Journal of Universal Computer Science, vol. 16, no. 6 (2010), 921-937
submitted: 15/9/09, accepted: 28/2/10, appeared: 28/3/10 © J.UCS

A General Framework for Multi-Human Tracking using Kalman Filter and Fast Mean Shift Algorithms

Ahmed Ali
(Graduate School of Advanced Technology and Science, University of Tokushima
Tokushima, Japan
ali@is.tokushima-u.ac.jp)

Kenji Terada
(Graduate School of Advanced Technology and Science, University of Tokushima
Tokushima, Japan
terada@is.tokushima-u.ac.jp)

Abstract: The task of reliable detection and tracking of multiple objects becomes highly complex for crowded scenarios. In this paper, a robust framework is presented for multi-Human tracking. The key contribution of the work is to use fast calculation for mean shift algorithm to perform tracking for the cases when Kalman filter fails due to measurement error. Local density maxima in the difference image - usually representing moving objects - are outlined by a fast non-parametric mean shift clustering procedure. The proposed approach has the robust ability to track moving objects, both separately and in groups, in consecutive frames under some kinds of difficulties such as rapid appearance changes caused by image noise and occlusion.

Keywords: Kalman Filter, Fast Mean Algorithm, Human Tracking
Categories: I.2.10, I.4

1 Introduction

Multiple object tracking is an extensively investigated subject in the field of visual surveillance [Collins, 00] . Its main complexity stems from the fact that observed data is usually contaminated with noise, missing observations or clutter. Such situations can be encountered for instance in blob-based object detection approaches relying on background subtraction [Collins, 00] & [Fuentes, 01], where under-segmentation, over-segmentation and false detections are frequent problems ultimately leading to failures when tracking is performed.

Background subtraction is one of the most common approaches for detecting foreground objects from video sequences. Recently, some statistical methods are used to extract change regions from the background. The Gaussian mixture model is the most representative background model. As an example, Stauffer and Grimson [Stauffer, 99] presented an adaptive background mixture model for real-time tracking. In their work, they modelled each pixel as a mixture of Gaussians and used an online approximation to update it. The Gaussian distributions of the adaptive mixture models were then evaluated to determine the pixels most likely from a background process, which resulted in a reliable, real-time outdoor tracker which can deal with lighting

changes and clutter. Some work used Gaussian mixture model incorporated with other methods [Lee, 03]& [Porikli, 03] . Others methods for background segmentation are discussed in [Bovik, 00] &[Cheung, 04]

Kalman models and Kalman filters are very important tools and often used for tracking moving objects. Kalman filters are typically used to make predictions for the following frame and to locate the position or to identify related parameters of the moving object. For example, Broida and Chellappa [Broida, 86] used the Kalman filter to track points in noisy images. In stereo camera-based object tracking, Beymer and Konolige [Beymer, 99] use the Kalman filter for predicting the object's position and speed in x - z dimensions. Rosales and Sclaroff [Rosales, 99] use the extended Kalman filter to estimate 3D trajectory of an object from 2D motion. A common approach to handle complete occlusion during tracking is to model the object motion by linear dynamic models or by nonlinear dynamics and, in the case of occlusion, to keep on predicting the object location until the object reappears. For example, a linear velocity model is used in Beymer and Konolige [Beymer, 99] and a Kalman filter is used for estimating the location and motion of objects. For the image segmentation problem, Mean-Shift Clustering is commonly used. Comaniciu and Meer [Comaniciu, 02] propose the mean-shift approach to find clusters in the joint spatial and color space. A recent article by Pulford [Pulford, 05] summarizes the techniques in widespread use and classifies them into 35 different algorithmic types as well as providing a comprehensive literature survey of this area.

A previous work for combing Kalman filter and Mean shift Algorithm was done by Comaniciu and Ramesh [Comaniciu, 00] employs the mean shift analysis to derive the target candidate that is the most similar to a given, target model, while the prediction of the next target location is computed with a Kalman filter. A recent work also combing Kalman filter and Mean shift Algorithm appeared in [Wang, 99] , as using Kalman filter to predict the possible position on the object in the next frame of the video image, and using the mean shift algorithm to search in this neighboring range.

The exhaustive search in the neighborhood of the predicted target location for the best target candidate is, however, a computationally intensive process. As a solution to this problem, this work proposes a robust multi-Human tracking framework showing a fast and good tracking performance in consecutive frames under some kind of difficulties such as rapid appearance changes caused by image noise and occlusion. The key contribution of this work is to use fast computation of mean shift algorithm for the cases when Kalman filter fails due to measurement errors caused by these above mentioned difficulties.

The paper is organized as follows: section 2 gives a detailed description for the tracking algorithm subsequences and also gives details on the proposed tracking algorithm scheme; The experimental results, the discussion of these results, and the performance evaluation of the proposed system are demonstrated in Section 3. Finally, the paper is concluded in section 4.

2 The Tracking Algorithm

Object tracking is the problem of estimating the positions and other relevant information of moving objects in image sequences. Two-frame tracking can be accomplished using correlation-based matching methods, optical flow techniques, or change-based moving object detection methods. The tracking algorithm can be briefly described in the next subsections.

2.1 Object Extraction from background

Evidently, before we start with tracking of moving objects, we need to extract moving objects from the background. We use background subtraction to segment the moving objects. Each background pixel is modelled using a mixture of Gaussian distributions. The Gaussians are evaluated using a simple heuristic to hypothesize which are most likely to be part of the "background process". Each pixel is modeled by a mixture of K Gaussians as stated in formula (1):

$$P(X_t)\sum_{i=1}^{K}\omega_{i,t}\eta(X_t;\mu_{i,t},\sum_{i,t}) \tag{1}$$

where X_t is the variable, which represents the pixel, and t represents time. Here K is the number of distributions: normally we choose K between 3 to 5. $\omega_{i,t}$ is an estimate of the weight of the i th Gaussian in the mixture at time t, $\mu_{i,t}$ is the mean value of the i th Gaussian in the mixture at time t. $\sum_{i,t}$ is the covariance matrix of the i th Gaussian in the mixture at time t. Every new pixel value X_t is checked against the existing K Gaussian distributions until a match is found. Based on the matching results, the background is updated as follows: X_t matches component i, that is X_t decreases by 2.5 standard deviations of the distribution, then the parameters of the i th component are updated as follows:

$$\omega_{i,t} = (1-\alpha)\omega_{i,t-1} + \alpha \tag{2}$$

$$\mu_{i,t} = (1-\rho)\mu_{i,t-1} + \rho I_t \tag{3}$$

$$\sigma_{i,t}^2 = (1-\rho)\sigma_{i,t-1}^2 + \rho(I_t-\mu_{i,t})^T(I_t-\mu_{i,t}) \tag{4}$$

Where $\rho = \alpha \Pr(I_t \mid \mu_{i,t-1},\sum_{i,t-1})$, α is the predefined learning parameter, $\mu_{i,t}$ the mean value of the pixel at time t, and I_t is the recent pixel at time t. The parameters for unmatched distribution remain unchanged. But $\omega_{i,t}$ will be adjusted using formula: $\omega_{i,t} = (1-\alpha)\omega_{i,t-1}$.

If X_t matches none of the K distributions, then the least probable distribution is replaced by a distribution where the current value acts as its mean value. The variance is chosen to be high and the a-priori weight is low [Stauffer, 99].

The background estimation problem is solved by specifying the Gaussian distributions, which have the most supporting evidence and the least variance. Because the moving object has larger variance than a background pixel, so in order to represents background processes, first the Gaussians are ordered by the value of $\omega_{i,t} / \left\| \sum_{i,t} \right\|$ in decreasing order.

The background distribution stays on top with the lowest variance by applying a threshold T, where

$$B = \arg\min_b \left(\frac{\sum_{i=1}^{b} \omega_{i,t}}{\sum_{i=1}^{K} \omega_{i,t}} \succ T \right) \tag{5}$$

All pixels X_t which do not match any of these components will be marked as foreground.

2.2 Connected component analysis

Once initialized, moving objects in the scene are segmented using a log-likelihood measure between the incoming frames and the current background model. The input video stream is low-pass filtered to reduce noise effects. A connected components analysis is then applied to the resulting image. Initial segmentation is usually noisy, so morphological operations and size filters are applied. The output of this segmentation process is shown in Figure 1.

At this stage we calculate the following features for each moving object region: bounding rectangle: the smallest rectangle that contains the object region. We keep record of the coordinate of the upper left position and the lower right position, what also provides size information (width and height of each rectangle). color: the mean R G B values of the moving object. center: this work use the center of the bounding box as a simple approximation of the centroid of a moving object region. velocity: defined as movement of number of pixels/second in both horizontal and vertical direction.

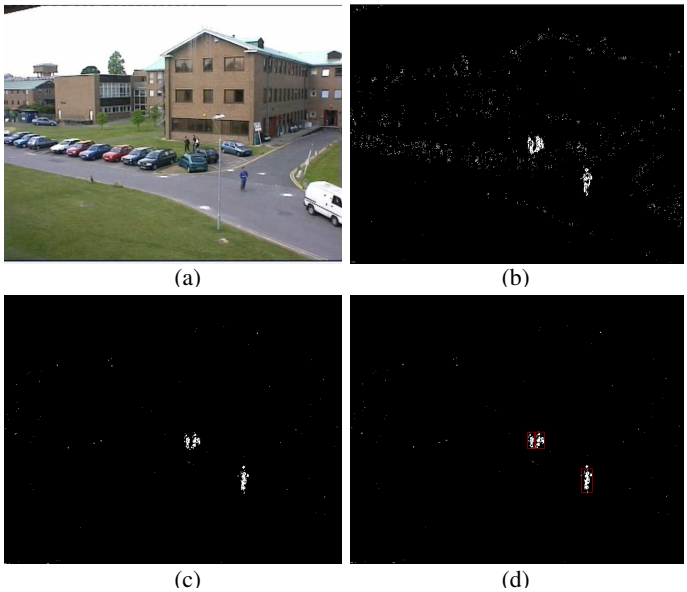

Figure 1: Change image detection; (a) the original image, (b) the difference image, (c) the image after applying morphological filters, and (d) the blob detection.

2.3 Tracking using Kalman Filter

A Kalman filter is used to estimate the state of a linear system where the state is assumed to be distributed by a Gaussian. Kalman filtering is composed of two steps, prediction and correction [Mohendar, 01] & [Bar-Shalom, 88]. The Kalman filter is a recursive estimator. This means that only the estimated state form the previous time step and the current measurement are needed to compute the estimate for the current state. In contrast to batch estimation techniques [Świątek, 07], no history of observations and/or estimates is required .

Kalman filter consists of five equations and it can be divide them into two groups: the update equations and the correct equations. The update equations are responsible for projecting forward the current state and error covariance estimates to obtain the priori estimates for the next time step. The correct equations are responsible for the feedback, in other words, for incorporating a new measurement into the priori estimate to obtain an improved posteriori estimate.

The state variable and observation of Kalman filter in this paper are object locations. Besides, we represented the rectangular window using center coordinate. Because the moving state changed little in the neighboring consecutive frames, we thought of the system as linear Gaussian one and the state parameters of Kalman filter were object location, its velocity, and the width of the rectangle which represent the width of a Human.

The state-space representation of the tracker used in the Kalman filter is given in Eq. (6).

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \\ x'_{t+1} \\ y'_{t+1} \\ W_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \\ x'_t \\ y'_t \\ W_t \end{bmatrix} + w_t \tag{6}$$

where, x_{t+1} and y_{t+1} are the predicted coordinates of the object and x'_t and y'_t are the velocities in the respective direction, W_t represents the width of the Human rectangle, Δt represents the time interval of state correction and w_t is the white Gaussian noise with diagonal variance Q.

The predicted coordinates and dimensions of the rectangle are used to locate the Human in the present frame. When the Humans are distinguished, the Kalman vector is updated using the measurement equation as shown in Eq. (7).

$$\begin{bmatrix} mx_{t+1} \\ my_{t+1} \\ mW_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-\Delta t} \\ y_{t-\Delta t} \\ x'_{t-\Delta t} \\ y'_{t-\Delta t} \\ W_{t-\Delta t} \end{bmatrix} + v_t \tag{7}$$

Where mx_{t+1} and my_{t+1} are the measured coordinates, the value mW_{t+1} is the measure width of Human at time t+1 and v_t is white Gaussian noise with diagonal variance R. The position, velocity and acceleration are updated based on the values obtained in the present frame and the data from the previous frame.

2.4 Difficulties using Kalman Filter for Tracking

The major problem with the tracking was that of the target detection. change-based detection technique relies on connected component analysis, which leads to poor object segmentation quality in the case of overlapping humans and/or shadows. So Measurements given by the detector are sometimes very noisy and Kalman prediction follows measurement and that may misled Kalman to predict the true position of the object. In order to handle this problem, fast mean shift algorithm is used. It is used to locate densities extrema, which gives clue that whether Kalman prediction is right or it is misled by wrong measurement. In case of wrong prediction, it is corrected with the help of densities extrema in the scene. Local density maxima in the difference image - usually representing moving objects - are outlined by a fast non-parametric mean shift clustering procedure as discussed in the next section.

2.5 Fast Mean Shift Computation

The mean shift algorithm is a nonparametric technique to locate density extrema or modes of a given distribution by an iterative procedure [Comaniciu, 99]& [Beleznai, 06]. Starting from a location x the local mean shift vector represents an offset to x′, which is a translation towards the nearest mode along the direction of maximum increase in the underlying density function. The local density is estimated within the local neighborhood of a kernel by kernel density estimation where at a data point a kernel weights $K(a)$ are combined with weights $I(a)$ associated with the data. For digital images sample weights are defined by the pixel in intensities at pixel location a.

Fast computation of the new location vector x' as in Eq. 8 can be performed as [Patrice, 99] :

$$x'' = \frac{\sum_a K''(a-x) ii_x(a)}{\sum_a K''(a-x) ii(a)} \tag{8}$$

where K'' represents the second derivative of the kernel K, differentiated with respect to each dimension of the image space, i.e. the x- and y-coordinates.

The functions ii_x and ii are the double integrals, i.e. two-dimensional integral images in the form of Eq. 9 and Eq. 10 respectively,

$$ii_x(x) = \sum_{x_i \prec x} I(x_i) x_i \tag{9}$$

and

$$ii(x) = \sum_{x_i \prec x} I(x_i). \tag{10}$$

Equation (9) and (10) represent the so-cold integral images [Viola, 01], primarily used for fast area sum computation in various application. If the kernel K is uniform with bounded support, its second derivative becomes sparse containing only four impulse functions at its corners. Thus, evaluating a convolution takes only the summation of four corner values in the given integral image.

The integral image - also called summed area table [Crow, 84] - provides an efficient means to compute area integrals for a given image. The integral image at a location (x, y) contains the cumulative sum of pixels located to the left and above (x, y) including the pixel (x, y):

$$I_{\text{int}}(x, y) = \sum_{x' \le x, y' \le y} I(x', y') \tag{11}$$

where $I(x, y)$ is an arbitrary image and $I_{\text{int}}(x, y)$ is the integral image, i.e., the two-dimensional cumulative distribution function of $I(x, y)$. Fast computation of the integral image can be performed in a single pass [20].

Using the integral image, the area sum of a rectangular region within the original image can he efficiently computed by the following step:

$$S_{Area} = I_{\text{int}}(x-1, y-1) + I_{\text{int}}(x+W-1, y+H-1) \\ -I_{\text{int}}(x-1, y+H-1) + I_{\text{int}}(x+W-1, y-1) \quad (12)$$

where S_{Area} is the area sum within the rectangle with upper left corner (x, y) with (w, h) parameters for width and height.

Efficient computation of the mean shift vector can be achieved by computing integral images of the first moments:

$$I^{x}_{\text{int}}(x, y) = \sum_{x' \le x,\, y' \le y} x' I(x', y'), \quad (13)$$

$$I^{y}_{\text{int}}(x, y) = \sum_{x' \le x,\, y' \le y} y' I(x', y'). \quad (14)$$

The above quantities can be computed in the same pass as the zeroth moment (Eq. 11). Using the integral images of the zeroth and the first moments the mean shift vector components applying a uniform kernel can he expressed as:

$$m_x(x, y) = \frac{S_{Area}(I^{x}_{\text{int}})}{S_{Area}(I_{\text{int}})} - x, \quad (15)$$

$$m_y(x, y) = \frac{S_{Area}(I^{y}_{\text{int}})}{S_{Area}(I_{\text{int}})} - y. \quad (16)$$

The numerator and denominator of the above expressions denote the area sums computed over the window rectangle $(W(y), H(y))$ using the pre-computed integral images for zeroth and the first $(I^{x}_{\text{int}}, I^{y}_{\text{int}})$ moments.

2.6 Mean Shift Clustering

The difference image is the outcome of a change detection process by forming the difference between the current image of an image sequence and a reference image representing the background. The difference image generated for a crowded scene usually contains large number of high-intensity peaks or modes. Our principal objective is to find individual modes representing possible human candidates.

The clustering step is facilitated by the use of a human size model $\{H(x), W(x)\}$, where H and W denote human height and width, respectively. This information is obtained by a simple calibration step.

The principal steps of mean shift clustering are performed analogously to the steps described in [Comaniciu, 02]:

1. The difference image intensity maximum is mapped to unit intensity and its entire range is scaled proportionally.
2. A sample set of n points $X_1 \ldots X_n$ is defined by locating local maxima. Local maxima are found by:
   a) Locating the global intensity maximum and adding it to the list of sample set;
   b) Resetting the difference image intensity around the found maximum within

a window of size $(0.5W(y),0.5H(y))$;

c) Repeating the maximum search of step (a) until the found maximum drops below a threshold T_1.

The points of the sample set are subsequently used in a mean shift procedure. The final result does not depend critically on T1. A very low value just increases the run time and generates more outliers which have to be eliminated by the model-based mode validation step.

3. The fast mean shift procedure is applied to the points of the sample set with a window size of $(H(X_i),W(X_i))$ according to the local size model. The mean shift procedure converges to the nearest mode typically within 3-4 iterations.

The mode seeking process delineates a path between the initial point of the sample set and the detected local mode candidate. Each mean shift iteration defines a point on the path, what we denote as a path-point PX. Thus, each detected mode candidate location has an associated set of path-points $\{PX_1,...,PX_n\}$, not including the mode itself.

When the mean shift offset vector is computed according to (8), the area sum (i.e. sum of pixel intensities) within the kernel (denominator of the expression in (8)) is also obtained.

The set of area sum magnitudes $\{S_1,...,S_n\}$ is useful to have since it provides information on the magnitude of the local density and it can be used in the occlusion handling step evaluating a given spatial configuration of kernels.

4. Given the finite size of the mean shift convergence criterion, detected mode candidate locations – obtained for the same peak of underlying density - might slightly deviate. Detected mode candidates are linked based on spatial proximity: all detected modes within a window of the size (W, H) are grouped together and a cluster center Y is obtained by taking the mean of linked candidate coordinates. Path-point sets belonging to grouped mode candidates are also merged, such as the sets of area sum magnitudes.

The merged set of path points is used to delineate the cluster: a bounding box representation of the basin of attraction is obtained by determining the spatial extrema of path points in x- and y-directions. At this stage the detected clusters can be considered as object candidates and probable outliers can be eliminated by imposing size constraints on the size of the attraction basins.

2.7 Schematic Description of the Tracking Algorithm

After extract moving objects from the background and applying connected component analysis to extract some features for the moving object such as Bounding box, color, center and velocity, then this work uses Kalman predictor to predict the next position of Human on the next frame. After this step we check whether Kalman predictor is misled or not. On this work we decided that if pixel difference between detected measurement obtained “object centroid” by connected component analysis method and Kalman filter predicted position is greater than a certain threshold, then Kalman predicted position is discarded and the motion region is detected by triggering fast mean shift algorithm and it guides Kalman filter, by finding actual motion region in

the scene, for more accurate prediction of next state of the object. Therefore with the help of this modification, the proposed tracking algorithm scheme can be finally described by the following diagram as seen on Figure 2.

2.8 Occlusion and Recovery

Occlusion is a common problem encountered in tracking multiple people. This might be due to complete or partial overlap of a person by another person or due to a rigid object occluding the person completely. To overcome the problem of partial occlusion, the proposed algorithm makes use of predicted values from the Kalman prediction algorithm to predict the next state "position " of each person as well as examine the predicted parameter , mW_{t+1} , the measure width of the person at time t+1 as if it is greater than a certain threshold "previously determined using the expected Human width" then the Kalman filter repredict the next position of the person based on the updated parameter mW_{t+1} .

3 Experiments and Results

This section show the results obtained by the proposed algorithm. The system was tested using Intel(R) Core Duo CPU device with 2.10 GHZ and 3GB of RAM and Window XP operation system, and Matlab R2007b are used. In addition, the image sequences consists of JPG images with 320x240 resolutions per frame. Some kind of experimental videos are used to evaluate the robust ability of the proposed method. Since the surveillance problem not affected by the silently low resolutions frames, this work try to make the system more efficient by using low resolution image sequences that consequently decrease the tracking time.

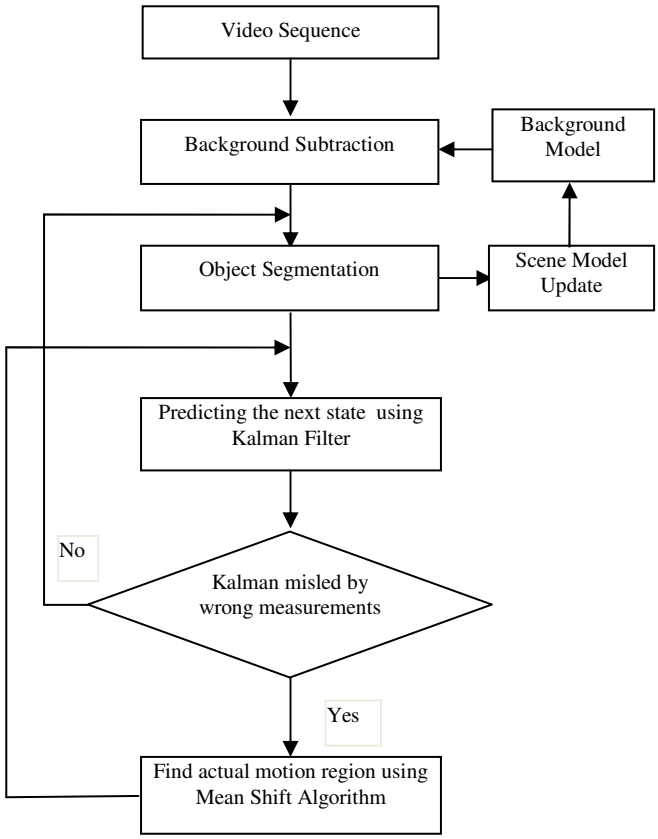

Figure 2: The proposed Tracking Algorithm scheme.

The system tested by two examples of outdoor scenes using PETS2000 and PETS2001 Image Sequences. The proposed system also tested using PETS2002 Image Sequences as an example of indoor scenes. The system efficiency affected by increasing the number of Humans to track per frame as assigning a Kalman filter to every one of multiple Humans. This means the more Kalman filters assign for Humans, the much time it take to track, the less efficiency the system will be.

In experiment one, the system was tested using outdoor image sequence, PETS2000 Image Sequences with 2687 total number of frames. The system detected 5453 moving Humans appeared partially or completely over the total video frames. The results show that system successfully tracked the objects by the help of fast mean shift tracker that precisely detected the true position of the Human and also show the ability to track Human in some complex situations as seen in some frames in Figure 3. In this figure, Humans appear to partially occluded with each other and The system precisely tracked them.

Moreover the system tested using another outdoor image sequence, experiment two, PETS2001 Image Sequences with 289 frames and the system detected 389 moving Humans appeared partially or totally over the frames. The result show good tracking as seen in Figure 4 and also show the ability to detect false alarms "a bird flying" that appeared in some frames and detected them as false alarms.

Figure 3: Image sequence one, the proposed system show the ability to track people in complex situation "The red rectangle show the location where the object detected and the green one show the predicted position by the system."

As an example of indoor image sequence, experiment three, the system was challenged using PETS2002 Image Sequences scenes with 1302 frames to test tracking Humans indoors. The system detected 3689 moving Human appear partially or completely on the video frames. The false alarm detected by the proposed system due to the shadow of Human due to mirrors as well as the luminance changed.

The proposed system track Humans both separately and in groups walking and interacting with each others on the hall of a shopping center. The most challenger of this experiment is the noises on the image sequence due to shadow of Humans due to mirrors as well as the occlusion of Humans with each others. The results seen on Figure 5 show the ability of the proposed system to track Humans indoors under these difficulties.

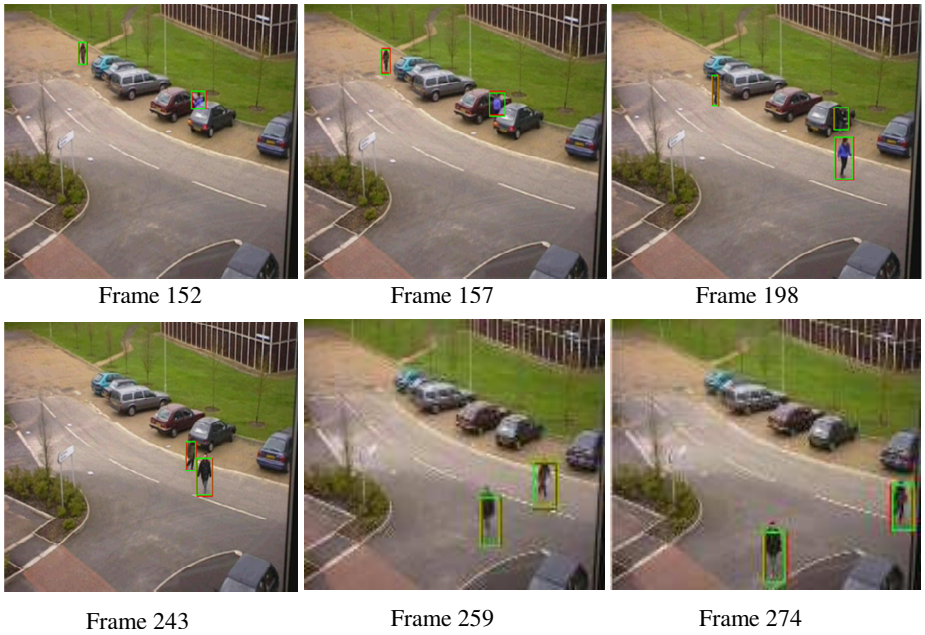

Figure 4: Image sequence two, Humans outdoors are successfully tracked using the proposed system.

In addition, to show the ability to deals with the problem of partial occlusion, the system successfully tracked a person partially occluded with a car as shown in Figure 6. As shown on the figure, the person tracked with the help of mean shift algorithm that precisely detected the correct motion region and Kalman filter that successfully predict the next position of the person as well as the correct prediction of mW_{t+1} , the measure width of the person at time t+1, that help finding the person who occluded with another object "a Car."

With all of these experiments, the overall results show that the proposed algorithm has its robust ability to track moving Humans in consecutive frames under some kinds of difficulties such as rapid appearance changes caused by image noise, and occlusion when compared to other robust methods such as mean shift algorithm.

Finally, the system evaluation is measured by calculating the detection and false alarm rate using Mean shift tracker based on [Comaniciu, 02] and the proposed system. The detection rate is calculated by dividing the true Humans detected by the proposed system by the valid moving Humans appears partially or completely over the video frames. The experiments show that results obtained by the proposed system are better than that obtained using Mean shift tracker as described in Table I. As described on the table the detection rate obtained using the proposed system is higher compared with that obtained using Mean shift algorithm.

Figure 5: An example for tracking Humans indoors, image sequence three. The proposed system tracks a group of people walking on the hall of a shopping center.

The missed detection by the proposed system referred some complex situations due to full occluded objects with each other since the proposed system deals only on the partial occlusion. Another reason for the missed detection by the proposed system is the image resolution, since the system deals with a low resolution video scenes to increase the accuracy of the system by decreasing the tracking time and that is useful for all surveillance problems.

Figure 6: An example show the accuracy of the system to track the object under occlusion.

	Mean Shift			**The Proposed system**		
	Seq. 1	**Seq. 2**	**Seq.3**	**Seq.1**	**Seq.2**	**Seq.3**
Valid moving Humans	5754	402	3922	5754	402	3922
Correct detection	5263	378	3601	5453	389	3689
Missed detection	491	24	321	301	13	233
False alarm	83	8	74	79	4	53
Detection rate	91.46%	94.02%	91.81%	94.76%	96.76%	94.05%
False alarm rate	1.44%	1.99%	1.89%	1.37%	0.99%	1.35%

Table 1: The performance evaluation for the proposed system compared with Mean Shift Tracker.

4 Conclusions

This paper presents a robust framework for multi- Human tracking. It includes a combination of Kalman filter and fast mean shift algorithm. Moving objects in the scene are segmented between the incoming frames and the current background model. The input video stream is low-pass filtered to reduce noise effects. A connected components analysis is then applied to the resulting image and a bounding box containing each object is drawn and some features are recorded. Kalman filter continuously predicts the next state of the object by predicting its centroid and velocity. For some kinds of difficulties such as rapid appearance changes caused by image noise and occlusion, Kalman prediction may be misled by wrong measurement. In that case, motion region is detected by triggering fast mean shift algorithm to find the actual motion region and to guide Kalman filter. The proposed method is an efficient multi-Human tracking algorithm by comparing with mean shift algorithm. The system implemented by the proposed method can track moving objects in real time as well as the system need to be more efficient to solve the total occlusion problem.

References

[Bar-Shalom, 88] Bar-Shalom, Y. and Foreman, T.: Tracking and Data Association. Academic Press Inc. 1988.

[Beleznai, 06] Beleznai C., Frühstück B., and Horst D.: Human tracking by fast mean shift mode seeking, Journal of Multimedia, vol. 1, no. 1, pp. 1-8, 2006.

[Beymer, 99] Beymer, D., and Konolige,K. :Real-time tracking of multiple people using continuous detection, In IEEE International Conference on Computer Vision (ICCV) Frame-Rate Workshop. 1999.

[Bovik, 00] Al Bovik: Handbook of Image and Video Processing, Academic Press, 2000.

[Broida, 86] Broida, T., and Chellappa, R. : Estimation of object motion parameters from noisy images, IEEE Trans. Patt. Analy. Mach. Intell. vol. 8, no. 1, pp. 90–99. 1986.

[Cheung, 04] Cheung, S. C. S., and Kamath, C.: Robust techniques for background subtraction in urban traffic video," SPIE Electronic Imaging: Video Communications and Image Processing, San Jose, CA, pp. 881-892, 2004.

[Collins, 00] Collins R., Lipton A., Kanade T., Fujiyoshi H., Duggins D., Tsin Y., Tolliver D., Enomoto N., and Hasegawa O.: A system for video surveillance and monitoring: VSAM final report, Carnegie Mellon University, Robotics Institute, In Tech. Rep. CMU-RI-TR-00-122000.

[Comaniciu, 99] Comaniciu, D., and Meer, P.: Mean Shift Anallysis and Applications, In IEEE Int. Conf. Computer Vision, Kerkyra, Greece, pp. 1197-1203, 1999.

[Comaniciu, 00] Comaniciu D., and Ramesh V.: Mean shift and optimal prediction for efficient object tracking, Proceedings of International Conference on Image Processing, vol. 3, pp.70-73, 2000.

[Comaniciu, 02] Comaniciu, D., and Meer, P.: Mean shift: A robust approach toward feature space analysis, IEEE Trans. Patt. Analy. Mach. Intell. vol. 24, no. 5, pp. 603–619, 2002.

[Crow, 84] Crow E: Summed-area tables for texture mapping, in proceedings of SIGGRAPH, vol. 18, pp. 207-212, 1984.

[Fuentes, 01] Fuentes L. M., and Velastin S. A.: People tracking in surveillance applications, In Proceedings of 2nd IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Kauai, Hawaii, 2001.

[Lee, 03] Lee, D.S., Hull, J.J., and Erol, B.: A Bayesian framework for Gaussian mixture background modeling. Proc. ICIP, Barcelona, Spain, Vol. 3, pp. 973–976, 2003.

[Mohendar, 01] Mohendar S. Grewal, Angus G. Andrews: Kalman Filtering Theory and Practice using MATLAB, second edition, John Wiley & Sons, Inc., 2001.

[Patrice, 99] Patrice Y. Simard, Leon Bottou, Patrick Haffner and Yann Le Cun, Boxlets: a Fast Convolution Algorithm for Signal Processing and Neural Networks, Advances in Neural Information Processing Systems, vol. 11, pp. 571-577, 1999.

[Porikli, 03] Porikli, F., and Tuzel, O.: Human body tracking by adaptive background models and mean-shift analysis. Proc. PETS-ICVS, Austraia, 2003.

[Pulford, 05] Pulford G. W.:Taxonomy of multiple target tracking methods, IEE Radar, Sonar and Navigation, vol. 152, no. 5, pp.291- 304, 2005.

[Rosales, 99] Rosales, R., and Sclaroff, S.,: 3d trajectory recovery for tracking multiple objects and trajectory guided recognition of actions, In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 117–123, 1999.

[Stauffer, 99] Stauffer, C., and Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. Proc. CVPR, Fort Collins, Colorado, USA, pp. 246–252, 1999.

[Świątek, 07] J. Świątek: Parameter Estimation of Systems Described by the Relation with Noisy Observations, Journal of Universal Computer Science, vol. 13, no. 2, 199-208, 2007.

[Viola, 01] Viola, P., and Jones, M.: Rapid object detection using a boosted cascade of simple features, In Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, Hawai, vol. 1, pp. 511-518, 2001.

[Wang, 08] Wang L., Hu S., and Zhang X.: Detecting and tracking of small moving target under the background of sea level, 9th International Conference on Signal Processing , pp. 989-992, 2008.