

Target Detection and Tracking for Video Surveillance

S.VASUHI¹, V.VAIDEHI²

Department of Electronics Engineering¹, Department of Information Technology²,
Anna University

Madras Institute of Technology Campus, Chennai.

INDIA

¹ vasuhisrinivasan@yahoo.com, ² vaidehivijay@gmail.com

Abstract— Target detection and tracking is an important problem in the automatic surveillance system. This paper proposes a Combined Gaussian Hidden Markov Model based Kalman Filter (CGHMM-KF) scheme for tracking people in multiple camera sensor network for monitoring and tracking of target (person/vehicle) in secured area. To detect the target under different illumination conditions, HMM with Mixture of Gaussians (MoG) is adapted. The MoG estimates the background and detects the foreground and the HMM modeling technique captures the shape of the desired object from the foreground. Finally, tracking of multiple targets is done by Kalman Filter (KF) with a bounding box, indicating the location of the person even with the motion in the background. The area of coverage can be extended dynamically using multiple cameras. The proposed approach provides better detection and tracking of person even in the presence of occlusion, target miss association and multiple persons in the environment.

Key-Words —HMM, Kalman Filter, MoG, Multi-camera, Multi-Target.

1 Introduction

Video surveillance has long been in use for surveillance and monitoring purpose in highly secured areas like banks, malls, industries etc. Traditionally, the video streams are monitored online by human operators and stored for future reference. Many deployed systems are able to reliably detect and track the movement of persons in indoor and controlled environment, with a unique identification mark like RFID tag [2]. Though this approach is fruitful for the controlled indoor environment, it will not be able to detect and track when the person is not wearing the RFID tag. To overcome this problem, the image based detection and tracking of the target is proposed.

Vision based multi-target tracking has been studied extensively in the literature and several algorithms are available in the literature to track people using camera images [1], [5], [7], [8], [20]. However, most of the existing methods still have severe limitations such as camera position, Noise image due to a poor quality image source, varying pose, illumination, moved background objects, shadows and occlusion conditions. Classifying multiple detected targets into human, vehicle or animal is yet another difficult problem and computationally expensive task. The first step in the detection of a target is the extraction of robust feature set for clearly discriminating target

of interest from the background. The problem of tracking people in real time for an interactive environment using [9], [21] background subtraction model, which models the background as static environment, but it requires more texture information for dense stereo reconstruction. The different features include template, color, contour, histogram of gradients [3], etc. of an object image.

The problems associated with automatic real time visual surveillance included tracking unwanted target rather than desired target, and changes in the background [4], occlusions, and the assumption that the background environment is a static model [5]. The distance metric learning reliably represented the similarity between different appearances of the object as well as the difference in appearance between the object and the background [1].

Moreover, scenes often included many other dynamic objects, fast changes in lighting, and complex object interactions [17] shadows and reflections that greatly influence the image. The techniques discussed in [4] assume that static and dynamic occlusions are rare and can be handled as short-term special cases. However in many real-world settings, it is not possible to place a camera in the ideal location that minimizes occlusions (like a very high overhead view). Hence a robust technique

for tracking is required to handle frequent and prolonged occlusions, to work in crowded areas with multiple views. [6], [7].

When surveillance was performed over a wide area, multi-camera techniques needed to provide correspondence between views and assumed that the appearance of a feature in one view will be similar to its appearance in another view. This assumption failed for widely separated views where the scene geometry and lighting could result in a lack of commonly observed features [8].

The person tracking was performed from offline data using background subtraction and multiple cameras; it introduced some delay in segmentation because sensor fusion is done by rendering foregrounds from multiple sensors image [9]. The region based stereo technique required background modelling and assumed that everyone in the scene is wearing uniquely coloured clothing [10] to perform the region based correspondence.

Kalman filtering was the first used filter for visual tracking. Since various extensions of the filter have shown much success [11], [3],[12],[22] for person tracking. When the state space is discrete and is made up of a finite number of states, Hidden Markov Model (HMM) [13],[18] can be applied for tracking.

Tracking of people with particle filter [14] demonstrates tracking in a cluttered office environment with two people but does not discuss the cost of rendering an image from a model per particle per time step [15]. Tracking methods are based on the visual hull techniques, which are sensitive to errors in foreground segmentation and are not suited for environments with many occlusions because the visual hull becomes loose and cannot resolve individuals [16].

To overcome the problems such as changes in the background, occlusion, colour, texture, size etc, in the existing multiple target tracking methods, a novel Combined Gaussian Hidden Markov Model and Kalman Filter (CGHMM-KF) is proposed in this paper. The paper is organized as follows. The proposed system overview is given in section 2. Section 3 explains the target detection and tracking. The implementation details and results are discussed in section 4. And the paper concludes in section 5.

2 System Overview

In this paper, a system that accomplishes the real-time simultaneous tracking of multiple targets in complex real time scenarios is proposed using Combined Gaussian Hidden Markov Model and Kalman Filter (CGHMM-KF) scheme. The proposed system uses background estimation to detect the

foreground objects and the presence of object is identified using Pseudo-2D Hidden Markov Model (P2DHMM) with Mixture of Gaussians (MoG) primitives. For tracking, Kalman Filter is used to predict the location of interested object (person) with Centre of Gravity (CoG) parameter. The flow diagram of proposed CGHMM-KF target tracking system in camera sensor network is shown in Fig. 1.

3 Target Detection and Tracking

The first step of the tracking is separation of the targets from the background. There are two approaches used for identification of the target. The identity of the target can be obtained manually by configuring the system or else by training a stochastic model using various images of the target, to automatically detect the target in the frames.

There are two common methods are in use for the detection of the target from the video, namely, motion segmentation and background subtraction. Motion segmentation is basically a threshold of the difference between the current image and the sequence images by assuming that the background does not change over successive frames. This method is easy and fast in many applications, but some problems appear when tracking multiple targets or when a target stops. The feature based background estimation for detection of a target involves two steps, namely, feature extraction and classification. A detection process is more efficient if it is based on the features that encode some information about the class to be detected.

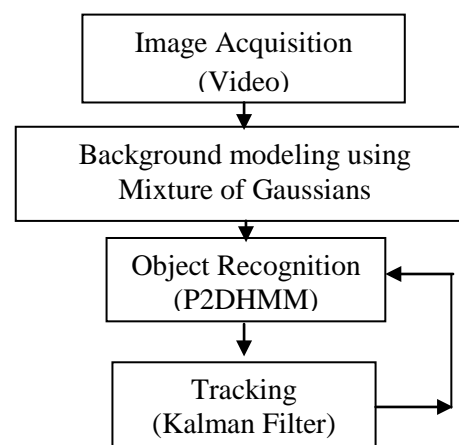


Figure 1. Flow diagram of proposed target tracking system from video images

3.1 Background modeling using Mixture of Gaussians

Though the background is static there may be some background variations due to lightening and

illumination changes, pose and view point variations. Hence, Mixture of Gaussians (MoG) approach is used for background modeling. From the video frame, background estimation based on MoG is used to detect the moving targets from the static background.

MoG is a probability density function of X, where X is a random process representing the pixel at time k. Here, the intensity features of each pixel $X_1, X_2, X_3, \dots, X_k$ modeled by mixture of M Gaussians [20]. Each pixel intensity in the frame is represented by a mixture (number) of Gaussian functions that sum together to form the probability of observing the current pixel value, which is given by the equation (1)

$$P(X_k) = \sum_{j=1}^M w_{j,k} \eta(X_k, \mu_{j,k}, C_{j,k}) \quad (1)$$

where, M is the number of distributions, in the proposed model three Gaussians distributions i.e M = 3 are considered.

$w_{j,k}$ is the weight parameter of the jth Gaussian component at time (frame) k and satisfy the constraint $\sum_{j=1}^M w_{j,k} = 1$

$\eta(X_k, \mu_{j,k}, C_{j,k})$ is the distribution of jth component, and it is represented by equation (2)

$$\eta(X_k, \mu, C) = \frac{1}{(2\pi)^{D/2} |C|^{1/2}} e^{-\frac{1}{2}(X_k - \mu)^T C^{-1} (X_k - \mu)} \quad (2)$$

where $\mu_{j,k}$ is mean at frame k and $C_{j,k}$ standard deviation of jth component.

The parameters of the MOG's model are the number of Gaussians M, the weight $w_{j,k}$ associated to the jth Gaussian at time k, the mean $\mu_{j,k}$ and the covariance matrix $C_{j,k}$. The parameters are initialized with $w_j = w_0, \mu_j = \mu_0, C_j = C_0$.

For an incoming new frame at times k+1, a match test is performed for each pixel and a pixel matches Gaussian distribution if the Mahalanobis distance is calculated using equation (3)

$$\sqrt{(X_{j,t+1} - \mu_{j,t+1})^T C_{j,t+1}^{-1} (X_{j,t+1} - \mu_{j,t+1})} \leq s\sigma_{j,t} \quad (3)$$

where, D is a constant deviation threshold.

After performing the match test for the newly observed pixel, if a match is found with one of the M Gaussians, the update is done using equations (4) to (6).

$$w_{j,k+1} = (1 - \psi)w_{j,k} + \psi \quad (4)$$

where ψ is a constant learning rate. The mean and variances are updated as follows.

$$\mu_{j,k+1} = (1 - \rho)\mu_{j,k} + \rho X_{k+1} \quad (5)$$

$$\sigma_{j,k+1}^2 = (1 - \rho)\sigma_{j,k}^2 + \rho(X_{k+1} - \mu_{j,k+1})(X_{k+1} - \mu_{j,k+1})^T \quad (6)$$

where $\rho = \psi \eta(X_k, \mu_{j,k}, C_{j,k})$

If the no match is found, then the Gaussian distribution with low probability is replaced with the new distribution. The current pixel value has initially high variance and low prior weight. The weight is updated as follows,

$$w_{j,k+1} = (1 - \psi)w_{j,k}$$

The MoG distinguish the pixel which was associated to the foreground and which was assigned to background. For the foreground detection, the Gaussian distributions are normalized, and the weight

to standard deviation ratio is given by, $\frac{w_j}{\sigma_j}$. The

background components which have the low variance and the high weight will stay at the top of the distributions.

The first B Gaussians distributions which exceed certain threshold Th are chosen as the background distribution,

$$B = \arg \min_b \left(\sum_{j=1}^b w_{j,k} > Th \right) \quad (7)$$

Where, b is the number of background components. The threshold Th is the background component weight threshold.

The remaining distributions are considered to represent a foreground. In order to determine if a new pixel is part of the background, compare it with the existing B Gaussian distributions in turn. If the pixel value is within a scaling factor of a background distribution's standard deviation, it is considered part of the background or otherwise, it is foreground [21]. The detected foreground objects are processed with a PCA based feature extraction method that is adopted from a face recognition system [23]. The result of the feature extraction is a two dimensional array of data and it is applied to P2DHMM for learning the structure of the human body. For target recognition, the P2DHMM is trained with MoG features. The desired target is recognized by proper training of P2DHMM.

3.2 Target Recognition using Hidden Markov Model

For recognition of a desired target among the foreground targets, P2DHMM is used. Pseudo 2-D HMMs (P2DHMMs) are extension of the 1-dimensional HMM in order to model 2-dimensional (2D) data. They are called pseudo, due to the fact that they are not real as 2D since it does not connect all possible states and state alignments of consecutive columns are calculated independently.

HMM is a doubly stochastic process with an underlying stochastic process that is not observable (hidden), but can only be observed through another set of stochastic processes that produce the sequence of observed symbols. To explain the elements and mechanism of HMM, consider there is n number of states in the model. At each time k, a new state is entered, based on a transition probability distribution which depends on the previous state (the Markovian property). After each transition, an observation output symbol is produced according to a probability distribution which depends on the current state [13].

Fig. 2 shows a Markov chain with 4 states. At any index of time, system undergoes a change of state (possibly back to the same state) according to a set of probabilities. A full probabilistic description of the above system requires specification of the current state and all previous states.

The stochastic process could be called an observable Markov model since the output of process is the set of states at each instant of time, where each state corresponds to a physical event. The Markov assumption states that the probability of the occurrence of pixel at time k depends only on occurrence of pixel at time k-1.

Each column of the image will be assigned to one of the super states and blocks in the column will be assigned as states [19].

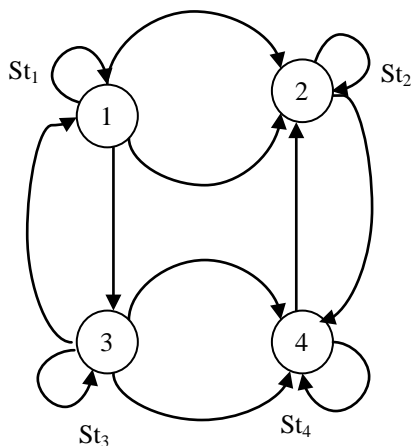


Fig. 2 A Markov chain with 4 states

The parameters associated with HMM are number of states, number of events, initial-state probabilities, state-transition probabilities and discrete output probabilities.

The set of states are denoted by $St = \{St_1, St_2, St_3, \dots, St_n\}$, where n is number of states.

State of HMM at instance k is denoted by q_k [24]. The state transition probability distribution A is given by $\{a_{ij}\}$, where a_{ij} denotes the probability of changing the state from St_i to St_j as given by the equation (8) and (9),

$$a_{ij} = P[q_{k+1} = St_j / q_k = St_i], 1 \leq i, j \leq n \quad (8)$$

$$\sum_{j=1}^n a_{ij} = 1 \quad (9)$$

The observation symbol probability distribution B is given by, $\{b_j(o)\}$,

where, $b_j(o) = P[o \text{ at } k / q_k = St_j], 1 \leq j \leq n$ is probability of observing the symbol o in state St_j at instance k.

$\pi = \{\pi_i\}$ is initial state distribution, where π_i gives probability of HMM being in state St_i at the instance $k = 1$. $\pi_i = P[q_1 = St_i], 1 \leq i \leq n$.

The above stochastic process could be called an observable Markov model since the output of the process is the set of states at each instant of time. The transition probabilities control the way, the hidden state at time k is chosen given the hidden state at time k-1. The output probabilities govern the distribution of the observed variable at a particular time given the state of the hidden variable at that time.

The P2DHMM is capable of modelling the target (human body/any target) with varying positions. To identify a person/target in the complex environment, the P2DHMM is trained with static images. The images are pre-processed with the edge detector and trained. The transition probabilities control the way the hidden state at time k is chosen given the hidden state at time k-1. The output probabilities govern the distribution of the observed variable at a particular time, given the state of the hidden variable at that time.

3.3 Tracking using KF

Kalman filter (KF) is a set of mathematical equations which provide an efficient computational solution to sequential systems. The filter is very powerful in several aspects: it supports estimation of past, present, and future states (prediction). KF is mainly utilized for occlusion detection and to solve target miss association problem. KF uses the output of the P2DHMM for tracking the detected person by the estimation of a bounding box trajectory indicating the location of the person within the video sequence. From the output of the P2DHMM, a bounding box is constructed around the detected person. The Centre of Gravity (CoG) of the person is calculated from the output of the P2DHMM using the Viterbi alignment [13][18].

The coordinates of the COG are denoted as x_s, y_s and the size of the bounding box denoted as w and h , serves as a measurement input to the KF.

The tracking procedure starts with the presentation of the first frame of the video sequence to the trained P2DHMM. Once the person is found among the moving targets, a Kalman Filter (KF) is used to achieve continuous tracking of the person. The CoG of the detected person by P2DHMM is computed to know the position of the person in the frame. The coordinates of the CoG, width and height of the bounding box are given as inputs to KF. The state and measurement equation for KF is given by equations (10) and (11) respectively.

$$x_k = Fx_{k-1} + Gp_{k-1} \quad (10)$$

$$Z_k = H x_k + v_k \quad (11)$$

where, x_k - State vector at time k (frame k)

F - State transition matrix

G - Control input matrix

p - Process noise

Q - Process noise covariance

$$Q = E\{p_k p_k^T\}$$

v_k - Measurement noise

R - Measurement noise covariance

$$R_k = E\{v_k v_k^T\}$$

Z_k - Actual measurement at time k , provides position of the person in the captured frame as COG coordinate and size of the bounding box. The state transition matrix F is given by equation (12)

$$F = \begin{bmatrix} 1 & 0 & T & 0 & 0 & 0 \\ 0 & 1 & 0 & T & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (12)$$

H - is Measurement matrix given by equation (13)

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (13)$$

The state vector constituted by the information about the detected foreground object includes position, velocity and size of the bounding box. The velocity information is computed from the position.

$$x = \begin{bmatrix} x_p & y_p & x_v & y_v & w & h \end{bmatrix}^T$$

The measurement matrix provides the information about the center of the detected object and size of the bounding box.

$$Z = \begin{bmatrix} x_p & y_p & w & h \end{bmatrix}^T \quad (14)$$

where, x_p and y_p - positional components in x and coordinate of Center of Gravity (CoG) for the detected person.

x_v and y_v are the horizontal and vertical velocity of CoG of person.

w and h are width and height of person [19] and T is the sampling time duration.

Any unexpected movement or deviation (maneuver) is modeled as a zero mean Gaussian system noise with covariance Q . Assume that the person moves with constant velocity between two consecutive frames $k-1$ and k . The measurement error is modeled as the error in the measurement which is the maximum distance at which a detected target is being sensed by the detection system.

Based on measurement, the KF constructs the state vector x and computes the predicted and the filtered state vectors according to the following equations from (15) – (20). The predicted state vector is

$$x_{k/k-1} = F_{k-1} x_{k-1/k-1} + G_{k-1} \quad (15)$$

The predicted error covariance is:

$$P_{k/k-1} = F_{k-1}P_{k-1/k-1}F_{k-1}^T + Q_{k-1} \quad (16)$$

Innovation (or) residual error is given by

$$e = (Z_k - H x_{k/k-1}) \quad (17)$$

Kalman gain is given by:

$$K_k = P_{k/k-1}H^T [HP_{k/k-1}H^T + R_k]^{-1} \quad (18)$$

Filtered state vector is given by

$$x_{k/k} = x_{k/k-1} + K_k(Z_k - Hx_{k/k-1}) \quad (19)$$

Filtered error covariance is given by:

$$P_{k/k} = (I - K_kH)P_{k/k-1} \quad (20)$$

Tracking of multiple people is done by Kalman Filter (KF) with the located CoG points in the frame of the entire video sequence. The filtered state vector x_k is used in return as input to the P2DHMM in order to improve the estimation of the measurement vector Z_k . This provides a cooperative feedback between the KF and the P2DHMM and vice versa.

Occlusion is defined as the lack of visual clues in the video. The ability of tracking algorithms to handle occlusion is crucial to provide a good estimate of the object state. Occlusion handling aims to reduce the effects of the lack of information on an object under occlusion. In the multiple person crossing scenarios, the output of the KF helps in tracking the specific person under occluded condition.

From the inputs obtained from frame $k-1$ and the measurement equation Z_{k-1} , the system will predict the state vector x_k . The state vector x_{k-1} is used to mark the position of the object in the images of the sequence. The predicted vector x_k is fed back as input to improve the estimation of vector Z_k in the next frame. That is to say the bounding box predicted by the KF is enlarged. The example of occlusion handling of KF is shown in Fig. 3.

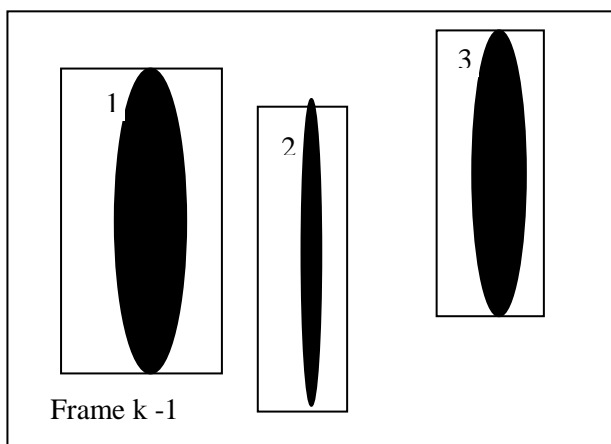


Fig. 3 Detected Blobs in the frame k -1

From the Fig. 3, it is observed that, there are three different persons are identified from the precious frame $k-1$ and labeled as target 1, 2 and 3 by the bounding box. Next consider the current frame k which is shown in Fig. 4.

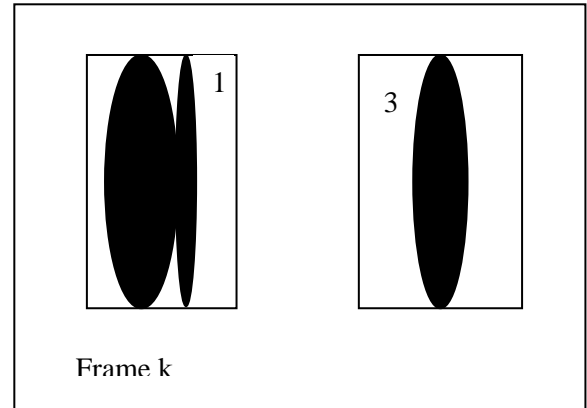


Fig. 4 Detected Blobs in the current frame k

The Fig. 4 shows that, only two blobs are identified in the current frame. The tracker labeled as target 1 and 3. Therefore, it is a need to identify whether target number 2 has disappeared or occluded. In this situation, the KF prediction is utilized for targets position identification in current frame. As can be seen, for target 3 only is in its bounding box, but target 1 and 2 are in bounding box of target 1 and since the tracker labeled this blob as target 1. From the above discussion conclude there is an occlusion between target 2 and target 1.

4 Results and Discussion

The system is tested in typical indoor and outdoor environments for handling various situations; background modeling, occlusion and target miss association. The system was initially tested in non-real time offline mode using videos in .avi and .mp4 formats. The videos run at 30fps speed and the frame size is 240*320 pixels.

The presence of a moving target is identified by background estimation of MoG. From the foreground the blobs are identified and the center point of the blob is identified as CoG. The bounding boxes are placed around the blobs to show the tracking action. The rectangular boxes surround the target that is being tracked. Boxes are colored differently for different targets for convenience. The output is saved as video in .avi format for future reference.

Consider scenario 1 with single person moving from right to left and left to right of camera view. The video is converted into number of frames. The

background estimation is performed MoG and the detected foreground for selected frames are shown in Fig. 5.

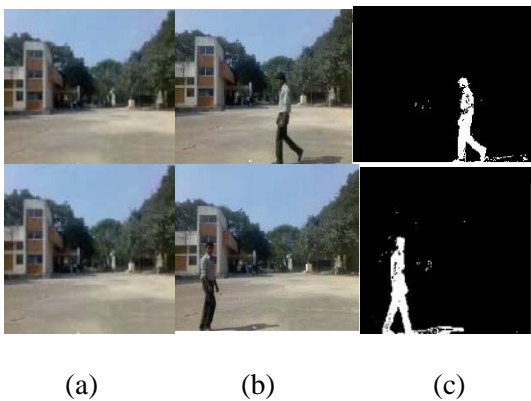


Fig. 5 Results of MoG for scenario 1

In Fig. 5, the first row shows the results of frame 90 and the second row shows the results of frame 120. Each row contains the estimated background (Fig. 5 a), original frame (Fig. 5 b) and the detected foreground (Fig. 5 c). The foreground contains blobs of the moving targets.

In scenario 2 with single person is considered. The captured image, background image, and the results of MoG are shown in Fig. 6 (a), (b), (c) respectively.

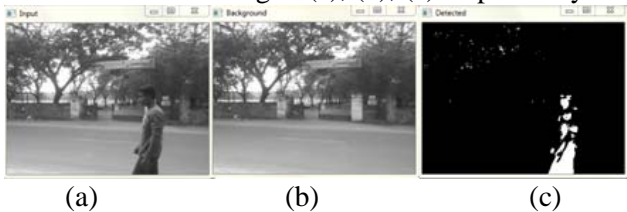


Fig. 6 Results of MoG for scenario 2

From Fig. 6 it can be inferred that, the output for background estimation using MoG is obtained and it identifies the moving target as blob. In Fig. 6 (a) shows the captured image in the scenario and the estimated background image are shown in Fig. 6 (b). The detected blob from the foreground is shown in Fig. 6 (c).

Consider the same scenario 2 with different moving targets in the background is considered. The Fig. 7 shows an assortment of targets like cars, persons walking; persons riding a bike etc detected using this method.

The detected targets are recognized using the target recognition techniques which are employed to segment out the desire target alone and accurately estimate the background even in the presence of moving targets. The moving targets are correctly identified to be a part of the foreground and not of the background. In Fig. 7 (a) the captured image in the scenario is given, from the captured image the background image shown in the Fig. 7(b). The

detected blobs in the foreground are shown in Fig. 7(c).



Fig. 7 Results of MoG for scenario 2 with multiple moving targets

From the scenario 2 shown in Fig. 7, it is observed that a person initially moves towards left side of the camera view and he takes a turn and move towards the right side of the scene. All the moving vehicles in the foreground are identified as blobs.

The single person tracking from scenario 2 is shown in Fig. 8. From the foreground, the desired targets are identified by training using P2DHMM. Tracking is performed by KF shown by red colour bounding box around the detected blob.



Fig. 8 Single person tracking in scenario 2

From Fig. 8, it is observed that the 2DHMM is trained only for the person, so the scooter with two people is not identified.

In the proposed system, it is also possible to track a single person in the presence of other moving persons by specifying the target number in the beginning and also able to track a single person in

presence of multiple people. Consider a scenario with three people is shown in Fig. 9. The column (a) and (b) in Fig. 9 shows the input and tracking output. The first row shows the background; in the second row person 1 is seen entering into the scenario. The third row represents the entry of person 2 in the scenario and, in the fourth row, person 3 comes into camera view of the camera.

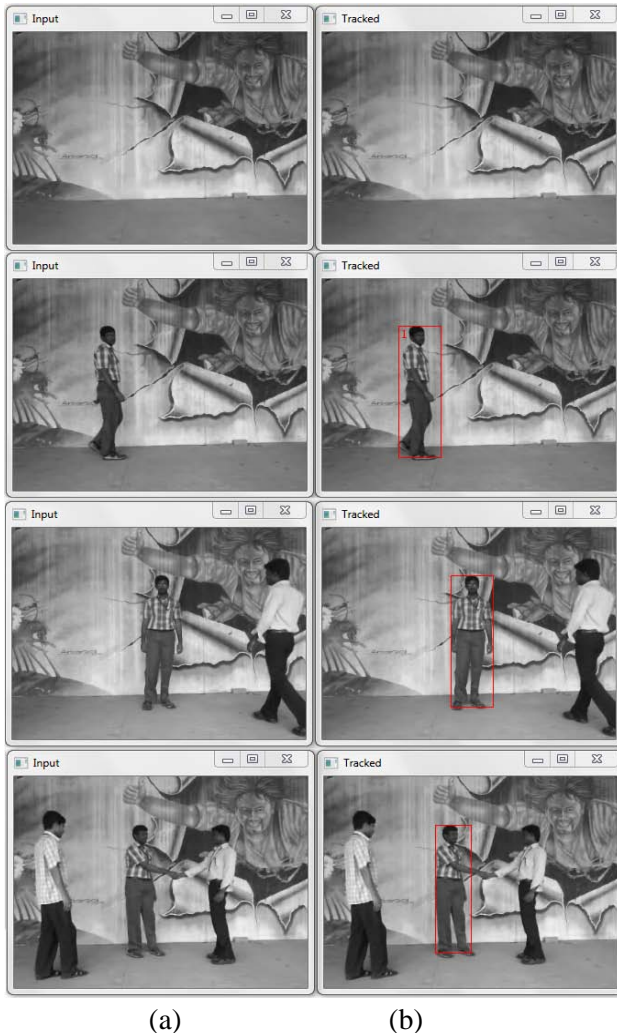


Fig. 9 Single person (1) tracking in multiple person scenario

From Fig. 9 it is observed that the system performs detection and tracking of a single person in a multiple person scenario.

Using the same scenario, the system effectively tracks all the three persons as shown in Fig. 10. Different colour bounding boxes are used to represent the multiple people tracking and the same colour is maintained throughout the entire video sequence as shown in Fig. 10.

The other common problem in person tracking is the erratic movements of the persons in the scenario. As a result, they may often be occluded by obstacles present in the field of vision. Consider a scenario, during the desired person tracking, due to the obstacles like pillar, tree or post etc., the targets are

partially occluded as shown in Fig. 11.

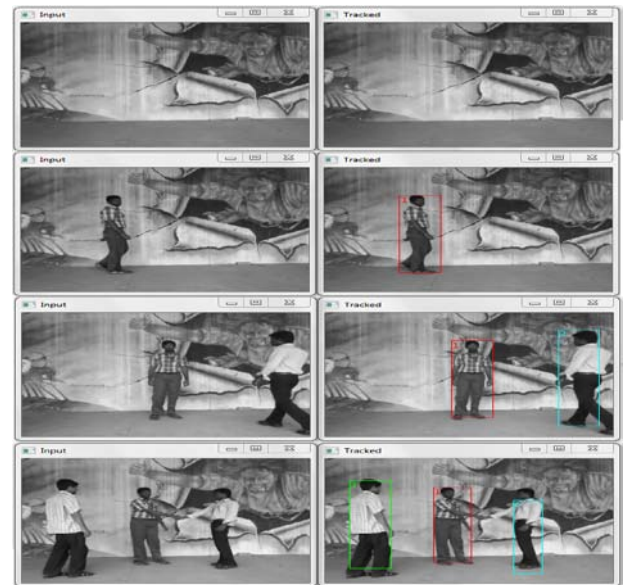


Fig. 10 Multiple person tracking in multiple person scenario

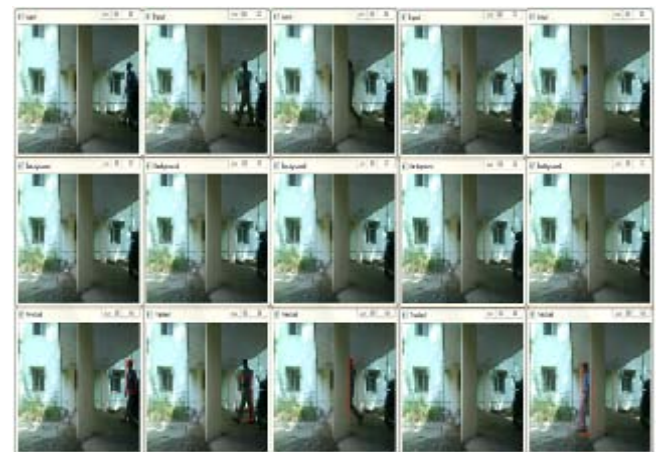


Fig. 11 Tracking of target in presence of a multiple occlusions

The above Fig. 11, the first represent the video frames obtained from camera. The second row gives the estimated background. The third row shows the tracking results obtained from the tracking system. From the results, it is observed that, the proposed system effectively perform tracking, when the target is occluded by the obstacles.

The other problem associated with video tracking is target miss association. Consider a scenario representing occlusion of two persons shown in Fig. 12, where the original and tracked outputs are given as left (a) and right (b) side frames.

In the tracked output frames, the person A is identified by red colour bounding box and the person B is identified by blue colour bounding box. When two similar targets (two persons) are simultaneously crossing a pillar and occluded by the pillar, after they come out, both the persons are exactly tracked and

identified. It is observed that the proposed system accurately track the multiple targets, even they occluded by obstacles.



Fig. 12. Target miss-association solutions

The tracking error of the proposed CGHMM-KF is compared with the existing Mixture of Gaussian – Kalman Filter (MoG-KF) and Frame Difference–Kalman Filter (FD-KF) methods as shown in Fig. 13. For this comparison, the frames 204 (first row) and 403 (second row) are considered.

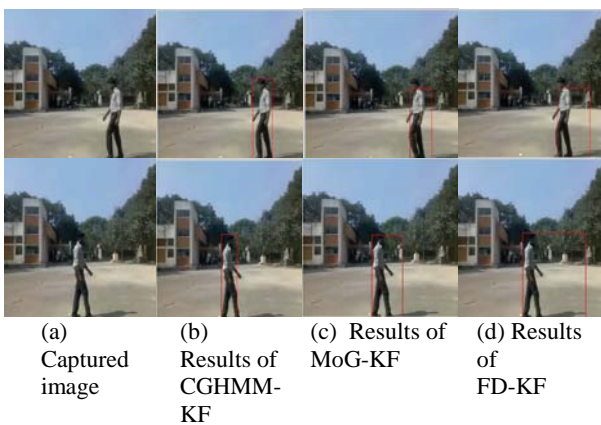


Fig. 13 Comparison of multiple algorithms for target tracking

From the results shown in Fig. 13 it is found that, the width and height of the bounding box is not accurate for MoG-KF and FD-KF because the system considers shadow effects. The proposed CGHMM-

KF exactly detects the blob and bounding box is constructed around the detected person.

For error comparison, the scenarios with increased number of person are considered. The tracking error performance is shown in Fig. 14. For each increased number of persons, 1000 frames are considered.

The percentage of error against number of targets is compared for CGHMM-KF, MoG-KF and FD-KF is shown in Fig.14. The Fig. 14 infers that the MoG-KF and FD-KF produces more percentage of tracking error than the proposed CGHMM-KF.

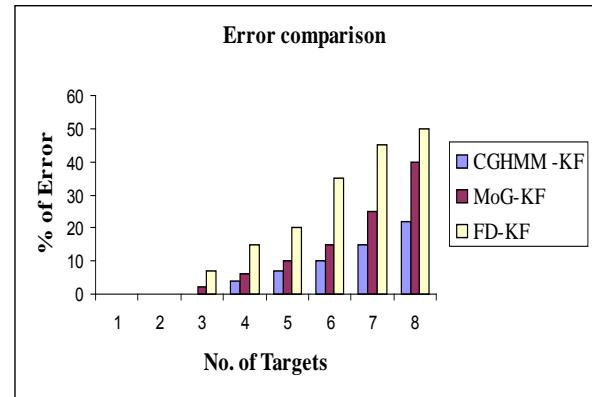


Fig. 14 Tracking error comparisons

5 Conclusion

In this paper a reliable and robust approach for tracking of multiple persons in arbitrary complex environments has been proposed and successfully modelled. The major contribution of this approach is that the proposed system uses the two stochastic modelling techniques (model-based and feature-based) using KF for tracking. The proposed Combined Gaussian Hidden Markov Model and Kalman Filter (CGHMM-KF) effectively overcomes the problems that arise due to pose variation, multiple occlusions and target miss-association. Hence the proposed system is capable of handling complex tracking problem and provides a solution for tracking specific persons in the presence of other multiple moving people/target. The proposed system is also compared with MoG-KF and FD-KF systems and the result shows that the proposed system works better even in the environment with multiple targets. In future, this work can be improved with multiple camera sensors for improved surveillance applications. The problems regarding background variations, camera motions, including panning, tilt, and zooming make the video frame coordinates with respect to the coordinate system can also be considered for further improvement.

References

- [1]. Grigorios Tsagakatakis and Andreas Savakis, "Online Distance Metric Learning for Object Tracking" *IEEE Transactions on Circuits And Systems For Video Technology*, Vol. 21, No. 12, pp. 1810- 1821, 2011.
- [2]. Bhargavi, R., Sri Ganesh, K., RajaSekar, M., Rabinder Singh, P., Vaidehi, V., "An Integrated System of Complex Event Processing and Kalman Filter for Multiple People Tracking in WSN" *IEEE International Conference on Recent Trends in Information Technology*, 2011, pp.890–895.
- [3]. Dalal, N., Triggs, B, "Histograms of oriented gradients for human detection" *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886-893.
- [4]. D. Forsyth and J. Ponce, *Computer Vision - A Modern Approach*, Prentice Hall, 2003.
- [5]. A. Iketani, A. Nagai, Y. Kuno, and Y. Shirai, "Detecting Persons on Changing Background", *Proc.s of ICPR*, Vol.1,1998, pp.74–76.
- [6]. M. Isard and J. MacCormick. Bramble, "A bayesian multiple- blob tracker", *IEEE International Conference on Computer Vision*, volume 2, 2001,pp. 34–41.
- [7]. Anurag Mittal and Larry S. Davis, "M2tracker: A multiview approach to segmenting and tracking people in a cluttered scene using region-based stereo," *Proc. on Computer Vision*, 2002, pp. 18–36.
- [8]. J. Kang, I. Cohen, and G. Medioni, "Soccer player tracking across uncalibrated camera streams", *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, 2003, pp. 1 – 8.
- [9]. John Krumm, Steve Harris, Brian Meyers, Barry Brumitt, Michael Hale, and Steve Shafer, "Multi-camera multi-person tracking for easy living," *IEEE International Workshop on Visual Surveillance*, 2000, pp. 1 – 8.
- [10]. T.Darrell, D. Demirdjian, N.Checka, and P. Felzenszwalb, "Plan-view trajectory estimation with dense stereo background models," *Proc.of the International Conference on Computer Vision*, 2001, pp. 1 – 8.
- [11]. T. Zhao, R. Nevatia, and F. Lv. "Segmentation and tracking of multiple humans in complex situations", *Proc. Computer Vision and Pattern Recognition*, 2001, pp.1208 - 1221.
- [12]. J. H. Piater and J. L. Crowley, "Multi-modal tracking of interacting targets using gaussian approximations", *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, 2001, pp. 1 – 8.
- [13]. L. R. Rabiner and B. H. Huang, "An Introduction to Hidden Markov Models", *IEEE ASSP Magazine*, 1986, pp. 4–16.
- [14]. Tatsuya Osawa, Xiaojun Wu, Kaoru Wakabayashi, and Takayuki Yasuno, "Human tracking by particle filtering using full 3d model of both target and environment," *Proc. Pattern Recognition*, 2006, pp. 25 – 28.
- [15]. Saad M. Khan and Mubarak Shah, "A multiview approach to tracking people in crowded scenes using a planar homography constraint," *Proc. Computer Vision*, Vol. 4, 2006, pp. 133 – 146.
- [16]. A. Lopez, C. Canton-Ferrer, and J.R. Casas, "Multiperson 3d tracking with particle filters on voxels," *Proc. Acoustics, Speech and Signal Processing*, Vol. 1, 2007 pp. 913–916.
- [17]. Tao Zhao, Nevatia, R, "Tracking multiple humans in complex situations", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.26, No.9, 2004, pp.1208 – 1221.
- [18]. L. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition", *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, Vol.77, 1989, pp 257 -286.
- [19]. Gerhard Rigoll, Stefan Eickeler, Stefan M'uller, "Person Tracking in Real-World Scenarios Using Statistical Method", *IEEE Fourth International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 342 - 347.
- [20]. Bouwmans.T, El Baf.F., Vachon,B, "Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey" *Recent Patents on Computer Science*, Vol.1, No.3, 2008, pp. 219-237.
- [21]. Yun-fang Zhu, "Moving Objects Detection and Segmentation Based on Background Subtraction and Image Over-Segmentation" *Journal of Software*, Vol. 6, No. 7, 2011, pp. 1361 – 1367.
- [22]. Mirabi, M. and Javadi, S. "People Tracking in Outdoor Environment Using Kalman Filter", *Proc. on Intelligent Systems Modelling and Simulation*, 2012, pp. 303 – 307.
- [23]. Samal, A. and Iyengar, P. A. "Human Face Detection Using Silhouettes", *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 9, No. 6, 1995, pp. 845-867.
- [24]. Thakoor, N. and Gao, J. "Hidden Markov Model based 2D Shape Classification", *Advanced Concepts for Intelligent Vision Systems, Springer LNCS*, Vol. 3708, 2005, pp. 60 – 67.