Advances in Image And Video Processing

TABLE OF CONTENTS

DISCLAIMER

All the contributions are published in good faith and intentions to promote and encourage research activities around the globe. The contributions are property of their respective authors/owners and the journal is not responsible for any content that hurts someone's views or feelings etc.

Local Visual Feature Detection and Description for Non-Rigid 3D Objects

[1]Kaiman Zeng, [1]Nansong Wu, [2]Lu Wang, and [1]Kang K. Yen

[1]Department of Electrical and Computer Engineering, Florida International University, United States;
[2]School of Science and Technology, St. Thomas University, United States;
kzeng001@fiu.edu; nwu001@fiu.edu; lwang@stu.edu; yenk@fiu.edu

ABSTRACT

Feature extraction is an essential step in various image processing and computer vision tasks, such as object recognition, image retrieval, 3D construction, virtual reality, and so on. Design of feature extraction method is probably the single most important factor in achieving high performance of various tasks. Different applications create different challenges and requirements for the design of visual features. In this paper, we explored and investigated the effectiveness of different combinations of promising local feature detectors and descriptors for non-rigid 3D objects. Different configurations of visual feature detectors and descriptors have been enumerated, and each configuration has been evaluated by image matching accuracy. The results indicated that the scale-invariant feature transform feature detector and descriptor achieved the best overall performance in describing local features of non-rigid 3D object.

Keywords: Feature extraction; local feature detector; Local feature descriptor.

1 Introduction

Feature extraction plays a decisive role in visual content-based image retrieval. A good feature should properly represent the image characteristics, be repeatedly detected in images that capture the same objects/scenes while under different imaging condition, and also be distinctive so that it could distinguish it from other similar images. Besides, an ideal feature should be robust to imaging variations, such as rotation, viewpoint changes, illumination changes and occlusions. There is no universal defined feature, since different problems and different types of applications often have different characteristics. When the application domain changes, it usually requires re-designing feature detector and descriptor to capture features and achieve high performance. A feature is referred to as an interesting point/region in an image. Interesting points/regions are visually salient. Design of feature extraction method is probably the single most important factor in achieving high performance of various computer vision tasks [1]. Given the large number of feature extraction methods researched in the literatures, which feature extraction method is the best for a given application? This question leads us to characterize the available feature extraction methods, so that the most promising methods could be sorted out. In this paper, we concentrated on 3D object under different viewpoint. In particular, we are interested in recognizing 3D objects whose shape is neither fixed nor known a priori. Previous work on object recognition has concentrated on rigid objects of known 3D shape to simplify the task [2, 3]. These approaches have difficulty in dealing with unstructured objects, and thus cannot be applied to more generic categories of objects. Non-rigid object is a significant challenge because of its large variation and deformation within the object

classes. The non-rigid deformation often observes large variation globally. Their local structures are somewhat more invariant to the changes. On that basis, our focus is on non-rigid 3D object recognition with local features.

Image local feature extraction usually consists of two stages: feature detection and feature description. A local feature commonly refers to a local pattern in an image that changes from its direct neighborhood in property or multiple properties of intensity, color, and texture simultaneously. Feature detection is algorithms that compute abstractions of image information and make local decisions at every image pixels whether there is an image feature of a given property type. The resulting features are subsets of the image domain, often in the form of isolated points, continuous curves or connected regions. Once the feature is detected, the local image patch around the feature is extracted and generated as the feature descriptor.

In this paper, the effectiveness of several promising local features on 3D non-rigid objects are explored and investigated. We configure different visual feature detectors and descriptors, and evaluate each configuration in detail. To the best of our knowledge, existing research on the comparison of visual feature detectors and descriptors are conducted for other computer vision tasks. In literature [3] the effectiveness of different visual feature detectors and descriptors are compared for mobile visual search of rigid product like books and CDs. The comparison study in literature [4] is focused on the visual object categorization. Neither of these comparisons targeted the effectiveness of 3D object recognition, the focus of this paper. The performance of different combination of visual feature detectors and descriptors on non-rigid 3D object has not been fully understood. The contribution of our work is filling this knowledge gap. Different combinations of detector and descriptor are enumerated and evaluated by the accuracy of image matching. This accuracy indicates how accurately the repeatable salient local features can be detected, described, and matched from one imaging condition to another.

The paper is organized as the follows. In Section 2, we have a literature review of classic and recent feature extraction techniques. Section 3 discusses the details of the researched feature detectors and descriptors. In Section 4, several experiments of different combination of feature detector and descriptor are conducted on the benchmark datasets. And their performances are compared in the forms of accuracy of image matching. Finally, we conclude comparison results with promising feature extraction techniques and discuss future works in Section 5.

2 The Existing Feature Extraction Techniques

Local feature, representing local patches of an image, has shown promise in many tasks of computer vision, such as image match, object recognition, image registration and so on. Feature detection is utilized as the initial step in local feature extraction algorithms. It is a classic research area in image processing and computer vision. And there are a variety of different types of features, e.g. edges, corners/keypoints, regions of interest and ridges. The corner/keypoint is treated as the same concept since a corner can be not only considered as an intersection of two lines, but also a point that has two different edge directions within a local window of the point. Likewise, a keypoint can be defined as a corner, line endings, a point of local intensity maximum or minimum, or a point on a curve where the curvature is local maximum. As a result, the corner/keypoint detection is mainly divided into edge-based method and gray density based method. Current research is focused on gray density based corner/keypoint detection, since a small degree variation of the target object leads to great difference in edge extraction, and the edge extraction is computationally expensive [5, 6]. Gray density based approach detects the corner/keypoint by calculating the curvature and gradient of points. Moravec

operator, Forstner operator, Harris operator and SUSAN operator are some of the examples. Harris operator [7] is the most classic detector among them. Mikolajczyk takes the scale space theory into consideration and proposes Harris-Laplace detector, which applies Laplace-of-Gaussian (LoG) for automatic scale selection [8]. It obtains scale and shape information and can represent local structure of an image. Lowe applies Difference-of-Gaussian (DoG) filter, an approximate to LoG, in the SIFT algorithm to reduce computational complexity [9]. Also, in order to increase the algorithm efficiency, Hessian Affine, FAST, Hessian-blobs, and MSER are further proposed. In [10], Mikolajczyk et al. extract 10 different keypoint detectors within a common framework and compare them for various types of transformations. Van de Sande extracts 15 types of local color features, and examines their performance on transformation invariance for image classification. Many detection methods are studied seeking a balance between keypoint repeatability and computational complexity [11].

After the keypoint detection, we compute a descriptor on the local patch. Feature descriptors can be divided into gradient-based descriptors, spatial frequency based descriptors, differential invariants, moment invariants, and so on. Among them, the histogram of gradient-based method has been wildly used. The gradient histogram is used to represent different local texture and shape features. The Scale Invariant Feature Transform (SIFT) descriptor proposed by Lowe is a landmark in research of local feature descriptor. It is highly discriminative and robust to scaling, rotation, light condition change, view position change, as well as noise distortion [9]. Since then, it has drawn considerable interests and a larger number descriptors based on the idea of SIFT emerges. SURF [12] uses the Haar wavelet to approximate the gradient SIFT operation, and uses image integral for fast computation. DAISY [13, 14] applies the SIFT idea for dense feature extraction. The difference is that DAISY use Gaussian convolution to generate the gradient histogram. Affine SIFT [15] simulates different perspectives for feature matching, and obtains good performance on viewpoint changes, especially large viewpoint changes. Since SIFT works on the gray-scale model, many color-based SIFT descriptors are proposed to solve the color variations, such as CSIFT, RGB-SIFT, HSV-SIFT, rgSIFT, Hue-SIFT, Opponent SIFT, and Transformed-color SIFT [11, 16, 17]. Most of them are obtained by computing SIFT descriptors over channels of different color space independently; therefore they usually have higher dimension (e.g. 3 × 128 dimension for RGB-SIFT) descriptors than SIFT. The color boosted SIFT introduced in [18] involves the amended color histogram factor based on RGB color space model into the SIFT. It retains sufficient color information and is robust to photometrical variations. Song et al. proposed compact local descriptors using an approximate affine transform between image space and color space [19]. Burghouts et al. performed an evaluation of local color invariants [20].

3 Local Feature Extraction for Non-rigid Object

In this section, we discuss the visual features considered in our work. The feature detectors include Harris, FAST, SIFT, SURF, and BRISK detectors. For the descriptions, the BRISK, SIFT, and SURF feature descriptors are considered. We choose these feature detectors and descriptors for the following reasons. First, Harris detector is the best-known operator around. The SIFT is the most widely used and successful detector developed in recent decade for different computer vision. The FAST, SURF, and BRISK detectors achieve a good balance between the detection performance and computation complexity. Second, the selected feature descriptors have the potential to handle the task of object recognition based on previous studies of other researchers. For instance, Chandrasekhar et al. [3], compared several feature descriptors for visual search application, and reported the SIFT feature descriptor as one of the promising one. The SIFT and SURF are concluded in Lankinen's work [4] as the top two reliable descriptors for visual object classification. The BRISK descriptor is considered in our work because of its big advantage in computation speed.

3.1 Harris detector

Harris detector, proposed by Harris and Stephens [7], is developed from the auto-correlation matrix, also called the second moment matrix. Given an image *I*, an approximation to the local auto-correlation matrix of *I* is computed at every pixel (*x*, *y*):

$$M(x,y) = \begin{bmatrix} \sum_{u,v} w_{u,v} \cdot I_x^2(x_r,x_y) & \sum_{u,v} w_{u,v} \cdot I_x(x_r,x_y)I_y(x_r,x_y) \\ \sum_{u,v} w_{u,v} \cdot I_x(x_r,x_y)I_y(x_r,x_y) & \sum_{u,v} w_{u,v} \cdot I_y^2(x_r,x_y) \end{bmatrix}$$

where I_x and I_y are the partial derivative of image $I(x,y)$ with respect to *x* and *y*. $(x_r, y_r) = (x+u, y+v)$ and *w*(*u*, *v*) is the weighting function. *w*(*u*, *v*) can be a constant or a Gaussian function $\exp(\dfrac{-(u-x)^2-(v-y)^2}{2S^2})$.

M presents the gradient distribution in a local neighborhood of an image pixel (*x*, *y*). The image pixel can be classified into three regions according to the eigenvalues λ_1 and λ_2 of M. If both λ_1 and λ_2 are small, the image pixel belongs to flat region. If λ_1 is far larger than λ_2 or vise versa, the image pixel is located in edge region. If both λ_1 and λ_2 are large and $\lambda_1 \gg \lambda_2$, the pixel is the corner in the image. In order to reduce the computation cost, Harris proposed a cornerness measure that derived from two eigenvalues:

$$c(x,y) = \lambda_1\lambda_2 - a(\lambda_1 + \lambda_2)^2 = \det(M(x,y)) - a \cdot [\text{trace}(M(x,y))]^2$$

where $c(x,y)$ denotes the cornerness measure, $\det(M(x,y))$ is the determinant of $M(x,y)$, and $\text{trace}(M(x,y))$ is the trace of $M(x,y)$. a is the experience constant, typically ranging from 0.04 to 0.06.

Then, non-maximum suppression is performed in a 3×3 or 5×5 neighborhood, and the local maxima of the corneress function forms the corner features of the image.

3.2 Features from Accelerated Segment Test

FAST is a high-speed corner detector developed by Rosten and Drummond [21]. The detection is performed on a discrete Bresenham circle around a candidate image pixel *p*. If there is a set of contiguous pixels at least nine on the circle around *p*, and they are all brighter or darker than the intensity of *p* by a pre-defined threshold *t*, then *p* is considered as a corner candidate. Besides, the algorithm is accelerated with a decision tree to reduce the number of pixels that need to be processed. Subsequently, the following score is computed at each corner candidate to remove the false candidates:

$$s(p) = \max(\sum_{q \in S_+} \left| I_q - I_p \right| - t, \sum_{q \in S_-} \left| I_q - I_p \right| - t)$$

where S_+ is the subset of contiguous pixels that are brighter than *p* by *t* on the circle. S_- is the subset of contiguous pixels that are darker than *p* by *t* on the circle. The corner candidates, who have an adjacent corner with a higher score, will be removed. Then, non-maximum suppression is applied to locate corner features.

3.3   Binary Robust Invariant Scalable Key Points

BRISK, proposed by Leutenegger et al. [22], is a binary local feature detection and description method with very high computational efficiency. The first step is to create a scale space pyramid, generally consisting of 4-layer octave images and 4-layer intra-octave images. Each octave is half-sampled from previous octave, and each intra-octave is down-sampled so that it is located between two octaves. Next, the FAST detector score s is computed at each octave and intra-octave to generate the keypoint candidates. Non-maximum suppression is then performed at each octave and intra-octave so that score s is the maximum within a 3×3 neighborhood; and score s is the largest among the scales above and below. These maxima are then interpolated using a 1D quadratic function across scale spaces and the local maximum is chosen as the scale for the feature found.

Given a set of the detected keypoints, the BRISK descriptor is constructed as a binary descriptor by simple brightness comparison tests. The brightness comparison test is performed on the samples in a pattern. This pattern is defined as N equally spaced locations on circles concentric with the keypoint.

3.4   Scale-Invariant Feature Transform

SIFT, introduced by Lowe [9], is a scale invariant feature detector with highly distinctive feature descriptor. In order to achieve scale invariance, a scale space pyramid of images is first built through convolutions of image I with differences of Gaussians (DoG) at different scales S:

$$DoG_{k,S}(x,y) = G(x,y,kS) - G(x,y,S) = \frac{1}{2p(kS)^2} \cdot e^{-\frac{x^2+y^2}{2(kS)^2}} - \frac{1}{2pS^2} \cdot e^{-\frac{x^2+y^2}{2S^2}}$$

Then, each sample is compared with its 3×3 neighbors at current layer I_n, as well as the 3×3 neighbors from layers above and below (I_{n-1} and I_{n+1}) at the same octave. These local extrema are considered as keypoints. Further, the keypoint location is refined by interpolating the sample points and its direct neighbors. Keypoints with low contrast and small ratio of principal curvatures are removed. Subsequently, the gradient magnitudes and orientations of the remaining keypoints are computed. The orientations are then weighted by a Gaussian window and the gradient magnitude, and the dominant orientations are sorted out from the histogram of the weighted orientations. If multiple dominant orientations exist at a keypoint, for every dominant orientation an additional keypoint are generated.

Now, the located keypoints have been assigned with orientations and scales. A local coordinate system can be defined to compute the SIFT descriptor. A new orientation histogram is computed within a 16×16 local window and then 4×4 sub windows. For each sub window, the orientation histogram is calculated with 8 bins and weighted again by a Gaussian window and corresponding gradient magnitude. This yields the SIFT descriptor of length 128 ($4 \times 4 \times 8$).

3.5   Speeded-Up Robust Features

SURF, designed by Bay et al. [12], is similar to SIFT with faster feature detection and description. SURF detector is developed from the determinant of the Hessian matrix:

$$H(x,y,S) = \begin{bmatrix} \frac{\partial^2}{\partial x^2} G(S) \cdot I(x,y) & \frac{\partial}{\partial x}\frac{\partial}{\partial y} G(S) \cdot I(x,y) \\ \frac{\partial}{\partial x}\frac{\partial}{\partial y} G(S) \cdot I(x,y) & \frac{\partial^2}{\partial y^2} G(S) \cdot I(x,y) \end{bmatrix}$$

It then employs box filters to approximate the second order Gaussian partial derivative for scale space analysis. The score in SURF is defined as:

$$s(x, y, S) = D_{xx}(S) \, \dot{} \, D_{yy}(S) - [0.9 D_{xy}(S)]^2 \, » \det(H(x, y, S))$$

where D_{xx}, D_{yy} and D_{xy} are the convolution of the image using box filters. Constant factor 0.9 is chosen to make the approximate solution closer to $\det(H(x, y, S))$. Then, a non-maximum suppression is performed in a $3 \, \acute{} \, 3 \, \acute{} \, 3$ neighborhood, and the resulted maxima are interpolated across scale spaces to localize the keypoints.

Once the SURF features are localized, the SURF description is addressed in two steps: first, extracting an orientation according to the information from a circular region around the keypoints; second, defining a square region oriented along the formed orientation, and computing the SURF descriptor from the square region. Specifically, the circular region in the first step is convoluted with Haar wavelet along *x* and *y* axes. The radius of the circular neighborhood is decided by the scale, at which the keypoint is detected. So do the sampling step and wavelet response. The wavelet response is then weighted with a Gaussian, and represented as a vector with response strength along *x* and *y* axis. The dominant orientation is determined by the sum of all responses within a rotating square window. Next, this orientation window is further split up to $4 \, \acute{} \, 4$ sub square windows, and the descriptor vector is defined as:

$$v = \left[\begin{array}{cccc} \sum d_x & \sum d_y & \sum |d_x| & \sum |d_y| \end{array} \right]$$

d_x and d_y denote the Haar wavelet responses in *x* and *y* directions for each sub square region. The generated descriptor vector has a length of 64 ($4 \, \acute{} \, 4 \, \acute{} \, 4$).

4 Experiments and Analysis

4.1 Data Set

In order to evaluate the performance of different feature detectors and descriptors, we conducted several experiments of image matching on the benchmark dataset of Oxford Dataset [23]. We also perform experiments on the benchmark dataset of Columbia Object Image Library - COIL 100 [24]. Figure 1 show typical images selected from these datasets. The Oxford dataset has been widely used for evaluating performance of local image descriptors. It contains image pairs under various image transformations, including scale, rotation, image blur, illumination, JPEG compression and viewpoint changes. The dataset also contains ground truth homographies corresponding to the image pairs. Figure 1 (a) shows some image pairs under different image transformations in this dataset. COIL 100 is a database of color images of objects. The objects are placed on a motorized turntable against a black background. The turntable is rotated through 360 degrees to vary object pose with respect to a fixed color camera. Images of the objects are taken at pose intervals of 5 degrees. This corresponds to 72 poses per object and the images are size normalized.

(a)   (b)

Figure 1. Typical images selected from the datasets: (a) Oxford dataset (b) COIL 100 dataset

4.2   Experimental Evaluation and Analysis

In this experiment we implement 5 feature detectors (Harris, BRISK, FAST, SIFT and SURF) and 3 descriptors (BRISK, SIFT, and SURF) in MATLAB. All combinations are evaluated except for the SIFT-BRISK, since the SIFT detector is not compatible with BRISK descriptor.

Table 1. Average Accuracy for Different Combinations of Feature Detectors and Descriptors

Detector	Descriptor		
	BRISK	SIFT	SURF
Harris	0.3351	0.3264	0.3018
BRISK	0.4288	0.4113	0.3907
FAST	0.4637	0.5021	0.4579
SIFT	N/A	0.5173	0.3725
SURF	0.411	0.4556	0.423

The average accuracy of image matching for every combination of feature detectors and descriptors are recorded in Table 1. The results show that the combination of SIFT-SIFT provides the most accurate matching features at matching rate of 0.5173. Following it, the combination of FAST-SIFT achieved comparable performance of matching rate 0.5021. With the same detector, SIFT descriptor and BRISK descriptor performs better than SURF descriptor generally, except for the case of SURF detector.

5 Conclusion

In this paper, we evaluated the effectiveness of different combinations of local feature detectors and descriptors for non-rigid 3D objects. We selected several classic and widely used visual feature detectors (Harris, BRISK, FAST, SIFT, and SURF) and descriptors (BRISK, SIFT, and SURF). The primary difference between this work and the comparison studies of other researchers is that they are targeted in different applications, so that face in different visual characteristics and raise new challenges. It was unclear which feature detection and description methods are best suitable for non-rigid 3D objects. Our evaluation results indicated that the SIFT achieved the best overall performance in describing image local features. This finding could benefit reshaping existing or ongoing other research work based on visual feature, such as non-rigid object visual search. We will use these

findings in the future to tune and design new visual features to improve object recognition accuracy and adapt to different applications.

REFERENCES

[1]   Trier, Ø.D., A.K. Jain, and T. Taxt, *Feature extraction methods for character recognition-a survey.* Pattern recognition, 1996. **29**(4): p. 641-662 %@ 0031-3203.

[2]   Knopp, J., et al. *Hough transform and 3D SURF for robust three dimensional classification*. in *Computer Vision–ECCV 2010*. 2010. Springer.

[3]   Chandrasekhar, V., et al. *Comparison of local feature descriptors for mobile visual search*. in *Image Processing (ICIP), 2010 17th IEEE International Conference on*. 2010. IEEE.

[4]   Lankinen, J., V. Kangas, and J.-K. Kamarainen. *A comparison of local feature detectors and descriptors for visual object categorization by intra-class repeatability and matching*. in *Pattern Recognition (ICPR), 2012 21st International Conference on*. 2012. IEEE.

[5]   He, X.-C. and N.H.C. Yung. *Curvature scale space corner detector with adaptive threshold and dynamic region of support*. 2004. IEEE.

[6]   Mokhtarian, F. and R. Suomela, *Robust image corner detection through curvature scale space.* Ieee Transactions on Pattern Analysis and Machine Intelligence, 1998. **20**(12): p. 1376-1381.

[7]   Harris, C. and M. Stephens. *A combined corner and edge detector*. 1988. Citeseer.

[8]   Mikolajczyk, K. and C. Schmid. *Indexing based on scale invariant interest points*. in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. 2001. IEEE.

[9]   Lowe, D.G., *Distinctive image features from scale-invariant keypoints.* International Journal of Computer Vision, 2004. **60**(2): p. 91-110.

[10]   Mikolajczyk, K. and C. Schmid, *A performance evaluation of local descriptors.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2005. **27**(10): p. 1615-1630.

[11]   Van De Sande, K.E.A., T. Gevers, and C.G.M. Snoek, *Evaluating color descriptors for object and scene recognition.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010. **32**(9): p. 1582-1596.

[12]   Bay, H., T. Tuytelaars, and L. Van Gool, *Surf: Speeded up robust features*, in *Computer vision–ECCV 2006*. 2006, Springer. p. 404-417 %@ 3540338322.

[13]   Tola, E., V. Lepetit, and P. Fua, *Daisy: An efficient dense descriptor applied to wide-baseline stereo.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010. **32**(5): p. 815-830

[14]   Winder, S., G. Hua, and M. Brown. *Picking the best daisy*. 2009. IEEE.

[15]   Morel, J.-M. and G. Yu, *ASIFT: A new framework for fully affine invariant image comparison.* SIAM Journal on Imaging Sciences, 2009. **2**(2): p. 438-469.

[16]   Behmo, R., et al., *Towards optimal naive bayes nearest neighbor*, in *Computer Vision–ECCV 2010*. 2010, Springer. p. 171-184.

[17]   Van De Weijer, J. and C. Schmid. *Applying color names to image description*. 2007. IEEE.

[18]   Zeng, K., N. Wu, and K.K. Yen, *A Color Boosted Local Feature Extraction Method for Mobile Product Search.* Int. J. on Recent Trends in Engineering and Technology, 2014. **10**(2): p. 78-84.

[19]   Song, X., D. Muselet, and A. Trémeau, *Affine transforms between image space and color space for invariant local descriptors.* Pattern Recognition, 2013. **46**(8): p. 2376-2389.

[20]   Burghouts, G.J. and J.-M. Geusebroek, *Performance evaluation of local colour invariants.* Computer Vision and Image Understanding, 2009. **113**(1): p. 48-62.

[21]   Rosten, E. and T. Drummond, *Machine learning for high-speed corner detection*, in *Computer Vision–ECCV 2006*. 2006, Springer. p. 430-443.

[22]   Leutenegger, S., M. Chli, and R.Y. Siegwart. *BRISK: Binary robust invariant scalable keypoints*. in *Computer Vision (ICCV), 2011 IEEE International Conference on*. 2011. IEEE.

[23]   *Oxford Affine Covariant Features dataset*. Available from: http://www.robots.ox.ac.uk/~vgg/research/affine/.

[24]   *Columbia University Image Library*. Available from: http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php.

Discovering Optimized Association Rules based on Image Content

Jyoti Jayprakash Deshmukh and Udhav Bhosle

Department of Electronics and Telecommunication Engineering, Rajiv Gandhi Institute of Technology, Mumbai, University of Mumbai, India;

jyotideshmukh11@gmail.com; udhavbhosle@gmail.com

ABSTRACT

Authors present the concept of image mining, an extension of data mining for discovering semantically meaningful information and image data relationship from a large collection of images. Association rule mining is the process of discovering useful and interesting rules, representing frequent patterns from large datasets, depends on user specified minimum support and confidence values. These constraints lead to exponential search space and dataset dependent minimum support and confidence values. The authors propose an optimization technique for overcoming these problems using multi-fitness function Genetic algorithm and constrained nonlinear minimization and minimax optimization method. Synthetic image set containing geometric shapes and standard MIAS medical image dataset are used to validate the proposed optimization algorithm.

Experimental results show that, Genetic algorithm generates more efficient, effective and strong association rules than constrained nonlinear minimization and minimax optimization method. Genetic algorithm achieves 50% and 90%, constrained nonlinear minimization and minimax optimization method achieves 22% and 74 %, reduction in association rules for synthetic image set and standard MIAS medical image dataset respectively.

Keywords: Image Mining; Association rule mining; Correlation measures; Apriori algorithm; mammogram; Genetic algorithm.

1   Introduction

Advanced image acquisition and storage technologies enabled the development of extensive image database. Every day, a large number of images are generated. Digital libraries are enriched with the addition of various images. Proper analysis of these images reveals useful information to the users. Humans find it difficult to discover underlying knowledge and patterns in large collection of images due to unavailability of effective tools. Image mining system extracts semantically meaningful information (knowledge) from image data automatically. Association rule mining is the process of finding useful and interesting rules from large datasets. These association rules depend on user specified minimum value of support and confidence. This leads to exponential search space and dataset dependent minimum support and confidence value.

Ji Zhang et al. [1] discussed image mining frameworks, current developments, issues in image mining, state-of-the-art techniques. C. Ordonez et al. [2] introduce data mining for knowledge discovery in image database. Authors concentrate on the problem of finding associations rules in 2-dimensional color images. Carson et al. [3] presents image representation to provide a transformation to a small

set of localized coherent regions from a raw pixel data in color and texture space. A framework to achieve higher retrieval efficiency using texture information is reported in [4]. Jawad Nagi et al. [6] proposed a method using morphological processing and seeded region growing algorithm for automated breast profile segmentation for region of interest (ROI) detection. The spatial arrangement of pixel intensities characterizes texture information. Texture is used as a visual feature to retrieve similar patterns from image database [7], [8]. Beyer et al. [9] proposed a technique to reduce number of features for increasing significance of each feature and thus to improve the discrimination accuracy. Agrawal et al. [11] first time discussed the problem of Association rule mining. From any market-basket type database, extraction of some useful and interesting rules is performed by a Pareto based genetic algorithm [16]. Manish Saggar et al. [17] optimize the association rules using Genetic algorithm and predict the rules which contain negative attributes. Wakabi-Waiswa et al. [18] proposed a multi-objective approach for generating optimal association rules using syntactic superiority and transactional superiority quality metrics.

The Authors proposed a method to find frequent significant patterns in a given collection of images using association rule mining. In this paper, association rules are found using Apriori algorithm. Among these rules, many are redundant and uncorrelated which give misleading information and hence exponential search space. So, the authors proposed an optimization method to get non-redundant, highly correlated and strong association rules. It includes optimization of association rules using Genetic algorithm and constrained nonlinear minimization and minimax optimization method.

The Proposed method is validated on synthetic image set containing geometric shapes and standard MIAS medical image dataset. It follows four steps. First step is pre-processing, image segmentation and extraction of objects. In second step, features are extracted from segmented object and feature vector is generated. Third step is formation of transaction database which is given as input to Apriori algorithm for generating association rules. In fourth step, generated association rules are optimized using two different optimization algorithms. First includes optimization of association rules using Genetic algorithm by refining them using multi-fitness function for the interesting measures such as Cosine, All-Confidence, Accuracy and Jaccard to get strong association rules. In second algorithm, the association rules are optimized using constrained nonlinear minimization and minimax optimization method.

The rest of this paper is organised as follows. Section two presents the proposed method. In section three, experimental results are discussed and section four comprises conclusion and scope for future work.
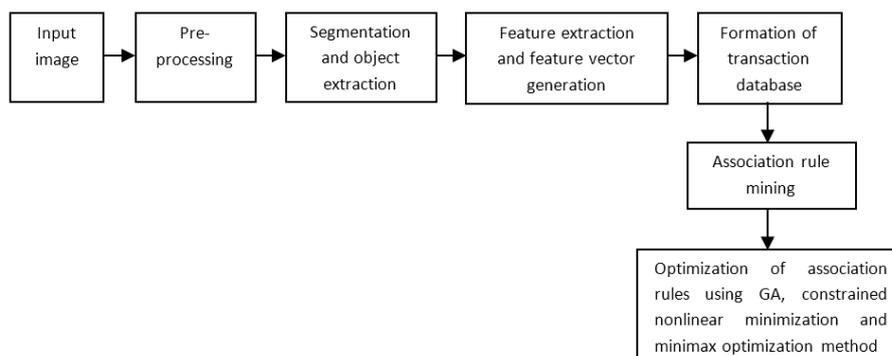
2 Proposed Method

Figure 1: Block diagram of the proposed method

Figure 1 shows the block diagram of the proposed method. Here the novelty is in optimization of association rules using Genetic algorithm and constrained nonlinear minimization and minimax optimization method. It gives non-redundant, highly correlated and strong association rules, and hence reduces search space and consequently improves the processing time of its future applications.

2.1 Image Pre-processing and Segmentation

In pre-processing step for synthetic images, median filtering is used to remove external noise introduced in images. The objects from each image are extracted using segmentation algorithm. The image is segmented into K regions using the spatial grouping of pixels. Using summary information, blob for each connected region having area greater than 2% of the image area is generated. These blobs are called objects [2].

For standard MIAS medical image dataset, median filtering is used to remove digitization noise introduced during image acquisition process. Figure 5(a) shows original mammogram image, figure 5(b) shows median filtered image. Radiopaque artifacts such as labels and wedges in mammogram images are removed using thresholding and morphological operations. Through experimentation, a global threshold with a value of T=100 is selected for transforming the grey-scale mammogram image into binary [0,1] format. Figure 5(c) shows thresholded image. Morphological operations such as dilation, erosion, opening, and closing are carried out on the binary images for suppression of artifacts, labels and wedges. Contrast enhancement is performed on the processed mammogram image. Figure 5(d) shows the resultant image obtained after applying contrast enhancement technique. Pectoral muscles are segmented by using the region growing technique. For region growing a seed is placed inside the pectoral muscle of input mammogram. Segmented image with region of interest (ROI) is shown in figure 5(e) [5], [6].

2.2 Feature Extraction and Feature Vector Generation

For synthetic images, features are extracted from the segmented objects and organized into feature vectors. Ten-dimensional feature vectors are produced which contain summary information about color, texture and area. Statistical texture features as average grey level, uniformity, average contrast, smoothness, third moment and entropy are used. For standard MIAS medical image dataset, Grey Level Co-occurrence Matrix (GLCM) is used to extract the features from the segmented images and are organized into feature vectors. A Co-occurrence matrix M (d, θ) is given by the relative frequency of occurrences of two grey level pixels i and j separated by d pixels in the θ orientation. Co-occurrence matrices are calculated for 0^0, 45^0, 90^0, and 135^0 directions, and for the distances 1, 2, 3, 4 and 5. In this process, 20 matrices of 16 by 16 integer elements per image are produced. For each matrix, seven features as presented in Table 1 are calculated. Feature vector of size 140 elements is produced to represent each image [7]. Also Table 1 gives grey level texture features positions in feature vector. Figure 2 describes the directionality used in GLCM.

2.3 Formation of Transaction Database and Association Rule Mining

For synthetic image set, objects in the input images are identified by segmentation algorithm and labeled using the image query processing algorithm. Similarity function is used to compare the current object with earlier identified and labeled object. Similarity measure is 1 if there is exact match for all desired features and it is 0, if the match becomes worse. If the match is found, same ID is assigned to current object else new ID will be created and assigned. The result of the above step is
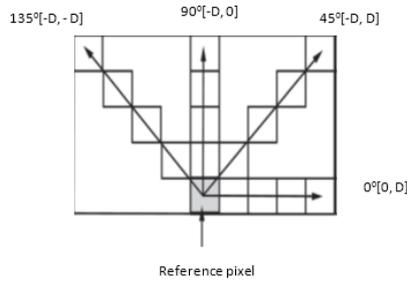
Figure 2: Directions used in GLCM matrix

Table 1. Texture features and their positions in feature vector [10]

Feature	Equation	Meaning	Position
Step	$\sum i \sum j\ P(i,j)$	Distribution	1-20
Variance	$\sum i \sum j\ P(i-j)^2 P(i,j)$	Contrast	21-40
Entropy	$\sum i \sum j\ P(i,j) \log(P(i,j))$	Suavity	41-60
Energy	$\sum i \sum j\ P(i-j)^2$	Uniformity	61-80
Homogeneity	$\sum_i \sum_j\ P(i-j)/(1+|i-j|)$	Homogeneity	81-100
3° Moment	$\sum i \sum j\ (i-j)^3 P(i,j)$	Distortion	101-120
Inv. Variance	$\sum_i \sum_j\ P(i,j)/(i-j)^2$	Inv. Contrast	120-140

a set of transactions, one for each image, which contains the object identifiers for the objects, contained in the image. Thus above created image transaction database has one entry per image containing object ID's. This is given as input to association rule mining algorithm to generate association rules. These object ID's appear in the association rule, showing correlation between these objects, present in the input images.

For standard MIAS medical image dataset, keywords of input mammogram images i.e. benign or malignant and feature vectors are used to build the transaction database. Transaction database has a transaction record for every input image and it is submitted to Apriori algorithm.

2.4 Association Rule Mining

Let R be the set of objects, called as items, and T be a set of data cases. *Association rule* is represented as A→B, where A, B⊂ R and A ∧ B=ϕ. A and B are called, the antecedent and the consequent of the rule respectively. Item set is a set of items, containing the antecedent and the consequent. Support and confidence gives the strength of an association rule. Support value gives how frequently a rule is applicable for a given data set. Value of confidence determines how frequently items in B appear in transactions that contain A [11]. These metrics are defined as,

$$\text{Support, S (A→B) = } P(A \cup B) \tag{1}$$

$$\text{Confidence, C (A→B) = } P(B|A) \tag{2}$$

The rule A→ B has support s in T, if s % of the data case in T contains both A and B. The rule valid for T with confidence c if c % of the data case in T that support A also support B. Association rule mining is to find all rules that have support and confidence value greater than some minimum threshold value for support and confidence, specified by the user.

A Brute-force approach determines the support and confidence of very possible rule for association rule mining. It is an expensive approach, as there exist many rules that can be extracted from a given data set. Let d be the items present in a data set, total number of possible rules extracted are,

$$R = 3^d − 2^{d+1} + 1 \tag{3}$$

Most of the rules are discarded after applying minimum support 20% and confidence 50%, making most of the computations waste. So, it is good to prune the rules early without finding support and confidence values. To improve performance of association rule mining algorithms, the support and confidence requirements should be decoupled [12]. To overcome this, Apriori algorithm [11] divides the problem into two tasks as: frequent item set generation and from the generated frequent item set, extraction of all the high-confidence rules. So, in proposed system, Apriori algorithm is used for mining frequent item sets for generating strong association rules.

2.5   Optimization of Association Rules using Genetic Algorithm

Association rule mining by using Apriori algorithm gives all rules in data which satisfy the minimum support and confidence threshold value. Information interpreted varies according to rule. Many times, rule having high value of support and confidence give conflicting or redundant information, which makes it uninteresting rule. Thus the confidence and support measures are insufficient for filtering out uninteresting association rules. To overcome this limitation, interestingness correlation measures are used for augmenting the support and confidence framework for association rules [14]. This leads to correlation rules of the form,

$$A \rightarrow B \text{ [support, confidence, correlation measures]} \qquad (4)$$

Multi-fitness function Genetic algorithm is used, to optimize association rules, generated in image mining process using Apriori algorithm. It includes interestingness correlation measures as fitness function rather than support and confidence to obtain strong association rules.

2.5.1   Outline of Basic Genetic Algorithm

Random population comprising n chromosomes is generated and the fitness function f(x) for each chromosome x in the population is evaluated. New population is created by repeatedly executing following steps until creation of the new population is finished [13].

In selection step, two parent chromosomes are selected according to their fitness, from a population. Further, crossover is performed between parents with a crossover probability, for creation of new offspring. If there is no crossover, produced offspring is an exact copy of parents. With a mutation probability, new offspring are mutated at each locus (position in chromosome) and placed in a new population. New generated population is used for a further processing of algorithm. When the end condition is satisfied, it stops there, and returns the best solution in current population. Then follow crossover step.

2.5.1   Operators in Proposed Genetic Algorithm

In proposed system, for selecting individuals with respect to fitness function, roulette wheel selection method is used. Single-point crossover method is used. In this crossover method one crossover point is selected, binary string is copied from beginning of chromosome to the crossover point of one parent and rest is copied from the second parent. Mutation gives a chance for flipping a gene within a chromosome. Crossover and mutation probabilities are empirically selected for the proposed algorithm, as 0.85 and 0.006 respectively.

Fitness Function: Multi-fitness function used in the proposed algorithm comprises interestingness correlation measures as All-Confidence, Cosine, Accuracy and Jaccard for refining the association rules. Many correlation measures are listed in the literature. However, we consider All Confidence,

Cosine, Accuracy and Jaccard for Multi-Fitness function. Interestingness correlation measures between two item sets, A and B, are defined as

$$\text{All-Confidence (A, B)} = \min [P \left(\frac{A}{B}\right), P \left(\frac{B}{A}\right)] \tag{5}$$

$$\text{Cosine (A, B)} = \frac{P(A \cup B)}{\sqrt{P(A).P(B)}} \tag{6}$$

$$\text{Accuracy (A, B)} = P(A, B) + P(\overline{A}, \overline{B}) \tag{7}$$

$$\text{Jaccard (A, B)} = \frac{P(A \cup B)}{P(A)+P(B)-P(A \cup B)} \tag{8}$$

Algorithm 1 gives optimized association rules generated using Genetic algorithm with interestingness correlation measures as fitness function.

Algorithm 1: Association rule optimization using Genetic algorithm

Input: n: total number of rules, S = 0, minimum Support threshold, minimum Confidence threshold and avg_opt_threshold_GA = average optimized threshold value using Multi-fitness function Genetic algorithm for Cosine, All-Confidence, Accuracy and Jaccard correlation measures.

Output: Strong association Rules

1. For every rule r_i, and $i \leq n$
2. Compute supp (r_i), Conf (r_i), Cosine(r_i), All-Confidence(r_i), Accuracy(r_i), Jaccard(r_i)
3. If Supp(r_i) > minimum Support threshold and Conf (r_i) > minimum Confidence threshold
   a. If Cosine(r_i), All-Confidence(r_i), Accuracy(r_i) and Jaccard(r_i) > avg_opt_threshold_GA
       i. Print 'Strong Association Rule'
       ii. S + +
   b. End
4. End
5. End For

2.6 Optimization of Association Rules using Constrained Nonlinear Minimization and Minimax Optimization Algorithm

Correlation measures such as cosine, all-confidence, accuracy and jaccard are nonlinear functions. So optimization of individual correlation measure is done by using constrained nonlinear minimization and minimax optimization method. Constrained nonlinear minimization is used to find a minimum of a constrained nonlinear multivariable function. Minimax optimization method is used to find a minimum of the worst-case value of a set of multivariable functions. Algorithm 2 gives optimized association rules generated using constrained nonlinear minimization and minimax optimization algorithm.

Algorithm 2: Association rule optimization using constrained nonlinear minimization and minimax optimization algorithm

Input: n: total number of rules, S = 0, minimum Support threshold, minimum Confidence threshold and avg_opt_threshold_CNMMOA = average optimized threshold value using constrained nonlinear minimization and minimax optimization algorithm for Cosine, All-Confidence, Accuracy and Jaccard correlation measures.

Output: Strong association Rules

1. For every rule r_i, and $i \le n$
2. Compute supp (r_i), Conf (r_i), Cosine(r_i), All-Confidence(r_i), Accuracy(r_i), Jaccard(r_i)
3. If Supp(r_i) > minimum Support threshold and Conf (r_i) > minimum Confidence threshold
   a. If Cosine(r_i), All-Confidence(r_i), Accuracy(r_i) and Jaccard(r_i) > avg_opt_threshold_CNMMOA
      i. Print 'Strong Association Rule'
      ii. S + +
   b. End
4. End
5. End For

3 Experimental Results and Discussions

To validate the proposed method, experiment is performed on two different data sets, carried on MATLAB environment.

3.1   Experiment I- The Synthetic Image Dataset

In this experiment, set of co-related synthetic images are selected. The objects from each image are extracted using segmentation algorithm and scale color selection is estimated. Ten-dimensional feature vectors are produced and these vectors contain summary information about color, texture and area. Statistical texture features as average grey level, uniformity, average contrast, smoothness, third moment and entropy are used. Using Expectation Maximization (EM) method, several clustering of feature vectors are produced and decided the best using Minimum description length principle. The image is segmented into K regions using the spatial grouping of pixels. Using summary information, blob for each connected region having area greater than 2% of the image area is generated. These blobs are called objects. Objects in the images are identified and labeled using the image query processing algorithm. Similarity function is used to compare the current object with earlier identified and labeled object. Similarity measure is 1 if there is exact match for all desired features and it is 0, if the match becomes worse. If the match is found, same ID is assigned to current object else new ID will be created and assigned to it. The result of the above step is a set of transactions, one for each image, which contains the object identifiers for the objects, contained in the image. This image transaction database is used as input to association rule mining algorithm to generate association rules.

10 representative images are considered for this experiment [2]. Figure 3 shows the original image on the left with different geometric shapes with white background. The images are the input data for our program.

By using image mining algorithm as discussed in section two, the various objects (blobs) present in these images are segmented. These blobs are labeled using similarity measure based on color and texture features. The same label is given to similar blobs based on their color and texture features. Color standard deviation is set to 0.29 and contrast standard deviation to 0.5 for object identification and matching. These parameters are tuned after several experiments. The results obtained are treated as transactions for association rule mining.
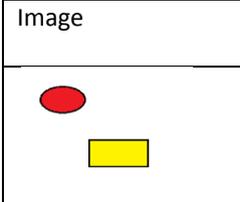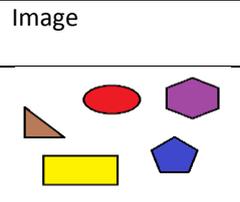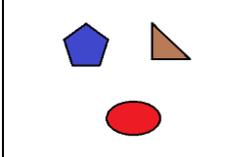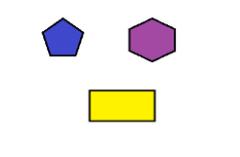
Image	Objects Identified	Image	Objects Identified
	-> 1,2		-> 1,2,3,6,7
	-> 1,3,6		-> 2,6,7
	-> 1,3,7		-> 6,7
	-> 1,2,3		-> 1,3,6
	-> 2,3,6,7		-> 1,2,3,6
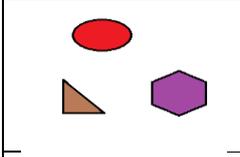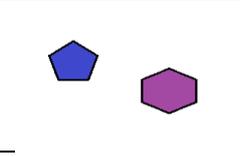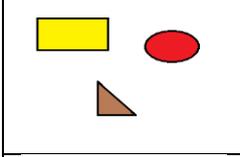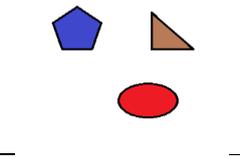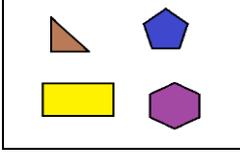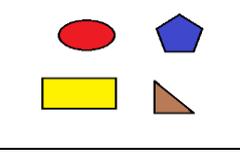
Figure 3: Images and objects identified [2]

Rules Generated:

Parameters are:

Support: 20%

Confidence: 70%

Number of records: 10

Number of Association Rules Generated: 18

1. {3} → {1}   S=60%   C=86%
2. {1}→ {3}   S=60%   C=86%
3. {6} → {3}   S=50%   C=71%
4. {3} → {6}   S=50%   C=71%
5. {1,6} → {3}   S=30%   C=100%
6. {3,6} → {1}   S=40%   C=80%
7. {7} → {6}   S=40%   C=80%
8. {2,7} → {6}   S=30%   C=100%
9. {2,3} → {1}   S=30%   C=75%
10. {1,2} → {3}   S=30%   C=75%
11. {2,6} → {3}   S=30%   C=75%
12. {2,3} → {6}   S=30%   C=75%
13. {6,7} → {2}   S=30%   C=75%
14. {2,6} → {7}   S=30%   C=75%
15. {1,2,6} →{3}S=20%   C=100%
16. {3,6,7} →{2}S=20%   C=100%

17. {2,3,7} →{6}S=20%   C=100%

18. {1,7} → {3} S=20%   C=100%

Let's analyze some of the above rules. The first rule {3} → {1} state that if there is triangle in the image then there is also a circle. This rule has support 60% and confidence 86% which are greater than minimum threshold of support and confidence. The rule {1, 6} → {3} means that if circle and pentagon is present in the image, then triangle is also present with support 30% and confidence 100%. The rule {1, 2, 6 } → {3} means that if circle, square and pentagon is present in the image, then triangle is also present with support 20% and confidence 100%.

3.1.1   Optimization of Association Rules using Genetic Algorithm

Correlation measures Cosine, All-Confidence, Accuracy and Jaccard are calculated for these rules, as listed in Table 2. Multi-fitness function Genetic algorithm is used to achieve strong association rules. Interestingness correlation measures such as Cosine, All-Confidence, Accuracy and Jaccard are selected empirically, as fitness function, as they show feasible results to refine the rules generated in image mining.

Table 2. Support, confidence and correlation measure

Rule No	Support	Confidence	Correlation Measure			
			Cosine	All-Confidence	Accuracy	Jaccard
1	60	86	0.8571	0.8571	0.8	0.75
2	60	86	0.8571	0.8571	0.8	0.75
3	50	71	0.7142	0.7142	0.6	0.555
4	50	71	0.7142	0.7142	0.6	0.555
5	40	100	0.7559	0.5714	0.4	0.5714
6	40	80	0.6761	0.5714	0.4	0.5
7	40	80	0.6761	0.5714	0.6	0.5
8	30	100	0.654	0.4285	0.3	0.4285
9	30	75	0.5669	0.4285	0.4	0.375
10	30	75	0.5669	0.4285	0.4	0.375
11	30	75	0.5669	0.4285	0.3	0.375
12	30	75	0.566	0.4285	0.3	0.375
13	30	75	0.6123	0.5	0.3	0.4285
14	30	75	0.67	0.6	0.3	0.5
15	20	100	0.5345	0.2857	0.2	0.2857
16	20	100	0.5773	0.333	0.2	0.333
17	20	100	0.5345	0.2857	0.2	0.2857
18	20	100	0.5345	0.2857	0.2	0.2857

It is observed that number of strong association rules generated depend on support, confidence and optimized value of interestingness correlation measures such as Cosine, All-Confidence, Accuracy and Jaccard. These correlation measures are linearly independent of each other. So optimization of individual correlation measure is carried out by using Genetic algorithm. By using Genetic algorithm, optimized values for Cosine, All-Confidence, Accuracy and Jaccard are listed in Table3. By using Multi-Fitness function Genetic algorithm, average optimized threshold value for correlation measures is 0.4. Rules satisfying average optimized threshold criteria are considered to be strong rules. Table 3 lists number of optimized association rules obtained by using interestingness correlation measures as fitness function of Genetic algorithm.

Table 3. Correlation measures and number of optimized rules by GA

Interestingness correlation measures as fitness function parameter of GA	Optimized value	No of optimized rules Using GA
Cosine	0.5741	11
all-confidence	0.4986	9
Accuracy	0.21	14
Jaccard	0.3619	14

Optimized association rules by using multi-fitness function Genetic algorithm are:

1. {3} → {1}   S=60% C=86%
2. {1} → {3}   S=60% C=86%
3. {6} → {3}   S=50% C=71%
4. {3} → {6}   S=50% C=71%
5. {1,6} → {3} S=30% C=100%
6. {3,6} → {1} S=40% C=80%
7. {7} → {6}   S=40% C=80%
8. {2,3} → {1} S=30% C=75%
9. {1,2} → {3} S=30% C=75%

3.1.2 Optimization of Association Rules using Constrained Nonlinear Minimization and Minimax Optimization Method

Correlation measures such as cosine, all-confidence, accuracy and jaccard are nonlinear functions. So optimization of individual correlation measure is done by using constrained nonlinear minimization and minimax optimization method. Optimized value for cosine, all-Confidence, accuracy and jaccard are listed in Table 4. For constrained nonlinear minimization and minimax optimization method, average optimized threshold value obtained for correlation measures is 0.225. Rules satisfying average optimized threshold criteria are considered to be strong rules. Table 4 lists number of strong association rules obtained by optimization of interestingness correlation measures using constrained nonlinear minimization and minimax optimization method.

Table 4. Correlation measures and number of optimized rules by constrained nonlinear minimization and minimax optimization method

Interestingness correlation measures	Optimized value	No of optimized rules
Cosine	0.1	18
all-Confidence	0.50	9
Accuracy	0.20	18
Jaccard	0.1	18

Optimized association rules by using constrained nonlinear minimization and minimax optimization method are:

1. {3}→{1}   S=60% C=86%
2. {1}→{3}   S=60% C=86%
3. {6}→{3}   S=50% C=71%
4. {3}→{6}   S=50% C=71%
5. {1,6}→{3} S=30% C=100%
6. {3,6}→{1} S=40% C=80%
7. {7}→{6}   S=40% C=80%
8. {2,7}→{6} S=30% C=100%
9. {2,3}→{1} S=30% C=75%

10. {1,2}→{3}   S=30%   C=75%

11. {2,6}→{3}   S=30%   C=75%

12. {2,3}→{6}   S=30%   C=75%

13. {6,7}→{2}   S=30%   C=75%

14. {2,6}→{7}   S=30%   C=75%

Figure 4 explains comparison of optimized association rules using Genetic algorithm, constrained nonlinear minimization and minimax optimization method. It shows that Genetic algorithm achieves better optimization of association rules than constrained nonlinear minimization and minimax optimization method. Thus, Genetic generates more efficient, effective and strong association rules than constrained nonlinear minimization and minimax optimization method.
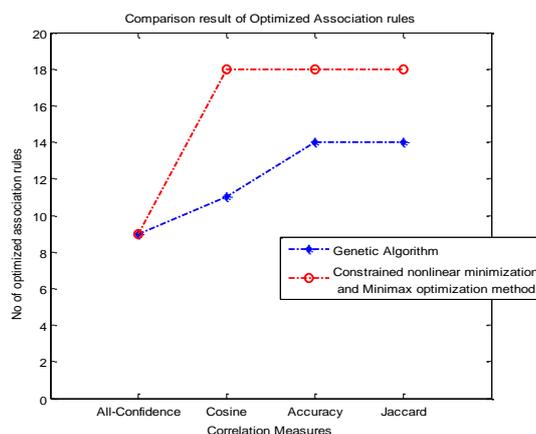
Figure 4: Comparison of optimized association rules using Genetic algorithm, constrained nonlinear minimization and minimax optimization method

3.2   Experiment II- Mammogram Image Dataset: MIAS

Mammography Image Analysis Society (MIAS) database is used to test proposed optimization method. The MIAS database is built by Suckling et al. [15], and is openly available for scientific research. MIAS dataset provide appropriate information based on types of background tissues, and the class of abnormalities present in the mammograms. The class of abnormality consists of normal-abnormal class, and again based upon the severity of abnormality; the abnormal class is divided into two sub-classes such as benign and malignant. The MIAS database contains 322 images, which are categorized into three according to tissue types like fatty, fatty-glandular and dense-glandular. Out of 322 images, 208 images are normal, 114 images are abnormal, and again among abnormal images the numbers of benign and malignant types are 63 and 51 respectively. All the abnormal images are considered for our experiment from this database. Results of segmentation step are shown in figure 5(a-e).

(a)   (b)   (c)   (d)   (e)
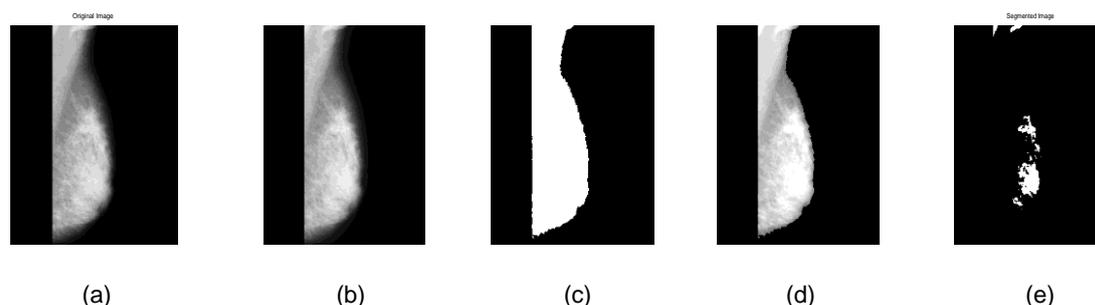
Figure 5: Mammogram segmentation process results; (a) original mammogram; (b) filtered image after noise removal; (c) Thresholded image; (d) image after contrast enhancement; (e) final segmented image

Textual feature are extracted from the segmented mammogram images (ROI) by using Grey Level Co-Occurrence Matrix (GLCM) method and these features are organized into feature vectors. Table 5 gives feature vector generated for MIAS database. Feature vector and the keyword of the input mammogram images i.e. benign or malignant are submitted to transaction database. Table 6 shows the transaction database for MIAS database, where first column presents class of image i.e. benign or malignant, is represented by 1001 and 1002 respectively. Second column onwards present feature value interval label. Unique label for each interval are assigned, which increases sequentially for next interval. The transaction representations of all the images in the training set are submitted to Apriori algorithm to generate association rules. At output, 14232 association rules are obtained by using Apriori algorithm for MIAS database. Examples of association rules mined are:

5,129 ->1001 i.e. Benign Image (Support=13% and Confidence=100%)

2, 25, 47 ->1002 i.e. Malignant image (Support=9% and Confidence=100%)

First rule explains that the image having the feature value interval label as 5 and 129 tend to be a benign image. Second rule explains that the image having the feature value interval label as 2, 25 and 47 tend to be a malignant image.

Next, multi-fitness function Genetic algorithm and constrained nonlinear minimization and minimax optimization algorithm are applied on all association rules generated by Apriori algorithm in earlier step to determine strong, effective and highly correlated association rules. For every rule support, confidence, and correlation measures as Cosine, All-Confidence, Accuracy and Jaccard are calculated. Average optimized threshold value for correlation measures Cosine, All-Confidence, Accuracy and Jaccard are 0.4 and 0.225 by using genetic algorithm and constrained nonlinear minimization and minimax optimization algorithm respectively. Rules satisfying average optimized threshold criteria are considered to be strong rules. Table 7 shows number of association rules obtained from Brute-force approach, Apriori algorithm and by both optimization algorithms.

Table 5. Feature vector generation

Image No. (1-114)	Features (1-140)											
	1	2	3	4	…	134	135	136	137	138	139	140
1	65792	65535	65278	65021	…	2.9111	3.07555	3.05777	3.08888	3.13777	3.16888	3.13777
2	65792	65535	65278	65021	…	0.59555	0.71111	0.82222	0.44000	0.54666	0.62666	0.66666
3	65792	65535	65278	65021	…	0.93777	1.07555	1.15111	0.74666	1.13333	1.50666	1.89777
4	65792	65535	65278	65021	…	3.66222	3.80444	3.89333	3.88888	4.01777	3.94666	3.73333
5	65792	65535	65278	65021	…	5.73777	6.27555	6.67111	4.16888	5.25777	6.02666	6.54666
6	65024	64770	64516	64262	…	3.84888	4.16888	4.34666	3.63555	3.68444	3.81333	4.07111
7	65792	65535	65278	65021	…	5.27555	5.89333	6.35111	3.56888	4.74666	5.43555	6.09333
8	65792	65535	65278	65021	…	2.43111	2.57333	0	0	0	0	0
9	65792	65535	65278	65021	…	2.96444	3.28000	1.19555	0.48888	0.60000	0.64888	0.65777
10	65792	65535	65278	65021	…	3.07555	3.05777	0	0	0	0	0
11	65792	65535	65278	65021	…	1.04000	1.20000	2.57333	1.82222	2.19111	2.70222	3.09777
12	65792	65535	65278	65021	…	3.10666	3.36888	3.28000	2.21777	2.89333	3.29333	3.53333
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
114	65792	65535	65278	65021	…	4.88888	5.35555	3.32444	4.49333	5.08000	5.28000	4.37333

Table 6. Transaction database

Image Class	Feature value interval label											
1001	2	26	50	74	120	144	168	170	194	218	242	266
1001	7	31	56	75	107	137	160	176	200	223	248	267
1001	11	31	56	85	112	137	160	179	202	223	248	277
1001	21	43	64	91	117	125	152	189	211	235	256	283
1001	20	45	70	95	99	123	147	188	211	237	262	287
1001	20	43	67	92	97	125	149	188	211	235	259	284
1001	19	43	67	94	100	125	149	187	211	235	259	286
1001	2	26	50	74	120	144	168	170	194	218	242	266
1001	8	31	56	75	105	137	160	176	202	223	248	267
1001	2	26	50	74	120	144	168	170	194	218	242	266
1001	15	37	62	89	103	131	154	183	207	229	254	281
1001	17	41	64	91	103	127	152	185	208	233	256	283
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
1002	13	34	59	88	111	134	157	181	205	226	251	280
1002	2	26	50	74	120	144	168	170	194	218	242	266

Table 7. Association rule optimization result

Dataset used	Brute-force approach	Apriori Algorithm	Optimization by constrained nonlinear minimization and minimax optimization algorithm	optimization by Genetic Algorithm
Synthetic image dataset	180	18	14	9
MIAS dataset	523250	14232	3702	1410

4 Conclusion

The authors proposed the association rule optimization method using two different optimization algorithms. First includes optimization of association rules using Multi-fitness function Genetic algorithm. Second is optimization of association rules using constrained nonlinear minimization and minimax optimization method. Proposed optimization algorithm is validated on synthetic image set containing geometric shapes and standard MIAS medical image dataset. Association rules referring to specific objects are found regardless of object position and object orientation. Interestingness correlation measures as Cosine, All-Confidence, Accuracy and Jaccard are linearly independent of each other. Using multi-fitness function Genetic algorithm these correlation measures are optimized.

Multi-fitness function Genetic algorithm generates more efficient, effective and strong association rules than constrained nonlinear minimization and minimax optimization method for image mining. The algorithm proved to be a critical approach in reducing number of redundant rules and complexity of system. Moreover, it can reduce the computation cost of image analysis and can easily be applied to other image analysis applications. Future scope includes, use of optimized association rules for mammogram classification.

ACKNOWLEDGEMENT

REFERENCES

[1]. J. Zhang, W. Hsu and M. L. Lee, *Image mining: trends and developments*. Journal of Intelligent Information Systems, 2002. 19(1): p. 7-23.

[2]. C. Ordonez and E. Omiecinski, *Discovering association rules based on image content.* Research and Technology Advances in Digital Libraries, 1999. Proceedings. IEEE Forum on, IEEE, 1999. p. 38-49.

[3]. C. Carson, S. Belongie, H. Greenspan and J. Malik, *Region-based image querying.* Content-Based Access of Image and Video Libraries, 1997. Proceedings. IEEE Workshop on, IEEE, 1997. p. 42-49.

[4]. M. Sahu, M. Shrivastava, *Image mining: a new approach for data mining based on texture.* IEEE International Conference on Computer and Communication Technology, 2012. p. 7-9.

[5]. R. Gonzalez and R. Woods, *Digital image processing*. Pearson Addison-Wesley Publications Co., Second Edition, March 1992.

[6]. J. Nagi, SA. Kareem, F. Nagi, SK. Ahmed, *Automated breast profile segmentation for ROI detection using digital mammograms.* Biomedical Engineering and Sciences (IECBES), 2010 IEEE EMBS Conference on, IEEE, 2010.

[7]. RM. Haralick, K. Shanmugam, IH. Dinstein, *Textural features for image classification.* Systems, Man and Cybernetics, IEEE Transactions on, 1973. 6: P. 610-621.

[8]. JC Felipe, AJ Traina, Jr C. Traina, *Retrieval by content of medical images using texture for tissue identification*. Computer-Based Medical Systems, 2003. Proceedings. 16th IEEE Symposium. IEEE, 2003. p. 175-180.

[9]. K. Beyer, J. Goldstein, R. Ramakrishnan, U. Shaft, *When is "nearest neighbor" meaningful?*. Database theory—ICDT'99, Springer Berlin Heidelberg, 1999. P. 217-235.

[10]. MX. Ribeiro, C. Traina, PM. Azevedo-Marques, *An association rule-based method to support medical image diagnosis with efficiency.* Multimedia, IEEE Transactions on, 2008. 10(2): p. 277-285.

[11]. R. Agrawal, T. Imieliński, A. Swami, *Mining association rules between sets of items in large databases*. ACM SIGMOD Record, 1993. 22(2): p. 207-216.

[12]. PN. Tan, M. Steinbach, V. Kumar, *Introduction to data mining.* Boston: Pearson Addison Wesley, Jun 2006.

[13]. S. N. Sivanandam and S. N. Deepa, *Introduction to genetic algorithm.* Springer Science and Business media, 2008.

[14]. J. Han, M. Kamber and J. Pei, *Data mining: concepts and techniques.* Elsevier, Jun 2011.

[15]. J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage, *The mammographic image analysis society digital mammogram database.* In IWDM, 1994. P. 211–221.

[16]. A. Ghosh and B. Nath, *Multi-objective rule mining using genetic algorithms*. Information Sciences, Elsevier, 2004. 163(1): p. 123–133.

[17]. M. Saggar, AK. Agrawa and A. Lad, *Optimization of association rule mining using improved genetic algorithms.* IEEE International Conference on Systems, Man and Cybernatics, 2004. p. 3725-3729.

[18]. P. P. Wakabi-Waiswa, V. Baryamureeba and K. Sarukesi, *Optimized association rule mining with genetic algorithms*. Natural Computation (ICNC), 2011 Seventh International Conference on, IEEE, 2011. 2: p. 1116-1120.

A Comparison of Different Clustering Methods for MIT BIH ECG Data

[1]**Sharathchandra Chilkuri** , [2]**M. Prabhakara Reddy, and** [3]**Ibrahim Patel**

[1,2]*Dept. of Biomedical Engineering, B.V. Raju Inst. of Tech., Narsapur Medak,(T. S) India;*
[3]*Dept. of ECE B.V. Raju Inst. of Tech., Narsapur Medak,(T. S) India*
sharathchandrachilkuri@gmail.com; prabhakarareddy.m@bvrit.ac.in; ptlibrahim@gmail.com;

ABSTRACT

Electrocardiogram can be occasionally or continuously measured from living Human beings. Regardless of their Disease, Now a day's physicians or doctors are suggesting to take ECG. These signals reflect the physiological processes and electrical activity of the Heart. Therefore, the study of ECG signals is essential for both medical applications and scientific studies for this purpose one requires best clustering method. It is difficult to provide a best clustering methods for the ECG signals because these categories may overlap, so that a method may have features from several categories. Nevertheless, it is useful to present a relatively organized picture of the different clustering methods.

Keywords: ECG; Clustering; MITBIH; QRS Detection; Filtering.

1   Introduction

The Electrocardiogram signal is generated by polarization and depolarization of the heart that occurs when pumping blood throughout the human body, and it can be recorded by contacting electrodes to the skin at specific locations on the body. It provides the valuable information regarding the cardiovascular diseases. Any abnormality in rhythm can provide useful information about the type of disease. in ecg QRS complex is a dominant of electrocardiographic signal. Its amplitude and time analysis, shape and appearance time of adjacent rhythms estimation can be used to diagnose a wide range of heart diseases. QRS complex is necessary for the determination of the heart rate, and as reference for beat alignment.Thus, the obvious problem is the precise definition of the occurrence time and other various parameters of QRS-complex.

Various methods for classification of arrhythmias have been developed by researchers and clustering technique is one of them. Although it is an unsupervised type of technique, it is advisable technique for analysis and interpretation of long term ECG Holter records. In this paper, we are tesing four clustering has been used for analysis.

2 Clustering

Clustering [1] can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. A loose definition of clustering could be "the process of organizing objects into groups whose members are similar in some way".A *cluster* is therefore a collection of objects which are "similar" between them and are "dissimilar" to the objects belonging to other clusters.

2.1   K-Means Algorithm

K-means (MacQueen, 1967)[1] is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \tag{1}$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the *n* data points from their respective cluster centers.

2.2   K-Medoids Algorithm

The K-means algorithm [1] is sensitive to outliers since an object with an extremely large value may substantially distort the distribution of data. How might the algorithm be modified to diminish such sensitivity? Instead of taking the mean value of the objects in a cluster as a reference point, a Medoid can be used, which is the most centrally located object in a cluster. Thus the partitioning method can still be performed based on the principle of minimizing the sum of the dissimilarities between each object and its corresponding reference point. This forms the basis of the K-Medoids method. The basic strategy of K Medoids clustering algorithms is to find k clusters in n objects by first arbitrarily finding a representative object (the Medoids) for each cluster. Each remaining object is clustered with the Medoid to which it is the most similar. K-Medoids method uses representative objects as reference points instead of taking the mean value of the objects in each cluster. The algorithm takes the input parameter k, the number of clusters to be partitioned among a set of n objects

2.3   Hierarchal Algorithm

These methods construct the clusters by recursively partitioning the instances in either a top-down or bottom-up fashion. These methods can be subdivided as following:

- Agglomerative hierarchical clustering — each object initially represents a cluster of its own. Then clusters are successively merged until the desired cluster structure is   obtained.
- Divisive hierarchical clustering — All objects initially belong to one cluster. Then the cluster is divided into sub-clusters, which are successively divided into their own sub-clusters. This process continues until the desired cluster structure is obtained.

The result of the hierarchical methods is a dendrogram, representing the nested grouping of objects and similarity levels at which groupings change. A clustering of the data objects is obtained by cutting

the dendrogram at the desired similarity level. The merging or division of clusters is performed according to some similarity measure, chosen so as to optimize some criterion (such as a sum of squares). The hierarchical clustering methods could be further divided according to the manner that the similarity measure is calculated

2.4 Fuzzy C-Means Algorithm

The most popular fuzzy clustering algorithm is the fuzzy c-means (FCM) algorithm. Even though it is better than the hard K-means algorithm at avoiding local minima, FCM can still converge to local minima of the squared error criterion. The design of membership functions is the most important problem in fuzzy clustering; different choices include those based on similarity decomposition and centroids of clusters. A generalization of the FCM algorithm has been proposed through a family of objective functions. A fuzzy c-shell algorithm and an adaptive variant for detecting circular and elliptical boundaries have been presented. Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method (developed by Dunn in 1973 and improved by Bezdek in 1981) is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 , 1 \le m < \infty \qquad (2)$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d- dimensional measured data, c_j is the d-dimension center of the cluster, and $\|*\|$ is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}} \qquad c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m}$$

This iteration will stop when $\max_{ij}\left\{\left|u_{ij}^{(k+1)} - u_{ij}^{(k)}\right|\right\} < \varepsilon$, where ε is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m.

3 Implementations

Step 1: Load the MIT-BIH ECG Database

Step 2: Convert the MIT-BIH ECG Database into MATLAB Readable Format

Step 3: Applied the IIR Butterworth filter along with notch filter on ECG Database

Step 4: The absolute slope [4] i.e. absolute value of the difference between two consecutive samples is calculated to enhance the signal in the region of QRS-complex. The absolute value of slope of the ECG signal is used as an important discriminating feature because absolute slope of the signal is much more in the QRS-region than in the rest of the region. Fig. shows the absolute slope of the filtered ECG signal.

Step 5: The various steps of four clustering algorithm four clustering algorithms as described in above section are followed in order to find the two cluster centers namely the QRS-cluster centre and the non QRS-cluster centre.

Step 6: After finding two cluster centers using four clustering algorithms, the slope curve shown in Fig. is scanned. The membership of slope, at a given sampling instant, is found. An output is 2 if a sample belongs to a QRS-cluster and output is 1 if it belongs to a non-QRS-cluster
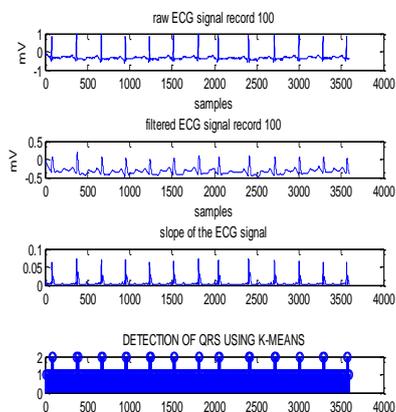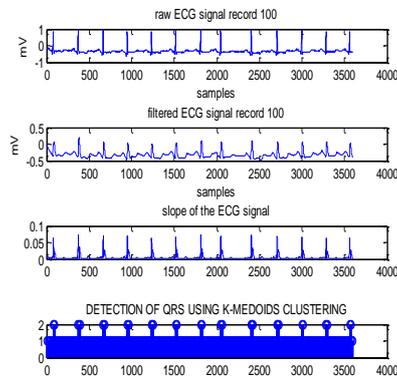
Figure 1: Implementation of K-MEANS

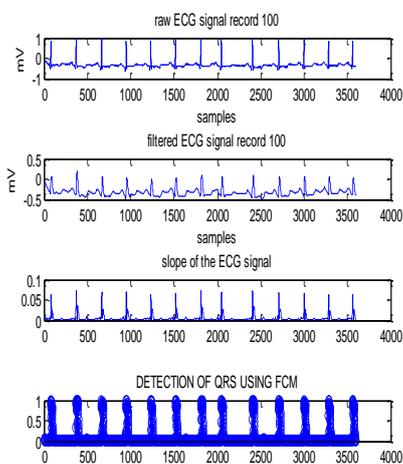**Figure 2: Implementation of K-MEDOIDS**

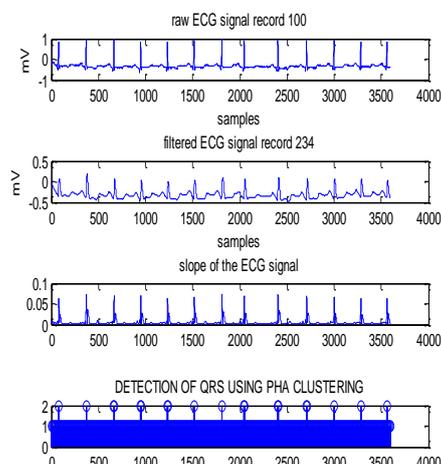**Figure 3: Implementation of FUZZY C-MEANS**

Figure 4:Implementation of HIERARICHAL

4 Result Analysis

In this paper we implemented four algorithms on MITBIH ECG Data. The four algorithms compared by four parameters i.e. Sensitivity, Specificity [6], Predictivity, Speed.

The above mentioned algorithms not only detect the QRS complexes of ECG, but also delineate them accurately detection is said to be true positive (TP) if the algorithm correctly discerns the QRS-complex and it is said to be false negative (FN) if the algorithm fails to detect the QRS complex. False positive (FP) detections are obtained if non QRS-wave is detected as a QRS-complex. The ECG signals used for analysis and detection in this work are a part of MIT-BIH Arrhythmia Database given on the website of MIT-BIH. The said algorithm is applied on total of 48 records from database.

 It is observed that, in the case of normal beats (i. e. for record number 100, 101, 102, 104, 105, 106, 107, 112,   113, 115, 117, 119, 121, 122, 123, 201, 202, 209,212, 213, 215, 217, 219, 220, 221, 222,   223,   228,   230,   231,   232,   234) and right bundle branch block (i.e. for record numbers118,124),[7] the results are encouraging and almost all the beats were detected successfully. Similarly, in the case of left bundle branch block also (i.e. for record numbers 111,207, 214), the total number of complexes detected are accurate and percentage range of Se and P+ is satisfactory. As the algorithm has been implemented in MATLAB working environment, therefore the part of the whole signal of each data set has been operated. In order to evaluate the accuracy of detection of QRS complex, three essential parameters: sensitivity Se and the positive predictivity P+(detection rate), specificity are used as listed in Table 1 These parameters describe the overall performance of the detector and their values are calculated as follows

$$Sensitivity = TP / (TP + FN)$$

$$Predectivity = TP / (TP + FP)$$

$$Specificity = TN / (TN + FP)$$

Using the above formula, Table 1 clearly shows the results of four methods in terms of sensitivity, specificity, predictivity is obtained for all 48 MIT-BIH Records. Also the percentage of false positive detection and false negative detection for all records are very less.

Table 1: Comparison of Different Algorithms

Clustering Method	Sensitivity	Specificity	Predictivity	Time(Sec)
K-Means	100%	100%	97%	0.41-0.44
K-Medoids	100%	100%	96%	0.81-1
Hierarchical	100%	100%	98.39%	0.22-0.23
Fuzzy C Means	100%	100%	96.39%	0.19-0.22

5 Conclusion

The four clustering methods have been comprehensively tested using the MIT BIH database covering wide variety of QRS complexes.

In above discussed four methods Sensitivity and specificity is approximately 100%

All Four methods got more than 95% detection rate or Predictivity.

 It is observed that hierarchal algorithm is suitable than K-Means, K-Medoids and FCM algorithm based on Predictivity for the data sets in MITBIH data base.

6 Future Scope

The project outcome is QRS and Non QRS clustering and comparison of various clustering methods. In future one can extend this project to multiple clusters like P, QRS, T, U waves. The information obtained from the above methods can be useful for ECG interpretation and analysis. For example Estimation of heart rate, HRV.

It is also possible to extend these methods for automatic annotation of ECG signal and diseases diagnosis not only for the ECG data but we can also use for other biological data such as radiological images, EEG, EMG, Genes data.

REFERENCES

[1]. A Tutorial on Clustering Algorithms; home.deib.polimi.it/matteucc/clustering/tutorial_html/index.html

[2]. www.physionet.org/physiobank/database/mitdb MIT Database.

[3]. Ms. Ananya, Dr.S.L.Nalbalwar,Swarali Seth, "Detection of QRS Complexes in ECG using Kmeans Algorithm" International Journal of Engineering Research & Technology (IJERT), Vol. 3 Issue 5, May – 2014

[4]. S. S. Meht. "Development of FCM based algorithm for the delineation of QRS-complexes in Electrocardiogram", 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), 12/2009

[5]. Sharathchandra C, M.Prabhakara reddy, Ibrahim patel, Rameshwar "Identification of QRS complexes using Hierarichal clustering Algorithm" IJIRCCE, Volume 4, Issue 2, Feb-2016

[6]. Sharathchandra C, M.Prabhakara reddy, Ibrahim patel, Rameshwar "Mark out of QRS complexes using fuzzy C-means Algorithm" IJAREST, Volume 3, Issue 2, Feb-2016

[7]. J. A. Hartigan and M. A. Wong (1979) "A K-Means Clustering Algorithm", Applied Statistics, Vol. 28, No. 1, p100-108

[8]. B.U. Kohler, C. Hennig, and R. Orglmeister, "The principles of software QRS detection," *IEEE Eng Biol. Mag*, vol. 21, pp. 42–57, 2002.

[9]. Matlab help, MATLAB MATHWORKS. http://www.mathworks.com

[10]. G.V. Van, K.V. Podmasteryev; Review of Algorithms Detection the QRS Complex based on machine Learning.

[11]. Jalil, Bushra, Olivier Laligant, Eric Fauvet, and Ouadi Beya. "Detection of QRS complex in ECG signal based on classification approach", 2010 IEEE International Conference