2010 Digital Image Computing: Techniques and Applications

Multiple Pedestrian Tracking using Colour and Motion Models

Zhengqiang Jiang*, Du Q. Huynh*, William Moran†, Subhash Challa† and Nick Spadaccini*
*School of Computer Science and Software Engineering
The University of Western Australia, Perth, Australia
Email: {jiang, du, nick}@csse.uwa.edu.au
†Department of Electrical and Electronic Engineering
The University of Melbourne, Melbourne, Australia,
Email: {b.moran, challas}@ee.unimelb.edu.au

***Abstract*—This paper presents a method that combines colour and motion information to track pedestrians in video sequences captured by a fixed camera. Pedestrians are firstly detected using the human detector proposed by Dalal and Triggs which involves computing the histogram of oriented gradients descriptors and classification using a linear support vector machine. For the colour-based model, we extract a 4-dimensional colour histogram for each detected pedestrian window and compare these colour histograms between consecutive video frames using the Bhattacharyya coefficient. For the motion model, we use a Kalman filter which recursively predicts and updates the estimates of the positions of pedestrians in the video frames. We evaluate our tracking method using videos from two pedestrian video datasets from the web. Our experimental results show that our tracking method outperforms one that uses only colour information and can handle partial occlusion.**

***Keywords*-Multiple pedestrian tracking; histogram of oriented gradients descriptor; colour histogram; Kalman filter**

I. Introduction

Tracking pedestrians in videos has many important computer vision applications including visual surveillance, people recognition, smart environments and human-machine interaction. Visual tracking of pedestrians allows the 2D trajectory of each pedestrian in the video images to be determined. The statistical data of such trajectories are useful for the planning of open space, localizations of entrance and exit doors, etc in buildings and offices. It also facilitates people counting and assessment of the duration of each person spent in the observed area.

Visual tracking of pedestrians consists of 2 main stages: pedestrian detection and association of the detected pedestrian windows [18]. Detecting pedestrians in images is a challenging process due to the variations in human appearance, body poses and viewing angles [10]. Many visual human detection methods have been proposed in the literature to tackle these issues. Wu *et al.* [17] use human body part detectors to detect multiple people in the presence of partial occlusion in cluttered scenes. Their body part and full body detectors are trained via a nested cascade method using edgelet features that embody both edge intensities and orientations. Their method can overcome partial occlusion, but it requires the images to be in high resolution. Dalal and Triggs [7] present an approach for finding pedestrians using the Histogram of Oriented Gradient (HOG) descriptor and a linear Support Vector Machine (SVM) classifier. Compared with the part based detector, their method can handle lower resolution images. In order to handle partial or even full occlusion, Senior *et al.* [14] build an appearance-based model with observation probabilities for all pixels in the foreground regions to detect pedestrians who are in complex interactions with each other in the image. They use an observation probability value to signify how likely each pixel is inside a foreground region. Although their appearance models for the pedestrians are updated in every frame, their method is not effective for detecting humans in long video sequences under illumination changes since their background model is not adaptive. Recently, Alahi *et al.* [2] use Stauffer and Grimson's [15] adaptive Gaussian mixture model for modelling the background variation in order to locate pedestrians as foregrounds pixel blobs. Their approach is therefore more robust against lighting changes in the scenes. Alahi *et al.*'s system requires the camera to be calibrated so as to map the detected pedestrians onto the ground plane.

Pedestrian tracking is a much more difficult problem as each pedestrian must be firstly detected from the video frames and then associated between video frames. In pedestrian detection, appearance models are not applicable as people wearing different colour clothes should have the same chance of being detected. However, in pedestrian tracking, the different colour clothing can be used to distinguish one person from another between video frames. Appearance models are thus important in pedestrian tracking. Among all the appearance models, colour supplies the most useful information. For instance, Wren *et al.* [16] integrate colour and 2D shape in their tracking methods. A Maximum Posterior Probability is used to help determine whether an image pixel belongs to a foreground region or not. However, their method cannot be extended to scenes that contain more than one person. Mason *et al.* [12], on the other hand, use colour histograms to track pedestrians. In their method, each video frame and the background reference image are first

978-0-7695-4271-3/10 $26.00 © 2010 IEEE
DOI 10.1109/DICTA.2010.63

divided into disjoint 40×40 blocks; colour histograms of the pixel blocks are then computed and compared using two different formulae, one of which being the Chi-squared measure of similarity. Corresponding pixel blocks between the video frame and the reference image that differ above a pre-defined threshold are classified as foreground/pedestrian blocks. Pérez *et al.* [13] combine colour histograms defined in the HSV colour space and a particle filter to track moving objects in video sequences. A reference colour histogram of the region of interest is constructed prior to the tracking process. The region whose colour distribution in each subsequent frame is then located via a particle filter. A merit of their method is that it can overcome partial and even complete object occlusion over several frames of the video sequence. A downside of it is the computational cost of the particle filter. More recently, Kang *et al.* [11] propose to use a shape model that is characterized by colour and edge histogram features represented in the polar coordinate system to track moving objects. In order to obtain the translation and rotation invariance, 8 control points are sampled around the reference circle containing the detected object/subject. The authors suggest to use measurement functions such as cross-correlation, Bhattacharyya distance, and Kullback-Leibler distance to determine whether the colour histogram models match between consecutive frames. A considerable disadvantage of their method is that it is incapable of dealing with occlusion.

Our research work reported in this paper shares some similarities with those reviewed above but also differs in some respects. Our approach employs colour histograms as the appearance model and combines with a state space tracking technique. To overcome the lack of spatial information of colour histograms and to minimize the computation cost, we construct two colour histograms, one for the lower and one for the upper parts of each human body and combine them into a single 4-dimensional colour histogram for pedestrian tracking. In term of computation cost, this is much more economical than Kang *et al.*'s approach [11] which uses many histograms at 8 different sampled points around each detected person. For people detection, we employ Dalal and Triggs's approach [6, 7] which involves using the HOG descriptors and an SVM linear classifier. Their people detection method, which we refer to as the *HOG human detector* throughout the paper, has been reported to be very robust for many different standing/walking poses of people in complex scenes. The state space tracking technique that we use is the Kalman filter. Although the Kalman filter has the limitation of being linear, for short video sequences, it is very robust and effective. It is also much cheaper computationally than the particle filter used by Pérez [13].

The paper is organized as follows. In Section 2, we describe our proposed visual human tracking method in detail, including the HOG human detector, 4D colour histograms and the Kalman filter. Experimental results will be shown in Section 3. Section 4 discusses a few issues in the tracking process. Finally, Section 5 gives conclusion and outlines future research direction.

II. Combining colour and motion in pedestrian tracking

A. Human detection

Since our pedestrian tracking algorithm uses Dalal and Triggs's [6, 7] method to detect humans in each video frame, for completeness of the paper, a detailed review of their method is given in this subsection.

The HOG human detector involves constructing a dense grid of oriented gradient histograms, which together form a so-called *HOG descriptor*. The construction of the HOG descriptors consists of several stages. Firstly, a colour normalization step is carried out to reduce the influence of shadows and illumination. The intensity gradient at a pixel is then computed for each colour channel. The next stage is to acquire a histogram of the gradient orientations at each spatial region or (cell). A range of adjacent cells are grouped into larger spatial blocks so as to normalize edge contrast and illumination. The last stage is to concatenate all of the normalized block vectors over each detection window to form a final window descriptor. Because the HOG descriptors are computed over a number of scales on overlapping windows in the image, they are vectors of 3780 elements. The high dimension of the HOG descriptors allows a simple linear SVM classifier to successfully classify detection windows into "non-pedestrian" and "pedestrian". We call each detection window that is classified as containing a pedestrian as the pedestrian window. To train their SVM classifier, Dalal [6] use a database of more than 500 images of pedestrian in many different standing/walking poses in complex scenes. These images include indoor and outdoor scenes taken at different times of the day. The HOG descriptors are known to contain redundant information because of the different detection window sizes (i.e. different scales) being used. So when the SVM classifier classifies a region as "pedestrian", a number of pedestrian windows of different scales are often associated with the region. The final step of the HOG human detector is therefore to perform "non-maximum suppression" and returns the pedestrian window of the best scale.

The HOG descriptors require a large amount of memory storage because of the high dimension of the descriptors. In addition, the HOG human detector requires the pedestrians to be in upright positions. For poses other than standing and walking, the performance of the HOG human detector deteriorates.

B. 4D colour histograms

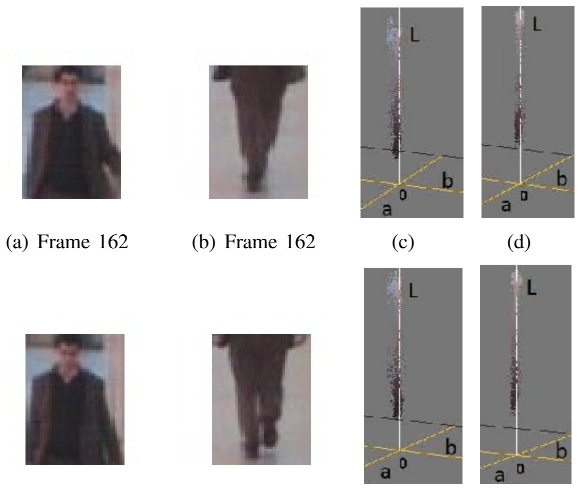
We use a 4-dimensional colour histogram to model the appearance of each detected person in the video frame. Each video frame is converted from the RGB colour space to the $L^*a^*b^*$ colour space via a nonlinear transformation. The

L^* component is related to human perception of brightness while the a^* and b^* components contain the chromaticity. We use the $L^*a^*b^*$ colour model because it is more closely related to human visual perception and is more suitable for handling scenes with shadows and illumination variations. For instance, when a bright red pixel is in shadow, it would appear to be in a dull red colour. This illumination change affects only the brightness part (L^*) of the colour. When computing the distance between two colours, one could use a smaller weight for the L^* component and a larger weight for the a^* and b^* components to minimize the effect of shadows.

For each pedestrian window returned by the HOG human detector, we extract the *bounding pedestrian window* that has the same centroid and is 20% smaller in width and height (the value 20% is determined empirically). A bounding pedestrian window is necessary in our system because the pedestrian windows often contain extra background regions that are needed for the computation of the HOG descriptors. We then halve each bounding pedestrian window into an upper subwindow and a lower subwindow and compute the colour histogram for each subwindow. Typically, if the pedestrian window corresponds to a false positive, it would be unlikely that a pedestrian window in a nearby position in the next video frame would be classified also as "pedestrian". As a result, pedestrian windows that are false positive would simply disappear gracefully in the Kalman filter tracking process. If the pedestrian window truly contains a pedestrian, then, assuming that the pedestrian is in an upright position (this is the same assumption adopted in the HOG human detector), the upper and lower subwindows would correspond to the upper and lower body parts. Concatenation of the two colour histograms produces a 4D colour histogram for each bounding pedestrian window. Figure 1 shows an example of a bounding pedestrian window containing a pedestrian detected in two video frames. It can be seen that the colour histograms look very similar.

An advantage of using 4D colour histograms is that spatial information is captured in the appearance model. Consider the situation where a person wearing a white jacket and black pants and another wearing a black jacket and white pants. If a single 3D colour histogram is extracted from each bounding pedestrian window, then the two colour histograms would look very similar, making it difficult to differentiate the two people. Furthermore, 4D colour histograms are more effective in handling self-occlusion and occlusion caused by obstacles or other moving pedestrians. For instance, if the lower body of a pedestrian is occluded in a video frame, then the 4D colour histogram may still match with that in the previous video frame because of the strong correspondence of the upper colour histograms.

The inverse distance between two colour histograms can be computed using the Bhattacharyya distance proposed by Bhattacharyya [3]. The related Bhattacharyya coefficient has



(a) Frame 162 (b) Frame 162 (c) (d)

(e) Frame 166 (f) Frame 166 (g) (h)

Figure 1. The colour histograms for the upper and lower bodies in the $L^*a^*b^*$ colour space for two video frames. Column 1: The upper subwindows of two bounding pedestrian windows; Column 2: The lower subwindows; Column 3: The colour histograms for the upper subwindows; Column 4: The colour histograms for the lower subwindows.

been widely used in the visual tracking literature, e.g. the mean shift tracker of Comaniciu *et al.* [5] uses the Bhattacharyya coefficient to track the model of a distribution. The colour histograms are normalized before their Bhattacharyya coefficients are computed. Given two colour histograms $\mathbf{h}_t^i$ and $\mathbf{h}_{t+1}^i$ represented as 4D tensors, their Bhattacharyya coefficient is given by

$$D_{t,t+1}^{i,j} = \frac{\boldsymbol{h}_t^i \cdot \boldsymbol{h}_{t+1}^j}{\|\boldsymbol{h}_t^i\| \|\boldsymbol{h}_{t+1}^j\|}, \quad (1)$$

where t denotes the frame index; indices i and j denote the i^{th} and j^{th} bounding pedestrian windows in video frames t and $t+1$ respectively. Bhattacharyya coefficients are in the range $[0,1]$. From this formula, we can see that a larger coefficient value indicates more overlap between the two 4D colour histograms. Therefore, by selecting an appropriate threshold, one can determine whether two bounding pedestrian windows correspond in two consecutive frames. An alternative is to use the $1-D_{t,t+1}^{i,j}$ value, where small values indicate closeness between the two 4D colour histograms.

C. The Kalman filter

The Kalman filter has been extensively applied in object tracking. Our tracking algorithm uses such a technique to predict and correct the human trajectories in video sequences. The Kalman filter involves a set of vector equations that model the update of the state from the current frame to the next one. We assume that the velocity of the tracked pedestrian is constant in the video frames, so the state vector is simplified and does not include the acceleration term. Such an assumption is a reasonable one in most visual surveillance applications. The state vectors in our system are denoted as $\mathbf{x}_t^i = (x_t^i, y_t^i, \dot{x}_t^i, \dot{y}_t^i, w_t^i, h_t^i, \dot{w}_t^i, \dot{h}_t^i, s_t^i)^\top \in \mathbb{R}^9$, where $(x_t^i,\, y_t^i)$ denotes the coordinates of the centroid and $(\dot{x}_t^i,\, \dot{y}_t^i)$ denotes the velocity of the i^{th} bounding pedestrian window in frame t; w_t^i and h_t^i are the width and height of that bounding pedestrian window; parameters $\dot{w}_t^i$ and $\dot{h}_t^i$ represent the rate of change of the bounding pedestrian window; s_t^i is the scale of the HOG descriptor returned by the HOG pedestrian detector for that window.

If we concatenate the state vectors that describe the positions and velocities of all the pedestrians together, then the entire state vector of our pedestrian tracker at each video frame has a variable dimension. However, it is easier to model the tracking process as having multiple Kalman filters, one for each pedestrian detected in the video frame. Following the standard Kalman filter process, we have

$$\begin{aligned} \mathbf{x}_t^i &= T_{t-1}^i \mathbf{x}_{t-1}^i + \mathbf{w}_{t-1} && (2) \\ \mathbf{z}_t^i &= H_t^i \mathbf{x}_t^i + \mathbf{v}_t, && (3) \end{aligned}$$

where T_{t-1}^i and H_t^i denote the state transition and the measurement matrices for the i^{th} person, respectively. The vectors $\mathbf{w}_{t-1}$ and $\mathbf{v}_t$ are noise terms which are assumed to be Gaussians with zero mean and covariance matrices Q_t and R_t. Vector $\mathbf{z}_t^i$ is the measurement for that person at time t. In our case, $\mathbf{z}_t^i$ is a vector in $\mathbb{R}^5$ containing the centroid, width and height, and scale of the i^{th} bounding pedestrian window. Matrix H_t^i is therefore a 5 $\times$ 9 projection matrix. Because of the constant velocity assumption, $T_t^i = T$ and H_t^i =H are for all i and t.

The Kalman filter has two stages: the *prediction* and the *correction* stages. In the prediction stage, the state vector $\mathbf{x}_t^i$ for the i^{th} person in the current frame is predicted using the corresponding state vector in the previous frames via a matrix T. In the correction stage, the observation vector for the i^{th} person in the current frame is used to update the state estimate and the error covariance matrix.

Before carrying out the Kalman filter model, correspondence problem for associating the i^{th} bounding pedestrian window at time t with the most likely observation at time $t+1$ is solved by comparing two 4D colour histograms between consecutive frames through the Bhattacharyya coefficient. The tracking process can be described as follows:

- The prediction stage involves two equations for predicting the state vectors and their associated error covariance matrices at time t given the measurements up to $t-1$.

$$\begin{aligned} \hat{\mathbf{x}}_{t|t-1}^i &= T\mathbf{x}_{t-1|t-1}^i && (4) \\ P_{t|t-1}^i &= TP_{t-1|t-1}^i T^\top + Q_t. && (5) \end{aligned}$$

- The correction stage updates the state vectors and their error covariance matrices with the current measurements.

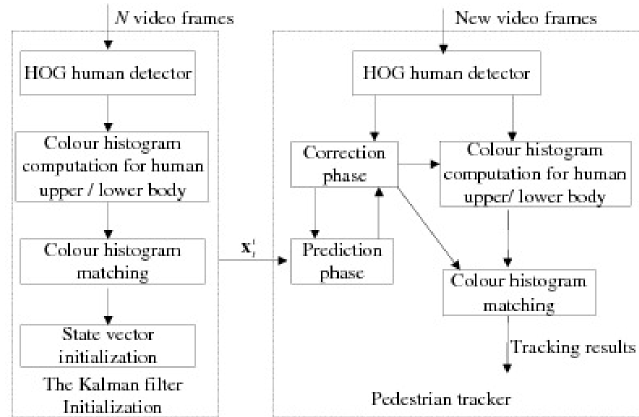


Figure 2. A block diagram of our people tracking system.

$$\begin{aligned} K_t^i &= P_{t|t-1}^i H^\top (HP_{t|t-1}^i H^\top + R_t)^{-1} && (6) \\ \hat{\mathbf{x}}_{t|t}^i &= \hat{\mathbf{x}}_{t|t-1}^i + K_t^i (Z_t^j - H\hat{\mathbf{x}}_{t|t-1}^i) && (7) \\ P_{t|t}^i &= (I - K_t^i H) P_{t|t-1}^i, && (8) \end{aligned}$$

where $\hat{\mathbf{x}}_{t|t}^i$ is the estimate of the state vector for the i^{th} bounding pedestrian window at time t, given the observation up to that time; $P_{t|t}^i$ is the error covariance matrix for the i^{th} bounding pedestrian window at time t indicating the accuracy of the estimated state vector, and K_t^i is the Kalman gain. The tracking process is recursive. The estimates computed from (7) are used as the state vectors on the right hand side of (4) for the next video frame.

D. System overview

Our system consists of several components as illustrated in Fig. 2. Human tracking starts after N video frames have been used for the Kalman filter initialization. In the left part of the diagram, the HOG human detector is firstly applied to the input video to obtain the bounding pedestrian windows. The 4D colour histograms are then computed and their Bhattacharyya coefficients are used for matching pedestrians between successive frames. The objective of this part of the block diagram is to construct the initial state vector for each pedestrian for the Kalman filter.

In the right part of the diagram, the HOG human detector is firstly applied to videos to obtain the bounding pedestrian windows. Simultaneously, the previous state vectors are fed to the prediction phase to predict the state vectors for the current frame. Then the correction phase updates these state vectors based on bounding pedestrian windows observed in the current frame. The 4D colour histogram computation and Bhattacharyya coefficient evaluation are similar to those for the left part of the diagram.

For a new pedestrian bounding window returned by the HOG human detector, we also use the steps of the left part of the diagram in N frames to initialise the new Kalman filter. If a existing pedestrian bounding window is not detected in several frames and its position is near the boundary of the field of the view from the camera, he/she is considered to walk outside the observed area. The Kalman filter for this pedestrian is then deleted.

III. Experimental results

Experiments were carried out on a PC with Intel 4 cores 2.67GHz processor and the pedestrian tracking program was coded using C++ and the OpenCV library.

We tested our pedestrian tracking method on the CAVIAR video database [1] and the CVLAB video dataset [9]. The first video sequence (Figs. 3 and 4) has 384×288 pixels per frame and 25 frames per second. The second video sequence (Fig. 5, 6 and 7) has the same frame rate but only 360×288 pixels per frame. In all our experiments, our video sequences are of 150 frames to 200 frames in length.

For the Kalman filter initialization stage, at least 20 video frames are required in order to get a reasonable estimate of the velocity term of the initial state vector for each bounding pedestrian window. In our experiment. we set N to 25 (see Fig. 2). Thus tracking starts at $t = 25$. The velocity term ($\dot{x}_t^i$, $\dot{y}_t^i$) was calculated by dividing the change of displacement of the bounding pedestrian windows between the first frame and the last one by N. The $\dot{w}_t^i$ and $\dot{h}_t^i$ terms were computed in a similar way. Assuming that the noise terms for all the components of the state vector are uncorrelated, the initial error covariance matrix for each pedestrian is defined to be $P_{0|0}^i$ = diag(σ_x^2, σ_y^2, $\sigma_{\dot{x}}^2$, $\sigma_{\dot{y}}^2$, σ_w^2, σ_h^2, $\sigma_{\dot{w}}^2$, $\sigma_{\dot{h}}^2$, σ_s^2). In all of our experiments, we empirically set $\sigma_x^2 = \sigma_y^2 = 0.7$, $\sigma_{\dot{x}}^2 = \sigma_{\dot{y}}^2 = 0.5$, $\sigma_w^2 = \sigma_h^2 = 0.7$, $\sigma_{\dot{w}}^2 = \sigma_{\dot{h}}^2 = 0.5$ and $\sigma_s^2 = 0.7$.

To illustrate the effectiveness of our pedestrian tracking method, we also conducted experiments that track pedestrians using colour histograms alone. Fig. 3 illustrates the human tracking result using the matching of colour histograms only. The HOG human detector may fail to detect the pedestrian when he/she is walking to a position that is sufficiently far away from the camera. As shown in Fig. 3, person $\sharp 1$ was detected in frames 25 to 46 but not in frame 47 and so could not be tracked by matching the colour histograms alone. The HOG human detector may sometimes return one large pedestrian window that encloses two pedestrians if one person is partially occluded by the other person. In frame 83, for instance, one pedestrian window containing persons $\sharp 3$ and $\sharp 4$ was returned when person $\sharp 4$ partially occluded person $\sharp 3$.

Figure 4 demonstrates continuous human tracking in a video sequence using a combination of colour and motion models. The combined method uses the Kalman filter to predict where the bounding pedestrian window of the missing person $\sharp 1$ was in frame 47. Then computing the colour histograms of the missing bounding window is to compare person $\sharp 1$'s colour histogram in the previous frame for assigning a probability value. To be robust in occlusion handling, we assign different weights to the colour distributions of the upper and lower pedestrian subwindows. Based on the observation that the human lower body is more likely to be occluded by other humans or obstacles in videos, the weight for upper colour histogram was set as 0.65 while the weight for lower colour histogram was set as 0.35. From Fig. 5, we can see that the proposed tracking method can deal with partial occlusion when two people merge into a group.

Using colour model only for human tracking is more likely to fail if two people wear similar clothing. As shown in Fig. 6, person $\sharp 1$ was incorrectly matched to person $\sharp 2$ and vice verse in frame 94 because their colour histograms look similar. This issue can be overcome by our combined method as illustrated in Fig. 7.

IV. Discussion

As our system depends on the human detector of Dalal [6], if the detector returns too many false positives or false negatives, the performance of our tracker would deteriorate. However, if the number of false positive pedestrian windows returned by the HOG human detector is sparse, then these windows would be discarded gracefully in our tracking process, as they are unlikely to have matching windows over time. On the other hand, if a pedestrian is detected in frame t by the HOG human detector but not in frame $t+1$ because of partial occlusion or poor contrast/resolution, then a false negative is said to occur in frame $t + 1$. In such a case, our tracking method can predict the state vector for this pedestrian at frame $t+1$, extract the 4D colour histogram and justify from the computed Bhattacharyya coefficient whether a match should be established.

For the case where a new pedestrian is detected by the HOG human detector, the pedestrian should not have a correspondence with any bounding pedestrian window from the previous frame. However, if its 4D colour histogram happens to match well with that of a previous bounding pedestrian window, then our tracking method would prevent the two windows to be associated with each other unless the motion model is not violated. That is, if the matching of the two pedestrian windows results in a drastic change of velocity, then the matching would be discarded.

Our algorithm vertically divides the bounding pedestrian window in two equal parts. To better deal with occlusion, the body segmentation method proposed by Elgammal *et al.* [8] can be employed. However, extra computational work would be required.

Colour histograms are known to be invariant to translation and rotation. They are also invariant to small changes of viewing angle. However, extremely crowded situations where pedestrians are grouped together and severely occluding each other would be a problem for pedestrian tracking.

Such a problem is not unique to our approach here. The main disadvantage of the Kalman filter is that it estimates the state vector for a linear system. Thus the Kalman filter is not suitable for dealing with cases such as when a pedestrian suddenly accelerates or changes direction.

V. Conclusion

We have presented a pedestrian tracking method that combines colour and motion models for video sequences captured by a fixed camera. Our proposed method can robustly track multiple pedestrians. Our experimental results show that our tracking method outperforms one that uses colour information alone. Furthermore, our method can deal with partial occlusion.

Our future work includes extending our tracking method to incorporate the interacting multiple model (IMM) method [4] which can better model complex motions of pedestrians in videos as it can predict the state parameters of a dynamic system with multiple models switching from one to another.

Acknowledgment

Jiang would like to acknowledge the financial support of a PhD scholarship from the ARC Linkage Project grant LP0883417 and Sensen Networks Pty Ltd.

References

[1] http://homepages.inf.ed.ac.uk/rbf/CAVIAR/.

[2] A. Alahi, Y. Boursier, L. Jacques, and P. Vandergheynst. A Sparsity Constrained Inverse Problem to Locate People in a Network of Cameras. In *16th International Conference on Digital Signal Processing*, pages 1–7, Jul 2009.

[3] A. Bhattacharyya. On a Measure of Divergence between two Statistical Populations defined by their Probability Distrbution. *Bulletin of the Calcutta Mathematical Society*, 35:99–109, 1943.

[4] H. Blom and Y. Bar-Shalom. The Interacting Multiple Model Algorithm for Systems with Markovian Switching Coefficients. *Automatic Control, IEEE Transactions on*, 33(8):780–783, Aug 1988.

[5] D. Comaniciu, V. Ramesh, and P. Meer. Real-time Tracking of Non-rigid Objects using Mean Shift. In *Proceedings. IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 142–149, 2000.

[6] N. Dalal. Finding People in Images and Videos. *Institut National Polytechnique de Grenoble*, Jul 2006.

[7] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893, Jun 2005.

[8] A. Elgammal and L. Davis. Probabilistic Framework for Segmenting People under Occlusion. In *Proceedings. Eighth IEEE International Conference on Computer Vision*, volume 2, pages 145–152, 2001.

[9] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua. Multicamera People Tracking with a Probabilistic Occupancy Map. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(2):267–282, Feb 2008.

[10] D. A. Forsyth, O. Arikan, L. Ikemoto, J. O'Brien, and D. Ramanan. Computational Studies of Human Motion: Part 1, Tracking and Motion Synthesis. *Foundations and Trend in Computer Graphics and Vision*, 1(2/3):77–254, 2005.

[11] J. Kang, I. Cohen, and G. Medioni. Object Reacquisition using Invariant Appearance Model. In *Proceedings of the 17th International Conference on Pattern Recognition*, volume 4, pages 759–762, 2004.

[12] M. Mason and Z. Duric. Using Histograms to Detect and Track Objects in Color Video. In *AIPR 30th Applied Imagery Pattern Recognition Workshop*, pages 154–159, 2001.

[13] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. *In European Conference on Computer Vision*, pages 661–675, 2002.

[14] A. Senior, A. Hampapur, Y. L. Tian, L. Brown, S. Pankanti, and R. Bolle. Appearance Models for Occlusion Handling. *Image and Vision Computing*, 24:1233–1243, 2006.

[15] C. Stauffer and W. Grimson. Adaptive Background Mixture Models for Real-time Tracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 246–252, 1999.

[16] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland. Pfinder: Real-time Tracking of the Human Body. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):780 –785, jul. 1997.

[17] B. Wu and R. Nevatia. Detection of Multiple, Partially Occluded Humans in a Single Image by Bayesian Combination of Edgelet Part Detectors. In *Tenth IEEE International Conference on Computer Vision*, volume 1, pages 90–97, 2005.

[18] A. Yilmaz, O. Javed, and M. Shah. Object Tracking: A Survey. *ACM Computing Surveys*, 38(4), Dec 2006.

(a) Frame 25 (b) Frame 47 (c) Frame 83 (d) Frame 129

Figure 3. Human tracking using the colour model only.

(a) Frame 25 (b) Frame 47 (c) Frame 83 (d) Frame 129

Figure 4. Human tracking using the colour model and motion model.

(a) Frame 33 (b) Frame 40 (c) Frame 49 (d) Frame 81

Figure 5. Human tracking in the presence of occlusion.

(a) Frame 25 (b) Frame 62 (c) Frame 94 (d) Frame 160

Figure 6. Tracking two people in similar clothing using the colour model only.

(a) Frame 25 (b) Frame 62 (c) Frame 94 (d) Frame 160

Figure 7. Tracking two people in similar clothing using the combined method.