Tracking Multiple People under Occlusion Using Multiple Cameras

Ting-Hsun Chang, Shaogang Gong and Eng-Jon Ong
Department of Computer Science
Queen Mary and Westfield College
London E1 4NS, UK
[cth|sgg|ongej]@dcs.qmw.ac.uk

Abstract

We describe a system for tracking multiple people with multiple cameras based on fusion of multiple cues. Face trackers are used to self-calibrate our system. Epipolar geometry and landmarks are employed to disambiguate the tracking problem. The correlation of visual information between different cameras is learnt using Support Vector Regression and Hierarchical Principal Component Analysis to estimate the subject appearance across cameras. The joint features of subjects extracted from multiple cameras are tracked and used as a model to re-track people once the subjects are lost tracking in the system. Results demonstrate that our system can deal with the occlusion.

1 Introduction

To be able to robustly track multiple people in an indoor environment is important in dynamic vision. Since the visual information from a single fixed camera is quite limited, there is a growing interest in obtaining more information by increasing the number of cameras [3,4,5,6,7,14,]. Tracking with multiple cameras not only increases the monitored area, but also helps to disambiguate in matching when subjects are occluded from a certain viewing angle.

One of the underlying issues in tracking with multiple cameras is to associate objects in different cameras. Compared to tracking with one camera, the correspondence problem of multiple cameras is necessarily more difficult, due to the need to use the image features obtained from different spatial coordinates [9]. Past work in cross-camera correspondence can be divided into two categories: geometry-based and recognition-based methods. Geometry-based method transforms the geometric features to the same spatial reference before matching is performed. Most proposed methods require explicit camera calibration [4,5,7]. Another popular assumption is that a person's feet are visible in the image and used as a geometric feature to establish cross camera correspondence [4,5,6]. This assumption is not always valid due to the occurrence of partial occlusions caused by objects in the environment. We attempt to build a system with two cameras located at un-constrained positions and without traditional calibration. The geometric information we used to build correspondence across cameras includes epipolar geometry and a landmark-based method.

Recognition based correspondence is actually a special case of object recognition. In order to associate subjects between two cameras, recognisable features are extracted from two camera views. Previous attempts [3,5,14] apply the colour information extracted from one camera as a model to match subjects in the other camera without any transformation. This could result in a wrong match because the apparent colour of a subject in two cameras might be "dissimilar". To address this problem we employ two methods, Support Vector Regression (SVR) and Hierarchical Principal Component Analysis (HPCA) to learn the correlation of visual information between two cameras. The system then uses this learnt correlation to transform the information between cameras. These two methods are then compared, and the better one is used in our system. We demonstrate that two apparent features which are colour and height of the subject work well in our tracking system. In addition, the CONDENSATION algorithm is employed to track the joint features of the subject extracted from two cameras. Two fixed cameras are used in our system, which are located at the corners of a wall and face the opposite wall in order to monitor the room from two different viewing angles. Fig. 1 shows the images captured from the two cameras. Once the system loses track in one camera or the tracking results become inconsistent, the subject's information is exchanged between two cameras to robustly re-track the subject. To do this, epipolar geometry, landmarks, apparent colour and apparent height of a subject are used to pass the subject information between cameras. If the target is lost in both cameras, the tracked joint features of the subject are used to re-track subjects.

The paper proceeds as follows. Multi-camera correspondence based on geometry and recognition features are discussed in Section 2 and Section 3 respectively. The use of CONDENSATION to track the features of a subject is also discussed in Section 3. Section 4 explains how to combine different cues and techniques. Experimental results are given in Section 5 and we conclude in Section 6.

2 Multi-Camera Correspondence Based on Geometry

To find the corresponding subjects in different cameras, the system first attempts to track subjects in each camera based on motion. In order to extract moving objects, background subtraction is first performed between a reference image and an incoming image followed by a suitable thresholding. The segmented pixels are then grouped into objects and tracked with Kalman filters. An object is then tracked by updating a rectangular window, called a *tracking window*, which circumscribes a segmented region. Once a subject is tracked in two cameras, they need to be associated between two cameras. In the following epipolar geometry and landmark method for geometrically constraining the subjects' positions are described.

Epipolar geometry is a fundamental projective constraint that exists between two camera views [10]. To get the epipolar geometry, a set of 3D corresponding points is necessary. A self-calibration method similar to the work of Azarbayejani and Pentland [3] is employed. We use skin colour [8] to extract the face of the subject from the moving blobs. The segmented region is tracked with a rectangular window referred as a *face box* (Fig. 1). The centroids of the tracked face boxes in two cameras are used to obtain the 3D corresponding points for self calibration and is also used as a subject's feature point to spatially locate its corresponding subject along the epipolar line in the other camera image. An association is made with the subject face box centroid closest to this epipolar line.

Fig. 1. Two views from the left and right cameras. A subject with face box, face box centroid (white dot), associated VA (big white box) (see Fig. 2). The epipolar line (black in right) is computed from the face box centroid of the subject in the left view.

Multiple vertical line landmarks are employed to further geometrically constrain the correspondence. To do this, the correspondence of these lines in two cameras must be established beforehand. We adopt the method proposed by Schmid and Zisserman [2] to build the line correspondence across cameras by employing epipolar geometry and correlation techniques. We introduce the following scheme to provide the system a geometric constraint on correspondence. For a vertical line landmark, there is a unique projective plane containing this line both in the world and the image as shown in Fig. 2 (left). The Field Of View (FOV) of a camera is partitioned into sub-spaces by the projective planes related to multiple lines in the world. We refer this sub-space as the Vertical Volume (VV) of a camera. The area in the image corresponding to a VV is called a Vertical Area (VA). An example is shown as in Fig. 1. Through the imaging process, any objects in a VV will appear in the corresponding VA in the image.

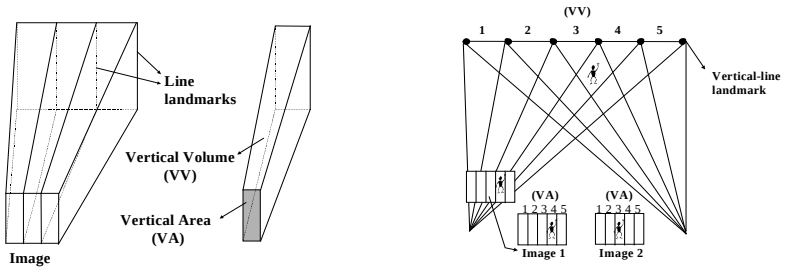

Fig. 2. Field of view (FOV) of a camera with the landmarks (left). Top view of the FOVs of two cameras (right).

For two cameras facing the same wall to monitor a same area, the top view of the FOV of two cameras together with their images are illustrated in Fig. 2 (right). The overlapping FOV is partitioned into multiple cells. For a subject appearing in the *n*th VA in image 1, its possible positions in image 2 are those VAs with numbers less than or equal to *n*, or even does not appear in image 2. We assume that subjects appear in the overlapping area in the FOV of both cameras. Therefore, for a subject in the *n*th VA in image 1, we can constrain its corresponding positions in image 2 to those VAs with numbers less than or equal to *n*. The same rule applies in image 2. Additionally, the cell position in the scene of a subject can be inferred (see Fig. 2) which is used to partition the room into some sub-areas for modelling the spatial change of visual information. Generally, the line landmarks are selected such that the areas of cells approximately equal to the size of a person (Fig.2 right). However, geometric method alone does not provide enough constraints to associate subjects across multiple cameras. In the next section, recognition-based appearance is described.

3 Multi-Camera Correspondence Based on Feature Mappings

In order to build correspondence across cameras, features of subjects need to be extracted from one camera view and transformed to a suitable value for matching the corresponding subjects in the other camera. Fig. 3 shows an example of the colour distribution sampled from one person's clothes in two different cameras and that the two clusters are well separated. Therefore, using the colour model obtained from one camera to associate the corresponding object in the other camera could result in a wrong match. It is also observed that the colour distribution of an object in one camera can change drastically over consecutive frames. This phenomenon of discontinuous colour shift is caused by the multiple illuminants in the environment and the digitisation effect of sequence capturing. For the apparent height of a subject, it is clear that the correlation of this information between two cameras is affected by the viewing geometry. Fig. 4 shows an example of the apparent height of a person extracted from two cameras. The mapping of the observations of two cameras can be discontinuous and non-linear especially when a sudden change in the apparent height occurs in one camera but not in the other.

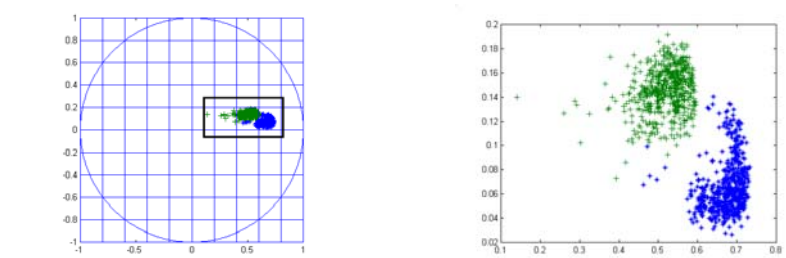

Fig. 3. The variation of one apparent colour in hue and saturation plane of the subject (Fig. 1) under different lighting and viewing conditions captured from two cameras over 200 frames and its enlarged plot (right). Only the mean of each colour sample in one frame is plotted .The dark dots '*' is the mean of each sample in an image frame of the right camera, and the light dots '+' are from the left camera.

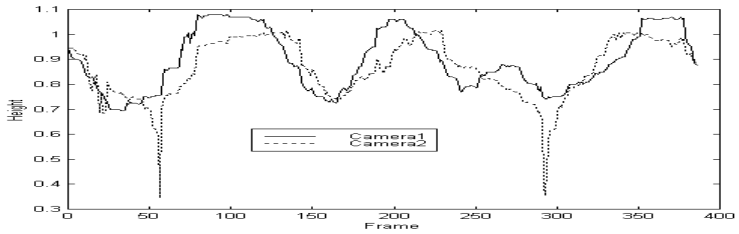

Fig. 4. The observed apparent height of a subject from two cameras over 380 frames when the subject walking around a room. The height measured from camera 2 drops seriously due to the partial occlusion of the lower body part by the table.

To obtain the correct features of the subjects between multiple cameras, the features can be first transform to a common space. The colour can be corrected using some colour calibration method for each camera [15] and the actual size of the subject can be used to help matching between two cameras [1]. However, the colour calibration is only

applicable in a constrained environment and getting the actual size requires stereo range information. We propose to model the non-linear mapping of the visual information between multiple cameras directly and use this model to estimate the feature value in one camera based on the observation of the other camera. Since the observed apparent colour and height varies with person's position, the spatial change of visual information need to be model. To partition the room into some sub-areas, we use the landmark method. Generally speaking, the larger the number of the cells, the stronger the correlation can be learnt. To learn the spatial change and correlation of the visual information between two cameras, SVR and HPCA are employed and compared. The better one is selected for building the correspondence across cameras in our system. The abstract representation of the subject appearance obtained from HPCA also provides the system a computation-inexpensive way to use the CONDENSATION algorithm.

To recognise subjects across cameras, the apparent colour and apparent height and position information are used. The colour distribution is modelled with Gaussian mixture models [8] and the apparent height is extracted from the segmented motion blob. The cell positions of the subjects obtained from the subjects' VA position in two cameras (Fig. 2) are also used as a part of the representation. Our blob representation is related to the work by Azarbayejani and Pentland [3]. It is extended to fuse different sensory data to offer a more complete object interpretation. The features extracted from each camera view are fused into a *joint feature vector* defined as follows:

$$\boldsymbol{V}_c = (x_1, \ x_2, \ \boldsymbol{G}_1, \ \boldsymbol{G}_2), \tag{1}$$

$$\boldsymbol{V}_h = (x_1, \ x_2, \ h_1, \ h_2), \tag{2}$$

where $\boldsymbol{V}_c$ and $\boldsymbol{V}_h$ are *joint colour feature vector* and *joint height feature vectors*, x_i is the VA position of the subject, $\boldsymbol{G}_i$ are the Gaussian model, and h_i is the measured height in camera i. The Gaussian model $\boldsymbol{G}$ consists of mean of the hue and saturation (m_H , m_S) and the covariance matrix.

3.1 Support Vector Regression for Feature Mappings

To estimate the features of a subject across multiple cameras, the non-linear mapping between cameras is formulated as regression estimation and learnt using Support Vector Regression [13]. In the training phase, a person wearing single colour clothes walks around a room, a single Gaussian is used to model the apparent colour, the apparent height of the subject is measured, and the cell positions are obtained from both cameras. Each variable of each feature of each camera needs a SVR to learn the estimation function and is trained independently. To learn the SVR for a feature variable, the subjects' VA position in two cameras, the appearance in the other camera and the observation of this variable are used. For example, n number of vectors $\boldsymbol{v_I} = (x_1, x_2, \boldsymbol{G}_1)_{\boldsymbol{i}}$, $i=1,2,...,n$, are used to train the mapping for m_{H2}, the mean value of hue of $\boldsymbol{G}_2$. After training, a set of Support Vectors $\boldsymbol{s_1}, \boldsymbol{s_2}, \ldots, \boldsymbol{s_p}$ can be obtained. Then, given an unknown observation $\boldsymbol{v} = (x_1, x_2, \boldsymbol{G}_1)$, one can predict m_{H2} in the camera 2 by:

$$m_{H2} = \sum_{\boldsymbol{j}=1}^{\boldsymbol{p}} \alpha_{\boldsymbol{j}} K(\boldsymbol{v}, \boldsymbol{s}_{\boldsymbol{j}}) + b, \tag{3}$$

where α_j is the coefficient of $\boldsymbol{s_j}$, K is a kernel function, and b is the bias. From the learnt mapping, the observed feature of a subject can be transformed to a correct value for the other camera to match the corresponding subject across cameras.

3.2 Hierarchical Principal Component Analysis Method

HPCA [12] is employed to learn the spatial change and correlation of the visual features between two cameras. The learning procedure for HPCA is as follows. Principal Component Analysis (PCA) is first performed on all the training joint feature vectors to remove the less significant components. This dimensionality-reduced space is referred as *global PCA space*. A *parameter space* is then given by the global PCA space into which all training data are projected. A clustering technique is then employed on the projected data. Each cluster is further represented by its principal components called *localised principal components*.

The missing data of a camera can be found by first forming a part of the joint feature vector based on the observations from the other camera, $\boldsymbol{v}_o$, and using an estimate for the missing part $\boldsymbol{v}_e$. This estimate can be the most recent observation, or mean of the previous observation. The synthesised vector, ($\boldsymbol{v}_o$, $\boldsymbol{v}_e$), is then used to find the *most probable point* on the learnt distribution in the parameter space. The synthesised vector is first projected into the global PCA space. After the closest cluster is found based on the Euclidean distance, the projected point, *P*, is constrained to the closest cluster using an approximation method using a *limiter* function [12] (see Fig. 5).

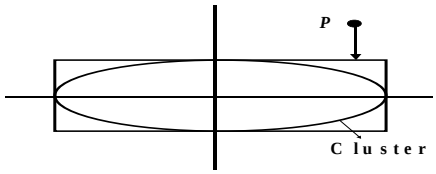

Fig. 5. The projected point is constrained to the closest cluster in the global PCA space. The rectangle corresponds to the threshold of the limiter function, and the axes are the localised principal components of the cluster.

After the most probable point is found in the closest cluster, this point is then projected back to the global PCA space and projected again to the original feature space. This projected back vector ($\boldsymbol{v}_o'$, $\boldsymbol{v}_e'$) is then used as a reference to find the missing data with a simple test criteria. If the difference of the part, $\boldsymbol{v}_o'$, of this recovered vector corresponding to the observation, $\boldsymbol{v}_o$, is smaller than a threshold, the part corresponding to the missing observation, $\boldsymbol{v}_e'$, is then used as the searched result, otherwise the same procedure is repeated for a certain number of times.

3.3 Comaprison of SVR and HPCA on Feature Mappings

Here we compare the estimates of visual information between cameras by using SVR and HPCA. The estimated value is also compared against the observation in the same camera. Table 1 shows the result for five example sequences. In each sequence, one different person is walking in the lab and the colour sample is taken and modelled with a Gaussian model in HS space for both cameras. The Gaussian model from the right camera is then used to predict the Gaussian of the apparent colour in the left camera. The average of the mean of the Gaussian for all frames in a sequence is shown and we omit the covariance matrix as its variation is not significant. It is observed that SVR outperform HPCA for all sequences, and the distance between the estimated mean by HPCA to the observed value is 294% of the mean by SVR. Fig. 7 shows the observed and the estimated height of the left camera based on the right camera by SVR and HPCA

for one sequence with one subject. The estimates by HPCA do not change for some consecutive frames. This means the searched results are the same due to the limiting effects in constraining a point to the nearest cluster. The mean error of the estimates using HPCA is 128% of the SVR for this sequence. The result from the five example sequences is the average mean error of using HPCA is 140% of the SVR.

	H (1)	S(1)	H (2)	S (2)	H (3)	S (3)	H (4)	S (4)	H (5)	S (5)
R Camera	16.23	0.54	27.26	0.31	52.66	0.32	164.36	0.45	215.62	0.12
L (HPCA)	10.74	0.60	18.61	0.50	40.09	0.41	163.47	0.39	229.82	0.22
L (SVR)	7.72	0.64	13.81	0.50	41.56	0.39	124.95	0.33	216.56	0.06
L Camera	7.16	0.66	11.49	0.51	44.51	0.42	130.96	0.36	218.78	0.06

Table. 1. Results from five example sequences. In each sequence, the hue and saturation observed from a person's clothes in the right and left cameras, and the estimated values for the left camera transformed from the observation in the right camera (by HPCA and SVR) are shown. The estimates from SVR are more accurate (closer to the observation) than HPCA for all sequences.

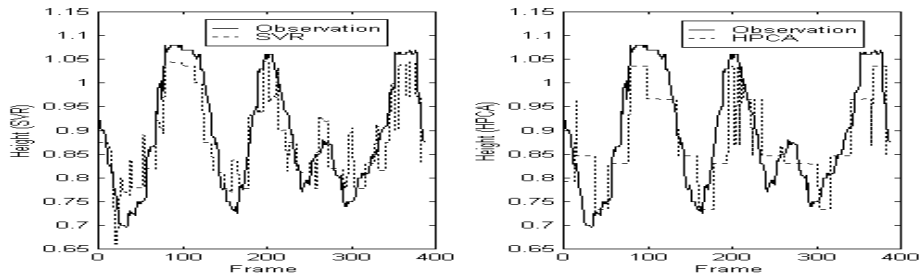

Fig. 6. Observed apparent height of a person in one camera and its estimates transformed from the observation in the other camera over one sequence using SVR and HPCA respectively. The height is measured from the pixel numbers of the subject image in the vertical direction and varies widely as the person walking around in the room. The mean error of the estimates using HPCA is 128% of the SVR for this sequence

3.4 Tracking Joint Feature using CONDENSATION

Having learnt the distribution of the training feature examples using HPCA, we then track the joint features in the global PCA space. To do this, the non-linearity and discontinuity of the feature dynamics which can cause irregular motion and sudden jump in the projection of the joint feature vectors into the global PCA space must be taken into account. To address this problem, the approach of modelling a matrix of transition probabilities between different subspaces in a global eigenspace [12] is employed to learn the dynamics. To track the joint features, we adopt the CONDENSATION algorithm [11].

To apply this algorithm, the sample prediction step is modified to use the transitional probability matrix, and the predicted joint feature vector is obtained by projecting the sample, $\boldsymbol{x}_i(t)$, back to the original joint feature space from the global PCA space. For updating the weighting, the observed data of each camera is first used to weight the corresponding part in the predicted joint feature vector separately. Then the product of the obtained values from two cameras is used to weight the sample with the assumption that observations of two cameras are independent. Once the tracking on a subject is lost, the system needs some features of the subject to recognise it. The tracked and predicted joint feature and position information are used to re-track the subject in both cameras.

4 Fusion for Tracking in Multiple Cameras

To robustly track multiple people, the system relies on multiple cues, and different views from different cameras. The tracking status is divided into two cases: (1) Occlusion presents in one camera but not in the other camera which referred to as One Camera Occlusion (OCO), or the tracking results are inconsistent in two camera. (2) Occlusion presents in both cameras referred to as Two Camera Occlusion (TCO).

For case 1, the information of the subjects is passed across cameras to build correspondence with the geometric and recognition-based methods. Four different cues, epipolar line, landmarks, apparent colour and height, are used for this case. From the discussion in section 3.3, the SVR is chosen to build correspondence across cameras for the recognition-based method. When occlusion is present in one camera the tracked subject motion from the Kalman filter is also used to disambiguate the hypothesised subjects. To take full advantage of the use of multiple cues, an association algorithm [6] to fuse the data from different sensors based on Bayes' theorem is used. For each cue, the decision on matching the hypothesised corresponding subject pairs between two cameras is made. The results from each cue are then fused together to make a final tracking report. For case 2, the method for case 1 is first used to associate the subjects across cameras. However, the system loses identities of the subjects due to the occlusion present in both cameras simultaneous. The system needs some features unique to each subject to recognise the subjects. The predicted joint features of the subjects which are tracked by the CONDENSATION algorithm are used to re-track the subjects.

5 Experiments

For learning the spatial change and the correlation of the visual information between cameras and predicting the visual features across cameras, the training data was obtained from 15 sequences. For SVR method, our system uses 12 SVR predictors to predict the apparent colour between two cameras and 2 SVRs for height. Each camera needs 6 SVRs to predict the Gaussian parameters and 1 SVR to predict the height based on the visual information observed in the other camera. For HPCA method, a 14 dimensional joint colour feature vector, $\boldsymbol{V}_c$, is used for representation colour information. It is also found that for colour estimation, the 5 principal eigenvectors accounted for about 80% of the variance in the eigensapce. This number was found to be sufficient for tracking. For the joint height, a 4 dimensional joint height feature vector, $\boldsymbol{V}_h$, is used, and 3 principal eigenvectors are selected. The system tracks at a rate of 0.5 Hz. The reason to slow down our system is due to the use of CONDENSATION for tracking joint features.

Here, we show the results from sequences with subjects and clothes that are not in our training sequences. The experiments show that our modelling of the spatial change and the correlation of the visual information can generalise to the 'unseen' information. An example of SVO is shown in Fig. 7 (a-c). In this case, the landmark method fails, the epipolar line method is less reliable, and the apparent height as well as the prediction by Kalman filter tracking in the left camera is dominant. Moreover, due to the colour of subjects' clothes being very similar, applying the colour model from the other camera to associate subjects fails. Our system succeeds to distinguish the hypothesised subjects by using the transformed colour model. In Fig. 7 (d-f), the system can robustly re-tracks multiple people even when occlusion are present in both cameras simultaneously. After subjects reappear, they are re-tracked using the joint features predicted by

CONDENSTATION. The apparent height is less reliable than the colour, since people are close in the world and the heights of two subjects are not very “dissimilar”. In this case, the Kalman filter fails since the subjects change direction during occlusion.

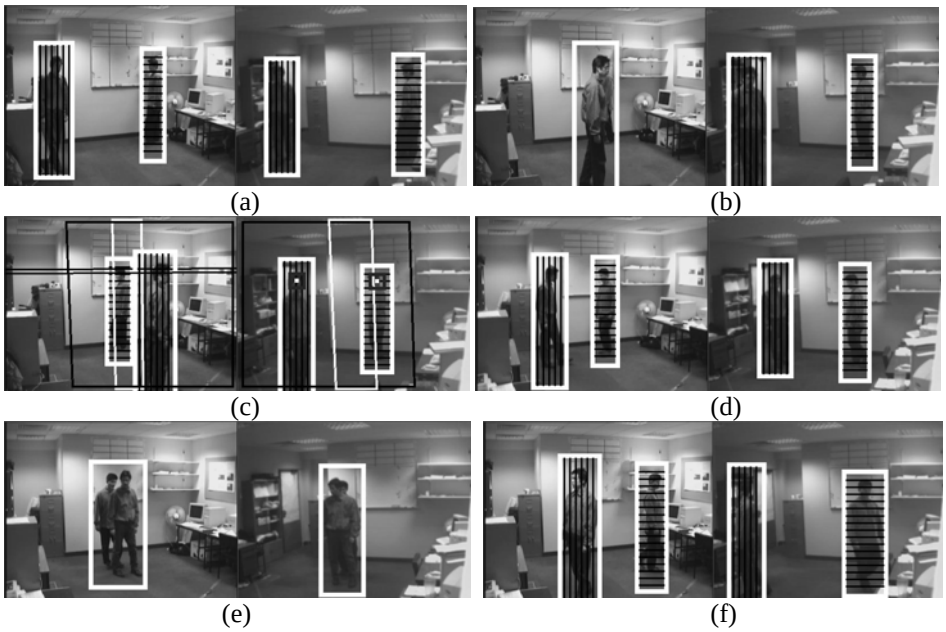

Fig. 7 (a-c) and (d-f) are two tracking events of OCO (One Camera Occlusion) and TCO (Two Camera Occlusion) respectively. (a) two subjects tracked and labelled with different marks in both cameras (the tracked subjects with same marked lines in two images correspond to a same person) before occlusion and (b) occlusion in the left camera where the tracking window without marked lines meaning the system fails to track subject with identity (c) shows the information of subjects are passed to the left camera by fusion multiple cues and the subjects are re-tracked and marked with correct lines. The big black box are the possible position for associating corresponding subject inferred by landmark method. (d) two subjects tracked in two cameras before occlusion (e) occlusion in both cameras simultaneously (f) subjects are re-tracked by the CONDENSATION algorithm.

6 Conclusion

We have presented a system for tracking multiple people in an indoor environment using multiple cameras. Epipolar and landmark-based methods are combined for tracking. Epipolar geometry and landmarks provide the system with geometric constraints on the position of hypothesised subjects when building correspondence across cameras. The spatial change and the correlation of the visual feature information which are apparent colour and height of the subjects between two cameras are learned using SVRs and HPCA. This learnt mapping function of the visual cues between cameras from two different cameras is then used to associate subjects across different cameras. The experimental results show that our model can generalise to novel sequences. We have demonstrated that the SVRs outperforms HPCA in transforming information between cameras. Therefore, SVRs is used in our system for building the correspondence across multiple cameras. This transformation process can also deal with colour inconstancy in a

multiple-camera setup by estimating a correct model of the subject's appearance across cameras, though it does not applies to all lighting conditions.

The vertical line landmark method fails to build correspondence when the position order of subjects in two cameras reverses. The assumption about the landmarks is that to get the 3D information of the subject in the scene, these line landmarks must be vertical in the world. However, there exist many vertical line landmarks from man-made objects in the indoor environment. Another limitation of the system is that the colour cue only works under the learnt lighting condition. To relax this problem by weighting the colour cue less when lighting condition changes, we plan to combine the different methods discussed in this paper with different weights according to their measurement condition to make the tracking system more robust.

References

[1] C. Eveland, K. Konolige, and R. C. Bolles, *Background Modelling for Segmentation of Video-Rate Stereo Sequences*, In CVPR, 266-271, 1998.

[2] C. Schmid and A. Zisserman, *Automatic Line Matching across Views*, In CVPR, 666-671,1997.

[3] A. Azarbayejani and A. P. Pentland, *Real-Time Self-Calibration Stereo Person Tracking Using 3-D Shape Estimation from blob Features*, M.I.T. Media Laboratory, Perceptual Computing Technical Report No. 363. 1996.

[4] K. Sato, T. Maeda, H. Kato, and S. Inokuchi, *CAD-based Object Tracking with Distributed Monocular Camera for Security Monitoring*, Proc. of 2nd CAD-based Vision Workshop. Champion PA. USA. 291-297,1994.

[5] S. Stillman, R. Tanawongsuwan, and I. Essa, *A system for Tracking and Recognising Multiple People with Multiple Cameras*, Georgia Institute of Technology, Graphics, Visualization and Usability Centre, Technical Report No. GIT-GVU-98-25, 1998.

[6] B. S. Rao and H. Durrant-Whyte, *A Decentralized Bayesian Algorithm for Identification of Tracked Targets*, IEEE Trans. on Systems, Man, and Cybernetics, Vol. 23. No. 6. 1683-1698, 1993.

[7] Q. Cai, J. K. Aggarwal, *Automatic Tracking of Human Motion in Indoor Scenes Across Multiple Synchronized Video Streams*, Proc. of 6th International Conference on Computer Vision, Bombay India, 356-362, 1998.

[8] R. Yogesh, S. J. Mckenna and S. Gong, *Segmentation and Tracking Using Colour Mixture Models*, Asian Conference on Computer Vision, Hong Kong, 607-614, 1998.

[9] J. K. Aggarwal and Q. Cai, *Human Motion Analysis: A Review*, Computer Vision and Image Understanding, Vol. 73. No. 3. 428-440, 1999.

[10]G. Xu and Z. Zhang, *Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach*. Kluwer Academic Publishers. 1996.

[11] M. Isard and A. Blake, *Condensation – Conditional Density Propagation for Visual Tracking*, Int. J. Computer Vision, 1998.

[12] T. Heap, *Learning Deformable Shape Models for Object*, Ph.D. thesis, School of Computer Studies, University of Leeds, UK, 1997.

[13] H. Drucker, C. Burges, L. Kaufman, A. Smola, and V. N. Vapnik, *Support Vector regression machines*, In Advances in Neural Information Processing Systems volume 9, The MIT Press, 1997.

[14] J. Orwell, P. Remagnino and G.A. Jones, *Multi-Camera Colour Tracking*, In CVPR, Colorado USA, June 1999.

[15] Y.C. Chang, J. F. Reid, *RGB Calibration for Colour Image Analysis in Machine Vision*, IEEE Transactions on Image Processing, Vol. 5, No. 10, 1414-1422, 1996.