International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 9 (2017) pp. 1848-1854


# Human path tracing using combined Color and Centroid feature in a Video

**Sunitha M R**
*Dept. of Computer Science & Engineering, AIT*
*Chikmagalur, Karnataka 577102, India.*

**H S Jayanna**
*Dept. of Information Science & Engineering, SIT*
*Tumkur, Karnataka, India*

**Ramegowda**
*BCE,*
*Shravanabelagolla, Karnataka, India.*

**Abstract**

Tracing the path of moving object has been done for videos with various methodologies to study the behavior of humans. The existing systems mainly focus on single feature and detect the objects by tracking the movement of feature selected. We propose a human path tracing method that uses low-level features like centroid location of traced humans by combining the color feature. Centroid location of each moving human in the current frame is computed using Kalman prediction technique based on previous locations of the object in earlier frames. Color features of each object are computed using histogram and match these color feature using Bhattacharyya coefficient method. Experimental result shows that the proposed method traces the path of tracked humans efficiently and also can handle occlusion.

**Keywords:** Gaussian mixture model, color histogram, multi-feature based tracking, centroid feature, Kalman prediction, Bhattacharyya coefficient.

## INTRODUCTION

Visual surveillance system has been used in many applications like traffic monitoring, shopping malls, military surveillance, and robotics [1]. The three stages in video analysis are: detecting objects which are moving, track their position with respect to scene and analyze their behavior.

Background subtraction is the simplest method to detect object in the scene by subtracting current frame from background model. The advanced method to generate the background model is Gaussian Mixture Model [2]. For every pixel, it maintains the mean and the variance of each pixel. It works well in different lighting condition. Other object detection methods are frame difference [3] and optical flow method [4].

Object tracking can be done using categories like model-based, region-based and feature-based [5]. Model-based method compares detected object in current frame with prior produced model. Region-based technique track object by object region in each frame. Feature-based method uses features like color, shape, texture, etc., of an object and matches them in successive frames.

In [6], model-based method was proposed to detect and track moving objects in the video stream. The method calculates best set of coherent motion regions with simple greedy method that detects and count similar objects moving in crowded scenes. Cognitive surveillance system was proposed which had learning phase to store objects behaviour. In later phase this stored information along with knowledge was considered for making automatic decision on crowd behavior [7].

Feature-based systems consider invariant feature such as color, texture, edges [8], [9], wavelet coefficient [10], spatial domain features [11] and Haar-like feature [12]. A feature-based method first identifies the feature and matches these selected features in each frame to track moving objects. Avidan used an adaptive ensemble of classifiers to distinguish objects from background. In this method, they have combined weak classifier into a strong classifier using AdaBoost technique [1].

CAMShift algorithm [13] is a simple, computational efficient method to track objects based on single hue. The major drawback of this method is with multi hue objects. This is overcome by combining multiple features of the object to improve tracking efficiency.

Many methods were introduced recently by combining multiple features to improve reliability and tracking performance. The integration of different clues into tracking framework has been discussed in [14], [15], [16]. Multi features like color and contour were integrated into particle filter using sampling technique [14]. Tracking scheme presented in [17] involved with fusion of local features and global appearance similarity.

A method proposed in [18] uses a particle filter to track humans using multi-features observation by exploiting skin and head-and-shoulder boundary at prior stage. It can track humans robustly in real-time videos.

The joint color-texture histogram to represent a target was presented in [19]. These representations were later applied to the mean shift. It also extracted another texture feature of detected object for representation. In this work they extracted only edge and corner features of the target region for robust tracking.

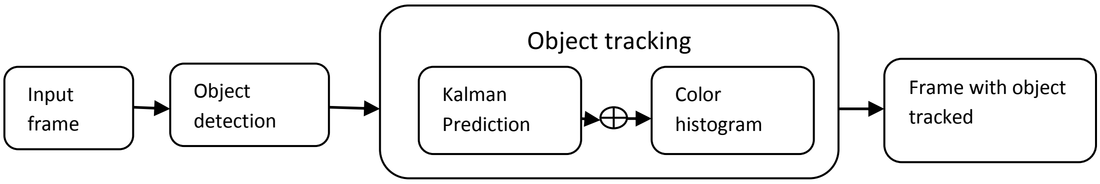


**Figure 1:** Block diagram of proposed system.

In [20] a method was presented to extract and adjust multiple feature spaces during tracking of an object. This improved the tracking performance by selecting features that best discriminate between object and the background. The method proposed in [21] combined both top-down features like faces, humans etc and bottom-up features like color, orientation, etc. to develop appearance model to track object. Selecting appropriate feature with distinct property, which complement its background gives uniqueness. Therefore, the object can be tracked easily in each frame [22].

The major challenge is identifying an appropriate feature to describe an object uniquely in the feature space and to distinguish from other object [23]. Some of the basic low level features like shape, texture, color, edge and motion can be used to describe an object. Features should be selected in such a way that it can sustain the challenges like change in illumination and motion blur.

The propose work focus on improving tracking results on optimization problem based on ideas proposed recently in the literature. Low-level features like color and centroid of moving objects are extracted and combined them together to

The rest of the paper is organized as follows. In Section 2, we propose the human tracking method based on combination of multi features. In Section 3, experimental results are presented. Finally, conclusions and future work are discussed in Section 4.

## PROPOSED METHOD

Many earlier works have shown that a combination of features could improve tracking performance [14], [15], [16]. This motivated to combine simple features like color and centroid of moving objects. The working principle of proposed method is given in the Figure 1. First frames are extracted from a given video. Each frame undergoes background subtraction to obtain the moving objects. For each detected objects two features like color and centroid are extracted. Later they are combined to predict and track the object. get better results as in [24]. The use of global information and complementary low-level feature solves the track loss issues to some extent compare to the single feature tracking.

### Object detection

Background subtraction is the simplest method used to detect moving objects in the scene. This method separates the foreground object from background in each frame. There are different approaches to extract the foreground from background using background subtraction method. Here, we have used Gaussian mixture model to construct the background model. Later, each frame is subtracted from the obtained background model to extract the foreground objects.

Simple Gaussian Mixture Model (GMM) proposed in [2] which is a successful technique that works well in challenging outdoor environments. A simple heuristic method is presented [25] which determine the intensity of the pixel which probably is a background. Later the pixel which does not match with these pixels is grouped as foreground. The mixture of $k$ Gaussian distribution [26] is as given below:

$$P(x_t) = \sum_{i=1}^{k} w_{i,t} \, N(X_t \mu_{i,t}, \Sigma_{i,t}) \quad (1)$$

Where

$$N(X_t \mu_{i,t}, \Sigma_{i,t}) = \frac{1}{(2pi)^{n/2}} \frac{1}{\Sigma_{i,t}^{\ 1/2}} \exp\left(-\frac{1}{2}(X_t - \mu_{i,t}) \Sigma_{i,t}^{-1} (X_t - \mu_{i,t})\right) \quad (2)$$

The first $B$ Gaussian distribution used as background of the scene is obtained as follow:

$$B = argmin\left(\sum_{i-1}^{b} w_{i,t} \; > T\right) \quad (3)$$

The $T$ is a threshold which is the minimum fraction of the background model. The other distributions are considered as foreground. For new frame at time $t+1$, a match is made for each pixel. A pixel matches a Gaussian distribution if the Mahalanobis distance is as follows:

$$((X_{t+1} - \mu_{t+1})^T) \sum_{i-1}^{b} (X_{t+1} - \mu_{t+1})^{0.5} < k * \sigma_{i,t} \quad (4)$$

Where $k$ is a threshold with value defined as $2.5$. Equation (4) yields two cases, either true or false. If it yields true, then use the following equation to update matched component of the $k$ Gaussians.

$$\sigma_{i,t+1} = (1-\rho)\sigma_{i,t}^2 + \rho(X_{x+1} - \mu_{x+1})\,(X_{x+1} - \mu_{x+1})^T \quad (5)$$

If none of the $k$ distribution component matched, then the least probability is replaced by distribution with current value as its mean, high weight variance and low weighted parameter.

### Tracking with multiple features

The proposed method to track human exploits on the fusion of centroid and color feature. It extracts centroid and color feature of each human in the frame. Later it merges these features and compares them in successive frame for tracking their position.

Tracking is divided into two stages, namely centroid matching using Kalman prediction and color histogram matching using Bhattacharyya coefficient. First centroid of the tracked object is computed and later, color histogram of each tracked object is calculated. The Centroid location of each object in current frame is tracked using kernel based prediction. Color histogram of each model is calculated and matched in current frame by using Bhattacharyya coefficient matching.

**Centroid feature**

Kalman method predicts the object's centroid location in each frame, and determines the likelihood of detection of each object with respect to its location tracking. Object motion in the current frame is an approximation of its motion in the previous frames. The Kalman filter is a recursive method with two stages. Method performs prediction and updating step in every iteration [27], [28]. The prediction step predicts the current centroid location of the moving objects based on earlier observation. The updating step calculates the present centroid location of the object and combines with the predicted location.

**Prediction state:** For each step $t$, Kalman method predicts current objects centroid location $x_t^-$ as shown below:

$$x_t^- = Ax_{t-1} + Bu_t \tag{6}$$

Where $x_{t-1}$ represents a vector of process state at time $t-1$, $A$ is a matrix which represents process transition state, $x_t$ is a control vector at time $t$ and $B$ converts control vector $u_t$ into state space.

**Update step:** After predicting location $x_t$, Kalman filter updates the measurement using following equation (7):

$$x_t = x_t^- + T_t\,(z_t - Hx_t^-) \tag{7}$$

Where $z_t$ contains two dimension with the form $[x_0,y_0,x_1,y_1,\ldots\ldots x_{n-1},y_{n-1}]$ with n different tracking algorithm. $H$ is a matrix defined as follow:

$$H = \begin{bmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \\ 1\,0\,0\,0 \\ 0\,10\,0 \\ \ldots\ldots. \\ 1\,0\,0\,0 \\ 0\,1\,0\,0 \end{bmatrix}$$

Where $R$ is noise covariance and $P_t$ is error covariance prediction as given below

$$P_t = A\,P_{t-1}\,A^T + Q \tag{9}$$

Where Q is process noise covariance.

**Color feature**

Color histogram of each detected object is calculated individually. For measuring two object similarities, Bhattacharyya coefficient matching technique is used [29]. When color histograms of two objects are compared using Bhattacharyya coefficient, the result obtained will be a value between *0* and *1*. This resulting value gives how well the two objects match each other. The value *1* represents similarity is high and similarity decreases as resulting value reaches *0*.

Let $h_1$ and $h_2$ be color histogram of two objects in two different frames. Then $S(h_1,h_2)$ gives similarity measure of two color histogram of two objects using Bhattacharyya coefficient method as given in equation (10).

$$S(h_1,h_2) = \sum\sum\sum(h_1 * h_2) \tag{10}$$

For two identical color histograms $h_1 * h_2 = 1$ gives a perfect match.

Later, both the color and centroid features are combined together to track the objects individually. The motion regions are usually around moving objects that are detected in the previous frame. Since in each frame objects are detected and tracked using feature sets, there is a greater opportunity to track objects with longer and uniform duration. Although sometimes object tracking may be lost or some may be tracked back again, but remaining objects will yield a coherent motion region with a high likelihood value.

The proposed method uses two features which complement and support each other to track coherent motion tracking region of the object by making the right selection. The trajectory of the human path is later drawn by plotting lines on location of the first position of object in the first frame to current position in current frame.

**Figure 2:** Tracking result of PETS 2009 view5 video Frames #12, #24, #36, #48, #60 and #72.

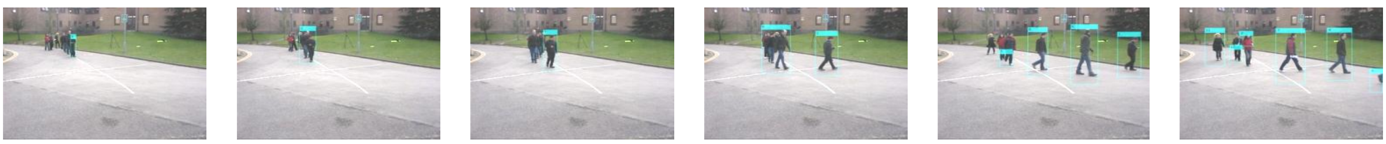

**Figure 3:** Tracking result of PETS 2009 view8 video Frames #12, #24, #36, #48, #60 and #72.


**Figure 4:** Tracking result of video2 with frame # 30, #40, #50, #60, #70 and #80.

## EXPERIMENTAL RESULTS

The proposed method is tested on a number of video sequences to evaluate the performance in terms of accuracy. Performance evaluation is done on standard data sets PETS 2009 available at [30] and in-house videos. The experimental result shows that the proposed method performs well in both in-door and out-door environment, when objects occluded and objects with similar color are present in the frames.

Simulated results for frame sequences with human tracking results gives suitability of the proposed method. The few experimental results on object tracking and trajectory of movement for both standard and in-house videos are given in following figures.

Figure 2 and 3 gives tracking result after applying the proposed method to PETS 2009 data set for view5 and view8 videos respectively. Figure 2 shows condition where humans are merged and still proposed method tracks the object as single object as in frame 36. As shown in Figure 3, the objects are occluded one behind another in the first few frames, and treated as a single object. Once they get separated one by one, they are tracked individually. Figure 4 gives tracking result of video with single human moving slowly.

Figure 5 gives the path of humans moving in the scene for PETS 2009 view1. As shown in the Figure 5, in frame #25 first human is able to trace even in occlusion condition. Proposed method also tracks object when they are detected partially. In figure 5 frames #64 shows that first human is further tracked after moved away from occlusion.

Figure 6 shows another case where humans are occluded partially and fully from one another and proposed method still able to trace them. Figure 7 shows how the proposed method works well for tracking humans with interhuman occlusion. Results show that proposed method works well in indoor video also as shown in Figure 8.

**Figure 5:** Path tracing of Video PETS 2009 view1 with frame #25, #50 and #64.

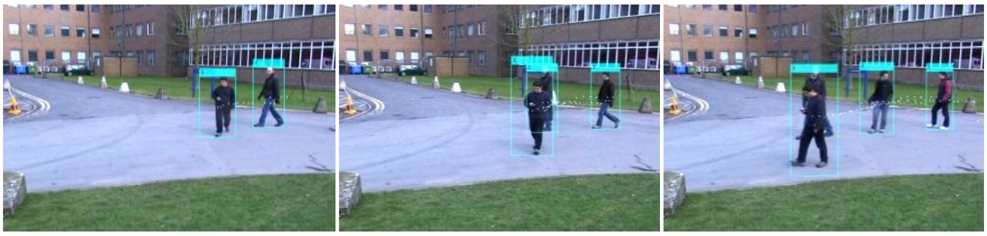

**Figure 6:** Path tracing of Video PETS 2009 view5 with frame #25, #35 and #45.


**Figure 7:** Path tracing of Video1 with frame #260, #290 and #370.


**Figure 8:** Path tracing of video3 with frame #5244, #5280 and #5450.

To evaluate the performance of the video segments, manually ground truth of moving objects were extracted and compared with the results obtained by proposed method [24]. Few parameters like True Positive (TP), False positive (FP) and False Negative (FN) are calculated.

Later, the performance metrics like Recall and Precision are calculated by using parameters like TP, FP and FN.

The Recall gives the percentage of matching true positive with ground truth as in equation (11).

$$\text{Recall} = \frac{\text{TP}}{(\text{TP+FN})} \tag{11}$$

Precision gives the percentage of true positive not matching with ground truth as in equation (12)

$$\text{Precision} = \frac{\text{TP}}{(\text{TP+FP})} \tag{12}$$

**Table 1:** Precision and Recall rate comparison of proposed method with CAMShift method.

| Method | PETS 2009 view1 | | PETS 2009 view5 | | Video1 | |
|---|---|---|---|---|---|---|
| | Recall | Precision | Recall | Precision | Recall | Precision |
| **CAM-shift** | 0.7892 | 0.9261 | 0.6955 | 0.9261 | 0.8922 | 0.8761 |
| **Our method** | 0.67235 | **0.9333** | 0.6428 | **0.9310** | 0.8281 | **0.9417** |

Both standard data set and publicly available data set were considered for evaluation purpose. Testing was done on several videos with objects moving in different speed, partial and full occlusions and in-door and out-door environment.

The precision and recall of CAM-shift method and proposed method for three different video are listed in Table 1. CAM shift is a state-of-the-art method that uses color feature to track moving objects. Table 1 shows that the precision rate for proposed method is better when compared to CAMShift method. However, the recall rate of the proposed method is poor, when compared to the other method. Our method has good performance in analyzing occlusion condition.

Based on results obtained, we can see that proposed method works well for tracking humans in both in-door and out-door environment. The method can also handle inter-object occlusion and background occlusion.

## CONCLUSIONS

In this paper we have proposed a method that traces human path based on combination of color and centroid feature. Objects are initially detected using Gaussian mixture model. Later these detected object's color and centroid features are extracted. The centroid of object in current frame is predicted by using Kalman filter method and color of object is computed and compared by using Bhatacharyya coefficient. We verify the proposed method effectiveness by conducting experiments on a wide range of video sequence recorded in both in-house and also with standard dataset. The result shows that proposed method performs well when compared to state-of-the-art method with different types of occlusion conditions. Table 1 show that the precision rate of our method is better than CAMShift method.

## REFERENCES

[1] S. Avidan, 2007, "Ensemble Tracking," IEEE Transaction Pattern Analysis and Machine Intelligence, 29(2), pp. 261-271.

[2] C. Stauffer and W. Grimson, 1999, "Adaptive background Mixture Models for real-time Tracking", Proc. IEEE Computer Vision and Pattern Recognition, Fort Collins, CO, USA, pp. 246-252.

[3] T. A. B. Wirayuda, K. A. Laksitowening, F. Sthevanie and R. Rismala, 2013, "Development methods for hybrid motion detection," Proc. IEEE International Conference of Information and Communication Technology (ICoICT)}, Bandung, Indonesia , pp. 218 – 222.

[4] M. Watanabe, N. Takeda and K. Onoguchi, 1996, " A moving object recognition method by optical flow analysis," Proc. 13th International Conference on Pattern Recognition, 1, Vienna, Austria, pp. 528 – 533.

[5] A. Yilmaz, O. Javed and M. Shah, 2006, "Object Tracking: A Survey," \emph{ACM Computing Survey}, 38(4), pp. 1-45.

[6] M. Cheriyadat, B. L. Bhaduri and R. J. Radke,2008, "Detecting Multiple Moving Objects in Crowded Environments with Coherent Motion Regions," Proc. 6th IEEE Workshop on Perceptual Organization in Computer Vision (POCV 08)in conjunction with IEEE CVPR08, pp. 1-8.

[7] S. Chiappino, P. Morerio, L. Marcenaro and C. S. Regazzoni, 2014, "Bio Inspired relevant Interaction Modelling in Cognitive Crowd Management," Journal of Ambient Intelligence and Humanized Computing, pp. 1-22.

[8] S. S. Intille and A. F. Bobick, 1997, "Closed World Tracking," Proc. 5th International Conference of Computer Vision, pp. 672-678.

[9] P. Tissainayagama, D. Sutter, Z. Tang, J. Xue, Y. Dai and J. Zheng, 2005, "Object Tracking in Image sequenced using Point Features," Pattern Recognition, pp. 105-113.

[10] S. M. youssef, M. A. Hamzs and A. F. Fazed, 2010, "Detection and tracking of multiple moving objects with occlusion in smart video surveillance systems," Proc. IEEE international conference on Intelligent Systems, London, United Kingdom, pp. 120-125.

[11] G. Pal, D Rudrapaul, S Acharjee, R Ray, S Chakraborty and N Dey, 2015, "Video Shot Boundary Detection: A Review," Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India CSI, 2, pp. 119-127.

[12] P. Viola, M. Jones and D. Snow, 2005, "Detecting Pedestrians using Patterns of Motion and Appearance," International Journal of Computer Vision, 63(2), pp. 153-161.

[13] D. Comanicium, V. Ramesh and P. Meer, 2000, "Measurement and classification of retinal vascular tortuosity," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2, pp. 142-149.

[14] M. Isard and A. Blake, 1998, "ICondensation: Unifying low-level and high-level tracking in a Stochastic Framework," Proc. European Conference on Computer Vision, pp. 893-908.

[15] D. Serby, E. K. Meier and L. Gool, 2004, "Probabilistic Object Tracking using multiple Features," Proc. 17th International Conference on Pattern Recognition (ICPR'04).

[16] O. Zoidi, A. Tefas and I. Pitas, 2013, "Visual Object Tracking based on local steering Kernels and Color Histograms," IEEE Transaction circuits and Systems for Video technology, 23(5), pp. 870-882.

[17] Z. H. Khan and I. Yu-Hua gu, 2010, "Joint Feature correspondences and Appearance Similarity for Robust Visual Object Tracking," IEEE Transaction on Information Forensics and Security, 5(3), pp. 591-606.

[18] T. Suwannatat, N. Indra-Payoong and K. Chinnasarn, 2015, "Robust human tracking based on multi-features particle," Proc. 12th International Joint Conference on Computer Science and Software Engineering (JCSSE), Songkhla, Thailand, pp. 12 – 17.

[19] Jifeng Ning, Lei Zhang, David Zhang and Chengke Wu, 2009, "Robust object tracking using joint color-texture histogram," Journal of Pattern Recognition and Artificial Intelligent, 23(7), pp. 1245-1263.

[20] R. Collins, Y. Liu and M. Leordeanu, 2005, "Online Selection of Discriminative Tracking Features," IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(1), pp. 1631-1643.

[21] H. Liu, Z. Liang and Q. Sun, 2014, "Robust Tracking with Discriminative Ranking Middle-level Patches," International Journal of Advanced Robotic Systems, 11(4), pp. 1-9.

[22] H. Yang, L. Shao, F. Zheng, L. Wangd and Z. Song, 2011,"Recent Advances and Trends in Visual Tracking : A review," Neurocomputing, Elsevier, 74, pp. 3823-3831.

[23] Anand Singh Jalal and Vrijendra Singh, 2012, "The State-of-the-Art in Visual Object Tracking," Informatica, 6, pp. 227-248.

[24] Sunitha M R, H S Jayanna and Ramegowda, 2014, "Tracking moving objects using combined color and centroid features," Proc. IEEE International Conference Computational Intelligence and Computing Research, Coimbatore, India, pp. 1-5.

[25] P. Kaewtrakulpong and R. Bowden, 2001, "An Improved Adaptive Background Mixture Model of realtime Tracking with Shadow Detection," Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, pp. 135-144.

[26] T. Boumans, F. E. Baf and B. Vachon, 2008,"Background Modelling using Mixture of Gaussians for Foreground Detection – a survey," Recent patents on computer science, pp. 219-237.

[27] M. Mehta, C. Goyal, M. Srivastava and R. Jain, 2010, "Real-time Object Detection and Tracking: Histogram Matching and Kalman Filter Approach," Proc. 2nd International Conference on Computer and Automatic Engineering, Singapore, pp. 769-801.

[28] G. Welch and G. Bishop, 2001, "An Introduction to the Kalman Filter," 2001, Proc. SIGGRAPH, pp. 19-24, 2001.

[29] Dubuisson, S, 22006, "The computation of the Bhattacharyya Distance between Histograms without Histograms," Proc. 2nd International Conference on Image Processing Theory Tools and Application IPTA, Paris, France, pp. 373-378.

[30] "Performance Evaluation of Tracking and Surveillance Workshop at CVPR 2009," 2009, http://www.cvg.rdg.ac.uk/PETS2009/