Video Object Tracking using Extreme Point Localization

P. Balakumar[1], P.Vimala[2]

[12](CSE, Mahendra Institute of Technology/ Anna University, India)

Abstract: *Despite the recent progress in both pixel-domain and compressed-domain video object tracking, the need for a tracking framework with both reasonable accuracy and reasonable complexity still exists. It is a method for tracking moving objects in SPIHT compressed video sequences using Blob tracking model. Built upon such a model, the proposed method works in the compressed domain and uses only the motion vectors (MVs) and block coding modes from the compressed bit stream to perform tracking. At each frame, the decision of whether a particular block belongs to the object being tracked is made with the help of the model, which is updated from frame to frame in order to follow the changes in the object's motion.*

Keywords- Compressed-domain video object tracking, SPIHT/AVC, Blob tracking.

I. INTRODUCTION

Human action analysis in images and videos can be either Action Recognition or Action Retrieval. Action recognition involves identifying the type of action in the video or image. This has its application in several fields like video surveillance, in training for human computer interface, in identifying sports events etc., Action retrieval involves retrieval of action subsequences that are similar to some query video clip and in case of images the retrieval of an image that is most similar to the query. This has a major technical challenge for the emerging field of content based retrieval of video or image from the World Wide Web (WWW). This is important because retrieval using keywords would sometimes make the search engines return irrelevant search results.

Recognition of human action is very simple for a human observer. But when it comes to the automation of the same, when we have to train a system, it is a difficult task and it becomes mandatory to develop a computational model [1]-[4]. In a video monitoring room there has to be a human observer to continuously monitor the actions of the human in the coverage of the surveillance camera. Human are always prone to error and lethargy as a result the rigid observation is sometimes loosened. So if there is a system that can identify undesirable actions and alert the human observer there cannot be any discrepancy in surveillance that too in areas where high level of security is mandatory. Another example for application of action recognition is in the field of content based image or video retrieval. Search for videos and images based on their contents give appropriate results than based on keywords and under such circumstances, it always requires the contents to be analyzed. Other applications of human action recognition are behavioral biometrics, Human Computer Interface, etc.

Human action analysis is an emerging field which holds its application in video surveillance [17]-[18]; Content based image or video retrieval. The term 'content' in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. Realizing the importance of human action analysis in this work a correlation based method for action analysis of human called the Elastic Sequence Correlation is approached. ESC is a special case of Dynamic Time Warping (DTW), which applies mostly to signal processing and Approximate Pattern Matching based on the edit distance, which applies mostly to computer and information sciences community. Given the query, comparison is done with a database and the amount of correlation is the parameter for action analysis be it action recognition or action retrieval.

II. PROPOSED SYSTEM

Unlike [1] here the query video or the test videos whose contents are to be analyzed are compared with all the images in the database one after the other. The database contains a set of videos each having different actions. The comparison here is done using three different algorithms.

The shape context feature is extracted from the first frame of both videos and shape distance is calculated for aligning one shape to other. The main difference between [1] and the proposed method is shape distance. From the shape distance of the first frame of both the videos the correlation between the two videos is computed cumulatively using any of the three algorithms, Elastic sequence Correlation, Dynamic Time Warping, and Approximate Pattern Matching. All the three algorithms have a very little variation. The first step in computing the shape context is the edge detection of a frame such that only the external contour of the human under action is obtained. From the external contour the feature points are extracted by sampling.

The log polar histogram of each of the sample points is computed and this entire vector, the shape context feature that demonstrates the relative positions of the entire sample points from each other. Using the shape context the Elastic Sequence Correlation [19], Dynamic Time warping, Approximate Pattern Matching algorithms are employed separately to analyze the action in the video sequence.

- Video Compression: SPIHT
- Video Tracking: Blob Tracking
- Dynamic Time Warping
- Approximate Pattern Matching
- Elastic Sequence Correlation

This system reduce the inference problem to matching a query to an existing set of annotated databases, a matching problem nicely solved by the proposed dynamic programming component which is the corner stone of our ESC framework.

Major contributions:

1) A new correlation-based framework (ESC) is proposed for action sequence analysis [19], which bears close connections to established work of DTW and Approximate Pattern Matching. By exploiting existing dedicated techniques developed for either DTW and edit distance, two ESC variants are further developed to address specific scenarios.

2) Examine the ESC in two important real-world applications and conduct extensive experiments where ESC is shown to perform competitively.

a) DYNAMIC TIME WARPING

Dynamic Time Warping is a widely used technique of time series analysis. It was originally developed for speech recognition, by comparison of query speech with a set of already recorded and stored database of speech of different utterances of a word or many. Other applications of DTW include handwriting and online signature matching, sign language recognition and gestures recognition, data mining and time series clustering (time series databases search), computer vision and computer animation, surveillance, protein sequence alignment and chemical engineering, music and signal processing. Extending the DTW for two dimensions applies to pattern recognition. Methods for quantitative comparisons between visual data can lend themselves easily to a wide range of applications such as text recognition, face recognition, action recognition etc., While one-dimensional warping between time series as performed by Dynamic Time Warping (DTW) is solved efficiently in the size of the input time series, warping in two dimensions poses a much more involved problem. The problem of warping in two dimensions has been shown to be NP-complete.

b) APPROXIMATE PATTERN MATCHING

Approximate Pattern Matching is the problem of finding a pattern in a text allowing errors i.e., insertions, deletions, substitutions of characters. A number of important problems related to string processing lead to algorithms for approximate string matching: text searching, pattern recognition, computational biology, audio processing, etc. Approximate two dimensional pattern matching has applications for instance, in computer vision (i.e. searching a sub image inside a large image) and OCR.

In three dimensions, the problem has applications in some types of medical data (e.g. MRI brain scans) and in biocomputing (e.g. detecting protein patterns on the surface of three dimensional virus reconstructions). For one dimension this problem is well-known, and is modeled using the edit distance. The edit distance between two strings A and B is defined as the minimum number of edit operations that must be carried out to make them equal. The allowed operations are insertion, deletion and substitution of characters in A or B. The problem of approximate string matching is defined as follows: given a text T of length n and a pattern P of length m, both being sequences over an alphabet $\sum$ of size ρ, find all segments (or "occurrences") in T whose edit distance to P is at most k where $0<k<m$. The classical solution is O (mn) time and involves dynamic programming. Given two images of the same size, the edit distance is the sum of the edit distance of the corresponding row images. This definition is justified when the images are transmitted row by row and there are not too many communication errors (e.g. photocopy images, where most errors come from the mechanical traction mechanism along one dimension only, or images transmitted by fax), but it is not appropriate otherwise. Using this model they define an approximate search problem where a sub image of size m×m is searched into a large image of size n×n, which they solve in $O(m^2n)$time using a generalization of the classical one-dimensional algorithm. However, for many other problems, the

KS distance does not reflect well simple cases of approximate matching in different settings.

c) ELASTIC SEQUENCE CORRELATION

Elastic Sequence Correlation is used to identify action subsequences from a database video that are similar to a given query video action clip [19]. ESC is a special case of two algorithms: Dynamic Time Warping (DTW) from the signal processing community and approximate pattern matching from the computer and information sciences community. The ESC algorithm is flexible in accommodating both the local and global feature representations.

ESC involves finding of correlation between two video sequences. It is assumed that each video sequence is captured by sampling in the time domain a stream of frames under a certain sample rate. The correlation between the query action and database action is plotted in a correlation matrix. The dynamic programming is used to identify the optimal solution efficiently from a search space defined by correlation matrix. Correlation is nothing but dependence. Dependence refers to the random variables which do not satisfy the mathematical conditions.

1. FEATURE EXTRACTION

The shape context at a point captures the distribution over relative positions of other shape points and thus summarizes global shape in a rich local descriptor. Shape contexts greatly simplify recovery of correspondences between points of two given shapes. Shape context leads to a robust score for measuring shape similarity once shapes are aligned. The shape context descriptor is tolerant to all common shape deformations. An image with n pixels is regarded as an n dimensional feature vector formed by concatenating the brightness values of the pixels. But a vector of pixel brightness values is quite an unsatisfactory representation of an object.

For image retrieval and shape similarity, there are several shape descriptors, ranging from moments and Fourier descriptors to Hands-off distance and the medial axis transform. The most comprehensive work on shape correspondence is the development of iterative optimization algorithm to determine point correspondences and image transformation jointly, where some generic transformation class like thin plate spline or affine can be used. The cost function that is being minimized is the sum of Euclidean distances between a point on first shape and the transformed second shape. The distances make sense only when there is at least a rough alignment shape. Joint estimation of correspondences and shape transformation leads to a difficult, highly non convex optimization problem, which is solved by deterministic annealing. Another work introduced a representation for matching dense clouds of oriented 3D points called spin image.

A spin image is a 2D histogram formed by spinning a plane around a normal vector on the surface of the object and counting the points that fall inside bins in the plane [17]. As the size of the plane is relatively small, the resulting signature is not as informative as a shape context for purposes of recovering correspondences. But this may have the tradeoff of additional robustness to occlusion. In another work the concept of order structure for characterizing local shape configurations is dealt where the relationship between points and tangent lines in a shape are used for recovering correspondences. The shape context is a very discriminative point descriptor, facilitating easy and robust correspondence recovery by incorporating global shape information into a local descriptor. Shape contexts greatly simplify the matching part leading to a very robust point registration technique. It is invariant to scale and translation and to a large extent robust to rotation and deformation.

2. SHAPE CONTEXT

To the best of our knowledge the most similar approach to our method is that of Serge Belongie, JitendraMalik, JanPuzicha [2]. The general ideas are similar but some important details are different.

Shape context begins by converting the edge elements of a sample into a set N feature points. These points can be internal or external contour. They need not correspond to key points such as maxima of curvature or inflection points. But the shape is sampled with roughly uniform spacing. Now consider the set of vectors originating from a point in the shape to all other points in the shape. These vectors express the appearance of the entire shape relative to the reference point. This set of N-1 vectors is a rich descriptor since as N gets large the representation of shape gets exact.

The full set of vectors as a shape descriptor is inappropriate since shapes and their sampled representation may vary from one instance to other. The distribution over relative positions is robust, compact and discriminative descriptor. For a point P on the shape, a coarse histogram of the relative coordinates of the remaining N-1 points is computed. This histogram is defined as the shape context of P.

The reference orientation for the coordinate system can be absolute or relative to a given axis. An attractive characteristic of the shape context is the invariance to common deformations. Invariance to translation is intrinsic to the shape context definition since everything is measured with respect to points on the object. To achieve scale invariance the radial distances is normalized by median distance between all pair of points in the shape. The shape context is intended to be a way of describing shapes that allows for measuring shape similarity and the recovering of point correspondences. The basic idea is to pick n points on the contours of a shape. For each point p_i on the shape, consider the $n - 1$ vectors obtained by connecting p_i to all other points. The set of all these vectors is a rich description of the shape localized at that point but is far too detailed.

The key idea is that the distribution over relative positions is a robust, compact, and highly discriminative descriptor. So, for the point p_i, the coarse histogram of the relative coordinates of the remaining $n - 1$ points,

$$h_i(k) = \#\{ q \neq p_i : (q - p_i) \in bin(k) \}$$

is defined to be the shape context of p_i. The bins are normally taken to be uniform in log-polar space. The fact that the shape context is a rich and discriminative descriptor can be seen in which the shape contexts of two different versions of the letter "A" are shown. It can be seen, since (d) and (e) are the shape contexts for two closely related points, they are quite similar, while the shape context in (f) is very different.

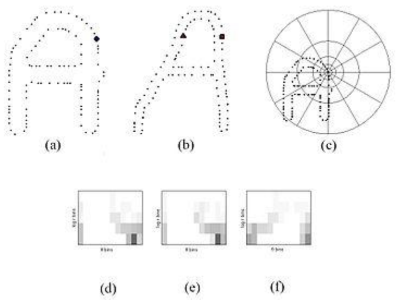

Fig 2.1 (a) the sampled edge points of shape1 Fig 2.1(b) the sampled edge points of shape2 Fig 2.1(c) log-polar bins used to compute the shape context Fig 2.1(d) shape context for the circle Fig 2.1(e) shape context for the diamond and Fig 2.1(f) shape context for the triangle.

In order for a feature descriptor to be useful, it needs to have certain in variances. In particular it needs to be invariant to translation, scale, small perturbations, and rotation. Translational invariance comes naturally to shape context. Scale invariance is obtained by normalizing all radial distances by the mean distance between all the point pairs in the shape although the median distance can also be used. Shape contexts are empirically demonstrated to be robust to deformations, noise, and outliers using synthetic point set matching experiments. Complete rotation invariance can be provided in shape contexts. One way is to measure angles at each point relative to the direction of the tangent at that point (since the points are chosen on edges). This results in a completely rotationally invariant descriptor. But this is not always desired since some local features lose their discriminative power if not measured relative to the same frame.

The shape contexts method for shape matching consists of the following steps,

- A set of points that lie on the edges of a known shape and another set of points on an unknown shape are randomly selected.

- The shape context of each point of step 1 is compared.

- Each point from the known shape to a point on an unknown shape is matched. To minimize the cost of matching, transformation (e.g. affine, thin plate spline, etc.) that warps the edges of the known shape to the unknown (essentially aligning the two shapes) is chosen. Then the point on the unknown shape that most closely corresponds to each warped point on the known shape is selected.

- The "shape distance" between each pair of points on the two shapes is computed. A weighted sum of the shape context distance, the image appearance distance, and the bending energy (a measure of how much transformation is required to bring the two shapes into alignment) are the parameters to compute shape distance.

- A nearest-neighbor classifier is used to compare its shape distance to shape distances of known objects, and hence identify the unknown shape.

3. SAMPLING SHAPE EDGES

The approach assumes that the shape of an object is essentially captured by a finite subset of the points on the internal or external contours on the object. These can be simply obtained using the canny edge detector and picking a random set of points from the edges. Note that these points need not and in general do not correspond to key-points such as maxima of curvature or inflection points. It is preferable to sample the shape with roughly uniform spacing, though it is not critical. The edge detection algorithm was chosen as canny edge detector because; the algorithm is adaptable to various environments. Its parameters allow it to be tailored to recognition of edges of differing characteristics depending on the particular requirements of a given implementation. The canny edge detection algorithm runs in 5 separate steps,

- **Smoothing**: Blurring of the image to remove noise.
- **Finding gradients**: The edges should be marked where the gradients of the image has large magnitudes.
- **Non-maximum suppression**: Only local maxima should be marked as edges.
- **Double thresholding**: Potential edges are determined by thresholding.
- **Edge tracking by hysteresis**: Final edges are determined by suppressing all edges that are not connected to a very certain (strong) edge.

4. COMPUTING SHAPE CONTEXT

Shape context is computed by counting the number of sample points in each of the bins in the log polar histogram. There are 5 radial bins and 12 angular bins and a total of 60 bins in all. So for each sample point the shape context is computed by having that sample point at the center of the log polar plot.

5. COMPUTING COST MATRIX

Considering two points *p* and *q* that have normalized *K*-bin histograms (i.e. shape contexts) $g(k)$ and $h(k)$. As shape contexts are distributions represented as histograms, it is natural to use the χ^2 test statistic as the shape context cost of matching the two points:

$$Cs = 0.5 * \sum_{k=1}^{K} \frac{[g(k)-h(k)]^2}{[g(k)+h(k)]}$$

The values of this range from 0 to 1. In addition to the shape context cost, an extra cost based on the appearance can be added. It could be a measure of tangent angle dissimilarity.

$$C_A=0.5*\left\|\begin{pmatrix}\cos\theta1\\ \sin\theta1\end{pmatrix} - \begin{pmatrix}\cos\theta2\\ \sin\theta2\end{pmatrix}\right\|$$

This is half the length of the chord in unit circle between the unit vectors with angles θ_1 and θ_2. Its values also range from 0 to 1. Now the total cost of matching the two points could be a weighted-sum of the two costs,

$$C=(1-B)Cs+BC_A$$

For each point p_i on the first shape and a point q_j on the second shape, the cost $C_{i,\ j}$ is calculated which forms the cost matrix.

6. FINDING MATCH THAT MINIMIZES COST

A one-to-one matching *pi* that matches each point *pi* on shape 1 and *qj* on shape 2 that minimizes the total cost of matching is needed.

$$H(\pi) = \sum C(ui, v\pi(i))$$

This can be done in $O(N3)$ time using the Hungarian method, although there are more efficient algorithms. To have robust handling of outliers, one can add dummy.

This would cause the matching algorithm to match outliers to a dummy if there is no real match. It shows the matching of two shapes, based on their shape context. From the results of matching the warping of one shape with respect to another is done.

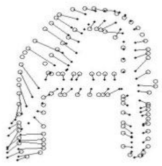

Fig 6.1Results of matching

The Hungarian Algorithm is used for finding an optimum match. Assuming that there are n jobs and n machines that Hungarian algorithm can be explained with the following steps.

- Let C = maximum value in the assignment matrix. Replace each cij with C − cij. This is done to convert the problem from a maximum assignment to minimum assignment one.
- From each row the row minimum values are subtracted.
- From each column the column minimum values are subtracted.
- Using as few lines as possible all the zeros in the matrix are covered. This is basically trial and error. Suppose k lines are used.
- If k < n, let m be the minimum uncovered number. Subtract m from every uncovered number. Add m to every number covered with two lines. Go back to the start of step 3.
- If k = n, go to step 4.

Starting with the top row, assignments are made downwards as you make assignments. An assignment can be made when there is exactly one zero in a row. Once an assignment it made, delete that row and column from the matrix. If an all n assignments cannot be made and all the remaining rows contain more than one zero, switch to columns. Starting with the left column, assignments are made rightwards. Iterate between row assignments and column assignments until as many unique assignments as possible is made. If still n assignments cannot be made and a unique assignment either with rows or columns cannot be made, make one arbitrarily by selecting a cell with a zero in it. Then unique row and/or column can be assigned.

7. COMPUTING SHAPE DISTANCE

Shape distance is the combination of three potential terms. Shape context distance, image appearance cost and bending energy. Shape context takes an important role. It is the symmetric sum of shape context matching costs over best matching points,

$$\mathrm{Dsc}(P,Q) = \frac{1}{n}\sum_{p\in P} \arg\min_{q\in Q} C(p,T(q)) + \frac{1}{m}\sum_{q\in Q} \arg\min_{p\in P} C(p,T(q))$$

Where *T* (·) is the estimated TPS transform that maps the points in *Q* to those in *P*.

Appearance cost: After establishing image correspondences and properly warping one image to match the other, one can define an appearance cost as the sum of squared brightness differences in Gaussian windows around corresponding image points:

$$\mathrm{Dac}(P,Q) = \frac{1}{n}\sum_{i=1}^{n}\sum_{\Delta\in Z^2} G(\Delta)[I_P \ (pi+\Delta) - I_Q \ (T(q(i)+\Delta)]^{\wedge}2$$

Where I_P and I_Q are the gray-level images, I_Q is the image after warping and G is a Gaussian windowing function.

Transformation cost: $D_{be}(P, Q)$ is the final cost which measures how much transformation is necessary to bring the two images into alignment. In case of TPS, it is assigned to be the bending energy. The nearest neighbor classifier (k-NN) can be used with shape distance.

III. DISCUSSION

Recognizing actions of human actors from video is an important topic in computer vision with many fundamental applications in video surveillance, video indexing and social sciences. From a computational perspective, actions are best defined as four-dimensional patterns in space and in time [20]-[21]. Video recordings of actions can similarly be defined as three-dimensional patterns in image-space and in time, resulting from the perspective projection of the world action onto the image plane at each time instant. Recognizing actions from a single video is however plagued with the unavoidable fact that part of the action are hidden from the camera because of self-occlusions. That the human brain is able to recognize actions from a single viewpoint should not hide the fact that actions are firmly four-dimensional, and furthermore, that the mental models of actions supporting recognition may also be four dimensional.

The human action recognition is classified into two major types. One finding a 3D kinematic model associated with the action and the other involves obtaining local features detected at different locations in video and these features are fed to the classifier as training input to identify the action [5]. They have developed an action recognition system of the latter type, building a probabilistic model for classification and classification is done using the Support Vector Machine (SVM) with Principal Component Analysis (PCA). Another technique for action recognition based on the shape properties extracted from the Poisson Equation which resulted in space time saliency and space time orientation.

Regard human actions as three dimensional shapes induced by the silhouettes in the space time volume. The approach is based on the observation that the human action in video generates a space-time shape in the space time volume. These space-time shapes contain both spatial information about the pose of the human figure at any time, as well as the dynamic information (global body motion and motion of the limbs relative to the body).The method utilizes properties of the solution to the Poisson equation to extract space-time features such as local space-time saliency, action dynamics, shape structure and orientation. These features are useful for action recognition, detection and clustering. The method is fast, does not require video alignment and is applicable in (but not limited to) many scenarios where the background is known. Several other approaches use information that could be derived from the space-time shape of an action.

IV.CONCLUSION

The Elastic Sequence Correlation is a correlation procedure based on DTW and Approximate Pattern Matching to find the similarity between two images. The videos chosen were those containing a specific human action like walking, running, jogging, boxing and hand waving. Based on the distance between the shape contexts, the video closely matching the query is identified. The results obtained prove to give good recognition of the action in the query in the case of ESC when compared to the other algorithms: DTW, Approximate Pattern Matching.

V. FUTURE ENHANCEMENT

As an enhancement to the algorithm, the discrepancies (due to factors like dress of the person, direction of human action etc.), in identifying the action will be reduced in future by making a few changes in the algorithm.

REFERENCES

[1] Sayed Hossein Khatoonabadi and Ivan V. Baji´c, *Senior Member, IEEE(2013)* , Video Object Tracking in the Compressed Domain Using Spatio-Temporal Markov Random Fields, IEEE transactions on image processing, vol. 22, no. 1, january 2013.

[2] Serge Belongie, JitendraMalik,JanPuzicha(2010), Shape Matching and Object Recognition Using Shape Contexts, IEEE Transc. vol.24,No.24.

[3] Christian K¨as and Henri Nicolas (2009), An Approach to Trajectory Estimation of Moving Objects in the H.264 Compressed Domain, Author manuscript, published in "PSIVT, Tokyo : Japon (2009)".

[4] J. K. Aggarwal, M. S. Ryoo1, The University of Texas at Austin, Electronics and Telecommunications Research Institute, Human Activity Analysis: A Review.

[5] Nazim Mir-Nasiri (2006),Camera–based 3D Object Tracking and- Following Mobile Robot, published inRobotics, Automation and Mechatronics, 2006 IEEE conference.

[6] Zoltan Kato, Ting-Chuen Pong, John Chung-Mong Lee (2009) , Color Image Classification and Parameter Estimation in a Markovian Framework, IEEE transactions on image processing,2009.

[7] Rick Szeliski1, Ramin Zabih2, Daniel Scharstein3, Olga Veksler4, Vladimir Klmogorov1, Aseem Agarwala5, Mashall Tappen6, and Carsten Rother1, A Comparative Study of Energy Minimization Methods for Markov Random Fields, Pattern Analysis and Machine Intelligence, IEEE Transactions on Volume:30, Issue: 6 2008.

[8] Sayed Hossein Khatoonabadi and Ivan V. Bajić, Compressed-Domain Global Motion Estimation Based on the Normalized Direct Linear Transform Algorithm, Circuits and Systems, 2000.

[9] Matthew Brand, Nuria Oliver, and Alex Pentland (1996), Coupled hidden Markov models for complex action recognition, Media Lab Perceptual Computing / Learning and Common Sense Technical Report 407.

[10] Marina Georgia Arvanitidou (1), Alexander Glantz (1), Andreas Krutz (1), Thomas Sikora (1) Marta Mrak (2), Ahmet Kondoz (2), Global motion estimation using variable block sizes and its application to object segmentation, Image Analysis for Multimedia Interactive Services, 2009.

[11] Yike Liu1, Noise reduction by vector median filtering, geophysics, VOL. 78, NO. 3 (MAY-JUNE 2013); P. V79–V86, 18 FIGS., 1 TABLE. 10.1190/GEO2012-0232.

[12] Chad Aeschliman, Johnny Park, and Avinash C. Kak, A Probabilistic Framework for Joint Segmentation and Tracking, Computer Vision and Pattern Recognition 2010.

[13] C. Solana-Cipres, L. Rodriguez-Benitez1 J. Moreno-Garcia2 L. Jimenez1 G. Fernandez- Escribano3, Real-time segmentation of moving objects in H.264 compressed domain with dynamic design of fuzzy sets, IFSA-EUSFLAT 2009.

[14] Marina Georgia Arvanitidou, Michael Tok, Andreas Krutz and Thomas Sikora, Short-term motion- based object segmentation, Multimedia and Expo (ICME), 2011 IEEE International Conference.

[15] Dorin Comaniciu, Visvanathan Ramesh, Peter Meer, Kernel-Based Object Tracking, IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol. 25, No. 4, April 2003.

[16] Stuart Geman, Donald Geman, Stochastic Relaxation, Gibbs distributions and the Baysian Restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI – 6, Jan 2009.

[17] Jang, D,S. Jang, S,W. , Choi, H,I. (Oct.2006) , 2D human body tracking with structural KalmanFilter, PatternRecognition,vol.35,no.10,pp.2041-2050.

[18] Lee,W.Ryu,S.W., Lee,S.J. (Feb.2008), Motion based object tracking with mobile camera, Electronics Letters, vol.34, no.3, pp.256-258.

[19] Li Wang, Li Cheng, Liang Wang. (June 2011), Elastic sequence correlation for Human Action Analysis, IEEE Transc. on image processing,vol.20 No. 6.

[20] M. Blank, L. Gorelick, E. Shechtman, M. Irani,, R. Basri, Actions as space-time shapes, in Proc. Int. Conf. Comput. Vis., 2005, pp. 1395–1402.

Balakumar P is currently working as Associate Professor and Head of the department of Computer Science and Engineering, Mahendra Institute of Technology. He received the Ph.D degree in Oct 2011.

Vimala P is currently pursuing her Master's degree in Computer Science and Engineering, Mahendra Institute of Technology.