Fuzzy Rule-based Classification of Human Tracking and Segmentation using Color Space Conversion

Sivabalakrishnan.M[1] and Dr.D.Manjula[2]

[1]Research Scholar, Department of CSE, College of Engineering, Anna University, Chennai, Tamil Nadu, India
sbkrishnanm@gmail.com
[2]Assistant Professor, Department of CSE, College of Engineering, Anna University, Chennai, Tamil Nadu, India
manju@annauniv.edu

ABSTRACT

In this study we are proposing a fuzzy-based rule system for tracking people by using simple tracking algorithms. A simple hue, saturation, and value (HSV) histogram-based color model was used to develop our system. We started by separating the moving object region from the background region by comparing the current frame with the constructed background image. In this paper, as a first step we obtained a motion image through the acquisition and segmentation of video sequences. In this situation where object shadows appear in the background region, a pre-processing median filter is applied to the input image to reduce the shadow effect, before identifying major blobs. The second step includes generating a set of blobs from detected varied regions in each image sequence. When objects are closer the segmented object gets rejected and is thereby detected as a single object. This error can be rectified by using fuzzy logic. .

KEYWORDS

Median filter, object tracking, background subtraction, rgb2hsv, fuzzy classification

1. INTRODUCTION

Tracking a moving object is one of the challenging tasks in computer vision such as visual surveillance, human computer interactions, etc. Tracking is becoming an important task in video surveillance especially in monitoring large-scale environments such as public and security sensitive areas. Typically, in video surveillance, an object of interest is identified and then monitored or tracked. People are classically the objects of interest in video surveillance applications, for example while walking through a secluded or security sensitive area. There is now an increasing interest in monitoring people in public areas, for example in shopping malls. When tracking objects of interest in a wide or public area, additional parameters are required to improve performance such as color of clothing [1], path and velocity of tracked object [2,3], and modeling set colors for tracked persons [4]. To obtain robust tracking of a target, a number of tracking methods are typically employed in order to overcome problems such as occlusion [1, 4, 5] and noise in surveillance videos. Tracking objects is performed in a sequence of video frames and it consists of two main stages: isolation of objects from background in each frame and association of objects in successive frames in order to trace them. Background subtraction techniques are mostly used for detection of motion in many real-time vision surveillance applications. In these approaches, difference between the incoming frame and the background image is considered to detect foreground objects. Background subtraction provides the most

DOI : 10.5121/ijaia.2010.1406

complete feature data, but it is extremely sensitive to dynamic scene alterations due to illumination changes and extraneous events. Researchers are now devoted to developing a robust background model in order to prevent falseness in motion detection caused by scene changes. The background model is periodically updated by using a combination of pixel-based methods and object-based methods. Unfortunately robust and efficient object tracking is still an open research issue. To start object tracking, trackers need to be initialized by an outer component [6]. Object tracking in image processing is usually based on a reference image of the object, or properties of the objects [7]. Unlike other background maintenance approaches, we do not design any extra procedure to update background model; instead, the background model method is periodically reinitialized to obtain the newest background scene. In many vision surveillance applications, the moving targets cast shadows that pose a challenge for accurately detecting them. . Moving cast shadows can cause object merging, object shape distortion, and even object losses (due to the shadow cast over another object). For this reason, detection of moving shadows is critical for accurate object detection in vision surveillance applications. In recent years, many algorithms have been proposed in the literature that deals with shadows. In this study, we use HSV histogram – based model to make a robust color-based tracker system using a median filter. Blobs are defined as a group of pixels that belong to a similar object in motion. They have proven to be a better feature cue than points, corners or ridges as they usually have a larger coverage area and total occlusion of the subject is more unlikely to happen. Rossi and Bozzoli [8] successfully used moving blobs to track and count people crossing the field of view of a vertically mounted camera. In a different approach with blobs, Bregler [9] represented each pixel in each motion image by its optical flow characteristics. A different, graph-based approach is described in [10], where the problem of merging and splitting of detected blobs is solved systematically. In ‘closed world tracking’ described in [11] our objective is to design a robust and stable tracking method that can be adjusted to varying environments and can be used for different applications, such as person tracking. In this paper, a novel human motion detection algorithm that uses a fuzzy rule-based classification scheme on moving blob regions is proposed.

This paper is organized as follows. Section 2 gives an outline of Related Works. Section 3 offers a description of the technique of Background Subtraction. Section 4 describes the Proposed System Models for Object Tracking. Section 5 documents the Experiments and Analysis. Finally, Section 6 presents the Conclusions of our study.

2. RELATED WORKS

Rossi and Bozzoli [8] successfully used moving blobs to track and count people crossing the field of view from a vertically mounted camera. In a different approach with blobs, Bregler [11] represented each pixel in each motion image by its optical flow characteristics according to certain features of the flow vector. The color spaces that are typically used in video tracking and surveillance are YCbCr [13] and HSV [14]. As highlighted in [15], the more efficient notion will be to detect the presence of a human being without having to pre-determine its body segments. Polana and Nelson [16] were among the first to champion the idea of using low-level visual features to track human motion. In their own words, they proposed a way to “get your man without finding his body parts”. The task of detecting human motion is incomplete without the classification phase to distinguish human movements from other motions belonging to animals and objects. With the emerging use of fuzzy logic in various applications, fuzzy-based classification schemes [17] have also proven to yield better accuracy rates than conventional shape-based [18] and motion-based [19] techniques.

3. TECHNIQUE OF BACKGROUND SUBTRACTION

Background subtraction is one of the most popular methods for novelty detection in video streams. Background Subtraction generates a foreground mask for every frame. This step is simply performed by subtracting the background image from the current frame. When the background view excluding the foreground objects is available, it becomes obvious that the foreground objects can be obtained by comparing the background image with the current video frame. It focuses on two major steps: First, to construct a statitistical representation of the background that is representative, robust to noise and sensitive to new objects; second, to build another statistical model called 'foreground' that represents the changes that take place on the scene. By applying this approach to each frame one can effectively track any moving object. Moreover, a background image can be elegantly used to determine the foreground objects by comparing the input frame with the background image and marking the differences as foreground objects. This technique is commonly known as background subtraction or change detection. It is the most popular approach in video surveillance applications because it is a computationally efficient technique and offers relative ease in obtaining background images for static surveillance cameras. [12]

In practice, camera noise and regions in which the object is of the same color as the background make the separation of foreground objects and background more difficult. There are a few post processing filters that can remove obvious errors like small clutter regions. Consecutive images are also used for reducing noise with the help of background subtraction. In such a case, we can safely assume that the moving object is a human. We can then detect moving human region in each image.

4. THE PROPOSED SYSTEM MODELS FOR OBJECT TRACKING

The proposed algorithm consists of five stages: image acquisition, RGB to HSV conversion, BitXOR operation, preprocessing and blob identification. Figure 1 shows the process flow of the proposed human motion detection algorithm. Each of these stages will be described in detail. We extract features in the RGB color space. Two feature variables, chromaticity and brightness distortions, are used to classify the foreground and background. The color model used here separates the brightness from the chromaticity components. The foreground and background classification is based on the following observation: Image blocks in the background should have little change in their color distortion.

The proposed rule base should be able to achieve a robust tracking system using sparse and erroneous data from the simple image processing algorithms. The following problems need to be solved:

Initialization: In this phase a region or object can be identified using simple heuristics.

Tracking: During initialization, as well as during tracking, more than one feature Candidate for the object can be found in a search region. The task is to select the most appropriate object and remove ambiguity.

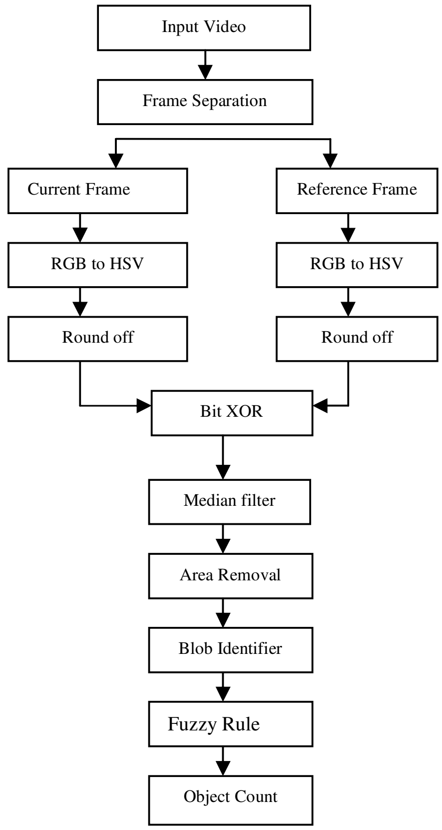


Figure 1. Proposed tracking system

Occlusions: Often during tracking blobs disappear due to occlusions. To achieve continuous tracking, occlusions need to be detected and movements need to be extrapolated until new data is valid.
Data loss: Using simple segmentation techniques, features may disappear in spite of the object being visible.

4.1 Image Acquisition and Segmentation

Image acquisition is the preliminary step in any motion-based vision application to obtain image frames from a stationary or moving camera, or from multiple cameras. Usually, a frame grabber is used to sub sample a sequence of video images at a specific frame rate before the actual processing begins. Video sequences can be of high frame rate (10-30*fps)* or low frame rate (< 10 *fps).* It is important to consider the type of video system used so that the appropriate segmentation method can be implemented. After the acquisition of image frames, image segmentation can be performed using background subtraction, depending on the frame rate of the video sequences.

4.1.1 Frame segmentation

For high frame rate sequences, the adjacent frame subtraction method is used since the change of motion between consecutive frames is very small. Thus, the difference image, $D(x, y, t)$ between an input frame and the next acquired frame after a fixed time interval c is given. The more widely known background subtraction technique is used for low frame rate sequences where the change of 1 *fps* denotes frames-per-second; and the standard measure of frame rate motion is larger. This method eliminates the stationary background, leaving only the desired motion regions. In some video systems, the acquisition process may be susceptible to erratic changes in illumination, reflection, and noise.

4.2 RGB to HSV Color space conversion

Color vision can be processed either by using RGB color space or HSV color space. RGB color space describes colors in terms of the amount of red, green, and blue that is present. HSV color space describes colors in terms of the hue, saturation, and value. In situations where color description plays an integral role, the HSV color model is often preferred over the RGB model. RGB defines color in terms of a combination of primary colors, where as HSV describes color using more familiar comparisons such as color, vibrancy and brightness. The color spaces that are typically used in video tracking and surveillance are YCbCr [13] and HSV [14].

HSV (M) returns an M-by-3 matrix containing an HSV color map. HSV, by itself, is the same length as the current figure's color map. The colors begin with red, pass through yellow, green, cyan, blue, magenta, and return to red. The map is particularly useful for displaying periodic functions. The mathematical relation between HSV and RGB is given by

$$H = \begin{cases} \theta & if\ B \leq G \\ 360^{0} - \theta & if\ B > G \end{cases} \tag{1}$$

$$Where\ \theta = cos^{-1} \frac{½ (R - G) + (R - B)}{[(R-G)^{2} + (R-B) (G-B)]^{1/2}}$$

$$S = 1 - [3/(R+G+B)]\ [min\ (R, G, B)] \tag{2}$$

$$V = 1/3\ (R+G+B) \tag{3}$$

Once the color space conversion is done for current and background frames, coefficients are then rounded and bitXOR mutually.

4.3 Pre-processing

The pre-processing stage consists of two tasks: morphological operations and blob identification, and is intended to prepare the motion image for the blob identification stage.

4.3.1 Morphological operations

The median filter is a nonlinear digital filtering technique, often used to remove noise. Noise reduction is a typical pre-processing step so as to improve the results. The median filter can eliminate the effect of input noise values with extremely large magnitudes.

Removal of motion at boundary – Pixels of the motion region that are located along the boundary are eliminated to avoid ambiguity of the region belonging to a possible moving object.

4.4 Fuzzy Classification

In the final stage, a fuzzy rule-based classification approach to classify the extracted major blob is proposed. The knowledge-base used here solves the problem of correspondence between the candidate regions to remove ambiguity, create reasonable extrapolation of movements and predict configurations of the model using a rule set. This rule set will prefer regions with a larger distance and prevent the tracking from finding near the head, prevents the head to be detected.

During tracking of the marker region, errors often occur due to imperfect tracking results, i.e. a region disappears or gets occluded, or the lighting level in the observed environment changes. It is necessary to update the model during tracking and to extrapolate the movements in case of occlusions or data loss and to determine the position in the next frame. If the image processing provides data, Another application is the overhead tracking of individuals and groups. Here, motion detection, luminance-level segmentation, and blob analysis are used together along with the rule-based approach that was earlier described. The process is as follows: Once the system identifies significant motion over at least five frames it will update its internal parameters for luminance-levels for the objects. This assures that the objects can still be tracked.

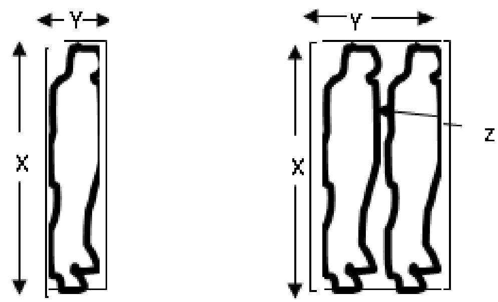


Figure 2. Silhouette of the user shown with Search regions

A fuzzy logic inference system is established to preserve the human body in frame *n* while detaching blocks that correspond to non-human artifacts. In the proposed method, we predict the bounding box of the human body in frame *n+1* from the statistics of previous frames. Our fuzzy logic inference system is based on the following observations:

(1) Each and every silhouette is identified.
(2) Bounding box is calculated for every individual.
(3) Person entering the focused screen is taken as an object and proceeds with an unique identity number until he leaves the frame.
(4) Continuous monitoring of the width and height of the silhouette.
(5) Counting is undertaken as soon as a new unique identifier number enters the frame.
(6) Counting memory of the unique identify number is resized and fuzzy is processed.
(7) Input is width and height of silhouette the bounding boxes.

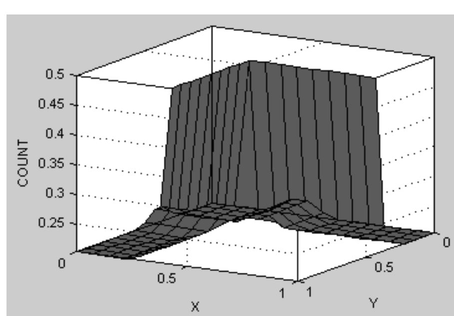
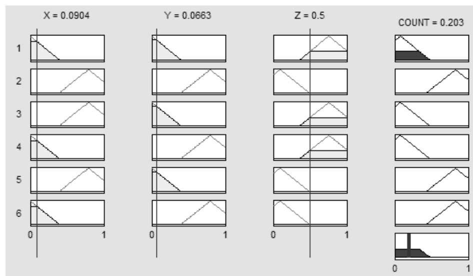


Figure 3. Surface viewer of our fuzzy system

The following six rules are used in our fuzzy logic in a MATLAB implementation. These are illustrated in Figure 3. Fuzzy rule is based with three input variables (*X*, *Y*, and *Z*) and one output variable *Count.*

1. If (*X* is CONSTANT) and (*Y* is CONSTANT) and (*Z* is FALSE) then (*COUNT* is ONE)
2. If (*X* is VARYING) and (*Y* is VARYING) and (*Z* is TRUE) then (*COUNT* is TWO)
3. If (*X* is VARYING) and (*Y* is CONSTANT) and (*Z* is FALSE) then (*COUNT* is ONE)
4. If (*X* is CONSTANT) and (*Y* is VARYING) and (*Z* is FALSE) then (*COUNT* is ONE)
5. If (*X* is VARYING) and (*Y* is CONSTANT) and (*Z* is TRUE) then (*COUNT* is TWO)
6. If (*X* is CONSTANT) and (*Y* is VARYING) and (*Z* is TRUE) then (*COUNT* is TWO)

Where *X* = width, *Y* = height, Z = holes



Figure~~.~~ 4. Membership function of six rules of fuzzy logic inference system

5. EXPERIMENTS AND ANALYSIS

5.1 Data acquisition and ground-truth

In our experiments we have used PET2006 Database, which consists of video sequences of human movements in various gait poses, non-human object motions and a combination of human and object movements, that have been constructed from our motion capture system. For validation, we tested the complete system in online human tracking experiments. No dataset was available for testing the complete tracking system, because of its dynamic nature. The tracking algorithm has been tested. We recorded all the experiments to extract their ground-truth manually for performance evaluation. MATLAB was used as the implementation tool.

In performance evaluation, each pixel in a background subtraction method of classification was determined to be true positive (TP)for a correctly classified foreground pixel, false positive (FP)for a background pixel that was incorrectly classified as foreground, true negative (TN) for a correctly classified background pixel, and false negative (FN)for a foreground pixel that was incorrectly classified as background. The different methods can be evaluated by the calculation of TP, TN, FP and FN. After every pixel had been classified into one of those four groups, the sensitivity and the specificity were calculated.

Sensitivity and specificity are statistical measures of the performance of a binary classification test. Sensitivity (also called recall rate in some fields) measures the proportion of actual positives, that are correctly identified. Specificity measures the proportion of negatives that are correctly identified. These two measures are closely related to the concepts of type I and type II errors. Sensitivity is defined in equation 4, and specificity is defined in equation 5

$$Sensitivity = TP / (TP + FN) \quad (4)$$
$$Specificity = TN / (FP + TN) \quad (5)$$

Sensitivity measures the proportion of actual positives that are correctly identified. Specificity measures the proportion of negatives that are correctly identified. These metrics are a measure of the accuracy or correctness of the tracking methodology that is being evaluated with respect to a reference point, known as the ground truth. Ground truth can be defined as reference or baseline data for determining the actual path of a tracked object. A correct track is considered true when the tracked point is within the boundary region of the tracked object.

For each object and each frame we now have a number of features (motion blobs, luminance blobs, and distance measures) that are combined using the fuzzy-based rating mechanism. The rule set uses several measures to rate candidate regions. For the tracking state these include: relative change in velocity, change in direction, blob size, luminance and distance from the last position. The result is a robust tracking of individuals that is stable even if lighting conditions differ.

5.2 Results

In Figures 1-3 we found that a) the input occurred without occulation. In Figures 1-3 we found that b) the corresponding human was tracked correctly with original bounding box and compared with ground truth. In Figures 1-3 we found that c) the input image occurred with occultation. In Figures 1-3 we found that d) the corresponding human was tracked correctly with predicted bounding box. Table 1 shows the number of objects before and after occultation. In order to avoid this situation, we implemented fuzzy rule based for the occulted frame. Based on the accurate prediction of human body position, the algorithm used a large temporal update window size unit in frames, for image blocks inside the bounding box while using a relatively small size for those outside the bounding box. The test cases showed the most frequent interaction patterns that have been observed and which had formerly led to errors in tracking. The fuzzy-based system is able to resolve all those dilemmas and can continuously track the user's body. A tracking system without the fuzzy interference system, did not offer the desired results. Hence, it was not possible to use the tracking results for the interaction of a user with a graphical interface. For this reason, we can only compare the performance of the two methods by stating that the improved one is efficient, whereas the other did not produce satisfactory results.

The second scenario showed an increase in tracking reliability after the fuzzy based method was applied, with a rule set adapted to the specific lighting conditions. The failure rate for critical cases was lower than seen with conventional tracking method.

By comparing the two tracking methods, it is very difficult to produce numbers that actually measure the problems we wanted to solve.

5.3 Discussion

Results show that our algorithm can handle and overcome the shortcomings of conventional tracking methods. This is because of using a repetitive target detection scheme and a motion prediction technique that do not rely on spatial proximity. Generally, according to the mean of distances, the location of the target is near the ground-truth. The target is usually localized within $1/4^{th}$ of the image diagonal from the image center. With the use of faster system frame rates, the results of tracking have improved significantly. In this case we will have more frames to process. When localization fails, it is because of similarity with or closeness to the color histogram of the target with other samples. The image resolution has an effect on the system frame rate and thus on tracking error. Table 1 shows the results of various frames before implementing fuzzy rule. Table 2 shows the results of various frames after implementing fuzzy rule. We can say that our maneuver towards fuzzy evaluation did improve stability. The fuzzy combination of motion-based tracking with luminance features did improve the tracking performance noticeably. Especially in difficult lighting conditions and situations with a patterned background it made tracking possible in situations where it was not possible earlier. Similarly, for the critical cases of first application, it made tracking possible where it was not possible earlier. There are no statistical results. These are based on mere human judgment, where it is stated that some critical problems have been solved.

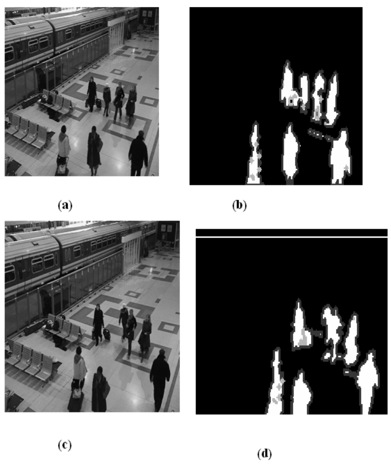


Figure 5. a) Original Image before occlusion
b) Silhouette extracted image before occlusion
c) Original image after occlusion
d) Silhouette extracted image after occlusion

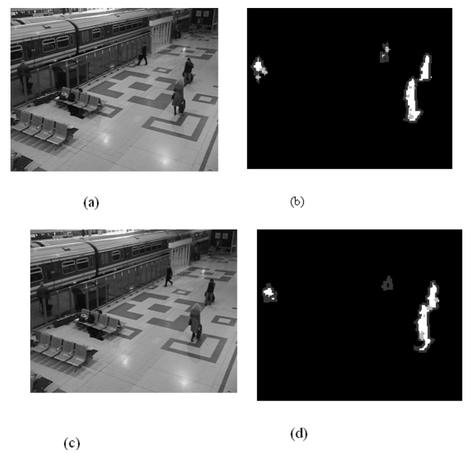


Figure 6. a) Original Image before occlusion
b) Silhouette extracted image before occlusion
c)Original image after occlusion
d) Silhouette extracted image after occlusion

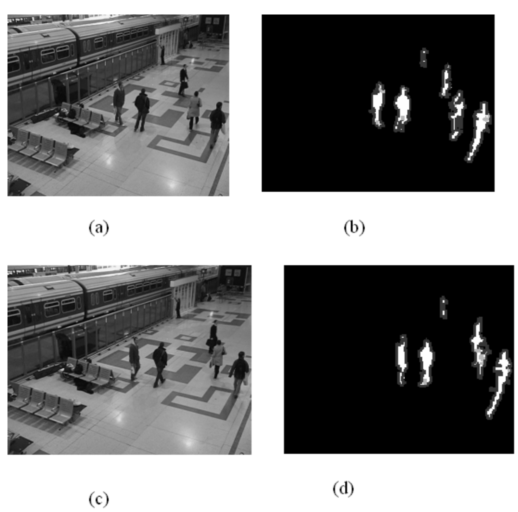


Figure 7. a) Original Image before occlusion b) Silhouette extracted image before occlusion
c) Original image after occlusion d) Silhouette extracted image after occlusion

S.I.No.	No. of objects (Before occultation)	No. of objects (After occultation)
1	7	6
2	4	3
3	6	5

Table 1: General system

S.I.No.	No. of objects (Before occultation)	No. of objects (After occultation)
1	7	7
2	4	4
3	6	6

Table 2: Proposed Fuzzy system

6. CONCLUSIONS

In this paper, we have proposed a fuzzy-based approach of combining several heuristics in image processing applications. The proposed method relies on a rule base that is used in a fuzzy inference system. For our study we built a background model and extracted classification features in RGB color space. To deal with the challenges of object extraction in dynamic environment, we fused high-level knowledge and low-level features and developed a fuzzy logic inference system. The results on several sequences show that this algorithm is efficient and robust for the dynamic environment, which contains new objects in it. We are currently working on making the prediction more accurate and creating a scheme to recover missing moving parts using known results from feature-based classification. We also intend to study the impact of the accuracy of results on the performance of future activity modeling and analysis. We believe that the proposed algorithm is not limited to the presented scenarios, but could be used as a general approach to introducing and adjusting heuristics in real-time image processing applications. For this reason, we are continuing our work on better future extraction to enable an intuitive tuning of the tracking system.

REFERENCES

[1] Bird, N.D., et al. (2005) “Detection of Loitering Individuals in Public Transportation Areas”, *IEEE Transactions on Intelligent Transportation Systems.*

[2] Bodor, R., B. Jackson, & N.P. Papanikolopoulos, (2003) “Vision-Based Human Tracking and Activity Recognition”, *11th Mediterranean Conference on Control and Automation.*

[3] Niu, W., et al. (2004) Human Activity Detection and Recognition for Video Surveillance”, *IEEE International Conference on Multimedia and Expo.*

[4] Iocchi, L. & R.C. Bollees occurred, (2005) “Integrated Plan-View Tracking and Color-based Person Models for Multiple People Tracking”, International *Conference on Image Processing*.

[5] Li, J., C.S. Chua, & Y.K. Ho, (2002) “Color-based Multiple People Tracking”, *7th International Conference on Control Automation, Robotics and Vision ICARCV.*

[6] Yunqiang, C. & R. Yong, (2006) “Real Time Object Tracking in Video Sequences”, *Signals and Communications Technologies, Interactive Video*, Part II, pp 67-88.

[7] Nummiaro, K., E. Koller-Meier, & L. Van Gool, (2002) "A Color-based Particle Filter", *Proc. First International Workshop on Generative-Model-Based Vision*, pp 53-60.

[8] Rossi, M., & A. Bozzoli, (1994), "Tracking and Counting People", *1st Int. Conf. on Image Processing*, pp 212-216.

[9] Masoud, O. & N. Papanikolopoulos, (June 1997) "A Robust Real-time Multi-level Model-based Pedestrian Tracking System", *Proc. ITS merica Seventh Annual Meeting.*

[10] Intille, S. J. Davis, & A. Bobick, (1997) "Real- time Closed-World Tracking", *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR '97).*

[11] Bregler, C. (1997) "Learning and Recognizing Human Dynamics in Video Sequences", *Proc. of IEEE CS Conf. on Comp. Vision and Patt. Recognition*, pp 568 -574.

[12] Sivabalakrishnan, M, & Dr. D.Manjula, "Adaptive Background Subtraction in Dynamic Environment using Fuzzy Logic", *International Journal on Computer Science and Engineering*, Vol. 2, pp 270-273.

[13] Chen, Y.J., et al. (2005) "The Implementation of a Standalone Video Tracking and Analysis System for Animal Behavior Measurement in Morris Water Maze", *27th Annual International Conference on Engineering in Medicine and Biology Society (IEEE-EMBS 2005), Shanghai.*

[14] Stern, H. & B. Efros, (2002) "Adaptive Color Space Switching for Face Tracking in Multi-Colored Lighting Environments", *Fifth IEEE Conference on Automatic Face and Gesture Recognition,* Washington, D.C.

[15] Aggarwal, J.K. & Q. Cai, (1997) "Human Motion Analysis: A Review", *Proceedings of IEEE Non-rigid and Articulated Motion Workshop*, pp 90-102.

[16] Polana, R. & R. Nelson, (1994) "Low Level Recognition of Human Motion (or how to get your man without finding his body parts)", *Proceedings of IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pp 77-82.

[17] Li, L., & M. K. H. Leung, (2001) "Fusion of Two Different Motion Cues for Intelligent Video Surveillance", *Proc. of IEEE Int. Conf. on Electrical and Electronic Tech. (TENCON)*, Vol. 1, pp 341-344.

[18] Lipton, A. J., Fujiyoshi, H., & R. S. Patil, (1998) "Moving Target Classification and Tracking from Real-time Video", *Proc. of IEEE Workshop on App .of Comp. Vis.*, pp 8-14.

[19] Cutler, R., & L. S. Davis, L. S. (2000) "Robust Real-time Periodic Motion Detection, Analysis, and Applications", *IEEE Trans. Patt. Anal. Mach. Intell,* Vol. 22, No. 8, pp 781-796.

M.Sivabalakrishnan is a Research Scholar in the Department of CSE, College of Engineering, Anna University, Chennai, India. He received his Master's degree from P.S.N.A College of Engineering, Anna University, Chennai, in 2004. His research interests include multimedia and image processing.

Dr. D. Manjula received her Master's and PhD degrees from College of Engineering, Anna University, Chennai, India in 1987 and 2004, respectively. She is currently an Assistant Professor in the Department of Computer Science and Engineering, Anna University, Chennai. Her research interests include NLP, Text Mining, Parallel processing, Grid computing, Database and multimedia.