International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 16 (2015) pp 36652-36657

Sign language recognition by hand gestures u7sing Shape Context Matching and Stochastic Context Free Grammar

V.J.Arulkarthick
Assistant Professor (SG), Department of ECE,
SNS College of Technology, Coimbatore, India
karthick.arul@gmail.com

S.Umamaheswari,
Associate Professor, Department of ECE,
Coimbatore Institute of Technology, Coimbatore, India
umamaheswari.cit@gmail.com

Abstract
A new approach for real time vision based sign language recognition using hand gestures is investigated in this paper. In the first step, hand posture is detected by calculating the similarity measures between query and database images. The measurement of similarity is calculated by solving for the correspondences between edge points of query and database hand posture images. Background of Query image is and the database images are stored before-hand with background elimination. Using these correspondences an aligning transform is estimated. The shape context is attached to every edge point that calculates the other points related to it. Thin-plate spline is used for estimating the transformation of edge points for the alignment and using this alignment postures, various degrees of freedom can be detected .Two similar hand postures will have similar shape context and the recognition is based on Fast Condensed Nearest Neighbourhood algorithm. From the sequence of detected hand postures the hand gesture is recognized through Stochastic Context Free Grammar (SCFG) which indicates a specific word in the sign language. The performance of the scheme is evaluated in terms of posture detection rate, process time for posture detection, gesture recognition accuracy and overall time for gesture recognition.

Keywords: Hand posture, Shape contexts, Thin-plate spline, Nearest Neighbourhood, Hand gesture, Stochastic Context Free grammar (SCFG).

Introduction
Efficient sign language recognition through hand gesture recognition is a challenging task in Human Computer Interaction. Sign language recognition makes the computers recognize the language visually from the real time hand gestures. Hand gesture is a significant concept of human hand movements which can be used as a fruitful and efficient medium of interaction between human and computer. Processing real time video images with the aim of recognizing hand gestures is one of the most attractive and highly used areas of research in the field of machine vision over the past decade [9].
Vision-based hand gesture recognition can be divided into two important methods: 1) appearance-based approaches and 2) 3-D hand model based approaches [1]. Appearance-Based Approaches use image features to model the visual appearance of the hand and compare these parameters with the extracted image features from the input video. 3-D hand model-Based Approaches rely on 3D kinematic hand models and try to estimate the hand parameters by comparing the input images and possible 2D appearance projected by the 3D hand model. 3D hand model-based approaches employ an estimation-by-synthesis strategy, and recover the hand parameters by aligning the appearance projected by the 3D hand model with the observed image features, and minimizing the discrepancy between them.
Hand Gestures involves hand motions with the physical movements of the fingers for the objective of conveying meaningful information or interacting with the environment. Hand Gestures are a powerful human-to-human medium of communication. However, vision-based hand gesture recognition is a challenging problem due to the complexity of hand gestures, which are rich in diversities due to high Degree of Freedom (DOF) involved by the human hand. This problem can only be addressed with powerful feature extraction method to increase the detection rate in hand posture detection.
This paper proposes a system for hand posture detection with a Fast Condensed Nearest Neighbourhood algorithm (FCNN) which has a Shape Context aligned by Thin Plate Spline (TPS) for hand posture detection. Using stochastic context-free grammar (SCFG) it builds a grammar which describes the structural information about hand gestures.
In this paper, Section II explains the overall structure in the proposed method. Section III elaborates on the extraction of features of hand postures through shape context method with align transform and detection of the same with Fast Condensed Nearest Neighbourhood algorithm. Section IV describes about hand gesture recognition using Stochastic Context Free Grammar and the conclusion and future enhancements are discussed in Section V.

The overall structure of the system
In this paper the hand gesture recognition method is applied for sign language recognition. The automatic sign language recognition is helpful for hearing impaired people who do not understand the sign language, and people who are physically

challenged. The overall structure can be divided into two parts, one detects hand postures from the images segregated from the video sequence whereas the other recognizes hand gestures from the sequences of detected hand postures.
The Fig.1 depicts the overall structure of the proposed system consisting of two parts. The first step of the first part of the proposed method is a simple algorithm for Key corresponding points between the hand postures. Hand postures are represented with the set of key corresponding points which is obtained from the sampling of edge detection output. The shape context is a shape descriptor used to determine the distribution of key corresponding points with respect to the given key corresponding point of a posture in the image.

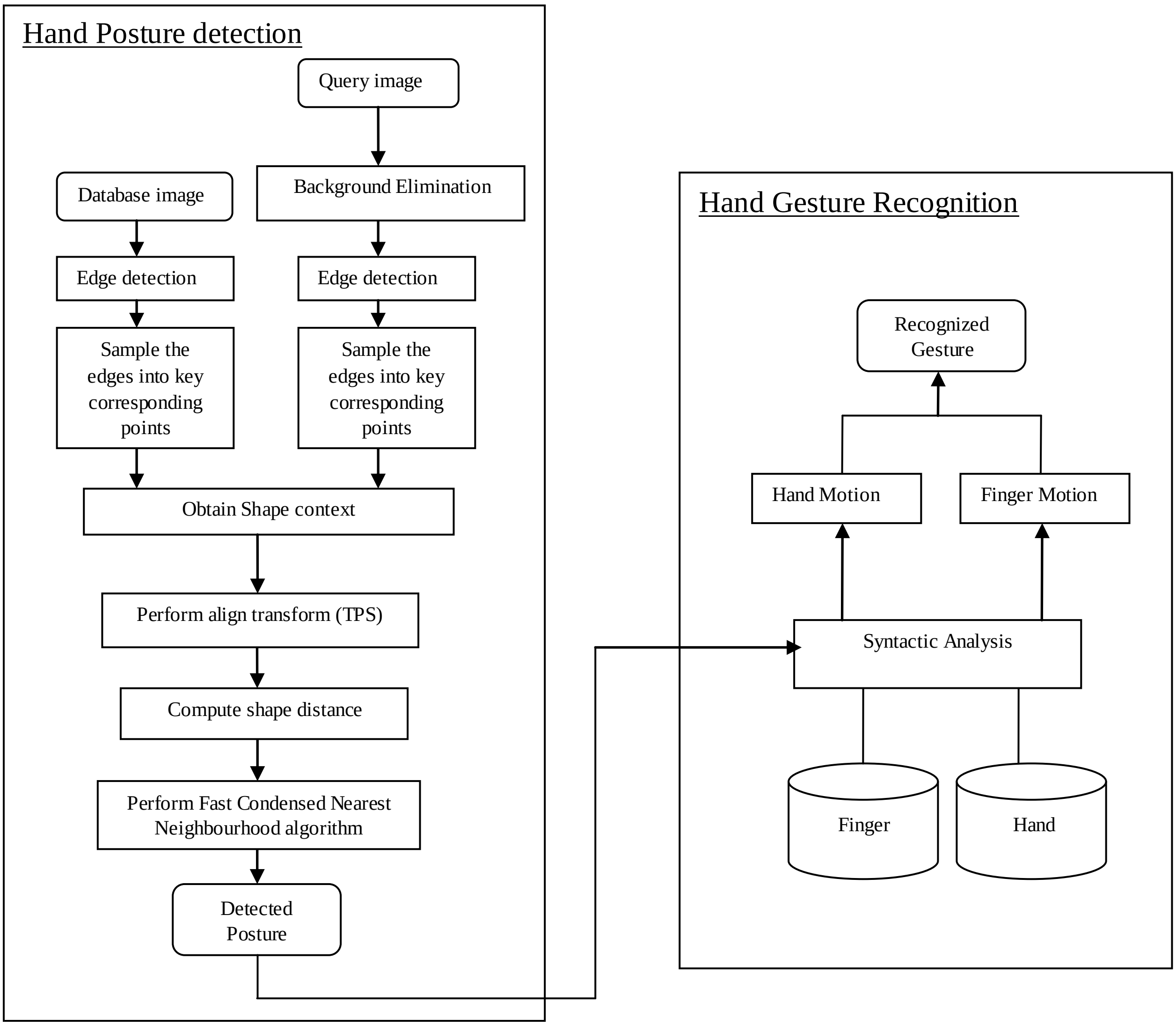


Fig.1.Structure of the proposed hand gesture recognition scheme

The second step of the first part is the aligning transformation to calculate the correspondence to the complete posture that maps the postures into different DOFs. Similarity is measured through magnitude of the aligning transform and sum of matching errors between the postures.
Third step of the first part is recognizing hand postures from the above similarity measures using Fast Condensed nearest neighbourhood algorithm which increases the speed of recognition.
Similarity measurement is computed for query image with data base image and the posture is recognized. Edge of a hand posture is detected through any edge detection method and 100 key corresponding points are selected on the edges of the hand posture to be recognized. The query image is background subtracted using pixel by pixel background subtraction method and the database images are taken without background.
The second part of the proposed system is hand gesture recognition. A hand gesture is an action composed of a sequence of hand postures. Hand gestures can be specified as building up by a group of hand postures in various ways of composition, just as phrases are built up by words. The complex hand gestures are described by using a small set of hand postures and grammatical rules. This part is

implemented with the help of stochastic context free grammar algorithm.

Posture detection

Hand posture is static hand pose and hand locations without any movements involved. Hand posture is detected using similarity measurement between database images and query image. The feature is extracted from the image using shape context aligned with thin plate spline and the posture is classified with Condensed Nearest Neighbourhood Algorithm.
Originally for the tasks like digit, character and object recognition the shape context approach was proposed (4). In a similar manner, each hand posture is assumed as infinite set of points on the 2D plane among which a finite set of points can be extracted to define the hand posture. The edges of the query image (Q) and the data base image (D) are obtained using canny edge detection method. Sampled edge points are obtained by sampling randomly along the contours of the edges.

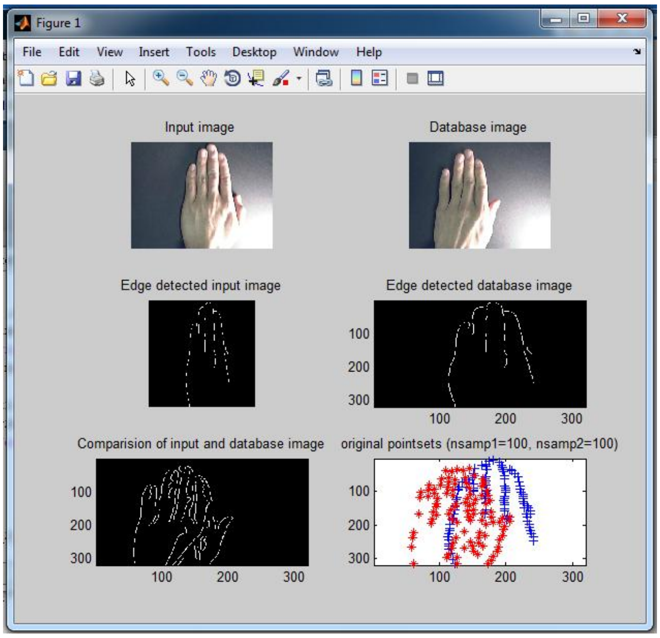


Fig.2.Steps to fix key corresponding points

Figure2 shows the step-by-step output in fixing the key corresponding points for posture detection. These sampled edge points are treated as vectors and are called key corresponding points or shape descriptors. The key corresponding points on the edges of these two images play a vital role in discriminating the hand postures. For each key corresponding point on the query image (Q_i), a finest matching key corresponding point on the data base image (D_i) is identified.
For each key corresponding point Q_i, a log polar histogram (LH) is computed and plotted based on the coordinates of the remaining points. This makes the descriptor more detailed in discriminating the posture. Histogram bins are used in plotting the log polar histogram. Five bins are used for Θ on the x axis and 12 bins are used for log r.

$$LH(k) = \#\{D \neq Q_i \; : Q - D_i \in bin(k) \qquad (1)$$

The detailed steps for shape context matching are implemented and the results for a palm posture are given fig.3 for sample. The cost matrix for the matching points are calculated using the equation

$$C_{ij} = \frac{1}{2}\sum_{k=1}^{K}\frac{[LH_i(k) - LH_j(k)]^2}{LH_i(k) - LH_j(k)} \qquad (2)$$

The one to one matching of key corresponding points is performed and the problem of assigning square matrix is solved with the Hungarian method. The dummy nodes can be added to make the cost matrix as square matrix when one-to-one matching is not exactly possible.
The information obtained from the shape context is not suitable to detect hand postures with varying Degrees Of Freedom (DOF). So the thin plate spline transform is applied as an align transform for realigning the computed shape context. The postures with different Degrees Of Freedom (DOF) can be realigned through this elastic transform and can be approximated to any of the posture.
The coarse alignment is estimated using translational and rotational offsets of arbitrarily selected edge points. This is iterated with all the possible parings and finally the maximum number of matched sample is reported. This coarse alignment is not sufficient in the classification of hand postures of various degrees of freedom because of elastic changes and inherent deformations in the acquired images.
Two separate TPS functions are used to model a coordinate transformation which yields

$$T(x, y) = (f_x(x, y), f_y(x, y)) \qquad (3)$$

a displacement field that maps any position in the first image to its interpolated location in the second image. TPS, a 2D spatial generalization of the cubic spline, is an effective tool for estimating the deformation in hand postures with various degrees of freedom based on shape context points. The brute-force minutiae matching technique to obtain a set of n matching minutiae pairs (shape context points) are,

$$M_1 = f(x_{1,i}, y_{1,i}) \qquad (4)$$

$$M_2 = f(x_{2,i}, y_{2,i}) \qquad (5)$$

where i=1,2,…..n. This function uniquely minimizes the `bending energy required to warp, resulting in the deformed image. A simple pixel averaging scheme is used to blend the images at the boundary. The composite image is then enhanced in order to obtain the mosaiced template. Figure3 shows aligned shape descriptors which help to detect posture with different degrees of freedom.

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 16 (2015) pp 36652-36657

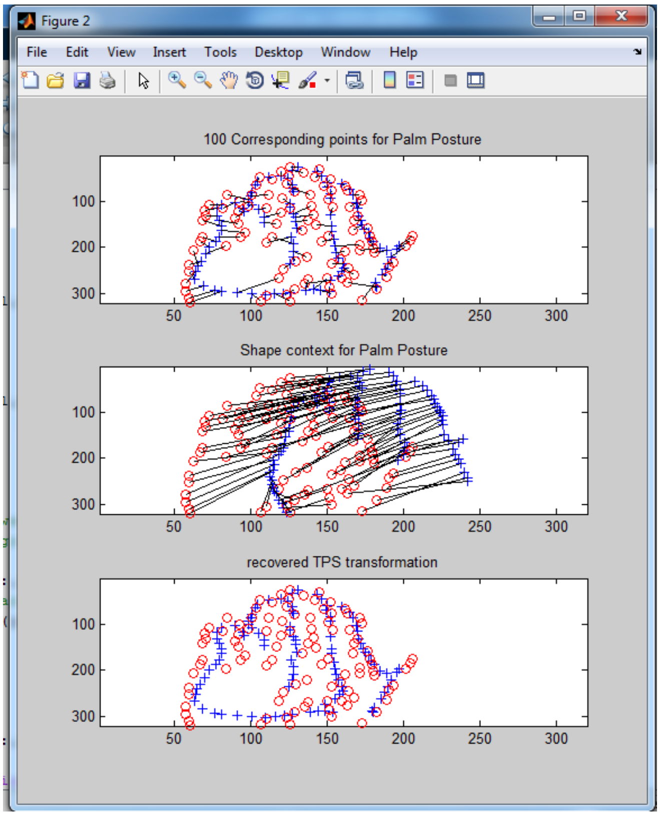


Fig.3.Alignment of shape context with Thin Plate Spline transform

The nearest neighbour decision rule assigns to an unclassified sample point, the classification of the nearest of a set of previously classified points. For this decision rule, no explicit knowledge of the underlying distributions of the data is needed.

Table1.Performance of the detection

Posture	Without Background subtraction		With Background subtraction	
	Detection Accuracy	Process Time (Seconds)	Detection Accuracy	Process Time(Seconds)
Palm	100%	1.528	100%	1.549
Fist	100%	2.139	100%	2.164
Two finger	98.3%	2.342	99.1%	2.365
Little finger	99.1%	1.836	100%	1.858

A strong point of the nearest neighbour rule is that, for all distributions, its probability of error is bounded above by twice the Bayes probability of error. In order to reduce space and time requirements, several techniques to reduce the size of the stored data for the NN rule are introduced as training set reduction, training set condensation, reference set thinning, and prototype selection algorithms. But, in order to manage huge amounts of data, methods exhibiting good scaling behaviour are definitively needed. In this work, an order independent algorithm for finding a training set consistent subset for the NN rule, called Fast Condensed Nearest Neighbourhood (FCNN) rule is applied.

Each hand posture is photographed by rotating camera for 2o in all the axis and totally 72 views per posture is collected. We prepared database sets by selecting number of equally spaced views for each hand posture and for the remaining views for test sets. The distance function of FCCN gives us a way to compare different similarity or distance measures. The detection accuracy and processing time are evaluated for various hand postures and listed in the Table1.

Gesture Recognition

A hand gesture is an action composed of a sequence of hand postures. The syntactic approach is appropriate to describe the hand gesture in terms of its constituent hand postures and provides the capability to describe a set of complex hand gestures. With the syntactic approach, hand gestures can be specified as building up by a group of hand postures in various ways of composition. The rules governing the composition of hand postures into different hand gestures can be specified by a grammar.

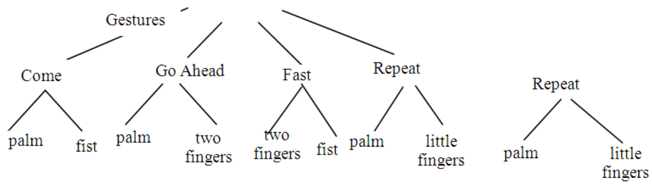


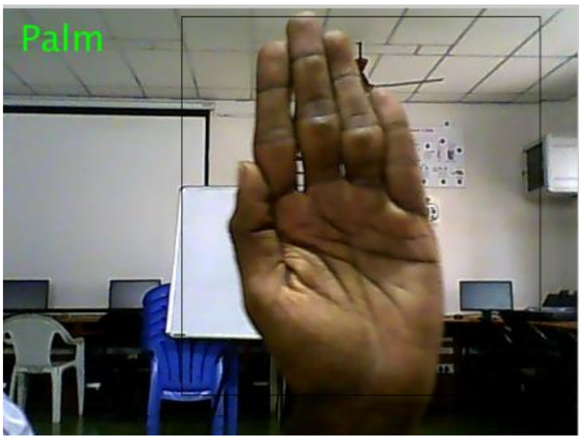
Fig.4.Words that are generated from different sequence of postures (gesture)

With the SCFGs, we just need to compute the probability of the pattern belonging to different classes (each class is described with its own grammar) and then choose the class with the highest probability value. When an exact parse is not possible, we can still perform an approximate parse and get the probabilities. The probabilistic information allows us to use statistical techniques not only to find the best matches but also to perform robust feature-detection that is guided by the grammar's probability structure.

The flexibility of the SCFG allows the user to easily change the grammar so that other gestures with different combinations of detected postures or more complex gestures can also be described. The assignment of the probability to each production rule can also be used to control the "wanted" gestures and the "unwanted" gestures.

Table2.Performance of the recognition

Hand gesture	Recognition Accuracy	Process Time(Seconds)
Come	100	4.02
Go Ahead	96.3	4.6
Fast	94.2	3.99
Repeat	93.2	4.13



5.(a)

5.(b)

5.(c)

5.(d)

Fig.5(a),(b),(c)&(d) Recognition results for sign language

Figure 5 shows samples of hand detection followed by gesture recognition results. The detection accuracy and processing time are evaluated for various hand postures and listed in the Table2. It is observed from the result that hand gestures are recognized with minimum error in reduced time. Our current implementation is in MATLAB, but a more efficient implementation using programming languages like C++ in OpenCV would lead to further reduction in the process time.

Conclusion

Hand posture detection is based on shape context matching with improved Condensed Nearest Neighbourhood algorithm and hand gesture is recognized by means of probabilistic finite state machine Stochastic Context Free Grammar from the detected series of hand postures. The experimental results of the proposed method demonstrate that the system correctly recognizes the hand postures and the performance is suitable for real time implementations. Future work is to create an efficient algorithm to recognize the sign language with multiple features like hand posture, facial expressions, head movement and lip movement for real time implementations.

References

[1]. Qing Chen, Nicolas D.Georganas, “Hand gesture recognition using haar-like features and a stochastic context-free grammar,” IEEE Transactions on Instrumentation and Measurement., Vol. 57, No. 8, 2008, pp. 1562-1570.

[2]. S.Belongie, J.Malik, J. Puzicha, “Shape Matching and Object Recognition Using Shape Contexts”, IEEE Transactions on Pattern Analysis And Machine Intelligence, Volume 20, No.24, April 2002.

[3]. Y.Lamdan,J.Schwartz, and H.Wolfson, “Affine Invariant Model-Based Object Recognition”, IEEE Transactions on Robotics and Automation, Volume 6,pp. 578-589,1990.

[4]. B.S. Manjunath and W.Y. Ma. “Texture features for browsing and retrieval of image data” IEEE Transactions on Pattern Analysis and Machine Intelligence”,(PAMI - Special issue on Digital Libraries), volume 18, no. 8, pp. 837-42, Aug. 1996.

[5]. Bharath Ramesh, Cheng Xiang, Tong Xeng Lee “Shape classification using invariant features and contextual information in the bag-of-words model”, Journal of Pattern Recognition-Elsevier, Volume 48, No.3, April 2015.

[6]. H. Chui and A. Rangarajan. “A new algorithm for non-rigid point matching”, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pages 44–51, June 2000.

[7]. R. H. Liang and M. Ouhyoung, “A real-time continuous gesture recognition system for sign language,” Proceedings of 3rd International Conference Automation Face Gesture Recognition, 1998, pp. 558–565.

[8]. Corradini, “Real-time gesture recognition by means of hybrid recognizers”, Proceedings of International Gesture Workshop Gesture Sign Language Human–Computer Interaction. New York: Springer-Verlag, 2001, Vol. 2298, pp. 34–46, Revised paper.

[9]. Mitra khaledian, Mohammad bagher Menhaj, “Real time Vision-based Hand gesture Recognition with Sift features”,TELKOMNIKA Indonesian journal of Electrical Engineering, Volume 15, No.1, July2015, pp.162-170.

[10]. F. L. Bookstein. “Principal warps: thin-plate splines and decomposition of deformations”, IEEE Transactions on Pattern Analysis and Machine Intelligence”, 11(6):567–585,June 1989.

[11]. S. Belongie, J. Malik, and J. Puzicha. “Shape context: A new descriptor for shape matching and object recognition”, In T. K. Leen, T. G. Dietterich, and V. Tresp, editors, Advances in Neural Information Processing Systems13: Proceedings of the 2000 Conference, pages 831–837, 2001.

[12]. V. J. Arulkarthick , D. Sangeetha , S. Umamaheswari ,“Sign Language Recognition using K-Means Clustered Haar-Like Features and a Stochastic Context Free Grammar”, European Journal of Scientific Research, Vol.78 No.1 2012, pp.74-84.

[13]. You Lei, Wang Hongpeng , Tan Dianxiong ,Wang Jue “A real-time hand gesture recognition algorithm for an embedded system”, IEEE conference on mechatronics and automation 2014.

[14]. Fred L. Bookstein,”Principal Warps: Thin-Plat Splines and the Decomposition of Deformation”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 11, No 6,June 1989.

[15]. V. J. Arulkarthick , S. Umamaheswari, “ Hand Posture Recognition using Shape Context Elastic matching and Fast Condensed NN algorithm” International Journal of Applied Engineering Research,Volume10 No.67,2015.

[16]. Shoa-Hu-Png,Deok-Hwan kim,Seok-Lyong Lee, Chin-Wan Chung, “A visual shape descriptor using sectors and shape context of contour lines”, Information Sciences 180, 2010 2925–2939, 26 April 2010.

[17]. Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., “Advances in Knowledge Discovery and Data Mining”. AAAI Press/Mit Press, 1996.

[18]. Berchtold, S., Keim, D. A., and Kriegel, “The X-tree: An index structure for high-dimensional data”, Proceedings of 22nd VLDB Conference ,1996.

[19]. Roussopoulos, N., Kelley, S., and Vincent, “Nearest neighbor queries”,Proceeding of ACM SIGMOD Conference on Management of Data 1995.

[20]. Callahan, P. B. and Kosaraju, S. R. “Algorithms for dynamic closest pair and n-body potential _elds”, Proceedings 6th ACM-SIAM Symposium Discrete Algorithms 1995.

[21]. Yao A. C. and Yao, F. F. “A general approach to d-dimensional geometric queries”, Proceedings of 17th Annual ACM Symposium on Theory Computations 1985.

[22]. De Berg M, van Kreveld M, Overmars M, and Schwarzkopf O, “Computational Geometry: Algorithms and Applications”, Springer-Verlag, Berlin, 1997.

[23]. White, D. A. and Jain R, “Similarity indexing with the SS-tree”, Proceedings of 12th IEEE International Conference Data Engineering 1996.

[24]. Arya S. and Mount D. M. “Approximate range searching”, Proceedings 11th Annual ACM Symposium on Computational Geometry 1995.

[25]. Feder T. and Greene D. H “Optimal algorithms for clustering”, Proceedings 20th Annual. ACM Symposium. Theory of Computations 1988, pp. 434-444.