBR systems are able to change over time the way of representing a case [6]. According Alain Mille, a case has to describe its context of use, which is very difficult to decide before any reuse and can change in time [22].
- Treatment: the use of the classic reasoning cycle is insufficient and inadequate in dynamic or emerging situations, unknown in advance.

In order to deal with this issue, we propose a Dynamic Case Based Reasoning based on a dynamic retrieve method, and we propose a dynamic retrieving method based on a complementary similarity measure, named Inverse Longest Common Sub-Sequence (ILCSS).

The rest of this paper is organized as follows: In the second section, we present a general introduction of intelligent tutoring system. In the third section we present a Multi-agents Case-Based Reasoning, and in the following part, we will propose the description of our approach in Case Based Reasoning and intelligent tutoring systems field: Incremental Dynamic Case-Based Reasoning founded on Multi-Agents System. In the next section, we present some development results of our system. Finally, we present a comparison between IDCBR-MAS system and other CBR system, and we will give the conclusion and our future work.

II. Intelligent Tutoring Systems

Intelligent Tutoring Systems (ITS) are computer systems designed to assist and facilitate the task of learning for the learner. It can personalize learning for learners, providing a less expensive solution for a diverse generation of learners. They have expertise in so far as they know the domain knowledge, how to teach (pedagogical knowledge) and also how to acquire information about the learner. We note that, the general architecture of Intelligent Tutoring Systems was represented in our articles [10]. Many researches have been designed and implemented in Intelligent Tutoring System, in order to assist a learner in his/her/their learning. There are, for example, tutors or teaching agents who accompany learners by proposing remedial activities [11]. There are also the agents of support to the group collaboration in the learning [7] encouraging, the learners participation and facilitating discussion between them. Other solutions are based on Multi-Agents System that incorporate and seek to make cooperation among various Intelligent Tutoring System [5]. The Baghera platform [32] exploits the concepts and methods of Multi-Agents approach. Baghera assists learners in their work solving exercise in geometry. They can interact with other learners or teachers. The teachers can know the progress of the learners work in order to intervene if needed. These tools of distance learning do not allow an individualized, continuous and real-time learners follow-up. They adopt a traditional pedagogical approach (behaviorist) instead of integrating the latest teaching approaches (constructivism and social constructivism [23], [30]). Finally, given the large number of learners who leave their training, the adaptation of learning according to the learners profile has become indispensable today.

Our contribution in these important areas is to design and develop an adaptable system that can ensure an automatic and a continuous monitoring of the learner. Moreover, our system is open, scalable and generic to support any learning subject.

III. Multi-Agents Case-Based Reasoning

A. Case-Based Reasoning

Case-Based Reasoning (CBR) is an artificial intelligence methodology which aims at solving new problems based on the solutions of similar past problems (past experiences) [14]. The solved problems are called source cases and are stored in a case-base (base of scenarios). The problem to be solved is called target case. A CBR is a combination of knowledge and processes to manage and re-use previous experience.

The Case-Based Reasoning cycle is composed of five steps as given at following figure (Fig. 1):

- Presentation: the current problem is identified and completed in such a way that it becomes compatible with the contents and retrieval methods of the case-base reasoning.
- Retrieve: The task of retrieve step is to find the most similar case or cases to the current problem in the case-base.
- Reuse: The goal of the reuse phase is to modify the solution of source case found in order to build a solution for the target case.
- Revise: The phase of revision is the step in which the solution suggested in the previous phase will be evaluated. If the solution is unsatisfactory, then it will be corrected.
- Retain: retaining the new experience and add it to the knowledge-base (case-base) [12], [1].

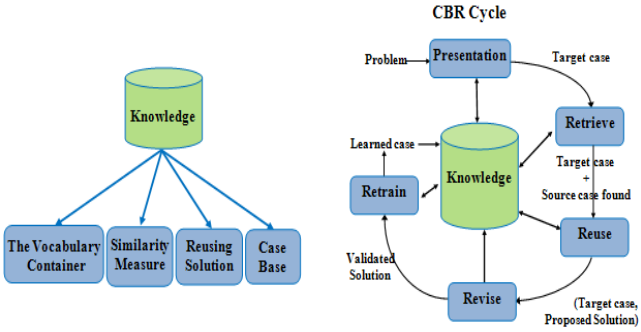


Fig. 1. The CBR components (Source [1], [12])

The systems based on the case-based reasoning can be classified into two categories [18]:

- Applications for static situation. For this type of system, the designer must have all the characteristics describing a case, in advance, in order to be able to realize its model. A data model of the field is thus refined through

an expertise in the field of application which can characterize a given situation. Thus, the cases are completely structured in this data model and often represented in a list (a: attributes, v: values). For example we have the system CHIEF [13] case base planner that builds new plans out of its memory of old ones. We do not exploit this type of CBR to develop our system, because in the approach for static situation, a problem must be completely described before starting the first step. Nevertheless in our situation, the learner traces (target case) evolve dynamically over time, so we must treat a dynamic situation with some particular features.

- Applications for dynamic situation. They differ when we compare them to static cases by the fact that they deal with temporal target cases (the situation), by looking for similar cases (better cases) based on a resemblance between histories (for more details on the subject, the reader may refer to [2], [18])). Several works relate to dynamic case based reasoning such as REBECAS [18] prediction of processes from observed behaviours, application to wildfire and SAPED [2].

B. Multi-Agents Case-Based Reasoning

The Multi-Agents System based on case based reasoning are used in many applications areas [25]. We can distinguish two types of applications (Table I):

- The Multi-Agents System in which each agent uses the case based reasoning internally for their own needs (level agent case based reasoning): This type is the first model that was applied in Multi-Agents CBR Systems. For this type of system, each agent is able to find similar cases to the target case in their own case base, also able to accomplish all steps of CBR cycle. For example we have the system ProCLAIM [29], MCBR [17] for distributed systems and CBR-TEAM [26] approach that uses a set of heterogeneous cooperative agents in a parametric design task (steam-condenser component design).
- The Multi-Agents System whose approach is a case based reasoning (level Multi-Agents Case Based Reasoning) : For this types of applications, the Multi-Agents Case Based Reasoning System distribute the some/all steps of the CBR cycle (Representation, Retrieve, Reuse, Revise, Retain) among several agents. The second category might be better than the first. Indeed the individual agents experience may be limited, therefore their Knowledge and prediction, so the agents are able to cooperate with other agents for a better prediction of the situation and they can benefit from the other agents capabilities. For example we have CCBR [21], RoBoCats [20] and S-MAS [24].

To our knowledge, no dynamic CBR cycle reasoning system exists.

We propose a system called Incremental Dynamic Case Based Reasoning-Multi-Agents System (IDCBR-MAS), able to find similar cases to the target case in their own case base. Our system is founded on 1) a dynamic cycle of case-based reasoning, and 2) a dynamic retrieving method based on a complementary similarity measure, named Inverse Longest Common Sub-Sequence (ILCSS) (for more details on the subject, the reader may refer to [10, 34]).

IV. Incremental Dynamic Case-Based Reasoning founded on Multi-Agents System

A. General architecture of our approach IDCBR-MAS

Our problem is similar to the CBR for dynamic situation. Indeed, the traces left by the learner during the learning session evolve dynamically over time; the case-based reasoning must take into account this evolution in an incremental way. In other words, we do not consider each evolution of the traces as a new target. The intelligent system (IDCBR-MAS) which we propose offer important features:

- It is dynamic. Indeed we must continually acquire new knowledge to better reproduce human behaviour in each situation.
- It is incremental; this is its major feature because the trace evolves in a dynamic way for the same target case.

The main benefits of our approach are the distributed capabilities of the Multi-Agents System and the self-adaption ability to the changes that occur in each situation. The system that we propose consists of the three layers components (as indicated in Figure Fig. 2:

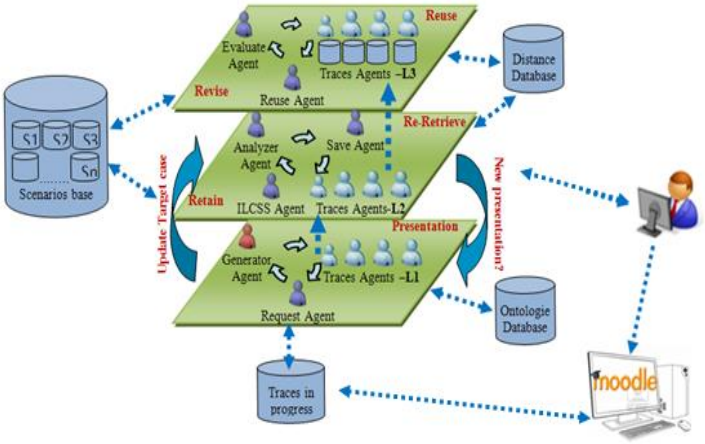


Fig. 2. General architecture of our approach

1) Presentation layer: The role of agents of this layer is to manage information arrived from the environment (the learner traces). This information feed the representation layer. The goal of this layer is to be both, a picture of the current situation being analyzed and to represent its dynamic evolution over time. The presentation layer contains the following agents:

- Request Agent: The role of this agent is to establish the link between the system and the environment. They feed the system with information from Distributed Information Systems (file traces). Also the goal of this agent is to check if there is any change in the traces file.
- Generator Agent: the role of this agent is to create and/or update the Traces Agents-L1: The Request Agent transmits the data received from environment to the Generator Agent. Two cases of figure are presented: if the Traces Agents-L1 (i) related to the learner i exist, then this last will be updated, else the Generator Agent creates a new Traces Agents-L1 (i).

- Traces Agents-L1: For each Lerner i we have a represented Trace Agent-L1(i). These agents will encapsulate the original traces of learners.

2) Interpretation Layer and storage: A set of agents allows the comparison between the current situation and past situations stored in the memory (scenarios). The Interpretation Layer contains the following agents:

- Traces Agents- L2: These agents contain the same information and data that have in the Trace Agents-L1 of the first layer. They differ by an abstraction of the data, originally described and managed by the Trace Agents-L1, that make it comparable to the past experiences stored in the memory.
- ILCSS Agent: The role of this agent is to evaluate in a continuously way the similarity between the current situation and past experiences based on the similarity measure ILCSS. The retrieve step of our system is based on this agent. The ILCSS Agent save the distances between the current situation and past experiences in Distance Table. It is responsible for reviewing these distances every time whenever necessary.
- Analyzer Agent: The goal of this agent is to check in a dynamic way if there is any change or update in Trace Agent -L2 (with the arrival of new information and data from the environment), then the Analyzer Agent asks ILCSS Agent to update Distance Table each time they have a change in the Trace Agent -L2, if not they asks the Request Agent if there is any change in traces file.

3) Prediction and Decision Layer: The role of agents of this layer is to predict the current situation by reusing past experiences selected by second layer. The choice of similar past experiences is evaluated by this layer, so one of these scenarios will be proposed to the learner. The layer contains the following agents:

- Traces Agents-L3: At this stage of reasoning the system adds a pointer to each agent the Traces Agents-L2. So the Traces Agents-L3 is identical to Traces Agents-L2 with a small difference, in fact for each Traces Agents-L2 we associate a list of similar scenarios through a pointer to the list of similar past experiences. The advantage of a pointer is that the list is not exhaustive and it changes dynamically over time following the change of the learner traces.
- Reuse Agent: The role of this agent is to predict future events of the situation by reusing the past experiences to the current situation.
- Evaluate Agent: The role of this agent is to evaluate the solution proposed by the Reuse Agent and to ensure that the similarity between the current situation and scenarios chosen by the Prediction layer is sufficient.
- Human Tutor or Human Agent: The human tutor is solicited if the system detects a learning situation requiring his intervention (failure to find one or more similar scenarios to the current situation).

B. From static to dynamic CBR cycle

We modify the CBR cycle in order to be able to handle dynamic situations and therefore we propose changes in the order of steps and a large change in the content of the steps of this cycle. In our approach the evaluation of the similarity between the current situation and similar past situations is a process continues. The retrieve step of the CBR cycle (as indicated in figure Fig.3) must take into account the change in the current situation in a dynamic way (in real-time). Our system will be able to repeat the retrieve step following the change of the current situation or whenever necessary.

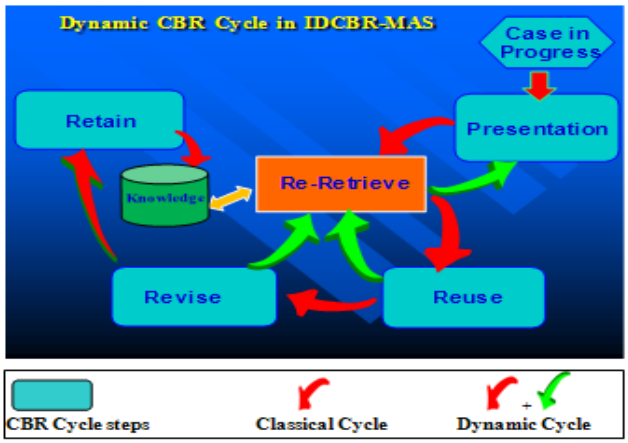


Fig. 3. Dynamic CBR Cycle in IDCBR-MAS

In addition, in our system the sequence of steps of the CBR cycle isn't important: in fact our system can stop each step in the CBR cycle and return to a previous step following the change of the current situation, and the order presentation – retrieve – reuse – revise – retain is not static or fixed, it can change and some steps can be re-run each time until the change in the situation.

Our agents are equipped with learning, communication and intelligence skills. They are able to stop the execution of the CBR cycle at a given step and time. They are able to re-run the different steps later following a change in the target situation. The highlight of our approach is that rerunning the retrieval step based our new dynamic similarity measure ILCSS. In each step CBR cycle of our approach we takes into account the previous results i.e. in time t_{i+1} we use the results in t_i. Therefore our CBR cycle takes into account the change of situation in a dynamic and incremental way.

1) *Retrieval steps :* Retrieval of previous case is one important step within the CBR paradigm. The success of retrieval step will depend on three factors: the case representation, case memory and similarity measure used to retrieve sources cases that are similar to the target case. There are two ways research for the sources case in dynamic situations:

- Research by evaluating similarity between the current situation and the already solved problems in a single dimension [18]. Several systems have been used this approach such as REBECAS [18] and SAPED [2].
- Research by evaluating similarity between the current problem and the already solved problems in a multiple dimension [2]. The multidimensional research, it is realized in a single step by taking into account all the parameters describing the current problem at the same

time. The multidimensional research is also used in several systems, such as CASEP2 [33].

2) State of the Art on Similarity Measures: Search for similar sources cases are based on the similarity measure. In this part, we present the principles similarity measures often used in case based reasoning, for more details on the subject, the reader my refer to [2] and to our articles [10], [34].

Biological Sequences Alignment: Dynamic Programming, is an important tool, which has been used for many applications in biology. It is a way of arranging the sequences of DNA, or protein to identify regions of similarity that may be a consequence of structural or functional relationships between the sequences. They are also used in different fields, such as natural language or data mining.

Minkowski distance: The Minkowski distance is a metric on Euclidean space which can be considered as a generalization of both the Euclidean distance.

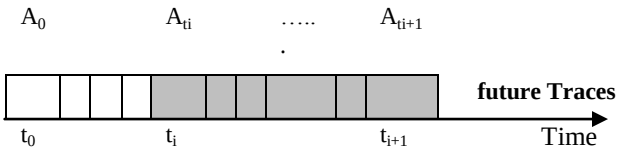
Longest Common Sub-Sequence (LCSS): the goal is to find the longest subsequence common in two or more sequences [31]. The LCSS is usually defined as: Given two sequences, find the longest subsequence present in both of them. A subsequence is a sequence that appears in the same order, but not necessarily contiguous. The main goal is to count the number of pairs of points considered similar when browsing the two compared sequences.

There are other similarity measures such as Dynamic Time Warping (DTW): The DTW algorithm is able to find the optimal alignment between two sequences. It is often used in speech recognition to determine if two waveforms represent the same spoken phrase. In addition to speech recognition, dynamic time warping has been successfully used in many other fields [2], such as robotics, data mining, and medicine.

TABLE I.
COMPARISON OF VARIOUS SIMILARITY MEASURES [2]

	Type	**Dimension**	**Length**
Biological Sequences Alignment	Symbolic	One-dimensional	Different
DTW	Digital	One-dimensional	Different
LCSS	Heterogeneous	Multidimensional	Different
Minkowski distance	Digital	One-dimensional	Same Length

3) Inverse Longest Common Sub-Sequence: The main goal of the retrieval phase in our system is to predict the behavior of the learner, by the reconciliation between the target trace and past traces or scenarios. The success of a case-based reasoning system depends primarily on the performance of the retrieval



step used and, more particularly, on similarity measure used to retrieve sources cases (scenarios) that are similar to the situation (traces in progress). Several research works have been focused on the similarity measure. Furthermore, these methods are not well suited when we compare two dynamic and heterogeneous sequences. In order to deal with this issue, we propose a complementary similarity measure entitled Inverse Longest Common Sub-Sequence an extension of the Longest Common Sub-Sequence measure.

In our system IDCBR-MAS the target case or target trace can be represented as a various actions of the learner (learner traces). It can be represented also as a collection of semantic features SF= (object, (qualification, value) +), we note object=O, qualification=Q and value=V, SF= (O,(Q,V)+), so the learner traces at time i, can be defined by the formula:

$$LT_i = \bigcup_{1 \le k \le i} SF_k$$

Where SFk = (O_k, ($Q_{k,1}$, V_1),…, ($Q_{k,d}$, V_d)) is a sequence of d+1 dimension. Finally the learner traces at time i+1 is a multidimensional sequence.

Let A and B two Traces with size n x d and m x d respectively, where:

A = (($O_{A,1}$, ($Q_{A,1,1}$, $V_{A,1,1}$),…, ($Q_{A,1,d}$, $V_{A,1,d}$), ($O_{A,2}$, ($Q_{A,2,1}$, $V_{A,2,1}$),…,($Q_{A,2,d}$, $V_{A,2,d}$)),….., ($O_{A,n}$, ($Q_{A,n,1}$, $V_{A,n,1}$),…, ($Q_{A,n,d}$, $V_{A,n,d}$)))

And

B = (($O_{B,1}$, ($Q_{B,1,1}$, $V_{B,11}$),…,($Q_{B,1,d}$, $V_{B,1,d}$), ($O_{B,2}$, ($Q_{B,2,1}$, $V_{B,21}$), …, ($Q_{B,2,d}$, $V_{B,2,d}$)),….., ($O_{B,m}$, ($Q_{B,m,1}$, $V_{B,m,1}$),…,($Q_{B,m,d}$, $V_{B,m,d}$))).

For a Trace A, let Tail(A) be the Trace:

Tail(A) = ($O_{A,2}$,($Q_{A,2,1}$,$V_{A,2,1}$),…, ($Q_{A,2,d}$, $V_{A,2,d}$)),….., ($O_{A,n}$, ($Q_{A,n,1}$,$V_{A,n,1}$),…, ($Q_{A,n,d}$, $V_{A,n,d}$))). Tail (A) it the trace A private their first vector.

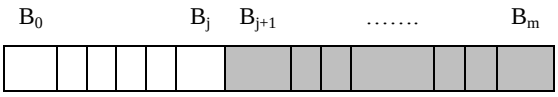
The goal is to count the number of pairs vectors considered similar when compared through the two traces. The similarity between two vectors (VA,i,1, VA,i,2, …, VA,i,d) from trace A, and (VB,j,1, VB,j,2, …, VB,j,d) from trace B it determined according to a threshold δ: if for each k =1,…,d $|V_{A,i,k} - V_{B,j,k}| \le \delta$. We also define an integer N, the parameter that will be able to control the temporal variance between two vectors of each of the traces in order to consider the two traces similar.

Let A and B two Traces, and given an integer N and a real number δ, we define the similarity measures between the two traces A and B, as follows recursive process: the process is initialized by comparing the two first vectors of traces (A, B). If any of the two traces is empty then the value of the similarity measure is equal to 0, and the process stops. Else if any of the two vectors traces are similar, then the similarity measure in this case is "1" more the similarity between the two traces deprived of their first vectors. Else the similarity is equal to the maximum of the similarity between a trace and the other private its first vector.

At the instant t=ti+1 the IDCBR-MAS system recovers the traces stored in the log file of server between the two instants ti and ti+1 and we have (A)t=ti+1 = Tail(A)t=ti+1 = A[ti,ti+1] see figure below.

Tail($A)_{t=ti+1}$ = $A_{[ti,ti+1]}$ = (($O_{A[ti+1,ti],1}$, ($Q_{A[ti+1,ti],1,1}$, $V_{A[ti+1,ti],1,1}$),…, ($Q_{A[ti+1,ti],1,d}$, $V_{A[ti+1,ti],1,d}$), ….., ($O_{A[ti+1,ti],n'}$, ($Q_{A[ti+1,ti],n,1}$, $V_{A[ti+1,ti],n',1}$),…, ($Q_{A[ti+1,ti],n',d}$, $V_{A[ti+1,ti],n',d}$))) avec 1<= i' <= n'

At the instant t=ti+1 it only remains the block B[j+1,m] of the B traces (block of the trace B have not yet been compared with the target trace), where Bj describe, the last common element between the two traces (A)t=ti and B at the instant t=ti.



$(B)_{t=ti+1} = Tail(B)_{t=ti+1} = B_{[j+1,m]}$
$Tail(B)_{t=ti+1} = B_{[j+1,m]} = ((O_{B[j+1,m],1}, (Q_{B[j+1,m],1,1}, V_{B[j+1,m],11}),\ldots,(Q_{B[j+1,m],1,d}, V_{B[j+1,m],1,d}), \ldots.., (O_{B[j+1,m],m'}, (Q_{B[j+1,m],m',1}, V_{B[j+1,m],m',1}),\ldots,(Q_{B[j+1,m],m',d}, V_{B[j+1,m],m',d})))$ with $1<= j' <= m'$ ($m'= m-j$)

The measure between the target traces A and the source trace B at time $t = t_{i+1}$ will depend on the results instantly $t = ti$ through the following recursive formula:

$$(ILCSS_{\delta,N,ti+1}(A,B))_{t=ti+1} = (ILCSS_{\delta,N,ti}(A,B))_{t=ti} + ILCSS_{\delta,N,ti+1}(A_{[ti,ti+1]}, B_{[j+1,m]})$$

With

$$ILCSS_{\delta,N,ti+1}(A_{[ti,ti+1]}, B_{[j+1,m]}) = \begin{cases} 0 \text{ if } A_{[ti,ti+1]} \text{ or } B_{[j+1,m]} \text{ is empty.} \\ 1 + ILCSS_{\delta,N,ti+1}(A_{[ti,ti+1]}, B_{[j+1,m]}) \text{ if } O_{A_{[ti,ti+1]}i'} = O_{B_{[j+1,m]}j'}, Q_{A_{[ti,ti+1]}i',k'} = Q_{B_{[j+1,m]}j',k'} \\ \left|V_{A_{[ti,ti+1]}i,k'} - V_{B_{[j+1,m]}j',k'}\right| \le \delta, \text{with } 1 \le i' \le n',\ 1 \le j' \le m', |i' - j'| \le N \text{ et } k' = 1 \ldots d. \\ \max(ILCSS_{\delta,N,ti+1}(Tail(A_{[ti,ti+1]}), B_{[j+1,m]}), ILCSS_{\delta,N,ti+1}(A_{[ti,ti+1]}, Tail(B_{[j+1,m]}))) \text{ else.} \end{cases}$$

The index of the last common element between the two traces $\mathbf{(A)_{t=ti+1}}$ and B at time $t = t_{i+1}$ is obtained using the following iterative formula.

J is initialized to 0;

And

$$\{j = j' \text{ if } O_{A_{[ti,ti+1]}i'} = O_{B_{[j+1,m]}j'}, Q_{A_{[ti,ti+1]}i',k} = Q_{B_{[j+1,m]}j',k'}, \left|V_{A_{[ti,ti+1]}i',k'} - V_{B_{[j+1,m]}j',k'}\right| < \delta \text{ and } j' > j$$

With $1 \le i' \le n'$, $1 \le j' \le m'$, $|i'- j'| \le N$ and $k'=1 \ldots d$

We define the distance between the two Traces A and B as follow:

$$D_{\delta,N,ti}(A,B) = 1 - \frac{ILCSS_{\delta,N,ti}(A,B)}{\min(n,m)}$$

Where $D_{\delta,N,ti}(A,B)$ verify the proprieties of the distance such as:

$$\begin{cases} D_{\delta,N,ti}(A,B) \ge 0 \\ D_{\delta,N,ti}(A,B) = 0 \text{ equal } A \approx B \\ D_{\delta,N,ti}(A,B) = D_{\delta,N,ti}(B,A) \end{cases}$$

$A \approx B$: A and B are two similar traces.

4) *Learner's Traces and case structure*: Based on the general definition of a trace given in [19], "a trace is a thing or a succession of things left by an unspecified action and relative to a being or an object; a succession of prints or marks which the passage of a being or an object leaves; it is what one recognizes that something existed; what remains of a past thing". In ITS literature, a digital trace is an observed collection, all structured information resulting from an interaction observation temporally located [22].

In our context, a digital trace is resulting from an activity observation representing a process interactional signature. Indeed, it is composed of the objects which are respectively located the ones compared to the others when observed and registered on a support. That means that a trace is explicitly composed of the structured objects and registered compared to a time representation of the activity traces. The structuring can be sequentially explicit (each trace observed is followed and/or preceded by another) or can also come from the temporal characteristic of the objects traces [19]. Indeed, the structuring depends on the type of the time representation and the time of the activity traces. We can distinguish two types of representations:

- They can be a temporal interval determined by two dates, (start and end of observation). In this case, the observed traces may be associated with an instant or an interval of time. Then we will be able to take into account chronological relationships between observations';
- They can be a sequence of unspecified elements (for example a sub-part of the whole of the set of integers). In this case, we will focus on the succession or the precedence of the trace observed (there is no chronological time).

In the current uses of the traces for the CEHL, collected situations are contrasted: from "we take what we have in well specified formats, what is called the logs" to "we scrupulously instrument the environment to recover the observed controlled and useful for different actors (learner and tutor). The first step consists of modeling the raw data contained in the log file. It is necessary to be able to collect the traces files containing at least, the following elements: time for the start date of the action, codes action which consists in codifying the learner's actions and learner concerned.

In our system IDCBR-MAS a target case or target trace is represented by a trace learner in progress when interacting with the Moodle platform. The cases sources are previous traces learners that are stored in database. The cases sources are traces left by the learners which followed the same training on the Moodle platform. The following figure shows the target trace structure:

Fig. 4. Learner traces when interacting with the platform Moodle

The trace can be written as follows: ((OA,1, (Q A,1,1, VA,1,1),…, (QA,1,d, VA,1,d), (OA,2, (QA,2,1,

VA,2,1),…,(QA,2,d, VA,2,d)),….., (OA,n, (QA,n,1, VA,n,1),…, (QA,n,d, VA,n,d)))

We developed a module in Moodle platform that can be the interface between the Moodle server and our IDCBR-MAS system. This module includes an xml file, which contains traces left by all learners in the Moodle log file and also contains the datalib file. The module uses the same Moodle database. The datalib file of Moodle platform has been modified in order to be able to record and save all traces of learners connected to the Moodle platform. The following figure shows the datalib file.

```
function insert_xml($log)
{
global $DB, $CFG;
if($log['userid']!=1 && $log['userid']!=0 && $log['userid']!=2)
{
$nomfichier=$CFG->dirroot."\xml\users.xml";
if(file_exists($nomfichier))
{
$user=$log['userid'];
  $nomfich=$CFG->dirroot."\xml\user".$log['userid']."_xml";
  $sql_request = "SELECT u.* $ctxselect FROM {user} u WHERE u.id=".$log['userid'];
$result = $DB->get_records_sql($sql_request, null, null);
if($result != null)
foreach($result as $us)
 $user=$us->firstname." ".$us->lastname;

 $sql_request = "SELECT MAX(l.id) as max $ctxselect FROM {log} l WHERE l.userid=".$l
$result3 = $DB->get_records_sql($sql_request, null, null);
if($result3 != null)
{

foreach($result3 as $l)
 $id=$l->max;
 }
 $sql_request = "SELECT c.* $ctxselect FROM {course} c WHERE c.id=".$log['course'];
$result2 = $DB->get_records_sql($sql_request, null, null);
if($result2 != null)
foreach($result2 as $c)
 $course=$c->fullname;
```

Fig. 5. datalib file, version IDCBR-MAS

In the next section, we present our Model based on AUML methodology. AUML or agent UML is a support notation for agent-oriented Multi-agents systems development. It consists in using the UML modeling language and extending it in order to represent agents, their behavior and interactions among them.

V. IDCBR-MAS SYSTEM MODELING

Our system IDCBR-MAS is composed of multiple interacting intelligent agents; it supports the specification, analysis, design and validation of our systems. We present the sequence diagram of the various interactions carried out between the various actors of the platform.

A. Presentation of the situation:

The presentation of the situation (learner's traces) by the platform is a task managed by several agents of the presentation layer of our system IDCBR-MAS. These agents are responsible for the update of the traces. The following sequence diagram illustrates the process of the situation presentation of the learner's traces.

Firstly the Request Agent addresses a request to server in order to retrieves the learner's traces left by the learner during the learning session and sending it to the Generator Agent, this last created/update the Traces Agents-L1: Two cases of figure are presented during the checking, if the Traces Agents-L1 (i) related to the learner i exists then the Traces Agents-L1 (i) will be updated, else the Generator Agent create a new Traces Agents-L1 (i) able to represent the learner i. the process will be re-run each time there is a change in the learner's traces.

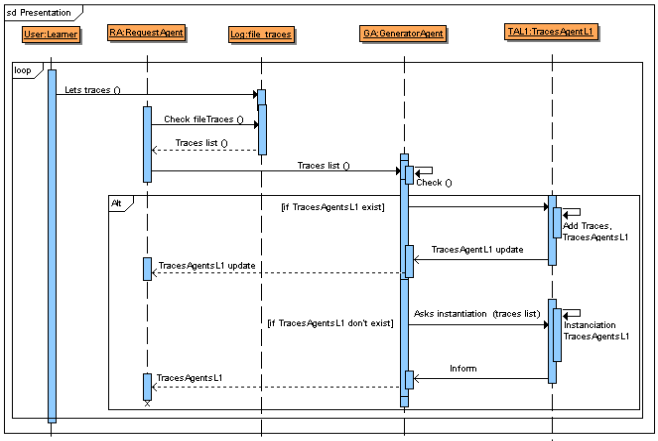


Fig. 6. The sequence diagram of the case presentation in IDCBR-MAS.

B. Interpretation of the situation

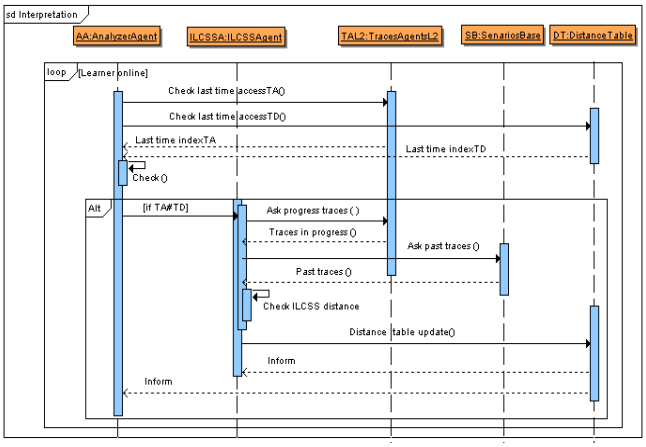


Fig. 7. The sequence diagram of the Interpretation case in IDCBR-MAS

Firstly the Analyzer Agent (AA) addresses a request to the Traces Agents-L2 and to the Distance Table in order to retrieves two chronological dates TA: the last update date in the traces file and DT: the last update date of the Distance Table. The Analyzer Agent check If TA= DT. If the two dates are not equal then the Analyzer Agent ask the ILCSS Agent to update the distance table which contains the distance between the current situation Traces Agents-L2 and the scenario stored in memory. This is based on the similarity measures ILCSS. The agent also asks periodically the Request Agent if there is any change in the learner's traces, whether the process will be re-executed.

First of all the Reuse Agent ask the Traces Agents-L3 to retrieve the current traces with the associated scenarios (the associated scenarios to the current traces are the scenarios that are very similar at learner's traces or target, based on the similarity measures ILCSS). Then the Evaluate Agent checks the Distance Table. If necessary the Reuse Agent asks the ILCSS Agent asks to check and update all distances between the current situation and scenarios stored in memory.

C. Prediction of the situation

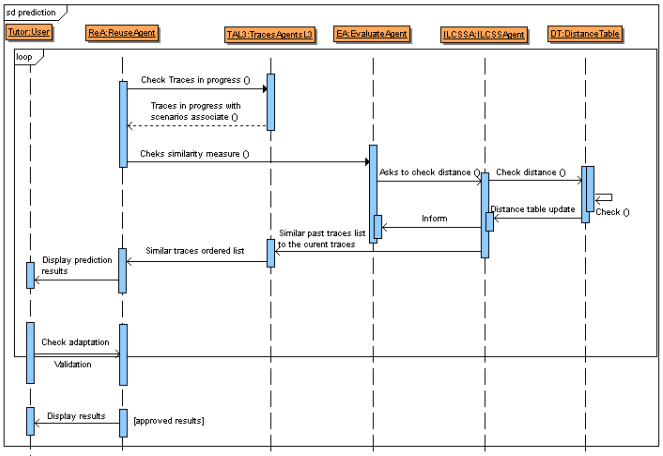


Fig. 8. The sequence diagram of the case prediction in IDCBR-MAS system

VI. IDCBR-MAS SYSTEM DEVELOPING

We developed our framework IDCBR-MAS based the JADE Agent Platform (Java Agent DEvelopment Framework). For the development of interfaces, we chose the languages Java, PHP and the tools EasyPHP, Apache, MySQL, phpMyAdmin.

A. Inter-Agent Communication in IDCBR-MAS

In order to supervise and control the communication and the IDCBR-MAS agents' behavior, we use Remote Monitoring Agent (RMA) of JADE platform. RMA is a graphical console for platform management and control. The RMA console is able to start other JADE tools. It a monitoring and debugging tool, made of a graphical user. It is able to displays the flow of interactions between agents in our IDCBR-MAS platform. The following figure shows the interactions between IDCBR-MAS agents.

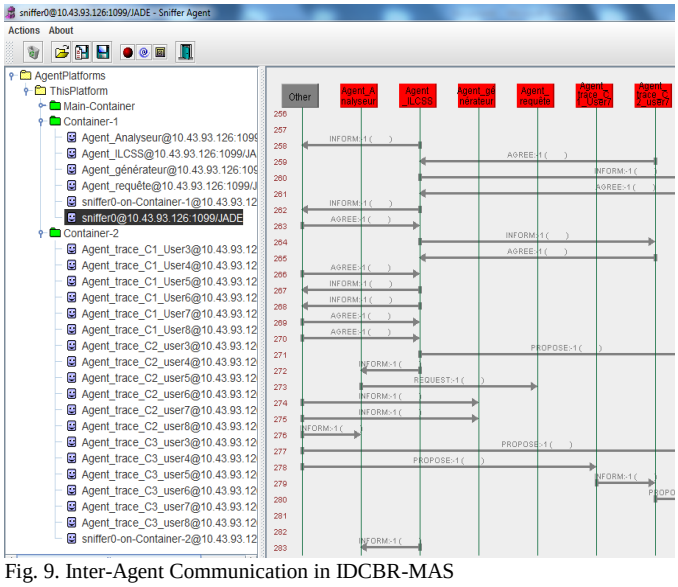


Fig. 9. Inter-Agent Communication in IDCBR-MAS

B. Monitoring the activity and communication between agents in IDCBR-MAS

This tool makes it possible to monitor the life cycle and communication of our agents: Sending and Receiving Messages by these agents. It is also possible to display the list of all the messages sent or received, completed with timestamp information in order to allow agent conversation recording and rehearsal. For example, the following figure shows the state as well as the transmitted/received messages for the ILCSS Agent of our IDCBR-MAS framework.

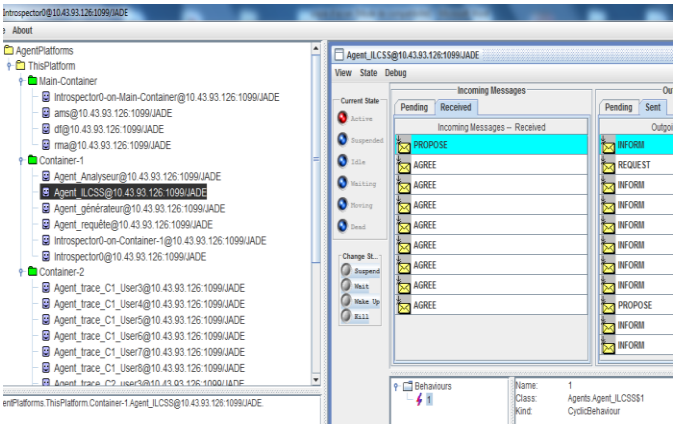


Fig. 10. Transmitted/received messages for ILCSS Agent

C. Distance between the target and previous traces

After the registration of a learner on the IDCBR-MAS platform, the learner will be able to run Moodle from our platform and subsequently launch a learning session. The tutor follows progressively the training of the learner.

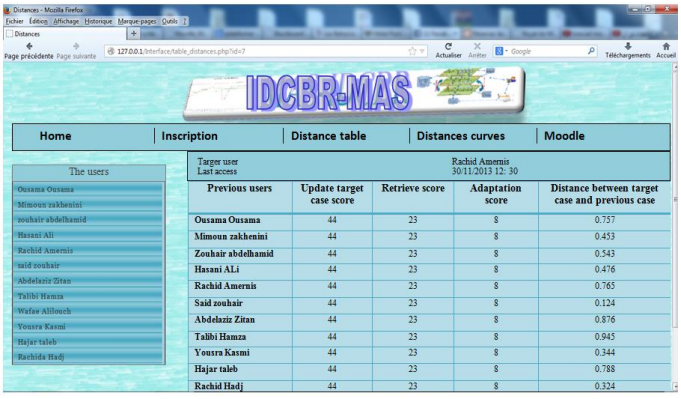


Fig. 11. Distance between the target and previous traces

All interactions, actions and productions of the learner are recorded on the log file in the Moodle database. Our system retrieves these traces through agents' interfaces permanently, and then they will be treated by the platform. In the figure we have a target case (traces left by target learner) and we have previous traces (traces left by previous learners). The update target case score present the number of update in the target case; the retrieve score present the number of re-retrieve of the previous cases very similar to target case by the agents of IDCBR-MAS platform. The distance between the target trace and past traces are calculated by the ILCSS Agent. These distances will be used as a key element in predicting of the

situation achieved by the adaptation agent. The system proposes to the tutor a list of the similar traces to the target trace in order to choose the best similar traces.

D. Distances curves between the target and previous traces

The following figure displays the distances curves between the target and previous traces in order to shows the distance between them, these curves are generated in real times starting from the results of retrieval phase. These curves display also the history of these distances. For Tutor, the distances curves present very important information about the change of the distances database. The Tutor will be able to take her decision and to choose the trace most similar to the target trace.

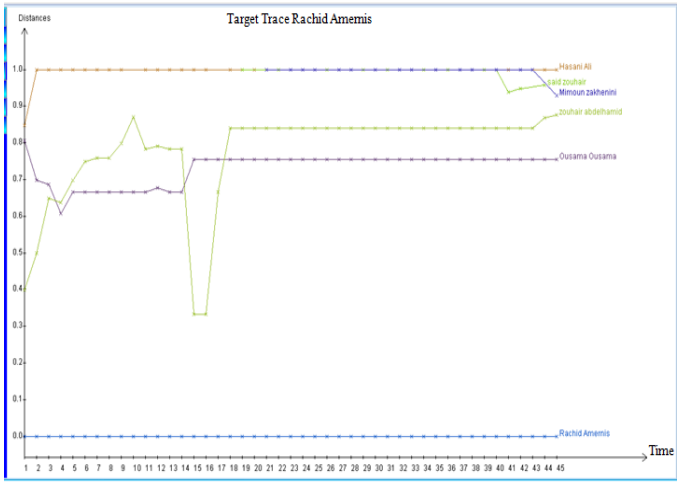


Fig. 12. Distances curves between the target traces and previous traces

VII. IDCBR-MAS & OTHER CBR SYSTEMS

Several researchers have focused on classical versus dynamic CBR architectures where target case are static versus dynamic, but all these systems have been used static CBR cycle. Consequently, the Incremental Dynamic CBR approach has been proposed as an appropriate alternative, which have demonstrated its efficacy. For example, In our approach the evaluation of the similarity between the target case and similar past cases is a process continues and the retrieve step of the CBR cycle take into account the change in the target case in real-time. Finally our system is founded on a dynamic retrieving method. The following table (Table II) shows a summary of the CBR systems.

TABLE II.
CBR SYSTEMS CLASSIFICATION

Target case	**CBR Cycle**	**Classical CBR Systems**	**CBR-MAS**	**CBR-Agent**
Static	Static	CHEF[13], CREEK [3], CASEY [15], RADIX [8]	CCBR[21], AMAL [27]	ProCLAIM [29]
Dynamic	Static	REBECAS [18], AuRA [16] , SAPED[2], CASEP2 [33], SBR[4]	CICLMAN [28], RoBoCats [20] , S-MAS [24]	MCBR [17], CBR-TEAM[26]
Dynamic	Dynamic			IDCBR-MAS

VIII. CONCLUSION AND FUTURE WORK

Our system allows connecting and comparing the current situation (target trace) to past situation (previous traces) that are stored in a database. The continuous analysis of information coming from the environment (learner's traces) makes it possible to suggest to various actors (learners and tutor) possible evolutions of the current situation.

The Multi-Agents architecture that we propose is based on three layers of agents with a pyramidal relation. The lower layer allows building a representation of the target case. The second layer implements a dynamic process: search for past situations similar to the current one. Finally, the prediction layer captures the responses sent by the second layer to transform them into actions proposed either by virtual tutor, or/and human tutor.

We have presented systems founded of Incremental and Dynamic Case Based Reasoning and we have also clarified that the CBR-based applications can be classified according to the study area: CBR for static situations and CBR for dynamic situations. In our situation, we have used a Dynamic system IDCBR-MAS, with a dynamic CBR cycle in order to push the limits of CBR cycle static. In fact, the current situation (target case) is a trace that evolves; the case based reasoning must take into account this evolution incrementally. In other words, it shouldn't consider each evolution of the trace as a new target case.

Our future work follows two different ways. First, we would like to use our framework in real experiment with e-learning platform of our university. Secondly, in the second part of our perspective, we will try to implement our approach in the field of Geographic Information Systems (GIS).

REFERENCES

[1] A. Aamodt and E. Plaza, "Case-Based Reasoning : Foundational Issues, Methodological Variations, and System Approaches", AICom-Artificial Intelligence Communications, IOS Press, Vol. 7: 1, pp. 39-59, 1994.

[2] A. Aich, Reconnaissance et prdiction de situations dynamiques : application l'assistance de personnes handicapes moteur, Thse de doctorat, Universit de Technologie de Troyes, 2007.

[3] A. Aamodt, Knowledge-Intensive Case-Based Reasoning and Sustained Learning. Proc. of the 9th European Conference on Artificial Intelligence, ECCBR04, Lecture Notes in Artificial Intelligence, Springer, 2004.

[4] C. Baptiste, Vers la gestion de scnarios pour la reconnaissance et l´anticipation de situations dynamiques , Universit de Technologie de Troyes, France, 2011.

[5] P. Brusilovski, "Distributed Intelligent Tutoring on the Web". 8th World Conference of Artificial Intelligence in Education. IOS Press. pp.482-489, 1997.

[6] A. Cordier, B. Mascret, A. Mille, Dynamic Case Based Reasoning for Contextual Reuse of Experience, Case-Based Reasoning Workshop, ICCBR 2010, Cindy Marling ed. Alessan-dria, Italy. pp. 69-78. 2010.

[7] G. Constantino, D. Suthers and J-I. Icaza, Designing and Evaluating a Collaboration Coach: Knowledge and Reasoning, Proceedings of the Artificial Intelligence in Education, AI-ED, J.D. Moore et al (Eds). The Netherlands: IOS Press. pp. 176-187, 2001.

[8] F. Corvaisier, Mille A., Pinon J.M., Radix 2, assistance la recherche d'information documentaire sur le web, In IC'98, Ingnierie des Connaissances, Pont--Mousson, France, INRIALORIA, Nancy, 1998.

[9] M. Ennaji, H. Boukachour, P. Grav, Une architecture Multi-Agent pour la pedagogie de la formation distance, In: MOSIM 2006, Rabat, Maroc.

[10] E. M. En-Naimi, A. Zouhair, B. Amami, H. Boukachour, P. Person and C. Bertelle, "Intelligent Tutoring Systems Based on the Multi-Agent Systems (ITS-MAS): The Dynamic and Incremental Case Based Reasoning (DICBR) Paradigm and the Inverse Longest Common Sub-Sequence (ILCSS) in the CEHL", The IJCSI, Vol. 9, Issue 6, No 3, pp. 112-121, November 2012.

[11] C. Frasson, L. Martin, G. Gouarderes and E. Aïmeur, "A distance learning Architecture Based on Networked Cognitive Agents", 4ème International Conference on Intelligent Tutoring Systems- ITS'98, San Antonio, USA, 1998, Lecture Notes in Computer Sciences 1452, Apringer Verlag, pp. 142-151.

[12] B. Fuchs, J. Lieber, A. Mille and A. Napoli, "Une première formalisation de la phase d'élaboration du raisonnement à partir de cas", Actes du 14ième atelier du raisonnement à partir de cas, Besançon, mars, 2006. [see url : http://www.lab.cnrs.fr/RaPC2006/programme.html].

[13] K.J. Hammond, CHEF: a model of case-based planning, Proc. Of AAAI86, Morgan Kaufman, pp. 267-271, 1986.

[14] J. Kolodner, , Case-Based Reasoning, Morgan Kaufmann, San Mateo, UCA, 1993.

[15] P. Koton, Using experience in learning and problem solving, PhD Thesis, Laboratory of Computer Science, Massachusetts Instite of Technology, Cambridge Massachusetts, USA 1988.

[16] M. Likhachev, M. Kaess, and R.C. Arkin, Learning behavioral parameterization using spatiotemporal case-based reasoning, The IEEE international conference on robotics and automation (ICRA02), Washington, D.C, 2002.

[17] D. Leake and R. Sooriamurthi, When two case bases are better than one: Exploiting multiple case bases, In ICCBR, pp. 321335, 2001.

[18] S. Loriette-Rougegrez S., "Raisonnement à partir de cas pour les évolutions spatiotemporelles de processus", Revue internationale de géomatique , vol. 8, nº 1-2, 1998, p. 207-227.

[19] J-C. Marty and A. Mille, , Analyse de traces et personnalisations des environnements informatiques pour lapprentissage humain, Edition Lavoisier, 2009.

[20] C. Marling, M. Tomko, M. Gillen, D. Alexander, D. Chelberg, Case-based reasoning for planning and world modeling in the RoboCup small size league, in: Workshop on Issues in Designing Physical Agents for Dynamic Real-Time Environments: World Modeling, Planning, Learning, and Communicating (IJCAI), 2003.

[21] L. McGinty and B. smyth, Collaborative case-based reasoning: Applications in personalized route planning. In ICCBR, pp. 362–376, 2001.

[22] A. Mille, , From case-based reasoning to traces-based reasoning, Annual Reviews in Control30(2):223-232, ELSEVIER, ISSN 1367-5788. 2006.

[23] J. Piaget, Psychologie et pédagogie, Paris: Denol-Gonthier, 1969.

[24] I. Pinzón Cristian, B. Javier, F. Juan, M. Juan, B. Corchado, S-MAS: An adaptive hierarchical distributed multi-agent architecture for blocking malicious SOAP messages within Web Services environments, Expert Systems with Applications Volume 38, Issue 5, May 2011, pp. 5486-5499.

[25] E. Plaza and L. Mcginty, , Distributed case-based reasoning, The Knowledge Engineering, Review, Vol. 00:0, 14.c 2005, Cambridge University Press, 2005.

[26] M. Prassad, V. Lesser, and S. Lander, Retrieval and reasoning in distributed case bases. Technical report, UMass Computer Science Department, 1995.

[27] S. Ontañón and E. Plaza, Learning and Joint Deliberation through Argumentation in Multi-Agents Systems, in AAMAS'07 Proceedings of the 6th international joint conference on Autonomous agents and Multi-Agents systems, Honolulu, HI, USA, May 14-18, 2007.

[28] L. k. Soh and J. Luo, Combining individual and cooperative learning for multiagent negotiations, In Int. Conf. Autonomous Agents and Multiagent Systems AAMAS03, 2003.

[29] P. Tolchinsky, S. Modgil, U. Cortes, and M. Sanchez-marre, Cbr and argument schemes for collaborative decision making. In Conference on computational models of argument, COMMA-06 (Vol. 144, pp. 71–82), 2006.

[30] L-S., Vygotski, Mind in society: the development of higher psychological processes, Harvard University Press, Cambridge, MA, 1978.

[31] M. Vlachos, K. Kollios, and G. Gunopulos, Discovery similar multidimensional trajectories. The 18th International Conference on Data Engineering (ICDE02), pages 673–684, San Jose, CA, 2002.

[32] C. Webber and S. Pesty, , Emergence de diagnostic par formation de coalitions - Application au diagnostic des conceptions d'un apprenant, Journes Francophones pour l'Intelligence Artificielle Distribue et les Systmes Multi-Agents, Hermes, Lille, pp.45-57, 2002.

[33] F. Zehraoui, Systèmes d'apprentissage connexionnistes et raisonnement partir de cas pour la classification et le classement de séquences, Thèse de doctorat, Université Paris13, 2004.

[34] ZOUHAIR Abdelhamid, El Mokhtar EN-NAIMI, Benaissa AMAMI, Hadhoum BOUKACHOUR, Patrick PERSON, Cyrille BERTELLE, Intelligent Tutoring Systems Founded on The Multi-Agent Incremental Dynamic Case Based Reasoning, IEEE CIST'12, Fez, Morocco, 2012.

Abdelhamid ZOUHAIR is a PhD student in Cotutelle between the Laboratory LIST, FST of Tangier, Morocco and the Laboratory LITIS, the University of Le Havre, France, since September 2009.

El Mokhtar EN-NAIMI is a Professor in Faculty of Sciences and Technologies of Tangier, Department of Computer Science. He is a member of the Laboratory LIST (Laboratoire d'Informatique, Systèmes et Télécommunications), the University of Abdelmalek Essaâdi, FST of Tangier, Morocco. In addition, he is an associate member of the ISCN - Institute of Complex Systems in Normandy, the University of Le Havre, France.

Benaissa AMAMI is a Professor in Faculty of Sciences and Technologies of Tangier. He was an Ex-Director of the Laboratory LIST (Laboratoire d'Informatique, Systèmes et Télécommunications), the University of Abdelmalek Essaâdi, FST of Tangier, Morocco.

Hadhoum BOUKACHOUR and **Patrick PERSON** are Professors in the University of Le Havre, France. They are members in the Laboratory LITIS (Laboratoire d'Informatique, de Traitement de l'Information et Système), The University of le Havre, France.

Cyrille BERTELLE: Professor in Computer Science, Complex Systems Modelling and Simulation. Member of the Research Laboratory LITIS at the University of Le Havre, Normandy, France (RI2C Team). Co-founder of ISCN - Institute of Complex Systems in Normandy. Vice-President of Research and Development at University of Le Havre, France

A Fault-Tolerant Mobile Computing Model Based On Scalable Replica

Meenakshi Sati[1],Vivek Vikash[2], Vishwanath Bijalwan[3] , Pinki Kumari[4],Manish Raj[5], Meenu Balodhi[6], Priya Gairola[7], Vijay Bhaskar Semwal[8]

[1 ,2,3] *NITTR Chandigrah*
[4] *Bansatali Vidya Peeth rajstan*
[5]*IIIT Allahabad*
[6,7] *Uttarakhand Techincal Univesity*
[8] *Siemens Information System, Banglore*

Abstract **— The most frequent challenge faced by mobile user is stay connected with online data, while disconnected or poorly connected store the replica of critical data. Nomadic users require replication to store copies of critical data on their mobile machines. Existing replication services do not provide all classes of mobile users with the capabilities they require, which include: the ability for direct synchronization between any two replicas, support for large numbers of replicas, and detailed control over what files reside on their local (mobile) replica. Existing peer-to-peer solutions would enable direct communication, but suffers from dramatic scaling problems in the number of replicas, limiting the number of overall users and impacting performance. Roam is a replication system designed to satisfy the requirements of the mobile user. Roam is based on the Ward Model, replication architecture for mobile environments. Using the Ward Model and new distributed algorithms, Roam provides a scalable replication solution for the mobile user. We describe the motivation, design, and implementation of Roam and report its performance. Replication is extremely important in mobile environments because nomadic users require local copies of important data.**

Keywords **— Fault-tolerant, Mobile Computing, Roam, Data Replictaion.**

I. INTRODUCTION

REPLICATION in mobile environments requires fundamentally different solutions than those previously proposed, because nomadicity presents a fundamentally new and different computing paradigm. Before mobile computing was a feasible reality, software designers simply didn't design their systems to be mobile-enabled. Computers were largely stationary, so there was no need to consider what would happen if it were easy for people to move around geographically. Consider the case of mobile-IP. When the Internet Protocol (IP) was developed, mobility was not feasible, and therefore was not considered in the design. Now that mobility *is* a reality, many researchers are trying to fix or redesign IP to support real mobility. Like the mobile-IP case, the replication system software must also be redesigned to account for mobile computing. *Roam* is built using the Ward Model, a new replication architecture that has been designed especially with mobility in mind, though of course it applies equally well to stationary environments [1]. The Ward Model provides a new replication paradigm that is neither strictly peer nor client-server; rather, it is a hybrid model of the two that allows everyone to be effectively a peer while maintaining good scalability in the number of replicas. Additionally, *Roam* provides a number of new distributed algorithms. For instance, *Roam* contains new and improved distributed algorithms for garbage collection and for dynamic management of the version vector, the main data structure behind most optimistic replication systems.

In this paper we present an approach for data replication that considers the mobility of clients: The data sets which have to be replicated to mobile clients depend often on dynamic parameters like location of the client or time. Therefore, we introduce a generic model for specifying fragments of the global database with respect to such parameters. The increasing availability of mobile devices and wireless communication technologies opens new applications which can improve services by providing real-time access to remote information sources. Examples of such applications are not only stock tickers or messenger applications but also mobile hospital information systems and car navigation systems. However, due to the mobile and wireless nature of these devices, several limitations have to be taken into account in developing and running such mobile information systems:

(1) The structure of the network as well as the availability of mobile nodes are subject to continuous changes.

(2) The bandwidth of the communication channels is – compared to traditional wired networks – relatively low and communication is much more expensive.

(3) The resources of mobile devices, e.g. memory and computing power, are strictly limited. For a mobile information system, the first two properties prohibit data access by only remote querying.

Thus, the mobile device should keep a copy of the data that is the subject of the current operations. The third issue restricts the data managed locally on the mobile device to a subset of the overall (remote) database. This raises the need for *semantic replication*: data that is relevant in a certain

DOI: 10.9781/ijimai.2014.267

situation has to be transferred to the mobile device and in case of local updates the modifications are propagated back to the main database. A main issue in this context is the reduction of the replicated data in order to meet the requirements of minimal resource consumption (storage space, transfer time or volume). A possible application is an advanced car driver information system. In addition to street maps stored on a local disk the mobile system retrieves further information about local (i.e., relevant to the current location and time) traffic spots, accidents, detours etc. from a federated database integrating these information from various sources and displays it to the user or uses it for other operations, e.g. route planning. When this information becomes irrelevant, it can be removed from the local database. This scenario could be extended to support updates – a driver reports an accident or the system detects a traffic jam automatically – as well as other kinds of parameters, for example the kind of car (truck, automobile). In this paper we present a replication approach which is based on semantic extensions of the global view. Using this approach, location-based services and services requiring replication depending on dynamic features can be easily (on the global view level) implemented. It is embedded into the context of mobility. While looking at related work we cannot consider all requirements caused by the special characteristics of mobile computing. Therefore, we focus on work related to replication in mobile information systems.

Here we present a replication system designed to satisfy the requirements of the mobile user. Roam is based on the Ward Model, a replication architecture specifically aimed at mobile environments. Together with the Ward Model and new, distributed algorithms such as improved garbage collection techniques and version vector management, Roam provides a scalable replication solution for the mobile user. We describe not only the design and motivation of such a system, but its implementation and performance as well.

Roam addresses scalability with a three-pronged attack. The Ward Model addresses issues of replica management, consistency topologies, and update distribution. Dynamic algorithms and mechanisms handle the scalability of the versioning information, required for consistency maintenance. Finally, we address consistency itself with new algorithms and mobile-friendly semantics. In summary, Roam is a comprehensive replication system for mobile and non-mobile users alike. With it, users can truly compute while mobile, paving the way for both improved user productivity and new and unseen research along mobile computing avenues. The particular point in the solution space is one which we believe addresses a real problem and has wide applicability, not just for replicating files in a mobile context but more generally for a wide range of replication-related problems, including but not limited to military cases, software development scenarios, distributed database problems like airline reservation systems, and general-purpose distributed computing.

The rest of paper is organized as follows: section 2 reviews the related work; section 3 discuss the basic ward model; section 4 presents the advanced ward model; section 5 shows the architectural design; section 6 presents the results; and section 7 conclude the paper.

II. A Brief Review of Related Work

Computing in the 1970s and 1980s meant using almost completely stationary machines. Designed to be situated in one location and rarely if ever moved, these machines were heavy, awkward, and clumsy. The resulting software architectures were adequate for the time, because these machines were too clumsy to move on any regular basis. Even the emerging “portable" machines at the end of the 1980s were not really portable; they could at best be described as \luggable." However, by the 1990s the hardware industry had made rapid progress in chip and LCD technology, as well as in general miniaturization. Machines emerged that were truly portable and as powerful as their stationary cousins. Making use of this new type of machine with its attendant portability, users became increasingly mobile. In 1996, approximately one third of the computers sold were mobile-enabled: that is, portable form-factor with communications capability.

Coda [2] is an optimistically replicated file system primarily using the client-server model. Coda provides replication flexibility akin to selective replication at the clients, but not at the replicated servers. The servers run a form of peer replication. Coda clients cannot directly inter-communicate due to restrictions of the client-server model. Use of this model dramatically simplifies the consistency algorithms, at the cost of limiting the system's utility for mobility. Coda is clearly superior in the low-bandwidth scenario, having greatly optimized communications and synchronization, especially in environments with weak connectivity [3]. Some of the same ideas could be applied in Roam, though additional research would be required to incorporate them into a peer model.

The Little Work project [4] is similar to Coda, but modifies only the clients, leaving the AFS [5] servers unaltered. Congestion caused by clients' slow links is reduced in a variety of ways, including client-side modifications of AFS, Little Work's underlying RPC, and other congestion avoidance and control methods. However, clients cannot directly communicate, hindering the usability of the system in dynamic, mobile environments.

The Bayou system [6] replicates databases rather than file system objects. Like Roam, it uses the peer-to-peer model. Unlike Roam, Bayou does not attempt to provide transparent conflict detection. Applications must specify a condition that determines when a conflicting access has been made, and must specify the particular resolution process.

Ficus [7] is one of the intellectual predecessors of Roam, and Roam therefore shares many of its characteristics. Both are based on a peer model, though Roam's Ward Model scales better than the Ficus traditional peer model. Both provide selective replication control. While each maintains consistency with a periodic reconciliation process, Ficus also uses best-effort propagation at update time. Ficus is aimed at a distributed Internet environment, and works well for its target. However, it is unsuitable for mobile use, and does not scale well.

Rumor [8] is the direct predecessor of Roam; much of Roam's implementation is directly based on modified Rumor code. Rumor is in turn a descendent of Ficus, and shares many

of its characteristics and problems. It is based on the traditional peer model, and relies upon periodic reconciliation to maintain consistency. While Rumor is better suited to a mobile environment than Ficus, its scaling properties are substantially the same.

An example of remote computing is illustrated in [9] applied to the geometry over internet. Resources such as executables, languages, packages, can be used from a remote computing system. They implemented a distributed system using geometry that outsources the computing-intensive tasks to remote servers that may be located in other universities or companies, linked to grids and clusters and so on. The architecture developed stresses the interoperability of the software, and a suitable high degree of decoupling between components hosted in various locations.

The definition of a level oriented business process methodology is enhanced in [10], which encourages the adaptation of the modeling notation to the modeling and technical knowledge shown by the expert. Their approach reduces the complexity found by domain experts and enables them to model their processes completely with a level of technical detail directly proportional to their knowledge. They also generate the codes on mobile plateform.

A simple client-server system architecture and algorithms is introduced in [11] for ubiquitous live video and VOD service support. The main features of the system are: efficient usage of network resources, emphasis on user personalization, and ease of implementation. The system supports many continuous service requirements such as QoS provision, user mobility between networks and between different communication devices, and simultaneous usage of a device by a number of users.

A framework for computation offloading is proposed in [12] for mobile cloud computing. Offloading of computationally intensive application parts from the mobile platform into a remote cloud infrastructure or nearby idle computers addresses this problem. They presented Mobile Augmentation Cloud Services (MACS) middleware which enables adaptive extension of Android application execution from a mobile client into the cloud.

Clustering and regression based technique is introduced in [13] for analyzing anonymized cellular network data to identify generally important locations, and to discern semantically meaningful locations such as home and work. They test it on arbitrary cellphone users, including those with low call rates.

Traditional client-server architectures are not capable of supporting the desired infrastructure and providing the required communication capabilities among replicas. While such functionality could have potentially been added to the client-server model, we expected that modifying a fundamentally client-server design to incorporate a rich communications structure would require such basic structural changes as to imply effectively starting from scratch. The goal is therefore to design a peer-based model that scales well. Replication services based on the traditional peer model, such as Rumar [8], AFS [5], and Ficus [14], all suffer from scaling problems. In response, we have designed the Ward Model. It provides a new and different form of the peer model, one based on a hybrid between client-server and peer solutions that clusters replicas into groups without affecting the underlying any-to-any communication capability between all system participants. Ward based model can also be integrated in other domains such as in fruit diseases [15], human activity recognition [16], PCA approaches [17], 3D model [18], semantic information [19], wireless sensor network [20, 21] and learning techniques [22].

III. BASIC WARD MODEL

We describe the basic Ward Model by defining wards, and ward sets. We then explain how the model maintains consistency and how the model supports mobility.

The key idea behind wards is to group volume replicas into containers that capture the notion of common or typical communication partners. For example, given four replicas in Los Angeles and four replicas in New York, the system would perform poorly if each replica in Los Angeles typically synchronized with a replica in New York. While the topology produces correct results, each machine pays the additional cost of long- distance communication, in terms of latency, efficiency, and price per byte. Additionally, at synchronization time, there is a greater chance of a failed network connection between Los Angeles and New York than between two local partners in either location, given that generally the long distance communication depends on having the local communication operational. The obvious improved approach is to have one candidate in Los Angeles communicate with one candidate in New York; these two candidates afterwards disseminate the information among their local colleagues.

Given that synchronization should typically occur with a local partner, both for economic and efficiency reasons, we would like to group replicas together to capture the notion of synchronization locality. We therefore build wards as a collection of “nearby" volume replicas, the details of which are discussed below. The ward members are only required to be loosely connected-continual, high-quality connectivity is not necessary.

Ward set

The ward set refers to the set of replicated data stored within the ward. In the basic Ward Model, the ward set is, by definition, equivalent to the entire volume, although the definition changes in the advanced model. Like the volume itself, the ward set is dynamic in character; it changes as the volume itself changes in response to new object creations and existing object deletions. It is the ward master's responsibility to synchronize its ward set with the ward sets from other wards. In the basic Ward Model, ward-set synchronization can be accomplished simply by contacting one other ward master: since all wards store the entire volume, all ward sets are identical. In the advanced model a more complex synchronization architecture will be utilized.

Maintenance of consistency

Consistency is maintained simultaneously both within each ward and among wards. In both intra- and inter-ward scenarios, the consistency topology refers to the communication pattern used between replicas. All of our consistency algorithms are topology-independent. For example, a quadratic message complexity results from an all-pairs reconciliation topology, but a ring (using a gossip-based transfer of information) [20-25] reduces the message complexity to a linear cost. A superior messaging plan avoids the quadratic cost when inter-site communication is available, but gracefully handles degraded communication. Using two separate topologies within the ward and between wards reduces the generality of the model and increases the requirement for special-case code. Although one could potentially identify hypothetical reasons why two topologies would be required, a topology that applies equally well to both scenarios is clearly a better choice for generality and simplicity. The adaptive ring allows rarely or never-communicating sites to share data by relying on third-party replicas to gossip on their behalf. For these reasons, it is an attractive topology to consider, and one that applies equally well both within and between wards.[25-30]

For instance, assume the ward contains replicas 2, 4, and 6. If replica 5 joins but only replica 4 is knowledgeable of 5's existence, then replica 6 will synchronize with replica 4, which 6 believes to be the next replica in the ring. However, as part of 6's synchronization it learns of 5's existence, so the ring automatically “heals" itself as the new information propagates. Since the underlying algorithms are topology independent, correctness is not affected by temporary spokes on the adaptive ring. The example is illustrated graphically in Figure 1. (Ward is used generally here, and also refers to the meta-ward that contains all ward masters.). Ward masters are illustrated as solid black replicas; wards are double circles surrounding the replicas. Arrows indicate synchronization paths, although the synchronization topology adapts itself to network topology.

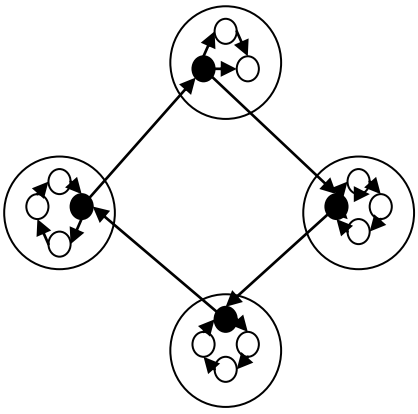


Fig. 1. The basic adaptive ring topology, both within and between wards.

Support for mobility

The model provides support for intra-ward mobility essentially for free, since everyone in the ward is a peer in the traditional sense. Intra-ward mobility occurs anytime the user is mobile within a restricted geographic area and is therefore only likely to encounter other machines from the same ward. For instance, moving within one's once from the desk to the couch demonstrates an intra-ward mobile action, as does moving around the once or within town to the local coffee shop. In all examples, the degree of motion is large enough to potentially change the set of “best" or most efficient communication partners, but not large enough to bring the user outside of his or her ward. For example, two colleagues in the once may never directly synchronize, typically relying on a set of third-party replicas to relay updates. However, if they meet at someone's home to discuss plans for the following day, they will usually want their machines to directly communicate.

IV. Advanced Ward Model

In the advanced Ward Model we introduce selective replication, the ability for a ward member to physically store only select portions of the complete volume. The details concerning how the reconciliation algorithms and controls handle selective replication are discussed in the next section. Here we describe the changes selective replication makes to the ward definitions, controls, and constructs.

Wards

The basic definition of a ward remains unchanged. The ward is still a collection of “nearby" volume replicas. However, all replicas need not store the same portions of the volume, which impacts both the ward set and synchronization topologies, as described in the following sections. The characteristics and abilities of ward members similarly remain unchanged. Specifically, any-to-any communication is still enabled between any two ward members. However, due to selective replication, any-to-any communication between all replicas may not always make physical sense for any particular layout of the file objects onto the set of volume replicas.

Ward set

The ward set is defined to be set of replicated data stored within the ward. In the basic Ward Model the ward set is equivalent to the volume. However, selective replication allows each ward member to store select portions of the volume, meaning that the set of data stored at each replica may be smaller than the entire volume. The ward set itself, therefore, may be smaller than volume, though of course it can never be larger. It is equivalent to the union of all replica sets for all replicas in the ward. The ward set changes dynamically as the set of data stored within the ward changes. For instance, if a given ward member uses the selective replication controls to locally add a new file system object, the ward set expands to include this new object. The ward set similarly decreases in size when replicas locally drop file replicas. Additionally, ward motion can change the ward set, as replicas move into and out of a given ward.

Ward masters

The ward master is responsible for the inter-ward synchronization of the entire ward set, and therefore must be able to identify the complete ward set. Use of both selective replication and ward changing, however, can change the ward set. Since any ward member can dynamically and optimistically change its replica set using the selective replication controls, and because new machines can at any time move into the ward (carrying with them their accompanying replica set), the ward set changes dynamically, lazily, and without global coordination. Without selective replication the ward set could be identified simply by naming the volume. With selective replication, however, the best mechanism for the ward master to identify its complete ward set is to individually name all objects in it.

Finally, we would like selective replication to help alleviate disk storage at the ward master. A ward member that wants to store a 100MB file should not necessarily force the ward master to also store that 100MB file solely for inter-ward synchronization purposes. We allow the ward master to store only a virtual replica-the equivalent to naming the object without storing _le data. When data is required for synchronization purposes, the ward master relays it via the physical data site.

Maintenance of consistency

Since each replica in the ward has a potentially different replica set, it follows that each ward master has a potentially different ward set. Therefore, in both intra- and inter-ward synchronization, we must use a more robust topology than a simple adaptive ring. We augment the basic adaptive ring to account for the differences between replica sets and ward sets. When the ward master is viewed as a “super-replica," the ward set appears exactly the same as a replica set for a super-replica, meaning that one approach can again be used in both intra- and inter-ward synchronization. The solution uses an adaptive ring for each file object, rather than one for the whole volume, and then coalesces multiple per-file rings into a single ring based on the intersection between replica.

Support for mobility

Selective replication introduces new difficulties for the ward motion algorithms. In the basic Ward Model, a machine moving into a new ward is guaranteed to have the same replica set as the new ward's ward set, because all participants store full volumes. Since the ward set is equivalent to the moving replica's replica set, it is straightforward for the new machine to integrate with the new ward. With selective replication, the mobile machine's replica set may differ from the new ward's existing ward set. As a result, the advanced Ward Model requires more rich and robust ward motion algorithms.

The ward master is responsible for the inter-ward synchronization of all intra-ward data. When a new machine enters the ward and brings with it data files not in the current ward set, there are two options: either the ward set expands to incorporate the new data objects, or it doesn't. In the former case the ward set changes, possibly causing changes at other ward masters since they must keep track of what is stored at the various ward masters to properly form their adaptive per-file rings. In the latter case the ward set remains unchanged, and there are no ripple effects affecting the other ward masters, but the mobile machine cannot synchronize all of its data completely within the new ward. Some of the data must be synchronized with another ward, most likely the original ward that the mobile machine came from. To properly decide which option is best, we must look at how physical motion actually occurs. Real mobility seems to occur in one of two modes:

visit: a temporary trip to a remote location, measured in hours or a small number of days

long stay: a longer stay at a remote location, perhaps for \a while," perhaps more permanently.

Each mode has accompanying expectations of cost and performance. Users expect a temporary mobile move to be lightweight and inexpensive: since they're not planning on staying very long, they don't want to pay a large up-front cost. Additionally, users will generally accept sub-optimal performance, given that they know the motion is temporary and the up-front cost is minimal. On the other hand, users moving for longer periods of time are generally willing to pay a more expensive up-front cost to gain better performance. Since they know they will remain at the remote location for a long time, they want good performance while there. The up-front cost is amortized over the length of the stay; users staying a short time do not gain significant benefit from the large cost, and are therefore generally unwilling to pay it. Real motion, of course, occurs over a continuous time spectrum, and doesn't always fall exactly into one of the two classifications. However, as a general paradigm the classifications seem to work fairly well, especially since positions in the middle of the spectrum can essentially be placed in either category.

V. System Architecture

As we discussed earlier, we are having several models to achieve services of Replication, but any of them is not able to fulfill our requirements. So we are using a new model, called Ward Model. Ward model is a hybrid model which combines the features of peer-to-peer and client server model as shown in figure 4. This model provides all the four basics qualities of replication system for mobility. In Ward Model, several new terms are used. These terms are Wards, Ward Member (MC), Ward Master (WM) and Ward set which we discussed already in previous sections. Figure 2 describes the ward model and Figure 3 shows the graphical views of ward models.

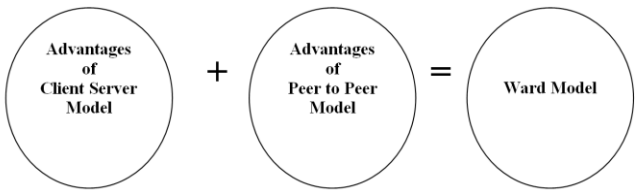


Fig. 2: Ward Model

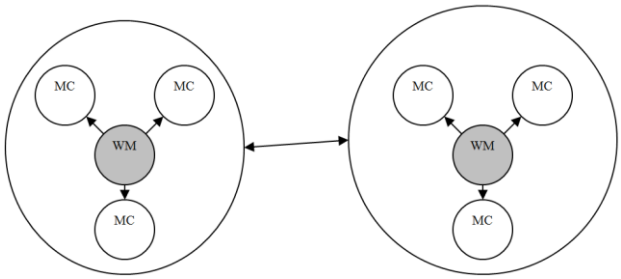


Fig. 3. Graphical view of Ward Model

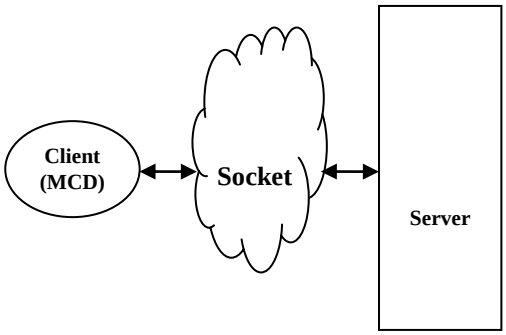


Fig. 4. Client server using socket.

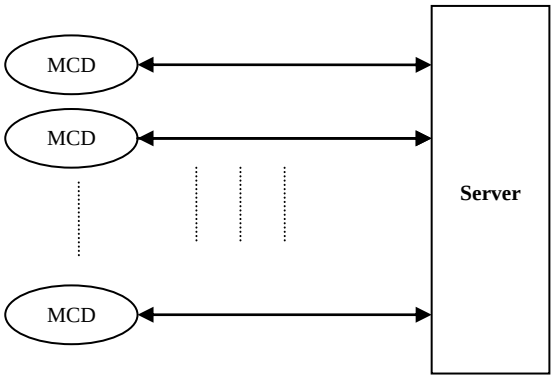


Fig. 5. Multiple Clients connected to Server.

Figure 4-9 shows the block architectures of the various phases of the approach such as, working of client server using socket, how multiple clients are connected to server, the process of client authentication, pinging of clients at server using hash table, receiving file from server and sending file to server.

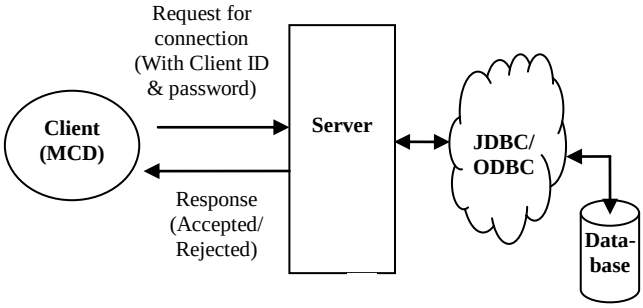


Fig. 6. Client Authentication

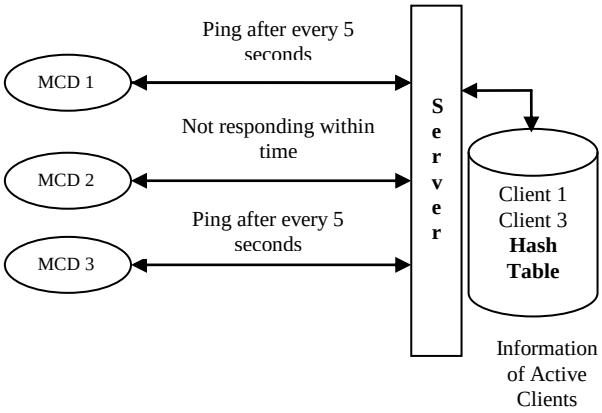


Fig. 7. Pinging of Clients at Server using hash table.

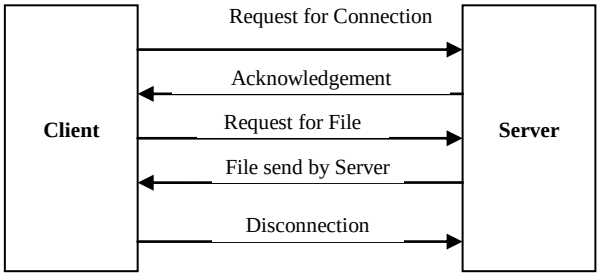


Fig. 8: Receiving file from server.

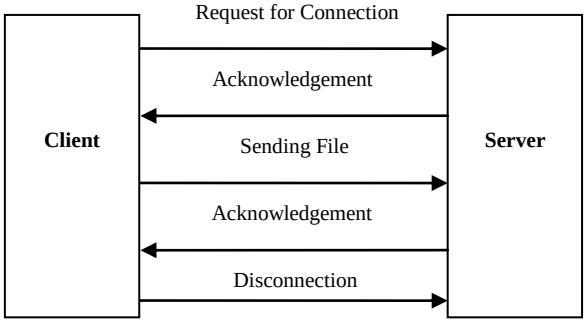


Fig. 9: Sending file to server.

Implementation Model: Implementation model is described with the help of following flow charts

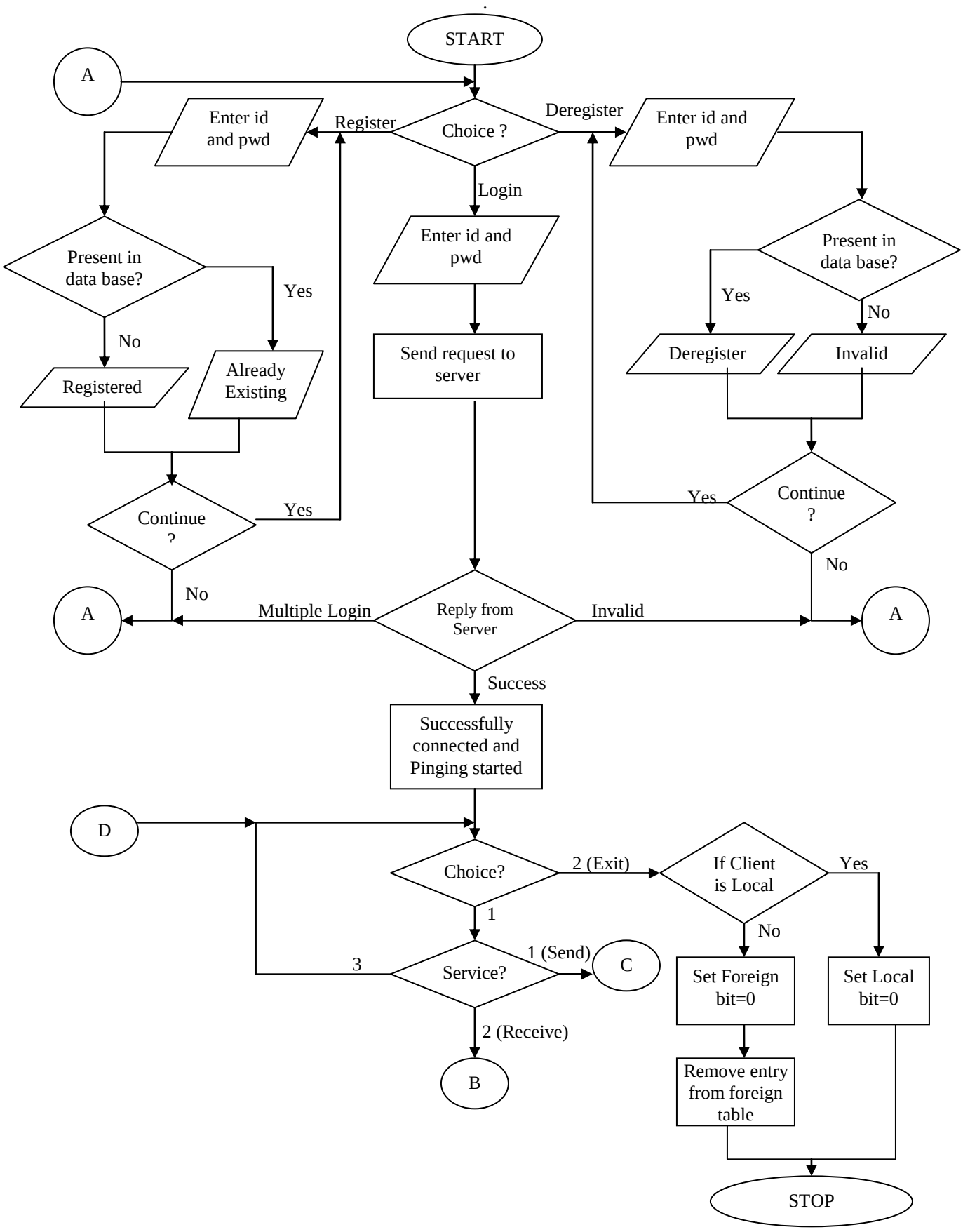


Fig. 10: Client Side Flow Chart

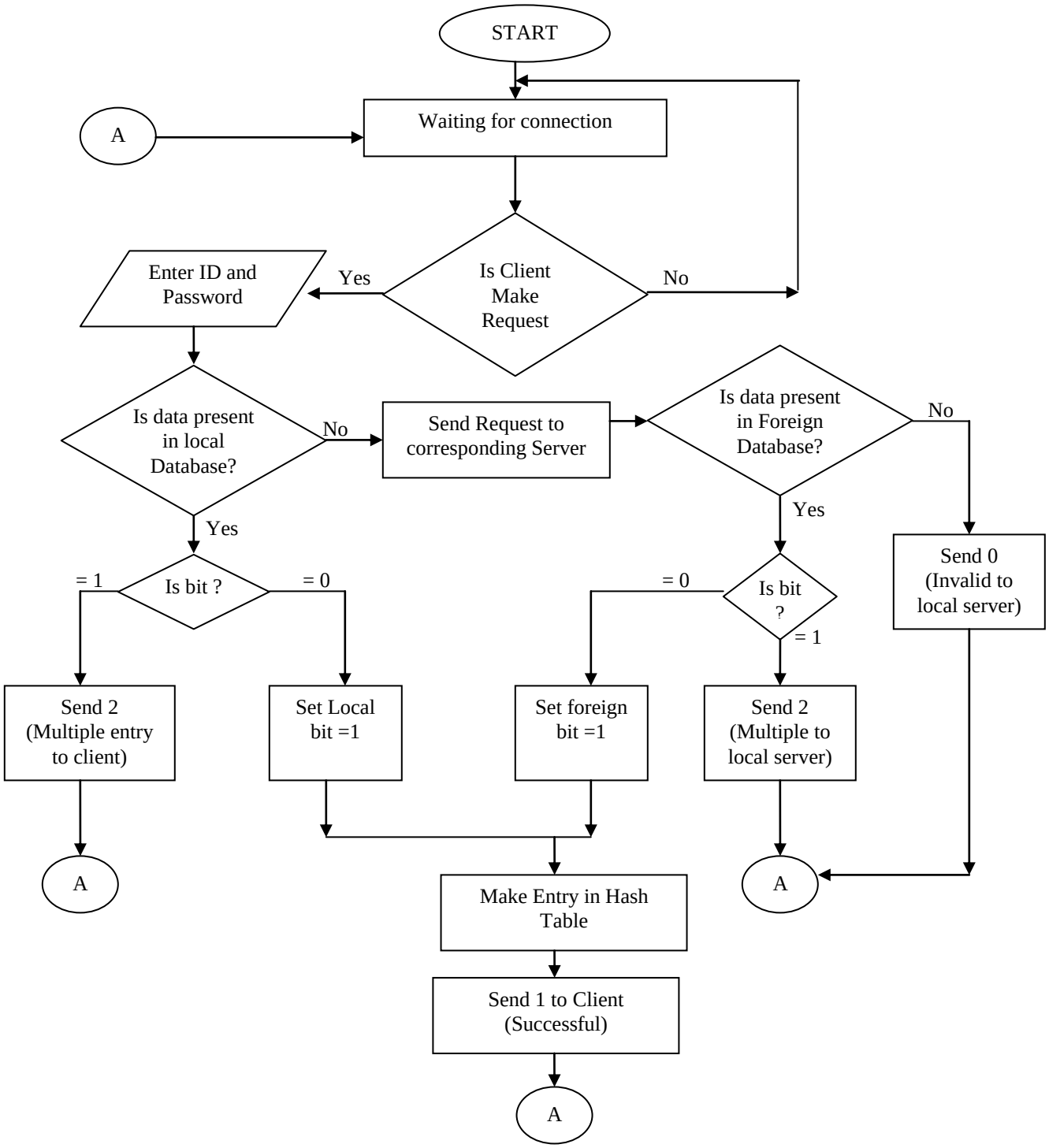


Fig. 11: Server Side Flow Chart

Figure 10 and Figure 11 describe the steps involved in the implementation of client and server side processes respectively with the help of scalable replica concepts while client made a request to the server.

VI. RESULTS

Database

We have used Microsoft- Access as Database. In this we have used following tables.

TABLE I
SERVER

Server ID	Server Name	Server Password
S1	Ser1	Pwd1
S2	Ser2	Pwd2

TABLE 2
SERVER1

Server ID	Server Name
M11	Pwd11
M12	Pwd12

TABLE 3
SERVER2

MCD_ID	MCD_PWD
M21	Pwd21
M22	Pwd22

Experimental Results

Figure 12 to 21 shows the output over command prompt while testing our proposed approach for different scenario. Figure 12 shows the registration of a new client. Client will enter the desired ID and password; if ID is available it will be assigned otherwise prompt for different ID. Figure 13 depicts the scenario of deletion of a registered ID. To delete the ID, password will be required.

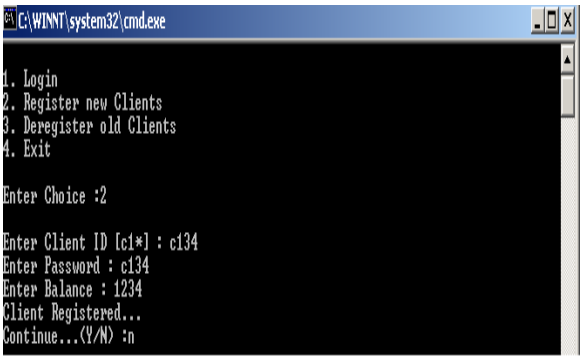


Fig. 12: Register new Clients

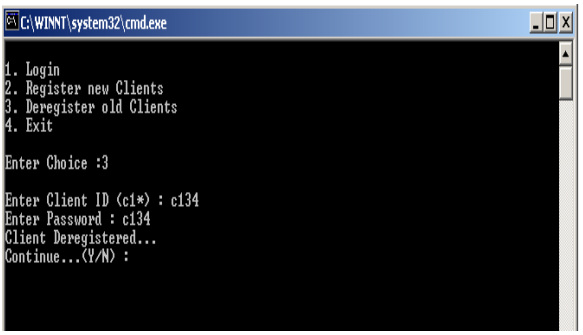


Fig. 13: De-registering old clients

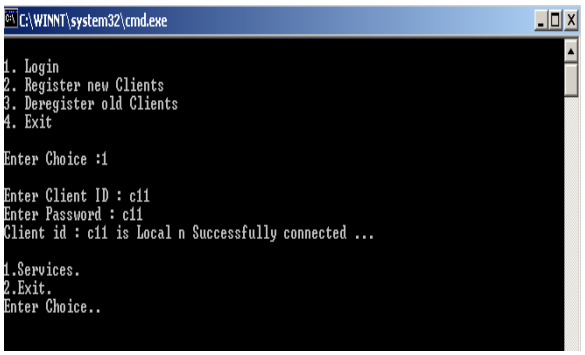


Fig. 14: Login

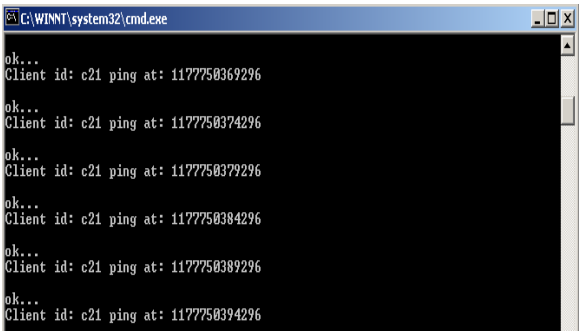


Fig. 15: Pinging

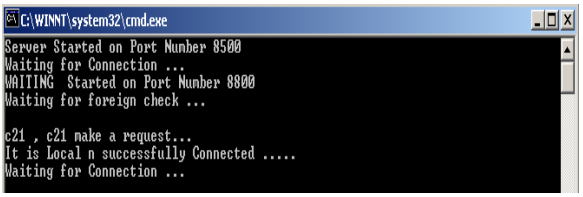


Fig. 16: Server Window

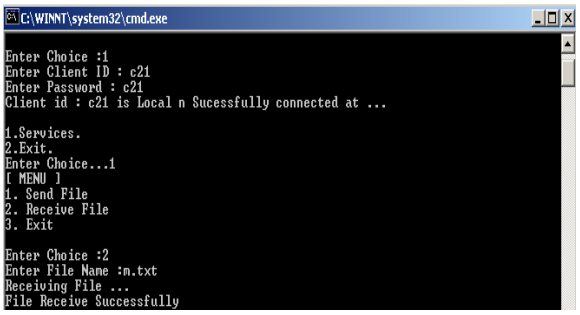


Fig. 17: Services used by clients

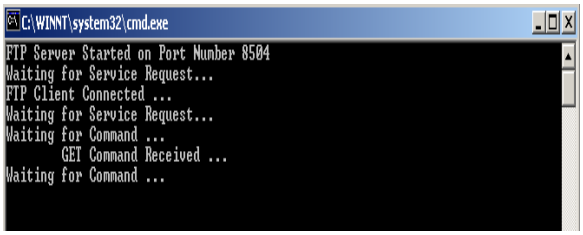


Fig. 18: FTP Server Window

Fig. 19: Connecting Foreign Client in VLR

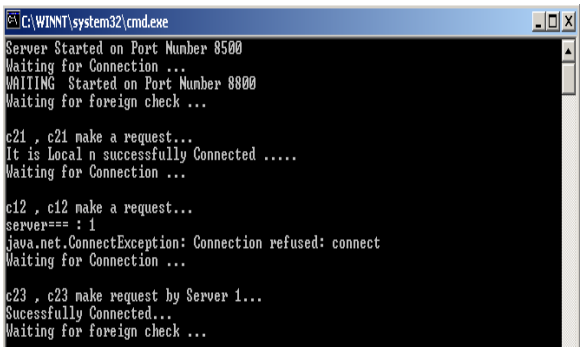


Fig. 20: Foreign Server Window

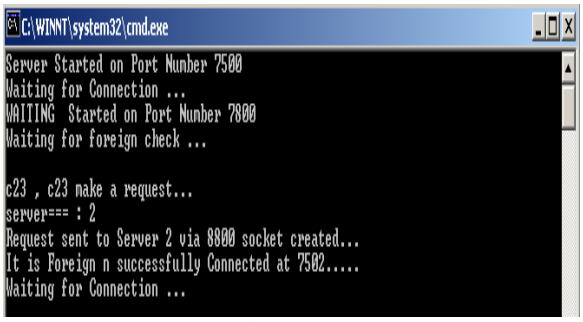


Fig. 21: Local Server Window (after connecting Foreign Client)

Figure 14 shows the steps of login; ID and password is required to login and after successful login some choices will be generated. Figure 15 is the case to test that any client is active or not. If it is pinging means it is active. Figure 16 shows the actions at the server side. At port number 8500, server is waiting for connection and if any request is made then it verifies its credential and after successful verification, client will be connected. Figure 17 shows the commands available for performing services to a client such as sending or receiving a file. Figure 18 shows the activities involved at the server side for file transfer to the clients. The example of connecting a foreign client is depicted in the Figure 19. Figure 20 shows the connection with a foreign server from different clients. Figure 21 shows the status of a local server window after connecting a foreign client. Foreign clients are connected at the port 7502 of local server.

From the experiments, it is deduced that scalable replica facilitates efficient client and server operations in the mobile computing which is purely based on the client and server architecture.

VII. Conclusion

In this paper we proposed a number of significant problems in large scale replication, and produced a scalable replication system for mobile environments. With the arrival of machines capable of supporting truly mobile computing came users wanting to access and update their data while mobile. If the user has more than one machine, for instance a laptop and a desktop, or if the data must be shared between multiple users, then the data must be replicated. Unfortunately, the existing replication systems are not mobile-compliant. Designed for stationary environments, they do not provide users with the abilities they require when mobile. Furthermore, we build into Roam's algorithms the ability to automatically handle and update outdated meta-data as part of the normal synchronization of user data. Roam makes extensive use of the optimistic approach in its management of the system. Roam is a successful piece of research, based not just on the performance analysis but also on real-world experience and use. It is our hope that Roam paves the way for real mobile use and future mobile computing research; simultaneously, we would like the underlying ideas and concepts to bear fruit and diseases and become used in other areas of computer science.

Refrences

[1] D. Ratner, P. Reiher, and G. J. Popek, “Roam: a scalable replication system for mobile computing,” in: Proceedings of the 10th International Workshop on Database and Expert Systems, 1999, pp. 96 – 104.

[2] M. Satyanarayanan, J. J. Kistler, P. Kumar, M. E. Okasaki, E. H. Siegel, and D. C. Steere, “Coda: A highly available file system for a distributed work station environment,” IEEE Transactions on Computers, vol. 39, no. 4, pp. 447-459, 1990.

[3] J. J. Kistler and M. Satyanarayanan, “Disconnected operation in the Coda file system,” ACM Transactions on Computer Systems, vol. 10, no. 1, pp. 3-25, 1992.

[4] P. Honeyman, L. Huston, J. Rees, and D. Bachmann, “The Little Work project,” in: Proceedings of the 3rd Workshop on Workstation Operating Systems, 1992, pp. 11-14.

[5] J. Howard, M. Kazar, S. Menees, D. Nichols, M. Satyanarayanan, R. Sidebotham, and M. West, “Scale and performance in a distributed file system,” ACM Transactions on Computer Systems, vol. 6, no. 1, pp. 51- 81, 1988.

[6] D. B. Terry, M. M. Theimer, K. Petersen, A. J. Demers, M. J. Spreitzer, and C. H. Hauser, “Managing update conicts in Bayou, a weakly connected replicated storage system,” in: Proceedings of the 15th Symposium on Operating Systems Principles, 1995, pp. 172-183.

[7] Jatav, Vinod Kumar, et al. "Wireless Sensor Networks: Attack Models and Detection." 2012 IACSIT Hong Kong Conferences, IPCSIT vol. 30 (2012)©(2012) IACSIT Press, Singapore. 2012.

[8] P. Reiher, J. Popek, M. Gunter, J. Salomone, and D. Ratner, "Peer-to-peer reconciliation based replication for mobile computers," in: Proceedings of the ECOOP Workshop on Mobility and Replication, 1996.

[9] Semwal, Vijay Bhaskar, K. Susheel Kumar, Vinay S. Bhaskar, and Meenakshi Sati. "Accurate location estimation of moving object with energy constraint & adaptive update algorithms to save data." arXiv preprint arXiv:1108.1321 (2011).

[10] M. Ferreira and M. Casquilho, "An Example in Remote Computing Over the Internet applied to Geometry," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 2, no. 3, pp. 38-43, 2013.

[11] J. Solis-Martinez, N. Garcia-Menendez, C. Pelayo G-Bustelo, and J. M. Cueva-Lovelle, "BPLOM: BPM Level-Oriented Methodology for Incremental Business Process Modeling and Code Generation on Mobile Platforms," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 2, no. 2, pp. 13-27, 2013.

[12] Neha Gandotra, Vishwanath Bijalwan "coexistence model of zigbee & ieee 802.11b (wlan) in ubiquitousnetwork environment" IJARCET, volume 1, Issue 4,June 2012.

[13] Vishwanath Bijalwan, Dr. Sanjay Singh "ANALYSIS &DESIGN OF JOINT PHY-MAC MODEL OF IEEE 802.15.4" IJSETR, volume 2, Issue 9, September 2013..

[14] S. Isaacman, R. Becker, R. Caceres, S. Kobourov,M. Martonosi, J. Rowland, and A. Varshavsky. Identifying important places in people's lives from cellular network data. In Proc. Int. Conf. on Pervasive Computing, San Francisco, Jun. 2011.

[15] R. G. Guy, "Ficus: a very large scale reliable distributed file system," Ph.D. dissertation, University of California, Los Angeles, June 1991.

[16] J. P. Gupta, N. Singh, P. Dixit, V. B. Semwal, and S. R. Dubey, "Human Activity Recognition using Gait Pattern," International Journal of Computer Vision and Image Processing, vol. 3, no. 3, pp. 31 – 53, 2013.

[17] K. K. Susheel, V. B. Semwal and R. C. Tripathi, "Real time face recognition using adaboost improved fast PCA algorithm," arXiv preprint arXiv:1108.1353 (2011).

[18] K. S. Kumar, V. B. Semwal, S. Prasad and R. C. Tripathi, "Generating 3D Model Using 2D Images of an Object," International Journal of Engineering Science, 2011.

[19] N. Singh, S. R. Dubey, P. Dixit and J. P. Gupta, "Semantic Image Retrieval by Combining Color, Texture and Shape Features," In the Proceedings of the International Conference on Computing Sciences, pp. 116-120, 2012.

[20] V. B. Semwal, V. B. Semwal, M. Sati and S. Verma, "Accurate location estimation of moving object in Wireless Sensor network," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 1, no. 4 , pp. 71-75, 2011.

[21] E. Felembane, C. Lee and E. Ekici, "MMSPEED: Multipath Multi-SPEED Protocol for QoS Guarantee of Reliability and Timeliness in Wireless Sensor Networks," IEEE Transactions on Mobile Computing, vol. 5, no. 6, pp. 738-754, 2006.

[22] S. R. Dubey and A. S. Jalal, "Adapted Approach for Fruit Disease Identification using Images," International Journal of Computer Vision and Image Processing, vol. 2, no. 3, pp. 44-58, 2012.

[23] Sanjay Singh, Vishwanath Bijalwan, "Design Of Wireless Sensor Network Node On Zigbee For Water Level Detection" Volume 3,Issue8, August 2013.

[24] Anchit Bijalwan, Vishwanath Bijalwan, "Examining the Crimninology using Network Forensic" 8th National Conference USCSTC ,Dec 2013.

[25] Vishwanath Bijalwan, Vinay Kumar, Pinki Kumari,Jordan Pascual," KNN based Machine Learning Approach for Text and Document Mining" International Journal of Database Theory and Application Vol.7, No.1 (2014), pp.61-70.

[26] Pinki Kumari and Vikas Pareek, RAKSHITA- A Novel web based Approach for Protecting Digital Copyrights Using Public Key Digital Watermarking and Human Fingerprints" International conference on methods and models in computer science (ICM2CS-2010).

[27] Pinki Kumari and Abhishek Vaish, "Instant Face detection and attributes recognition" in International Journal of Advanced Computer Science and Applications (IJACSA - ISSN 2156 5570):2011

[28] Pinki Kumari and Abhishek Vaish,"A Comparative study of Machine Learning algorithms for Emotion State Recognition through Physiological signal", Advances in Intelligent Systems and Computing, Vol.236-Springer; ISBN 978-81-322-1601-8

[29] Pinki Kumari and Abhishek Vaish "Brainwave's Energy feature Extraction using wavelet Transform" proceeding of IEEE SCEECS, 2014, MANIT, Bhopal ISBN: 978-1-4799-2526-1.

[30] Semwal, Vijay Bhaskar, K. Susheel Kumar, Vinay S. Bhaskar, and Meenakshi Sati. "Accurate location estimation of moving object with energy constraint & adaptive update algorithms to save data." arXiv preprint arXiv:1108.1321 (2011).

Meenakshi Sati presently working as Assistant Professor, India. She is B.Tech from Graphic Era University Dehradun and pursing M.Tech from NITTR Chandigrah. Her major research work Interest in Image Processing, Sensor Network, Design and Analysis of Algorithm, Data Structure.

Vivek Vikash is presently working as Software Engineer, Honeybell, India. He completed his M.Tech form Indian Institute of Information Technology, Allahabad, his major research work Interest in Image Processing, computer sensor network & Pattern Recognition.

Vishwanath Bijalwan is working as assistance professor at Uttarakhand Technical University Dehradun. He is person with lot of potential. His research interest is wireless sensor network, WiMax, Wi-Fi, machine learning, Informtion Retrieval.So far he has published 4 high quality researches paper and work on many MHRD funded project.

Pinki Kumari received her M.Tech in Computer Science and Engineering from Banasthali University, Rajasthan, and her Ph.D in Information Security. Her research interests include Pattern recognition, Cloud Computing, Big Data and Distributed Database.

Manish Raj is currently pursuing his Ph.D fm IIITA.. He received his M.Tech from IIITA India. He has more than 5 years of research and teaching experience. He supervised 5 PG and 21 UG students. He has 2 research papers in reputed international journal and conference.

Meenu Balodhi is currently pursuing her fm Uttaranchal technical university. His research area biometric identification and wireless sensor network

Priya Gairola is currently pursuing her fm Uttaranchal technical university. His research area biometric identification and wireless sensor network, Wi-Max

Vijay Bhaskar Semwal has been severed as a PCC member in several International conferences and Journals. He was also head for 5th IEEE conference on big data at COER Roorkee. He is carrying more than 5 year of academic and industry experience. He served for various top organizations like Siemens, Newgen etc. Earlier he received his M.tech in wireless sensor domain and was awarded by gold medal. His major research work interest in Wireless Sensor Network, Artificial Intelligence, Image Processing, Computer Network &Security, and Design & Analysis of Algorithm, Machine Learning, and Information Retrieval Soft Computing.

Business and Social Behaviour Intelligence Analysis Using PSO

Vinay S Bhaskar[1], Abhishek Kumar Singh [2], Jyoti Dhruw[3], Anubha Parashar[4], Mradula Sharma[5]

[1]*Human Resource Manager, Electro Equipment, Roorkee, India*
[2]*IIIT Allahabad*
[3]*Chhatrapati Shivaji Institute of Technology (CSIT, Durg, Chhattisgrah*
[4]*Maharshi Dayanand University, Rohtak*
[5]*MNNIT Alllahabad*

***Abstract* — The goal of this paper is to elaborate swarm intelligence for business intelligence decision making and the business rules management improvement. The paper introduces the decision making model which is based on the application of Artificial Neural Networks (ANNs) and Particle Swarm Optimization (PSO) algorithm. Essentially the business spatial data illustrate the group behaviors. The swarm optimization, which is highly influenced by the behavior of creature, performs in group. The Spatial data is defined as data that is represented by 2D or 3D images. SQL Server supports only 2D images till now. As we know that location is an essential part of any organizational data as well as business data: enterprises maintain customer address lists, own property, ship goods from and to warehouses, manage transport flows among their workforce, and perform many other activities. By means to say a lot of spatial data is used and processed by enterprises, organizations and other bodies in order to make the things more visible and self-descriptive. From the experiments, we found that PSO is can facilitate the intelligence in social and business behaviour.**

***Keywords* — PSO, Map, Artificial Intelligence, Geography, Optimization.**

I. INTRODUCTION

SWARM describes a behavior of an aggregate of animals of similar size and body orientation [1]. Swarm intelligence (SI) is based on the collective behavior of a group of animals. Collective intelligence emerges via grouping and communication, resulting in successful foraging (the act of searching for food and provisions) for individual in the group, for examples Bees, ants, termites, fishes, birds etc. The perform the following sequence of activity in group: Marching of ants in an army, Birds flocking in high skies, Fish school in deep waters, Foraging activity of micro-organisms.In the context of AI, SI systems are based on collective behavior of decentralized, self-organized systems [2]. Typically made up of a population of simple le agents interacting with one another locally and with their environment causing coherent functional global pattern to emerge. Distributed problem solving model without centralized control. Even with no centralized control structure dictating how individual agents should behave, local interactions between agents lead to the emergence of complex global behavior [3]. Swarms are powerful which can achieve things which no single individual could do.

An intelligent technology is the duplication of human thought process by machine. It learn from experience, interpreting ambiguities, rapid response to varying situations, applying reasoning to problem-solving and manipulating by applying knowledge, thinking and reasoning [4]. Different from traditional optimization technique, evolutionary computation techniques work on a population of potential solutions (points) of the search space. The most commonly used population-based evolutionary computation techniques is PSO [5]. It is a cost optimized solution. Organizations generate and collect large volumes of data, which they use in daily operations. Yet despite this wealth of data, many organizations have been unable to fully capitalize on its value because information implicit in the data is not easy to distinguish. However, to compete effectively today, taking advantage of high-return opportunities in a timely fashion, decision-makers must be able to identify and utilize the information. These requirements imply that an intelligent system must interact with a data warehouse and must interface with decision support systems (DSS), which are used by decision-makers in their daily activities. There is a substantial amount of empirical evidence that human intuitive judgment and decision-making can be far from optimal, and it deteriorates even further with complexity and stress. Because in many situations the quality of decisions is important, aiding the deficiencies of human judgment and decision-making has been a major focus of science throughout history [6] [7]. Disciplines such as statistics, economics, and operations research developed various methods for making rational choices. More recently, these methods, often enhanced by a variety of techniques originating from information science, cognitive psychology, and artificial intelligence, has been

DOI: 10.9781/ijimai.2014.268

implemented in the form of computer programs as integrated computing environments for complex decision making. Such environments are often given the common name of decision support systems (DSS) [20] [21].

The development and deployment of managerial decision support system represents an emerging trend in the business and organizational field in which the increased application of Decision Support Systems (DS DOI: 10.9781/ijimai.2014.268 Intelligent Systems (IS). Decision Support Systems (DSS) are a specific class of computerized information system that supports business and organizational decision-making activities [18] [19]. A properly designed DSS is an interactive software-based system intended to help decision makers compile useful information from raw data, documents, personal knowledge, and/or business models to identify and solve problems and make decisions. Competitive business pressures and a desire to leverage existing information technology investments have led many firms to explore the benefits of intelligent data management solutions such as Particle Swarm Optimization (PSO). This study proposes a new PSO (SPSO)-model based on product mix model for optimizing Constraint values as well as objective function. The formulations of the objective function for the minimization problem. This technology is designed to help businesses to finding multi objective functions, which can help to understand the purchasing behavior of their key customers, detect likely credit card or insurance claim fraud, predict probable changes in financial markets, etc. Keywords: Linear problem, Intelligent System, particle swarm optimization, simplex method. The Sql Server Spatial Database is designed to make the storage, retrieval, analysis and manipulation of spatial data easier and natural to users. Once we have data we can perform any operation easily like retrieve all data related to our concern, manipulate it as par requirement [17]. Spatial data is the main need for graphic visualization to make useful result about. It is useful to guess of localization i.e. longitude and latitude. Now we have Microsoft SQL Server 2008 which is including simple feature graphical representation of location data as location in the map by firing query for location in query results area which was not available in the previous version of Microsoft SQL Server i.e. Microsoft SQL Server 2000 and 2005. This visualizer works with a geography column in the query results and in graph by plotting location data and if multiple location column in appearing in as a query result we can select one to visualize [22] [23] [24]. For example suppose we have a location data then we have a choice to flexible plot the diagram of the map either it is rectangle or any other polygon shape for projection purpose on graph but we are assuming out projection should not be overlaid with default projection. By this technique we can easily and very convenient overlay a graph on map if we have the table of map outline data which we can use to do UNION ALL between the row set and the row set that have the map location [8]. Spatial data is useful in lots of cases because every customer has its own address. We usually think of addresses as street, city, state, country and ZIP code and in other words spatial data can use for finding the exact position of real world entity like suppose we have the database of customer where we are storing address which not merely containing which we are supposing street, city or state but actually it is showing the latitude and longitude. As technically by address we mean part of territory which can draw by polygon for convenient we are assuming single point not a polygon and by this we can guess about longitude and latitude for the answer of query like here:

- Find the nearest branch of bank for client info.
- And who is the representative or concern person for that particular client.
- And we can also retrieve the client information within a particular boundary i.e. how many clients are there in particular organisation of any business.

We can guess client information related to position within the range and outside so it is not just maintain the branch location of our office but also allowing to put information of client related to us. To make whole system and process convenient, reliable, to speedup of process, robust by storing data and analyzed data and get information and by exchanging information we can assure all the above mention benefit in our work. As we know that reliability, speed and robustness is only the required feature; here we will consider. Another important aspect or we can say the further advancement of spatial techniques is the visualization of spatial data using maps. The methods used for retrieval problem [9] [10], human activity and face recognition problem [11] [12], location estimation, [13] and scalable replica estimation problem [14] can also be integrated with proposed PSO based approach. After all, location information is all about maps, and to paraphrase a common saying, a map is certainly worth 1,000 words. A better choice is to use a visualizer that provides map overlay by default [15] [16]. SQL Server Studio 2008 use a map overlay for showing spatial data. Some other approaches are also used for similar task [25] [26] [27] [28]. The Advantages of proposed method are as follows: (1) Adaptability - Self-organizing, (2) Robustness - Ability to find a new solution if the current solution becomes invalid, (3) Reliability - Agents can be added or removed without disturbing behavior of the total system because of the distributed nature, (4) Simplicity, and (5) No central control. The rest of the paper is organized in further four section: section 2 discuss some background concepts which are used in our problem such as spatial data, visulizers and spatial index; section 3 introduced the proposed methodology of business intelligence using the concept of particle swarm optimization; section 4 shows some practical application of introduced work with result; finally section 5 conclude the paper with future remark of the paper.

II. Background Concepts

A. Spatial data

Spatial data is data which is use for finding the position of

the real world entity like we have to find the position of sea sore, restaurant, hotel, tourist palace, historical important location and some territory. In spatial database is the combination of all the data types, statics and indexing of location. For fast accessing, the location information from spatial database done by spatial function and spatial indexing. We can retrieve it through Sql.

```
DECLARE @addr nvarchar(256) = 'Some sample address, City, State, Zip';
DECLARE @addr_as_xy nvarchar(30);
DECLARE @g geography;
SET @addr_as_xy = dbo.Geocoder(@addr);
SET @g= geography::STPointFromText(@addr_as_xy, 4326);
```

Fig. 1. Code for location initialization

We have an instance of a location type with a sql variable declaration (@p position) or a column of table, and we have a number of way to initialize data type let's us take an example we have geographical data type who is using to show the instance of location point, the easier way to do it by use of STPointfromText() method of the location type. The STPointFromText() method is using for SRID (spatial reference identifier) with other feature textual representation of point that is POINT(x,y) in open Geospatial boundary. SRID identifier is the spatial reference system for use of either shape of earth whether flat or round mapping and it is enough to know the MapPoint geocoder Web Service uses coordinates of GPS to related to SRID 4326 so for location initialization our code will look something like in Figure 1:

B. Visualizers

Microsoft SQL Server 2008 provides a feature to visualize location result in either query result area or window. Microsoft SQL Server 2008 shows the spatial data as a graph plot in query result area to represent the longitude and latitude related to geography and we can select one column if we have lot of spatial column appearing in same time by query result. And column to display should be in SQL Server binary format, the ToString() method or STAsText() method not work with visualizer.

C. Spatial Indexes

After growing our enterprise to calculate the position between client and every entrepreneur and every salesman and every client might be too slow. Microsoft SQL Server 2008 has also spatial indexing. And here spatial indexing is based on ordinary B-tree index to make it faster like as relational indexes in SQL Server 2000 and 2005.If we have geographical data type, then we are dividing the entire globe into hemisphere and projecting each hemisphere onto a plane. And if we have geometrical data type, because we are specifying our own rectangular coordinate system, we can specify the boundaries that our spatial index covers the whole area. To return to the customer system, you could define a spatial index on the geog column in your customer table with the following data definition language (DDL):

```
CREATE SPATIAL INDEX cust_geog_idx
ON dbo.customer(geog)
GRIDS =(LEVEL1=HIGH,LEVEL2=HIGH,LEVEL3=HIGH,LEVEL4=HIGH)).
```

III. METHODOLOGY

A. Swarm inspired methods

Particle Swarm Optimization (PSO): PSO is a population based stochastic optimization technique developed by Eberhart and Kennedy in 1995. It is inspired by social behaviour of flocks of birds and school of fish. It is a set of agents (similar to ants), search in parallel for good solutions and co-operate through the pheromone-mediated indirect method of communication. They belong to a class of meta-heuristics. These systems started with their use in the Travelling Salesman Problem (TSP). They have applications to practical problems faced in business and industrial environments. The evolution of computational paradigm for an ant colony intelligent system (ACIS) is being used as an intelligent tool to help researchers solve many problems in areas of science and technology.

Particle Swarm Intelligent Systems: Originated with the idea to simulate the unpredictable choreography of a bird flock with Nearest-neighbour velocity matching, Multi-dimensional search, Acceleration by distance, and Elimination of ancillary variables. PSO shares many similarities with Genetic Algorithms (GA). The system is initialized with a population of random solutions (called *particles*) and searches for optima by updating generations. Each particle is assigned a randomized velocity. Particles fly around in a multidimensional search space or problem space by following the current optimum particles. However, unlike GA, PSO has no evolution operators such as crossover and mutation. Compared to GA, the advantages of PSO are that it is easy to implement and there are few parameters to adjust. Each particle adjusts its position according to its own experience and the experience of a neighboring particle. Particle keeps track of its co-ordinates in the problem space which are associated with the best solution/ fitness achieved so far along with the fitness value (*pbest*← *partcle best*). Overall best value obtained so far is also tracked by the global version of the particle optimizer along with its location (*gbest*). There are two versions exist (according to acceleration): (1) Global - At

each time step, the particle changes its velocity (*accelerates*) and moves towards its **pbest** and **gbest** and (2) Local - In addition to **pbest**, each particle also keeps track of the best solution (**lbest/nbest** – neighbour best) attained within a local topological neighbourhood of the particle. The acceleration thus depends on **pbest, lbest**, and **gbest**.

B. Problem solution

Conceptual framework of sense making (Psychological Systems): A psychological system can be thought of as an "information-processing" function. We measure psychological systems by identifying points in psychological space. Usually the psychological space is considered to be multidimensional. A swarm is a large number of homogenous, simple agents interacting locally among themselves, and their environment, with no central control to allow a global interesting behavior to emerge. Swarm-based algorithms have recently emerged as a family of nature-inspired, population-based algorithms that are capable of producing low cost, fast, and robust solutions to several complex problems. This indirect type of interaction is referred to as stigmergy, which essentially means communication through the environment.

"Philosophical Leaps" Required:

i. Individual minds = a point in space
ii. Multiple individuals can be plotted in a set of coordinates
iii. Measuring the individuals result in a "population of points"
iv. Individuals near each other imply that they are similar
v. Some areas of space are better than others Location.

Applying Social Psychology: Individuals (points) tend to move towards each other and influence each other. This is why; individuals want to be in agreement with their neighbors. Individuals (points) are influenced by their previous actions/behaviors and the success achieved by their neighbors. Figure 2 illustrates the working algorithm of particle swarm optimization. First the particle is initialized randomly. Then for each particle a fitness value is calculated. If the current fitness value is better than previous personal best then personnel based is set to current fitness value. The global best is also updated with the best fitness value. After updating the personnel best and global best, particle velocity is calculated using equation (a) of Figure 3 and [particle position is also updated using equation (b) of Figure 3. These steps are repeated until the convergence criteria is met (i.e. optimized solution or maximum iteration).

IV. PRACTICAL WORK AND RESULTS

Human being is a social animal and one of intelligent creation of god. The Human being whose behavior is strongly inspired by and govern my group activity like some animal fish schooling, bird flocking swam optimization. Whose behavior is influence by group .As there is a population which has their own knowledge of intelligence and best fit of cost of their daily need but if some trend can show them a different need and arise some tread which can motivate or can given some general scene about our prediction. Particle Swarm is an evolutionary computation based technique. There is a substantial amount of empirical evidence that human intuitive judgment and decision-making can be far from optimal, and it deteriorates even further with complexity and stress. Because in many situations the quality of decisions is important, aiding the deficiencies of human judgment and decision-making has been a major focus of science throughout history. Disciplines such as statistics, economics, and operations research developed various methods for making rational choices. More recently, these methods, often enhanced by a variety of techniques originating from information science, cognitive psychology, and artificial intelligence, has been implemented in the form of computer programs as integrated computing environments for complex decision making. Such environments are often given the common name of decision support systems (DSS).

```
For each particle
   Initialize particle
END

Do
   For each particle
      Calculate fitness value
      If the fitness value is better than its peronal best
      set current value as the new pBest
   End

   Choose the particle with the best fitness value of all as gBest
   For each particle
      Calculate particle velocity according equation (a)
      Update particle position according equation (b)
   End
While maximum iterations or minimum error criteria is not attained
```

Fig. 2. Working algorithm of PSO

```
Equation (a)
v[] = c0 *v[]
    + c1 * rand() * (pbest[] - present[])
    + c2 * rand() * (gbest[] - present[])
  (in the original method, c0=1, but many
   researchers now play with this parameter)

Equation (b)
present[] = present[] + v[]
```

Fig. 3. Velocity updating of a swarm in PSO

A PSO based algorithm is developed to define the bi-level pricing model. Experiments illustrate that thus PSO based algorithm can achieve a profit increase for buyers or vendors, if they are treated as leader under some situation (see Figure 4).

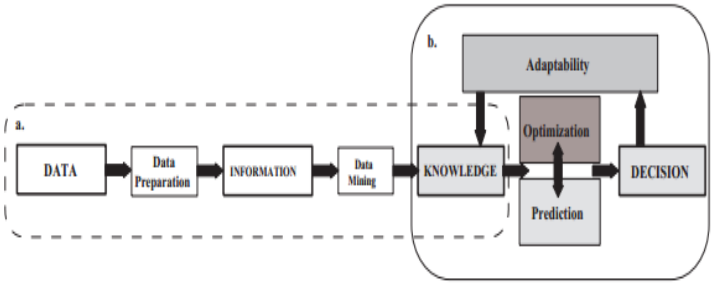


Fig. 4. PSO based Algorithm

Figure 5 shows points representing a set of more than 700 cities from the Mondial database in the SQL Server Management Studio visualizer.

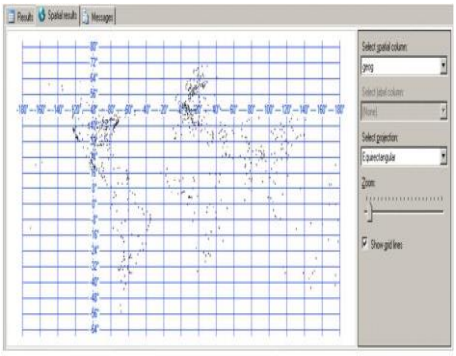


Fig. 5. Output of SQL Server query (A)

Following are the two SQL server queries and its output.
The SQL Server query (A): (see output in Figure 5)
SELECT geog, name
FROM Mondial.dbo.city
WHERE geog IS NOT NULL
The SQL Server query (B): (see output in Figure 6)
SELECT geo, name
FROM Mondial.dbo.cityname
WHERE geo IS NOT NULL
UNION ALL
SELECT geo, cntry_name
FROM SpatialSamples.dbo.cntry

An even better choice is to use a commercial or shareware visualizer that provides map overlay by default, as shown in Figure 5. We see that the Spatial Results tab in MS SQL Server 2008 shows a rowset of more than 600 points with a map overlay.

A. Application to Common Table Expression (CTE)

A CTE can be thought of as a temporary result set and are similar to a derived table in that it is not stored as an object and lasts only for the duration of the query. A CTE is generally considered to be more readable than a derived table and does not require the extra effort of declaring a Temp Table while providing the same benefits to the user. However; a CTE is more powerful than a derived table as it can also be self-referencing, or even referenced multiple times in the same query.

B. Application to Common Table Expression (CTE)

A CTE can be thought of as a temporary result set and are similar to a derived table in that it is not stored as an object and lasts only for the duration of the query. A CTE is generally considered to be more readable than a derived table and does not require the extra effort of declaring a Temp Table while providing the same benefits to the user. However; a CTE is more powerful than a derived table as it can also be self-referencing, or even referenced multiple times in the same query.

The basic syntax structure for a CTE is shown below:
WITH MyCTE
AS (SELECT EmpID, FirstName, LastName, ManagerID
FROM Employee
WHERE ManagerID IS NULL)
*SELECT **
FROM MyCTE
Building a Recursive CTE

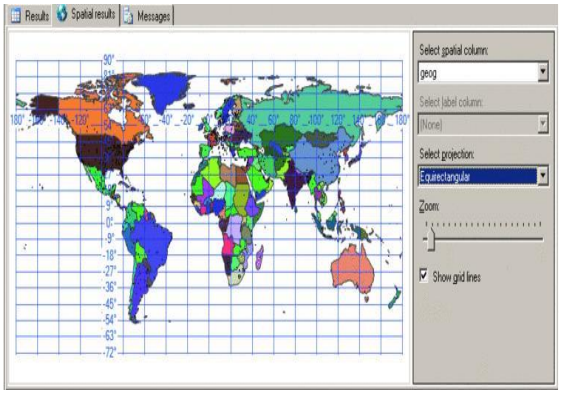


Fig. 6. Output of SQL Server query (B)

In the following examples, we will show how to harness the power of a recursive CTE query by fulfilling a common business requirement, retrieving hierarchical data. By the time the final query is complete you will be able to easily determine how many levels from the top executive each employee is. A recursive CTE requires four elements in order to work properly.

1. Anchor query (runs once and the results 'seed' the Recursive query)
2. Recursive query (runs multiple times and is the criteria for the remaining results)
3. UNION ALL statement to bind the Anchor and Recursive queries together.
4. INNER JOIN statement to bind the Recursive query to the results of the CTE.

The syntax structure is as follows,
WITH MyCTE
AS (SELECT EmpID, FirstName, LastName, ManagerID
FROM Employee
WHERE ManagerID IS NULL
UNION ALL
SELECT EmpID, FirstName, LastName, ManagerID
FROM Employee
INNER
JOIN MyCTE ON Employee.ManagerID = MyCTE.EmpID
WHERE Employee.ManagerID IS NOT NULL)
*SELECT **

FROM MyCTE.

V. CONCLUDING REMARKS

The above literature covers the spatial data and tools that are used to tackle the visualization aspect of spatial database. The tools include SQL SERVER 2008. In addition, the different concepts that these tools use are also mentioned. The effective and advanced applications can be developed using the features of SQL SERVER Spatial like ship tracking system and city bus management system.

REFERENCES

[1] A. Janecek, T. Jordan and F. B. Lima-Neto, "Agent-Based Social Simulation and PSO," *Advances in Swarm Intelligence*. Springer Berlin Heidelberg, pp. 63-73, 2013.

[2] J. Kennedy, "Particle swarm optimization," In *Encyclopedia of Machine Learning*, Springer US, pp. 760-766, 2010.

[3] M. G. Epitropakis, P. P.Vassilis and N. V. Michael, "Evolving cognitive and social experience in particle swarm optimization through differential evolution," *IEEE Congress on Evolutionary Computation* (CEC), 2010.

[4] P. Soni and M. Choudhary, "Hybrid Computational Intelligence for Optimization Based On PSO and DE," *International Journal of Advanced Research in Computer Engineering & Technology*, pp-265, 2012.

[5] H. Ahmed and G. Janice, "*Swarm Intelligence: Concepts, Models and Applications*," Technical Report, 2012-585. Queen's University, Kingston, Ontario, Canada K7L3N6, 2012.

[6] S.R. Dubey and A.S. Jalal, "Robust Approach for Fruit and Vegetable Classification", *Procedia Engineering*, 38, pp. 3449-3453, 2012.

[7] A. E. Yilmaz, "Swarm Behavior of the Electromagnetics Community as regards Using Swarm Intelligence in their Research Studies," *Acta Polytechnica Hungarica*, 2010.

[8] S.R. Dubey and A.S. Jalal, "Detection and Classification of Apple Fruit Diseases Using Complete Local Binary Patterns", In *Third International Conference on Computer and Communication Technology*, pp. 346-351, 2012.

[9] J. P. Gupta, N. Singh, P. Dixit, V. B. Semwal, and S. R. Dubey, "Human Activity Recognition using Gait Pattern," *International Journal of Computer Vision and Image Processing*, vol. 3, no. 3, pp. 31 – 53, 2013.

[10] M. Sati, V. Vikash, V. Bijalwan, P. Kumari, M. Raj, M. Balodhi, P. Gairola and V.B. Semwal, "A fault-tolerant mobile computing model based on scalable replica", *International Journal of Interactive Multimedia and Artificial Intelligence*, 2014.

[11] K.K. Susheel, V.B. Semwal and R.C. Tripathi, "Real time face recognition using adaboost improved fast PCA algorithm," *arXiv preprint arXiv*:1108.1353 (2011).

[12] K.S. Kumar, V.B. Semwal, S. Prasad and R.C. Tripathi, "Generating 3D Model Using 2D Images of an Object," *International Journal of Engineering Science*, 2011.

[13] N. Singh, S.R. Dubey, P. Dixit and J.P. Gupta, "Semantic Image Retrieval by Combining Color, Texture and Shape Features," *In the Proceedings of the International Conference on Computing Sciences*, pp. 116-120, 2012.

[14] V.B. Semwal, V.B. Semwal, M. Sati and S. Verma, "Accurate location estimation of moving object in Wireless Sensor network," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 1, no. 4 , pp. 71-75, 2011.

[15] S.R. Dubey and A.S. Jalal, "Adapted Approach for Fruit Disease Identification using Images," *International Journal of Computer Vision and Image Processing*, vol. 2, no. 3, pp. 44-58, 2012.

[16] S. Singh and V. Bijalwan, "Design Of Wireless Sensor Network Node On Zigbee For Water Level Detection", Vol. 3, No. 8, 2013.

[17] A. Bijalwan and V. Bijalwan, "Examining the Crimninology using Network Forensic", *National Conference USCSTC*, 2013.

[18] V. Bijalwan, V. Kumar, P. Kumari and J. Pascual, "KNN based Machine Learning Approach for Text and Document Mining", *International Journal of Database Theory and Application*, Vol. 7, No. 1, pp. 61-70, 2014.

[19] P. Kumari and V. Pareek, "RAKSHITA- A Novel web based Approach for Protecting Digital Copyrights Using Public Key Digital Watermarking and Human Fingerprints", *International conference on methods and models in computer science*, 2010.

[20] P. Kumari and A. Vaish, "Instant Face detection and attributes recognition" *International Journal of Advanced Computer Science and Applications*, 2011.

[21] P. Kumari and A. Vaish, "A Comparative study of Machine Learning algorithms for Emotion State Recognition through Physiological signal", *Advances in Intelligent Systems and Computing*, Vol. 236, 2013.

[22] P. Kumari and A. Vaish, "Brainwave's Energy feature Extraction using wavelet Transform", proceeding of IEEE SCEECS, 2014.

[23] V.B. Semwal, K.S. Kumar, V.S. Bhaskar and M. Sati, "Accurate location estimation of moving object with energy constraint & adaptive update algorithms to save data", *arXiv preprint arXiv:1108.1321*, 2011.

[24] J.P. Espada et al., "Virtual Objects on the Internet of Things", *International Journal of Interactive Multimedia and Artificial Intelligence*, 2011.

[25] R.G. Crespo, S.R. Aguilar, R.F. Escobar and N. Torres, "Dynamic, ecological, accessible and 3D Virtual Worlds-based Libraries using OpenSim and Sloodle along with mobile location and NFC for checking in", *International Journal of Interactive Multimedia & Artificial Intelligence*, Vol. 1, No. 7, 2012.

[26] K.S. Kumar et al, "Sports Video Summarization using Priority Curve Algorithm", *International Journal on Computer Science & Engineering*, 2010.

[27] R.G. Crespo et al., "Use of ARIMA mathematical analysis to model the implementation of expert system courses by means of free software OpenSim and Sloodle platforms in virtual university campuses", *Expert Systems with Applications*, Vol. 40, No. 18, pp. 7381-7390, 2013.

[28] S.R. Dubey, P. Dixit, N. Singh, and J.P. Gupta, "Infected fruit part detection using K-means clustering segmentation technique", *International Journal of Artificial Intelligence and Interactive Multimedia*, Vol. 2, No. 2, 2013.

Vinay S. Bhaskar has been severed as a PCC member in several International conferences and Journals. He is graduated from IIMS pune and serving birla tyre as HR&PR manager. Previously he served the organization named ITc, Malva Machine Tool etc.

Abhishek Kumar Singh is presently working as Software Engineer, Honeybell, India. He completed his M.Tech form Indian Institute of Information Technology, Allahabad, his major research work Interest in Image Processing, computer sensor network & Pattern Recognition.

Jyoti Dhruw presently working as Assistant Professor, India. She is B.Tech from *CSIT, Durg, Chhattisgrah* and pursing M.Tech from NITTR Chandigrah. Her major research work Interest in Image Processing, Sensor Network, Design and Analysis of Algorithm, Data Structure.

Anubha Parashar received her B.Tech in Computer Science and Engineering from *[4]Maharshi Dayanand University, Rohtak*. Her research interests include Pattern recognition, Cloud Computing, Big Data and Distributed Database.

Mradula Sharma is currently working as assistant professor at JERC, Jaipur and she completed her M.tech. from MNNIT Allahabad. Her majors are data structure, business intelligency.

Robust Lossless Semi Fragile Information Protection in Images

Pushkar Dixit[1], Nishant Singh[2], Jay Prakash Gupta[3]

[1]Faculty of Engineering and Technology Agra College, Agra, India
[2]Poornima Institute of Engineering and Technology, Jaipur, India
[3]Infosys Limited, Pune, India

***Abstract* – Internet security finds it difficult to keep the information secure and to maintain the integrity of the data. Sending messages over the internet secretly is one of the major tasks as it is widely used for passing the message. In order to achieve security there must be some mechanism to protect the data against unauthorized access. A lossless data hiding scheme is proposed in this paper which has a higher embedding capacity than other schemes. Unlike other schemes that are used for embedding fixed amount of data, the proposed data hiding method is block based approach and it uses a variable data embedding in different blocks which reduces the chances of distortion and increases the hiding capacity of the image. When the data is recovered the original image can be restored without any distortion. Our experimental results indicate that the proposed solution can significantly support the data hiding problem. We achieved good Peak signal-to-noise ratio (PSNR) while hiding large amount of data into smoother regions.**

***Keywords* – Data Security, Lossless Data Hiding, Semi Fragile, Encryption, Decryption, Internet Security, Steganography, Information Protection**

I. INTRODUCTION

ONE of the most important issues arising out of the wide spread use of internet is the security of information. Cryptography has long been used to maintain the secrecy of the data. There are a number of algorithms to encrypt and decrypt a message. However, sometimes it is not enough to keep a secret message. It is essential that except for the intended receiver of the message, no one should even come to know that any communication is taking place.

Nowadays there has been a rapid development in the internet and its technology, the individual prefers internet as the primary medium for communication between one parts of the world to another. There are many possible ways to transmit the data across the internet: via e-mails, chats, video streaming, video calling, etc. Internet has made the transmission of the data very simple, fast and accurate. Internet has both its merits and demerits as the intended users can view the information and one who is not intended can also do. Thus in order to make it visible only to the intended users, we must have some method such that people who are not intended may be prevented from viewing information directly. Thus we can say that one of the main problems with sending the data over the internet is 'security threat' posed, in other words personal or confidential data can be stolen or hacked in many ways. Therefore it becomes very important to take data security into consideration. It is one of the most essential factors that need attention during the process of sending or receiving of data.

Before proceeding further it is necessary to understand the conceptual difference between cryptography and steganography. Cryptography conceals information by encrypting it into cipher text and transmitting it to the user using an unknown key, whereas steganography hide the cipher text into a seemingly invisible image or other formats. The word steganography is of Greek origin which means "covered or hidden writing" [1]. Steganography is the art and science of writing hidden messages in such a way that no one, apart from the sender and intended recipient, suspects the existence of the message, a form of security through obscurity. It is therefore a book on magic. It is emerging in its peak because it does not attract anyone by itself.

Encrypting data into some form has been the most popular approach for protecting information, but this protection can be breached with enough computational power. An alternate approach to encrypting data is hiding it by making this information appear to be something else. This way only intended user can receive a true content. In particular, if the data is hidden inside an image then everyone but our intended users or the person it is meant for can view it as a picture which is transmitted. At the same time he/she could still retrieve the true information while the unintended people would view it only as an image.

Data hiding has its application in various areas due to the image being the most common digital media transmitted over the internet. Thus practically, it is very difficult for an unwanted user to masquerade the information that is transmitted over the channel by checking each and every image as it is a very time consuming job. Thus it is quite a nice option to choose images to hide the data and send them over internet and the receiver can easily extract the information from image.

DOI: 10.9781/ijimai.2014.269

Two important properties of steganographic technique are perception and data hiding capacity. Steganography generally exploits human perception because human senses are not trained to look for file that has hidden information inside it. Therefore steganography disguises information from people who try to hack it. Data hiding capacity is the amount of information that can be hidden in the cover object. The cover object means the image that we use for embedding the data and the stego object means the image obtained after embedding the data into cover object.

The different types of steganography techniques are substitution, transform domain, spread spectrum, statistical and distortion techniques and cover generation techniques. Substitution techniques replace the least significant bit of each pixel in the cover file with bits from the secret document. The transform domain techniques hide secret information in the transform space (like frequency domain) by modifying the least significant coefficient of the cover file. Most of the research work done in the area of transform domain embedding is focused on taking the advantage of redundancies in Discrete Cosine Transformation (DCT). Spread spectrum techniques spread hidden information over different bandwidths. Even if the parts of the message are removed from several bands, there would still be enough information present in other bands to make out the message. Statistical techniques change several statistics of the cover file and then split it into blocks where each block is used to hide one message bit. The most obvious limitation to these techniques is that the cover image must be very largely compared to the secret information it is going to carry fixed payload over the image. We can hide large amount of information in multiple files but it could lead to suspicion. Therefore, it is very important to use only one image file to hide the entire secret information.

Each application using information hiding technique has different requirements depending on the purpose of the application. Generally, there are four issues that we encounter while designing the algorithm: perceptibility, payload, security, robustness, and they are common to most applications. Because there are tradeoffs existing between those requirements, it is very challenging to design an algorithm that satisfies all the four requirements.

Hiding a fixed amount of data in the image may give a uniform distribution of data but it makes the image more suspicious for the attackers as the changes in the image are visible. We need to check the image by the means of threshold and make sure it is able to adapt to the amount of data that we want to embed into it. If we do not perform this step then it may result in loss of information or poor embedding of data. Also we have to select such an image which does not have large sharp details.

We present a simplified embedding algorithm based on difference expansion, which is capable of minimizing the distortion of the stego-image presents in the traditional difference expansion algorithms. The main principle underlying the proposed framework is blocks and centralized difference expansion. In the framework, the original cover image is partitioned into continuous non-overlapping blocks. The bits embedded in each block depend on its block size and the image complexity. A new approach is employed to find the image complexity of each image block, and all the blocks are categorized into three levels according to their block intensity values. Finally varying amounts of data are assigned to image blocks at different intensity levels. Although there are three types of blocks in the embedding procedure, only 1 bit is required to record these three blocks. This way, the proposed method can reach a higher hiding capacity while maintaining good visual quality of the stego-image. Our major concern is that the image should not show any visual effects and it carries as much data as possible. Some other approaches can also be utilized with to enhance the algorithm [2-9].

This paper is outlined as follows: section 2 provides some of the core concepts used in image steganography and presents a survey of efforts done by researchers in the past to address this problem; section 3 describes the framework for the scanning the image as well as embedding and extracting the data; section 4 discusses the result and compare it with classical approach; finally section 5 summarizes the novelties, achievements and limitations of the proposed solutions and indicate some future directions.

II. Literature Review

In this section, we focus on the previous work done by several researchers in the area of data hiding, steganography and steganalysis. Data hiding and steganography can be seen as instances of the image security. People have been resorting steganography or information hiding since Greek times. However, digital steganography is a relatively new research field. Since being undetectable is one of the essential requirements for steganographic applications, steganography and steganalysis techniques are evolving in competition with each other.

The aims of improving the original DE (data embedding) proposed by researchers are twofold: first is to make the embedding capacity as high as possible, second is to make the visible distortion as low as possible. To achieve high embedding capacity, the reviewed schemes adopted three different approaches: (i) simplifying the location map in order to increase its hiding capacity, (ii) embedding payload without location map, and (iii) expanding differences more than once which allows more data to be embedded. Meanwhile, the visual quality may be enhanced by: (i) using a predefined threshold T, (ii) selecting smooth areas to embed data, and (iii) using sophisticated classification functions. However, there is a tradeoff between distortions and embedding capacity. If distortion is minimized, lesser data can be embedded. On the

other hand, if the embedding capacity is increased, it results in low visible quality.

Most of the researchers in the field of data hiding or image steganography have considered capacity and robustness as a key for their approach. Some of them have considered both and some of them have considered them individually. In most of the techniques, fragile images are used which is of no use after the extraction of data and it can't be restored to the original state. The major part of the research attention has been paid to the perception part of the topic rather than that of capacity.

Recently, Li et al. [10] proposed a reversible data hiding (RDH) scheme based on two-dimensional difference-histogram modification by using difference-pair-mapping (DPM). First, by considering each pixel-pair and its context, a sequence consisting of pairs of difference values is computed. Then, a two-dimensional difference-histogram is generated by counting the frequency of the resulting difference-pairs. Finally, reversible data embedding is implemented according to a specifically designed DPM. Where, the DPM is an injective mapping defined on difference-pairs. It is a natural extension of expansion embedding and shifting techniques used in current histogram-based RDH methods.

Faragallah [11] proposed quadruple difference expansion-based reversible data hiding method for digital images which is characterized by two aspects. First, reversible data hiding scheme is enhanced to exhibit data hiding in color palette images. Second, the embedding level is improved by using quadruple difference expansion to guarantee the embedding of 2-bit data into color images. But they have not considered the level of the details present in the image and hide 2-bit at each place (i.e. smoother and non-smoother ares) which is a drawback of this approach.

In the this section, we discuss several approaches used by researchers [12-22] with the aim of being aware to the latest research carried out our focus is on those related to the formulated problems in this paper.

A. Literature Survey

The word steganography is originally derived from Greek words which mean ''Covered Writing''. It has been used in various forms for thousands of years. In the 5th century BC, Histaiacus shaved a slave's head, tattooed a message on his skull and the slave was dispatched with the message after his hair grew back [23-25]. In Saudi Arabia at the King Abdulaziz City of science and technology, a project was initiated to translate some ancient Arabic manuscripts into English on secret writing which are believed to have been written 1200 years ago. Some of these manuscripts were found in Turkey and Germany [26].

Color palette based steganography exploits the smooth ramp transition in colors as indicated in the color palette. The LSBs here are modified based on their positions in the palette index. Johnson and Jajodia [23] were in favour of using BMP (24 bit) instead of JPEG images. Their next-best choice was GIF files (256-color). BMP as well as GIF based steganography apply LSB techniques, while their resistance to statistical counter-attacks and compression are reported to be weak. BMP files are bigger as compared to other formats which render them improper for network transmissions. However JPEG images were avoided at the beginning because of their compression algorithm which does not support a direct LSB embedding into the spatial domain.

One of the earliest methods to discuss digital steganography is credited to Kurak and McHugh [27]. They proposed a method which resembles embedding into the 4 LSBs (least significant bits). They also examined image downgrading and contamination which is now known as image based steganography. Provos and Honeyman [24], at the University of Michigan, scrutinized three million images from popular websites looking for any trace of steganography. They have not found a single hidden message. Embedding hidden messages in video and audio files is also possible. Examples exist in [28] for hiding data in music files, and even in a simpler form such as in Hyper Text Markup Language (HTML), executable files (.EXE) and Extensible Markup Language (XML) [29].

Vleeschouwer et al. [30] solved the problem of salt-and-pepper noise artifact by using a circular interpretation of bijective transformation. The proposed algorithm guarantees the coherence of the transformation interpretation and, consequently, ensures total reversibility. To improve the performance of Fridrich et al.'s scheme in terms of message bits, Celik et al. [31] presented a high capacity, low distortion reversible data embedding algorithm by compressing quantization residues. Images can be obtained after a quantization process and then the CALIC lossless compression algorithm is used to get the compressed residues. The remainder of the compression space is used to hide the secret message. In addition, Ni et al. [32] utilizes zero or minimum point of histogram. If the peak is lower than the zero or minimum point in the histogram, it increases pixel values by one higher than the peak values to lower than the zero or minimum point in the histogram. While embedding, the whole image is searched thoroughly. Once a peak-pixel value is encountered, if the bit to be embedded is '1' the pixel is added by 1, else it is kept intact. The algorithm essentially does not follow the general principle of lossless watermarking. The advantages of this algorithm are (i) it is simple, (ii) it always offers a constant PSNR 48.0dB, (iii) distortions are quite invisible, and (iv) its capacity is high. The disadvantage is that the algorithm is time consuming because it searches the image several times.

Tian suggested multiple-layer embedding in order to achieve larger embedding capacity [33]. For example, the second layer embedding would take place in the orthogonal

direction, where the difference image is obtained by performing integer Haar wavelet transform on the embedded image in column direction. If the capacity of the two-layer embedding is still insufficient for the payload, a third layer embedding is needed. One performs integer Haar wavelet transform in row direction again and repeats the embedding operation. Such a process continues until the total embedding capacity is large enough for the payload. However, multiple-layer embedding results in some unexpected problems. First, image quality (in terms of peak signal-to-noise ratio (PSNR)) drops greatly after the first layer embedding due to the use of large differences. Second, the new difference image has smaller embedding capacity than its predecessor. Each layer-embedding progressively decreases the correlation not only in the embedding directions but also of the neighborhood.

In [34], a lot of secret data bits are hidden in a vector. After the difference of that vector is expanded the difference expansion by generalized integer transform to make it work for more than two pixels per vector with k-1 bits of secret data hidden into k pixels. However when it is pixel pair difference expansion or difference expansion for more than two pixels, there is additional information to save this keeps track of the characteristics of a vector. In this case, a location map is needed because it records the characteristics of a vector.

Maniccam and Bourbakis [35] presented a lossless image compression and information hiding scheme. In their methodology, they have performed both lossless compression and encryption schemes which are based on known SCAN patterns generated by the SCAN methodology. This SCAN is a formal language based two-dimensional spatial accessing methodology which can efficiently specify and generate a wide range of scanning paths or space filling curves. This algorithm has lossless image compression and encryption abilities. The only advantage of simultaneous lossless compression and strong encryption makes the methodology very useful but the drawback of the methodology is that compression-encryption takes longer time.

Paulson [36] reported that a group of scientists at Iowa State University were focusing on the development of an innovative application which they call ‘‘Artificial Neural Network Technology for steganography (ANNTS)’’ aimed at detecting all present steganography techniques including DCT, DWT and DFT. The inverse discrete Fourier transform (iDFT) encompasses round-off error which renders DFT improper for steganography applications.

Abdelwahab and Hassaan [37] proposed a data hiding technique in the DWT domain. Both secret and cover images are decomposed using DWT (1st level). Each of which is divided into disjoint 4 X 4 blocks. Blocks of the secret image fit into the cover blocks to determine the best matches. Afterwards, error blocks are generated and embedded into coefficients of the best matched blocks in the HL of the cover image. But the extracted payload is not totally identical to the embedded version as the only embedded and extracted bits belong to the secret image approximation while setting all the data in other sub images to zeros during the reconstruction process.

In [38], authors used a spatial domain technique in producing a finger print secret sharing steganography for robustness against image cropping attacks. The logic behind their proposed work is to divide the cover image into sub-images and compress and encrypt the secret data. The resulting data is then sub-divided in turn and embedded into those image portions. To recover the data, a Lagrange Interpolating Polynomial is applied along with an encryption algorithm. The computational load was high, but their algorithm parameters, namely the number of sub-images (n) and the threshold value (k) were not set to optimal values leaving the reader to guess the values. Data redundancy that they intended to eliminate does occur in their stego-image.

Lin et al. [39] created a method to restore the marked image to its pristine state after extracting the embedded data. They achieved this by applying the pick point of a histogram in the difference image to generate an inverse transformation in the spatial domain. The example shown in their hiding phase section might not be sufficient to verify the accuracy of the algorithm. Some questions remain unanswered such as what happens when we have two peak points instead of one? On which criterion will we base our selection? It is very likely that after the subtraction process we will have some values that collude with the peak value which confuses the extraction of the embedded data.

Wu and Shih [40] presented a GA-based algorithm which generates a stego-image to break the detection of the spatial domain and the frequency-domain steganalysis systems by artificially counterfeiting statistical features. This is the first paper of utilizing the evolutionary algorithms in the field of steganographic systems. Time complexity, which is usually the drawback of genetic based algorithms, is not discussed in this paper. They have only mentioned that the process is repeated until a predefined condition is satisfied or a constant number of iterations are reached.

Raja et al. [41] used wavelet transforms that map integers to integers instead of using the conventional wavelet Transforms. This overcomes the difficulty of floating point conversion that occurs after embedding. Some other approaches also can be employed to improve the performance [42-48].

III. Proposed Methodology

A good image steganography approach aims at concealing the highest amount of data in an image while maintaining its imperceptibility so that its visual quality is not hampered or least affected. The least significant bit scheme is one of the simplest and easily applicable data hiding methods, where bits of secret data are directly embedded in the least significant

bits of each image pixel. In traditional data embedding schemes, the exact original image cannot be recovered after data embedding. Compared with loss prone embedding methods, reversible data embedding methods embed a fixed payload into a digital content in a reversible fashion. After embedding, the image changes very little or looks no different.

Another obvious feature of reversible data embedding is the reversibility, that is, when the digital content has been used for the purposed it was embedded, one can extract the embedded data and restore the original content. There are a number of challenges that must be addressed to perform data hiding in images. The issues that we must keep in mind while designing the algorithm are perceptibility, payload, security and robustness. We must maintain a trade-off among all these and find a better solution to the problems encountered in data hiding.

Steganography techniques aim at secretly hiding data in a multimedia carrier such as text, audio, image or video, without raising any suspicion of alteration to its contents. The original carrier is referred to as the cover object. In this work, we mainly focused on image steganography. Therefore, the term cover object now becomes cover image. Figure 1 illustrates a basic information hiding system in which the embedding technique takes a cover image and a secret image as inputs and produces as output a stego image. Receiver side carry out the extraction process to retrieve the secret message from the stego image sent over the communication links to the receiver.

The proposed approach is comprised of three steps as shown in Figure 1. In the first step secret message is generated and an image is being selected. If the image is too small for the data then another image is selected to make sure that the data gets embedded into the image. In the second step data embedding process is carried out. In the last step data is being extracted by the extraction algorithm and the original image and protected information are recovered.

A. Message generation and image selection

The basic step in data hiding is that first of all we should have a data or secret message to be hidden in any form i.e. it may be in the form of text or any other form. After getting the data we change its form to some digital form (i.e. binary). The algorithm to convert the message into binary array form is as follows:

1. Read and store the characters of a message in an array A.
2. Do for 1 to length of A
   a. Convert each character into its decimal value.
   b. Convert those values from decimal to binary value.
   c. Store these values in an array A'.
3. Store the length of A' in L.

After we successfully converted the secret message into binary form, we know the total number of bits of the message that we need to embed into the image. Next we select an image and scan it to calculate the amount of data it can carry. If the message data is larger than the hiding capacity of the image then select different image otherwise continue with the same. The image that we are using here is called the cover image and the image obtained after embedding the data is known as the stego-image. Now, we have the data and the image in which the data is to be embedded.

B. Data embedding algorithm based on Difference Expansion

There are lots of redundancies in a digital image. If we change some pixel values to some extent, the appearance of the picture is similar to the original one. So, data embedding can be obtained by changing some pixel values in an image. The DE technique [33] reversibly embeds one bit data into two integers, which is explained in Figure 2. The DE technique uses the difference between two pixel values to embed one bit. Assuming there are two grayscale values x = 206, y = 201, we reversibly embed one bit b = 1. First the integer average and difference of x and y are computed,

$$l = \lfloor (x+y)/2 \rfloor \quad (1)$$

$$h = x-y \quad (2)$$

where the symbol $\lfloor \rfloor$ is the floor function meaning "the greatest integer less than or equal to". Next the difference value h is represented in its binary representation.

$$h = 5 = 101_2 \quad (3)$$

Then embedding bit b is appended into the binary representation of h after the least significant bit (LSB), and the new difference value h' is obtained.

$$h' = 101b_2 = 1011_2 = 11 \quad (4)$$

Mathematically, this is equivalent to

$$h' = 2 * h + b = 2* 5 + 1 = 11 \quad (5)$$

Finally the new grayscale values are computed, based on the new difference value h' and the original integer average value l, and new two pixel values x = 209, y = 198 are obtained. After finishing this process, one bit is embedded into the two pixel values.

$$x' = l + \lfloor (h'+1)/2 \rfloor = 203 + \lfloor (11+1)/2 \rfloor = 209 \quad (6)$$

$$y' = l - \lfloor h'/2 \rfloor = 203 - \lfloor 11/2 \rfloor = 198 \quad (7)$$

This method is focused on the data rather than the image as integrity of data is more important. In this technique, we have a text message that is to be hidden in the image and an image that is available in which the data is to be embedded. We consider that the image is of size 'M x N'. We use the image in our computations as grey scale image. Further we need to consider a block of particular size let it be 'm x n'. We also

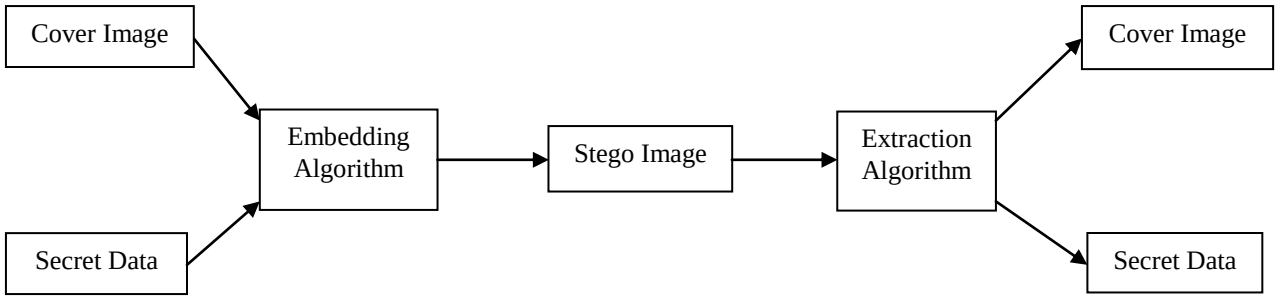


Fig. 1. Data Hiding and Extraction Process

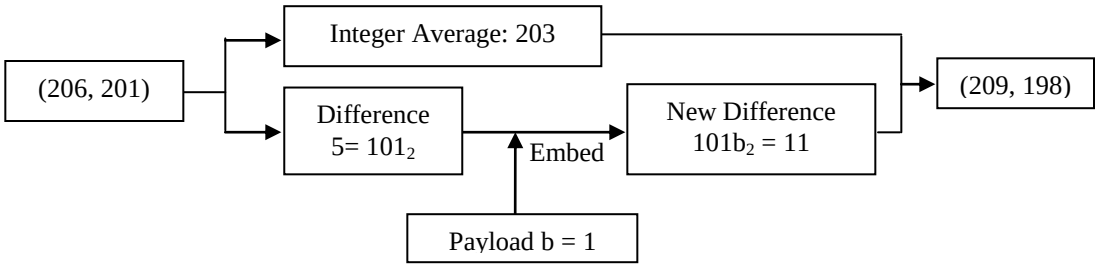


Fig. 2. An example of data embedding by Tian (2003) [33]

select a threshold value 'T' for keeping the image distortion free.

The image is divided into various non-overlapping blocks of size 'm x n'. For each block we have various components in the block. We arrange them in increasing order, select the mid value and subtract all the values from this mid value. Some values give us positive result some give negative. In the next step, we select the maximum difference of all the values and compare it with a threshold in order to decide the amount of data that we can embed into that block. The blocks are named as a, b and c. if a block belongs to type 'a', it means this block is located in a rather smooth area since the difference values are all very small. In this case, it is a very suitable block to hide more secret data bits here 3 bits of data may be embedded. Block 'b' belongs to the region of sharp detail or edges which can acquire only 1 bit of data and in the last block 'c' no secret data is embedded as to avoid distortion. We also use a record bit in order to identify the block after the embedding thus we set the record bit as 1 for all data embedded into block 'a' and record bit '0', for all data embedded into block 'b'. For block 'c', it is not needed. The input to this algorithm is an image and a secret message binary array threshold value. The output is the marked stego image. The process is carried out as shown in Figure 3.

1) Secret Data Embedding Algorithm

1. Segment the cover image into non overlapping blocks of size m x n
2. Label the components as $v_0, v_1, v_2, \ldots\ldots\ldots v_k$ where k= mn-1
3. Find v_{mid} and hence compute differences as $d_0, d_1, d_2, \ldots\ldots\ldots d_{m-1}, d_{m+1}, \ldots. d_{k-1}$
4. Define threshold 'T' and 'd_{max}'.
5. Find the type of block (i.e a, b and c) using T and divide the image into smooth regions and edge regions on the basis of threshold and difference in pixel values.
6. We embed large data in smooth regions and less data in non-smooth regions. The conditions are as follows:
   a. If $d_{max} < T/8$, then it belongs to block 'a' and three bits can be embedded into each block with record bit being 1.
   b. If $T/8 \leq d_{max} < T$, then it belongs to block 'b' and one bit can be embedded to each block with record bit being 0
   c. If $d_{max} \geq T$, then it belongs to the block 'c' and it is a non-embeddable block.
7. Output: an image with data embedded (i.e. stego image).

Data embedding process is illustrated in Figure 3 and can be summarized as follows: first of all non-overlapping blocks of a particular size are extracted from the image; then for a particular block, all the pixel values are sorted and mid value are subtracted from each pixel value. Maximum difference value is selected to decide the type of the block because the magnitude of the local differences can adequately describe the edges of the local regions of the image. According to the details of the edges using maximum difference value and a threshold value we decide the type of the bloc and further the number of bit to be embedded in that block. We are not embedding any information in that block which have very high level of edge details to avoid any degradation in the image.

C. Secret Data and Cover Image Extraction

In this method, now, we have a stego image containing secret information. We consider that this image is also of size 'M x N'. In the next step, we need to consider a block of particular

size let it be 'm x n' which is same as it was for the embedding part otherwise we would not be able to extract the data from it. We have prior knowledge of threshold value 'T' and we use its same value here also. The image is divided into various blocks of size 'm x n'. For each block we have various components in the block. We arrange them in increasing order and select the mid value then subtract all the values from the mid values. Some values give us positive result while some give negative then we select the maximum difference of all the values and compare it with the threshold in order to decide the amount of data we can extract from that block. If a block belongs to type 'a', it means that this block is located in a rather smooth area since the difference values are all very small. In this case, it must contain 3 bits of concealed data in it. We check the record bit, if it is '1' we need to extract the data by subtracting the last 3 bit from the values of the component. Then with the help of the d_{max} we restore the image value to what it was earlier. On the other hand if the block is of type 'b' it contains only 1 bit of data and we extract this data in the same way as we extracted for the block 'a'.

In this situation we only need to extract the LSB (Least Significant Bit) and check the record bit being '0'. For the blocks of type 'c', there is no secret data embedded. The input to this phase is a stego-image and a threshold value and the output is original restored cover image and the secret data. During the extraction process we also notice that we are able to recover the original cover image same as it was before. Thus, this suggests that there is no noise further added during the embedding other than the secret data embedded (see Figure 4).

1) Secret Data Extraction Algorithm

1. Segment the stego image into non overlapping blocks of size m x n
2. Label components as $v_0, v_1, \ldots\ldots\ldots v_k$, where k= mn-1
3. Find v_{mid} and hence compute differences as $d_0, d_1, d_2, \ldots\ldots\ldots d_{m-1}, d_{m+1}, \ldots. d_{k-1}$
4. We already know threshold 'T' and 'd_{max}'.
5. Find the type of block (i.e a, b and c) using T and divide the image into smooth regions and edge regions on the basis of threshold and difference in pixel values.
6. We extract large data from smooth regions and less from non smooth regions. The conditions are as follows
   a. If $d_{max} < T/8$, then it is block of type 'a' therefore three bits are to be extracted from each block with record bit being 1.
   b. If $T/8 \leq d_{max} < T$, then it is 'b' type block and one bit can be extracted from each block with record bit 0.
   c. If $d_{max} \geq T$, then the block belongs to type 'c' and no data can be extracted from such block.
7. Restore cover image with the extracted secret data.

The hidden information extraction approch is described in the Figure 4 and can be summrized as follows: first of all again the image is divided into non-overlapping blocks and for each block the pixel intensities are sorted and subtracted with the mid value in order to find the type of the block using the maximum difference value (i.e. the number of bits to be extracted from a particular block), it should be noted that the maximum difference value will be same as it was previously because intensity differences are invarint to the monotonic intensity change caused by the embedding bits; three, one or none number of bits are extracted according to the type 'a', type 'b' and type 'c' of the block respectively; and finally the information and image both are recovered.

Our approach is able to embed more data in smoother images keeping the distortion not visible at all. The most important fact as well as advantage of using this approach is that we are able to restore the original image back to its initial state (i.e. the state in which it was earlier before data embedding) thus our approach makes sure that no additional noise is added into the image other than the. This is one of the features of the DE (Difference Expansion) technique that the image can be restored to its previous state.

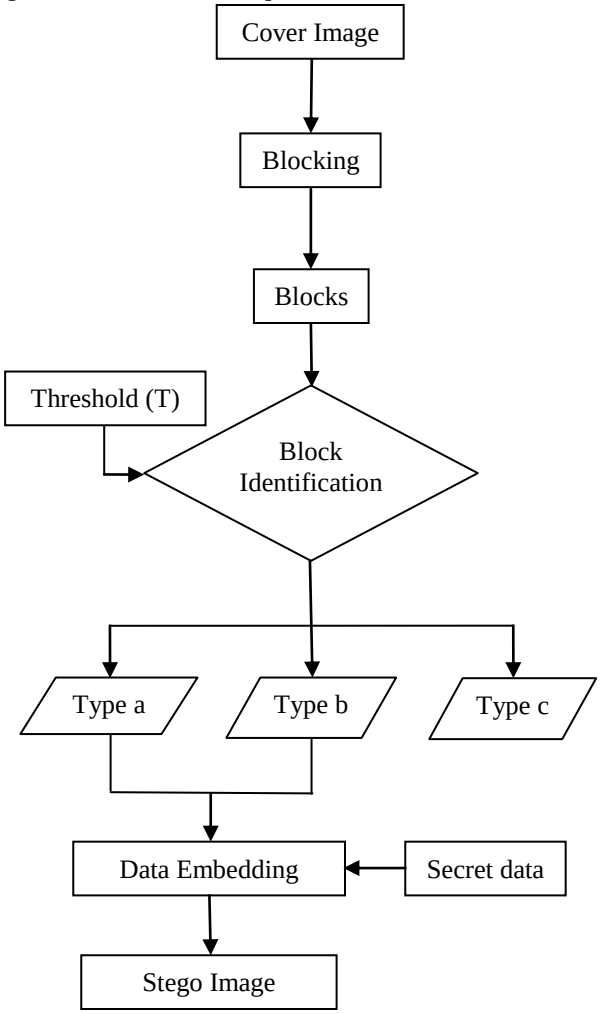


Fig. 3. Data Embedding Flowchart

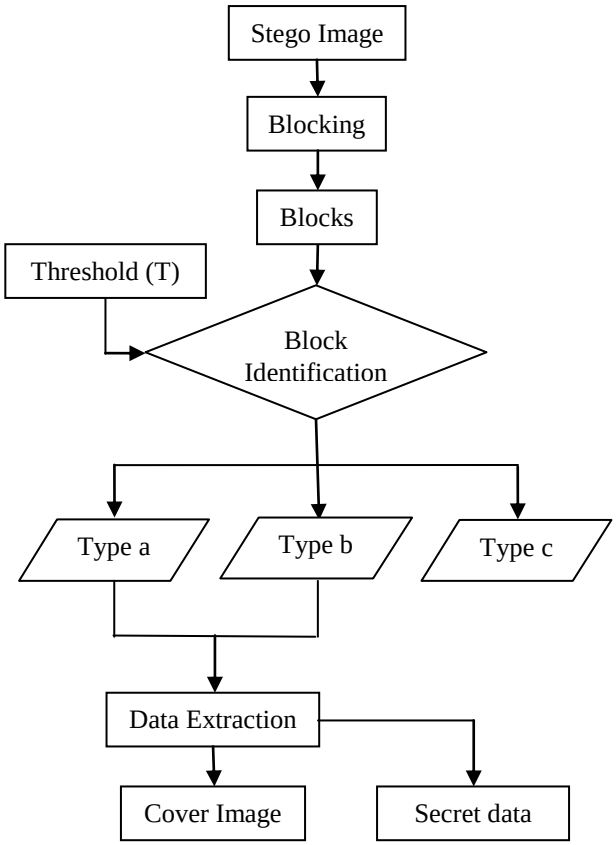


Fig. 4. Data Extraction Flowchart

IV. Results and Discussions

This section analyses the various aspects of the proposed method. Embedding is not prone to any visual attacks as the changes are hardly visible in the image. In addition to this we use an image that has smoother regions and hence it can acquire more data as compared to the one with less smooth region. Here we test our method for the correctness and performance. The PSNR (peak signal to noise ratio) is used to measure the quality of stego image when compared with the cover image. It is calculated using equation 8, where MSE is mean square error given by equation 9.

$$PSNR = 10 \log_{10} \frac{(2^n - 1)^2}{MSE} \qquad (8)$$

Where, MSE is given by

$$MSE = \frac{\sum (\hat{x} - x)^2}{A} \qquad (9)$$

The quality of the image is higher if the PSNR value of the image is high. Since PSNR is inversely proportional to MSE value of the image, the lower MSE value yields higher PSNR value. It means, the better the stego image quality the lower the MSE value will be. Using variable data to be hidden in the image we test the images for the PSNR and the Bit per Pixel (BPP). Where BPP is given by

$$BPP = \frac{n}{MXN} \qquad (10)$$

Where n is the number of bits to be embedded in the image while M and N are the dimensions of the image.

A. Data Set Used

In order to evaluate our proposed, we have used a data set of 20 images (i.e. 20 cover images). All the images used are gray scaled images of size 512 X 512. Figure 5 shows the 20 cover images used in this paper for the experiments. The simulation for the experiment was set up and carried out on a Windows XP Professional with 1.8 GHz dual core processor and 1 GB of RAM.

The proposed approach is implemented on the publicly available MATLAB R2009b. We have further referred the images of the data set as 01.jpg, 02.jpg, 03.jpg…..20.jpg in the Figure 5.

B. Experimental Results

We have obtained the stego image after data embedding (payload size of 61376 bit with BPP of 0.2341) in all cover images of the dataset and recovered all the images after extraction of data as it is shown in Figure 6. First column shows the original input cover images, second column depicts the stego image obtained after data hiding and third column shows the recovered images obtained after data extraction.

In Figure 6, we can see that there is nearly no visual difference among input, stego and recovered images it means our embedding algorithm is robust to noise and able to recover original image accurately. Figure 7 illustrates the quality of stego image with the size of payload data embedded. It shows the graph between PSNR and BPP for all images of the dataset used. From this figure it is clear that if size of the data increases (i.e. BPP increases) quality of the stego image degrades because the value of PSNR decreases.

We are able to achieve high embedding capacity with keeping PSNR in range and also the visual quality of the image is not getting too distorted to visualize as it is depicted by Table 1. This table listed the value of the PSNR obtained for each images after data embedding of varying sizes. The value of the PSNR is highest for 13.jpg in each case of payload data because 13.jpg has largest smoother area (we can see it in the 13th image of Figure 5) and the value of the PSNR is lowest for 14.jpg in each case of payload data because 14.jpg has lowest smoother area(we can see it in the 14th image of Figure 5).

(a) 01.jpg (b) 02.jpg (c) 03.jpg (d) 04.jpg (e) 05.jpg

(f) 06.jpg (g) 07.jpg (h) 08.jpg (i) 09.jpg (j) 10.jpg

(k) 11.jpg (l) 12.jpg (m) 13.jpg (n) 14.jpg (o) 15.jpg

(p) 16.jpg (q) 17.jpg (r) 18.jpg (s) 19.jpg (t) 20.jpg

Fig. 5. 20 cover images of the data set used in this paper

(a) 01.jpg

(b) 02.jpg

(c) 03.jpg

(d) 04.jpg

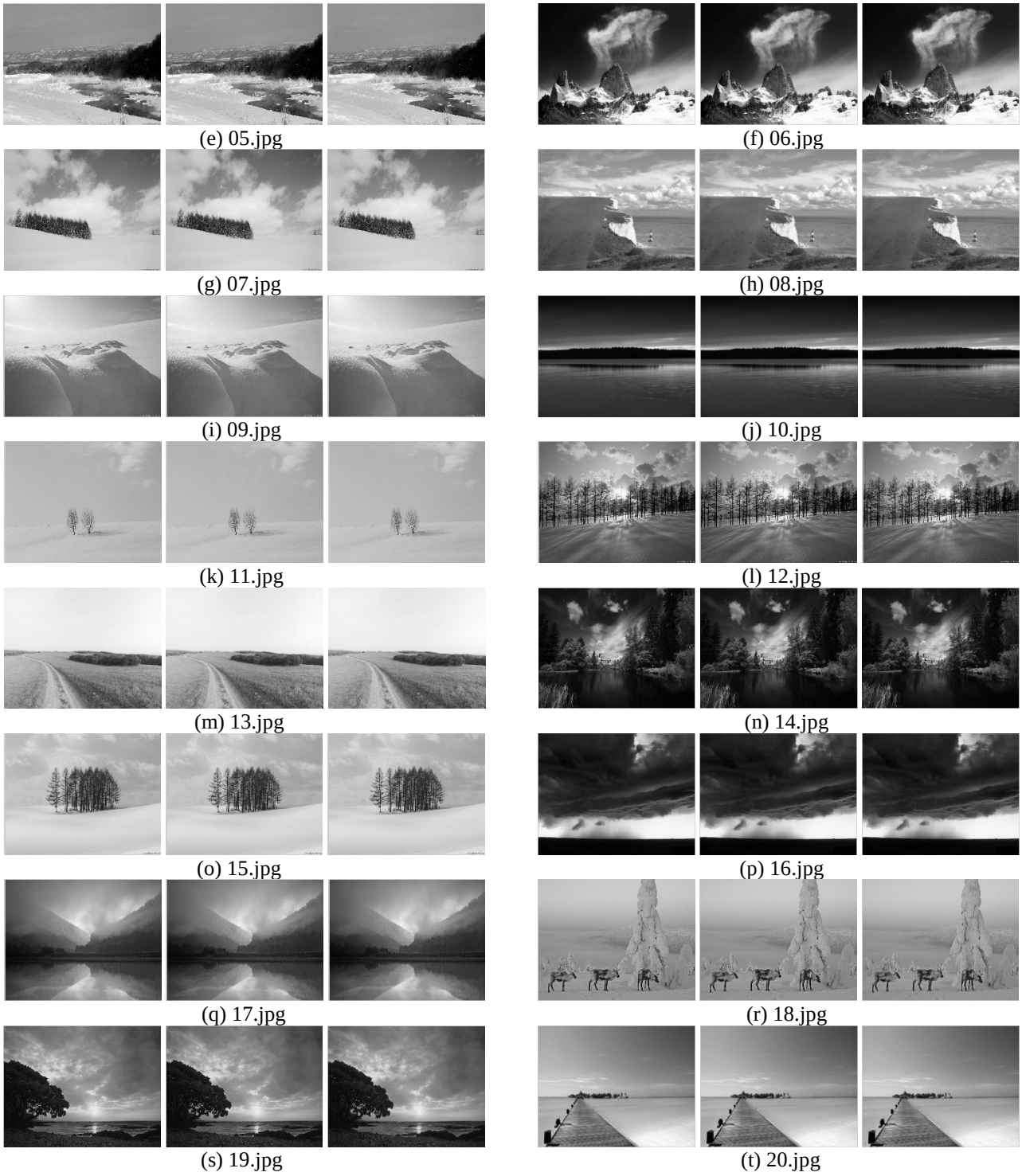


Fig. 6. Cover image (1st column), stego image (2nd column) and restored image (3rd column) for input images.

TABLE I.
RESULTS AFTER TESTING ALL THE IMAGES OF THE DATA SET

Image	Payload size in bits and their corresponding Bit Per Pixel for the images used for testing					
	14320	**29896**	**45008**	**61376**	**75304**	**93568**
	.0546	**.1140**	**.1717**	**.2341**	**.2873**	**.3569**
01.jpg	43.76	40.24	38.36	36.92	35.70	34.38
02.jpg	43.26	39.76	37.36	36.35	35.59	33.53
03.jpg	43.56	40.37	39.36	37.17	34.63	33.55
04.jpg	43.91	40.57	38.55	36.78	35.53	34.15
05.jpg	42.54	39.42	37.59	36.24	35.37	34.31
06.jpg	43.08	38.66	36.69	35.21	34.26	33.42
07.jpg	42.71	39.38	37.72	36.49	35.56	34.40
08.jpg	45.87	42.37	38.99	37.04	35.88	34.57
09.jpg	44.66	41.75	39.97	38.68	37.63	36.39
10.jpg	49.67	46.23	44.44	42.95	41.10	38.14
11.jpg	42.38	39.12	37.31	35.98	35.10	34.17
12.jpg	41.46	38.07	35.78	34.01	32.94	31.85
13.jpg	49.67	46.55	44.74	43.43	42.54	41.60
14.jpg	39.92	36.20	34.18	32.89	32.08	31.18
15.jpg	41.96	38.98	37.38	36.20	35.36	34.19
16.jpg	44.98	40.35	38.24	37.18	36.79	36.41
17.jpg	48.01	44.62	42.31	40.52	39.21	37.70
18.jpg	40.21	38.53	37.23	36.22	35.41	34.53
19.jpg	45.19	41.98	39.95	38.20	36.98	35.36
20.jpg	48.88	45.64	43.83	42.40	41.32	40.23

Table 2 shows the embedding capacity of proposed approach against classical approach using difference expansion. We compared our method with the Tian's approach [33] because it is a benchmark and widely adopted for comparison purposes. Moreover this is method which applies the local differences into consideration for data embedding and we expanded this method using multiple types of embedding blocks. From this table it is clear that the embedding capacity of proposed approach is higher than the classical approach for all the images used for test purpose. The hiding capacity is highest for 13.jpg and lowest for 14.jpg using proposed approach because 13.jpg has large smoother area and 14.jpg has large fine edges. It should be noted that by embedding the data bits the stego image is not same as the original image and some degradation can be seen using the PSNR in Figure 7 and Table 1 but the amount of degradation is less such that it can't be observed visually (see the input and stego images of Figure 6). The amount of degradation is less using our approach because we have not embedded any bit in blocks having more details but we embedded more number of bits in blocks having smooth details. Moreover, if the number of blocks having less details in the image are more then the amount of data using our approach will be more without loss of visual effect (see the number of bits embedded for each images in Table 2). Our experimental results indicate that the proposed solution significantly support the data hiding problem as well as it has higher hiding capacity than earlier approaches. Our algorithm embeds the amount of data according to the details of the image. More data can be attached with less detailed areas of the image (i.e. smoother part) and less data can be attached with the fine detailed areas.

TABLE II.
COMPARISON OF HIDING CAPACITY OF IMAGES BY PROPOSED AND CLASSICAL APPROACH BASED ON THE DIFFERENCE EXPANSION [33]

Image	Capacity by Difference Expansion(in bits)	Capacity by Proposed Approach(in bits)
01.jpg	121421	224966
02.jpg	125815	247795
03.jpg	104685	192327
04.jpg	135254	258903
05.jpg	137055	235224
06.jpg	101839	174839
07.jpg	122550	234142
08.jpg	124331	229485
09.jpg	131545	248200
10.jpg	92291	179724
11.jpg	132201	227796
12.jpg	128108	227957
13.jpg	105920	259619
14.jpg	139650	165627
15.jpg	130403	241323
16.jpg	99237	193444
17.jpg	119110	232369
18.jpg	127461	238708
19.jpg	130935	240258
20.jpg	114030	212238

V. CONCLUSIONS AND FUTURE DIRECTIONS

An image processing based approach is proposed and experimented in this paper for information protection. The proposed approach is comprised of mainly three steps. In the first step, message generation and image selection is performed using the data that is to be embedded. In the second step, we embed the data into the corresponding image. First, image is partitioned into the different number of non-overlapping blocks and then we embedded different number of bits in different blocks of the image according to the types of the block. Three types of blocks are considered in this paper according to the intensity of details of that block i.e. smooth, average and dense and embedded three, one and none bits respectively. In the third step, we extract the data and restore the image to its initial state. An approach similar to embedding one is also employed here to know the type of block. According to the type of block the bits are extracted. We have

used the difference expansion technique to embed the data into the image and to extract the data from the image. The major advantage of using difference expansion technique is that now a large amount of data can be embedded into the image and there is no visible effect on the image, moreover the image can be restored to its initial state, thus our approach shows that no noise is added to the image except the data which is extracted in the extraction phase. Our experimental results indicate that the proposed solution can significantly support the data hiding problem. The future work includes the consideration of RGB color images and videos for information hiding.

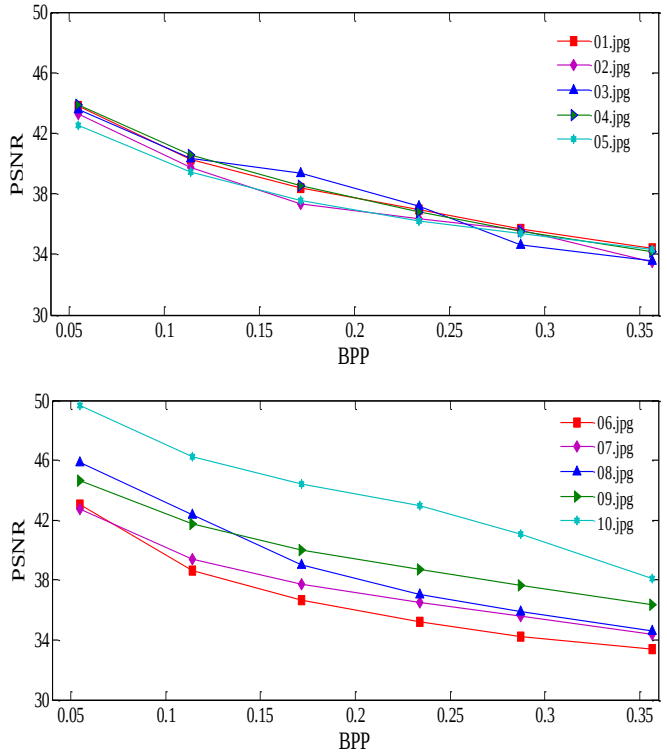
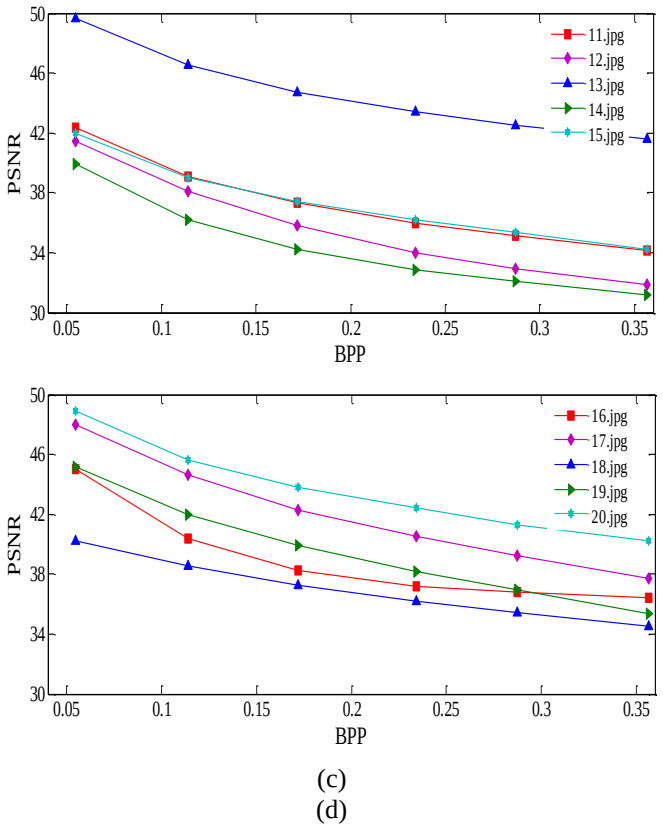


Fig. 7. Graph between PSNR and BPP for (a) images 1-5, (b) images 6-10, (c) images 11-15, and (d) images 16-20



REFERENCES

[1] J.C. Judge, "*Steganography: past, present, future*," SANS Institute publication, 2001.

[2] S. R. Dubey and A. S. Jalal, "Adapted Approach for Fruit Disease Identification using Images," *International Journal of Computer Vision and Image Processing*, Vol. 2, No. 3, pp. 44-58, 2012.

[3] V.B. Semwal, K.S. Kumar, V.B. Semwal and M. Sati, "Accurate location estimation of moving object with energy constraint & adaptive update algorithms to save data," arXiv preprint arXiv:1108.1321, 2011.

[4] W.L. Romero, R.G. Crespo and A.C. Sanz, "A prototype for linear features generalization," *International Journal of Interactive Multimedia & Artificial Intelligence*, Vol. 1, No. 3 2010.

[5] H. Bolivar, A. Pacheco and R.G. Crespo, "Semantics of immersive web through its architectural structure and graphic primitives," *International Journal of Interactive Multimedia & Artificial Intelligence*, Vol. 1, No. 3, 2010.

[6] S.J.B. Castro, R.G. Crespo and V.H.M. García, "Patterns of Software Development Process," *International Journal of Interactive Multimedia & Artificial Intelligence*, Vol. 1, No. 4, 2011.

[7] K.K. Susheel, V.B. Semwal and R.C. Tripathi, "Real time face recognition using adaboost improved fast PCA algorithm," arXiv preprint arXiv:1108.1353, 2011.

[8] S. R. Dubey and A. S. Jalal, "Robust Approach for Fruit and Vegetable Classification," *Procedia Engineering*, Vol. 38, pp. 3449-3453, 2012.

[9] S. R. Dubey, P. Dixit, N. Singh, and J.P. Gupta, "Infected fruit part detection using K-means clustering segmentation technique," *International Journal of Artificial Intelligence and Interactive Multimedia*. Vol. 2, No. 2, pp. 65-72, 2013.

[10] X. Li, W. Zhang, X. Gui and B. Yang, "A Novel Reversible Data Hiding Scheme Based on Two-Dimensional Difference-Histogram Modification," *IEEE Transactions on Information Forensics and Security*, Vol. 8, No. 7, pp. 1091 – 1100, 2013.

[11] O.S. Faragallah, "Quadruple Difference Expansion-Based Reversible Data Hiding Method for Digital Images," *Information Security Journal: A Global Perspective*, Vol. 21, No. 5, pp. 285-295, 2012.

[12] S. R. Dubey, N. Singh, J.P. Gupta and P. Dixit, "Defect Segmentation of Fruits using K-means Clustering Technique," In *the Proceedings of the Third International Conference on Technical and Managerial Innovation in Computing and Communications in Industry and Academia*. 2012.

[13] N. Singh, S. R. Dubey, P. Dixit and J.P. Gupta, "Semantic Image Retrieval Using Multiple," In *proceeding of the 3rd International Conference on Technical and Managerial Innovation in Computing and Communications in Industry and Academia*. 2012.

[14] S. R. Dubey and A. S. Jalal, "Detection and Classification of Apple Fruit Diseases Using Complete Local Binary Patterns," In *proceeding of the Third International Conference on Computer and Communication Technology*, pp. 346-351, 2012.

[15] S. R. Dubey and A. S. Jalal, "Species and variety detection of fruits and vegetables from images," *International Journal of Applied Pattern Recognition*. Vol. 1, No. 1, pp. 108-126, 2013.

[16] N. Singh, S. R. Dubey, P. Dixit and J.P. Gupta, "Semantic Image Retrieval by Combining Color, Texture and Shape Features," In *the Proceedings of the International Conference on Computing Sciences*, pp. 116-120, 2012.

[17] J. P. Gupta, N. Singh, P. Dixit, V. B. Semwal and S. R. Dubey, "Human Activity Recognition using Gait Pattern," *International Journal of Computer Vision and Image Processing*, Vol. 3, No. 3, pp. 31 – 53, 2013.

[18] K.S. Kumar, V.B. Semwal, S. Prasad and R.C. Tripathi, "Generating 3D Model Using 2D Images of an Object," *International Journal of Engineering Science*, 2011.

[19] V.B. Semwal, V.B. Semwal, M. Sati and S. Verma, "Accurate location estimation of moving object in Wireless Sensor network," *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol. 1, No. 4 , pp. 71-75, 2011.

[20] K.S. Kumar, S. Prasad, S. Banwral and V.B. Semwal, "Sports Video Summarization using Priority Curve Algorithm," *International Journal on Computer Science & Engineering*, 2010.

[21] R.G. Crespo, S.R. Aguilar, R.F. Escobar and N. Torres, "Dynamic, ecological, accessible and 3D Virtual Worlds-based Libraries using OpenSim and Sloodle along with mobile location and NFC for checking in," *International Journal of Interactive Multimedia & Artificial Intelligence*, Vol. 1, No. 7, 2012.

[22] C.J. Broncano, C. Pinilla, R.G. Crespo and A. Castillo, "Relative Radiometric Normalization of Multitemporal images," *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol. 1, No. 3, 2010.

[23] N.F. Johnson and S. Jajodia, "Exploring steganography: seeing the unseen," *IEEE Computer*, Vol. 31, No. 2, pp. 26–34, 1998.

[24] N. Provos and P. Honeyman, "Hide and seek: an introduction to steganography," *IEEE Security and Privacy*, Vol. 1, No. 3, pp. 32–44, 2003.

[25] P. Moulin and R. Koetter, "Data-hiding codes," *Proceedings of the IEEE*, Vol. 93, No. 12, pp. 2083–2126, 2005.

[26] S.B. Sadkhan, "Cryptography: current status and future trends," In *the Proceedings of the IEEE International Conference on Information & Communication Technologies: From Theory to Applications*, pp. 417–418, 2004.

[27] C. Kurak and J. McHugh, "A cautionary note on image downgrading," In *the Proceedings of the 8th IEEE Conference on Computer Security Applications*, pp. 153–159, 1992.

[28] C. Hosmer, "Discovering hidden evidence," *Journal of Digital Forensic Practice*, Vol. 1, No. 1, pp. 47–56, 2006.

[29] J.C.H. Castro, I.B. Lopez, J.M.E. Tapiador and A.R. Garnacho, "Steganography in games: a general methodology and its application to the game of Go," *Computers and Security*, Vol. 25, No. 1, pp. 64–71, 2006.

[30] C.D. Vleeschouwer, J.F. Delaigle and B. Macq, "Circular interpretation of bijective transformations in lossless watermarking for media asset management," *IEEE Transactions on Multimedia*, Vol. 5, No. 1, pp. 97–105, 2003.

[31] M.U. Celik, G. Sharma, A.M. Tekalp and E. Saber, "Lossless generalized-LSB data embedding," *IEEE Transactions on Image Processing*, Vol. 14, No. 2, pp. 253–266, 2005.

[32] Z. Ni, Y.Q. Shi, N. Ansari and W. Su, "Reversible data hiding," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 16, No. 3, pp. 354–362, 2006.

[33] J. Tian, "Reversible Data Embedding Using A Difference Expansion," I*EEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No. 8, pp. 890–896, 2003.

[34] A.M. Alattar, "Reversible watermarking using the difference expansion of a generalized integer transform," *IEEE Transactions on Image Processing*, Vol. 13, No. 8, pp. 1147–1156, 2004.

[35] S.S. Maniccam and N.G. Bourbakis, "Lossless image compression and encryption using SCAN," *Pattern Recognition*, Vol. 34, No. 6, pp. 1229–1245, 2001.

[36] L.D. Paulson, "New system fights steganography, News Briefs," *IEEE Computer Society*, Vol. 39, No.8, pp. 25–27, 2006.

[37] A.A. Abdelwahab and L.A. Hassaan, "A discrete wavelet transform based technique for image data hiding," In *the Proceedings of the 25th National Radio Science Conference*, pp. 1–9, 2008.

[38] V.M. Potdar, S. Han and E. Chang, "Fingerprinted secret sharing steganography for robustness against image cropping attacks," In *the Proceedings of the IEEE Third International Conference on Industrial Informatics*, pp. 717–724, 2005.

[39] C.C. Lin, W.L. Tai and C.C. Chang, "Multilevel reversible data hiding based on histogram modification of difference images," *Pattern Recognition*, Vol. 41, No. 12, pp. 3582–3591, 2008.

[40] Y.T. Wu and F.Y. Shih, "Genetic algorithm based methodology for breaking the steganalytic systems," *IEEE Transactions on Systems, Man, and Cybernetics—part B: cybernetics*, Vol. 36, No. 1, pp. 24–31, 2006.

[41] K.B. Raja, S. Sindhu, T.D. Mahalakshmi, S. Akshatha, B.K. Nithin, M. Sarvajith, K.R. Venugopal and L.M. Patnaik, "Robust image adaptive steganography using integer wavelets," In *the Proceedings of the Third International Conference on Communication System Software and Middleware*, Bangalore, India, pp. 614–621, 2008.

[42] S.R. Dubey, "*Automatic Recognition of Fruits and Vegetables and Detection of Fruit Diseases*", M.Tech Thesis, GLAU Mathura, India, 2012.

[43] M. Sati, V. Vikash, V. Bijalwan, P. Kumari, M. Raj, M. Balodhi, P. Gairola and V.B. Semwal, "A fault-tolerant mobile computing model based on scalable replica", *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol. 2, 2014.

[44] V.B. Semwal, S.A. Katiyar, P. Chakraborty and G.C. Nandi, "Biped Model Based on Human Gait Pattern Parameters for Sagittal Plane Movement", *CARE*, 2013.

[45] V.B. Semwal, A. Bhushan and G.C. Nandi, "Study of Humanoid Push Recovery Based on Experiments", *CARE*, 2013.

[46] S.R. Dubey and A.S. Jalal, "Automatic Fruit Disease Classification using Images", *Computer Vision and Image Processing in Intelligent Systems and Multimedia Technologies*, 2014.

[47] D.D. Agrawal, S.R. Dubey and A.S. Jalal, "Emotion Recognition from Facial Expressions based on Multi-level Classification", *International Journal of Computational Vision and Robotics*, 2014.

[48] S.R. Dubey and A.S. Jalal, "Fruit Disease Recognition using Improved Sum and Difference Histogram from Images", International Journal of Applied Pattern Recognition, 2014.

Pushkar Dixit is currently working as an Assistant Professor in Computer Science and Engineering, FET Agra College, Agra, India. He received his MTech in Computer Science from GLA University Mathura, India and BTech in Computer Science from FET Agra College Agra, India. He has 3 year of teaching and research. His research interests include image processing, network security and computer networks.

Nishant Singh received his MTech in CSE in 2012 from GLA University Mathura, India and BTech in CSE in 2010 from B.S.A. College of Engineering and Technology Mathura, India. He has 1 year of teaching and research experience and currently, he is working as an Assistant Professor in Department of Computer Engineering and Applications, Poornima Institute of Engineering and Technology, Jaipur, India. His research interests include image processing, computer vision and pattern recognition.

Jay Prakash Gupta received his MTech in CSE in 2012 from GLA University Mathura, India and BTech in CSE in 2010 from B.S.A. College of Engineering and Technology Mathura, India. He is working as a System Engineer in Infosys Limited, Pune, India. His research interests include image processing, computer vision and pattern recognition.

Bring research out of the lab, into real-life scenarios

- BIG DATA
- CLOUD COMPUTING
- COMPUTER VISION
- ROBOTICS
- INNOVATIONS IN ARTIFICIAL INTELLIGENCE
- INTELLIGENT SECURITY SYSTEMS
- INTELLIGENT MOBILE TECHNOLOGY
- SEMANTIC WEB
- IJIMAI.ORG

YOUR BEST RESEARCH PARTNER

imai solutions

Advanced Research Systems

http://www.imai-solutions.com

•Addresses your research and development issues in a way which develops your reputation and sharpens your competitive edge

•Our own global experience enables us to understand your business needs and empowers us to ask the right questions that lead to meaningful results.

•Opens up new and surprising creative avenues that can nourish and energise your organisation's design thinking

•Enables you to build a relationship with a extended network of researchers.

Paseo de la Castellana 95, Fl 15 - TorreEuropa. 28046 Madrid , Spain. T:+34914187921/F:+34914186999. info@imai-solutions.com

www.elasticbox.com
info@elasticbox.com
blog.elasticbox.com
Twitter: elasticbox

elasticbox

Deploy applications, not servers.

Challenges In Cloud Computing

Automation

- How long to deploy an application?
- What version do I use?
- How do I upgrade applications?

Portability

- How do I change providers?
- What is being used?
- How much does it cost?

Auto-Scaling

- Can my application auto-scale?
- How do I configure auto-scaling?

Disaster Recovery Planning

- Can my application tolerare faults?
- How do I recover my system?

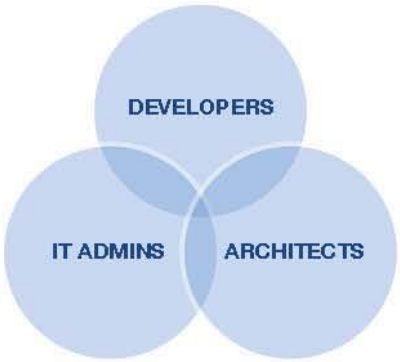
ElasticBox Solution

Automation

- ✔ Automatic Deployments
- ✔ Automatic Configuration
- ✔ Disaster Recovery

Runtime Environment

- ✔ Application Scaling
- ✔ Fault Tolerance
- ✔ Resource Clean-up
- ✔ Replication



Framework Design

- ✔ Architecture Policies
- ✔ Versioning
- ✔ Platform Management

Infrastructure Control

- ✔ Cost Analysis
- ✔ Policy Management
- ✔ Traceability

amazon web services™ · Windows Azure · hp · rackspace · openstack · EUCALYPTUS