MULTI-CAMERA CORRESPONDENCE BASED ON PRINCIPAL AXIS OF HUMAN BODY

Min Hu, Jianguang Lou, Weiming Hu, and Tieniu Tan

National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
P.O.Box 2728, Beijing, 100080, P.R.China
{mhu, jglou, wmhu, tnt}@nlpr.ia.ac.cn

ABSTRACT

Multi-camera correspondence of moving people is a relatively new issue in computer vision. To cope with it, we propose a simple but effective method based on the principal axis of human body. We apply the method to real video sequences in outdoor environments. The experimental results have demonstrated the efficiency of the proposed method.

1. INTRODUCTION

People tracking in multiple cameras is essentially a multi-camera correspondence problem, i.e. correspondence between people observed from different cameras at the same time. This is a relatively new issue in computer vision and some approaches have been proposed to deal with it. By what types of features are employed, the existing approaches can be divided into two major categories.

1.1. Region-based methods

Region-based methods use the feature of human regions to establish correspondence in multiple views. Color is a popular region cue to generate correspondence between the views [1]-[3]. However, color correspondence across multiple cameras is highly unreliable. For one thing, it relies on the dissimilarity color of human clothes. For another, lighting variations and viewpoint difference also lead to the same person being seen with different colors in different cameras.

1.2. Point-based methods

Point-based methods are to match feature points across views by geometric constraints. By the types of geometric constraints, feature point correspondence methods can be organized into two sub-classes: 1) *3D strategies*. In [4]-[5], centroid is taken as the feature point and correspondence is established by comparing the projected 3D centroids in the world coordinate system. In [6], the correspondence is established by using a set of feature points in the middle line on upper human body based on the epipolar constraint. Correspondence using 3D strategies has brought the burden of camera calibration. Furthermore, feature points extracted often do not correspond to the same physical 3D point. 2) *2D strategies*. In [7], field of view lines are used to establish the correspondence. However, the method assumes the feet of people to be visible. [8] provides the transfer error based on the homography constraint to correspond a pair of centroids in two views. Either using 3D or 2D strategies, point-based methods are easily influenced by noise. The performance of the correspondence decreases if only part of a person is visible due to occlusion or poor detection results. So, exact motion detection is required for these methods.

1.3. Our method

Bearing in mind the weakness of the existing methods outlined above, we use principal axes of human bodies to solve the multi-camera correspondence problem. The principal axis is the symmetric axis of a human body. The homographies that align the ground plane in different views are computed to generate geometric constraints for matching. Then the matching is performed by the maximization of the correspondence likelihoods reflecting the correspondence similarity of principal axis pairs. In contrast to region-based and point-based methods, the proposed principal axis-based method has the following desirable features: (a) The method does not need camera calibration except the homography between the ground plane in one camera to the other. (b) Principal axes are more robust to noise or poor motion detection. (c) The method can accurately localize human positions nd thus generate accurate trajectories of people.

The reminder of the paper is organized as follows. Section 2 describes how principal axes are used to establish correspondence of people across multiple cameras. Section 3 focuses on the principal axis detection methods under different situations. Section 4 shows experimental results on the PETS 2001 database. Finally, conclusion is given in Section 5.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

2. MULTI-CAMERA CORRESPONDENCE BASED ON PRINCIPAL AXES

Assume that camera i observes M people with principal axes $L_1^i, L_2^i, \cdots, L_M^i$ at time t, and camera j observes N people with principal axes $L_1^j, L_2^j, \cdots, L_N^j$ at time t. The correspondence problem is to find a set of solutions $\{m, n\}$ which maximize the correspondence likelihoods:

$$\{m,n\} = arg\max_{s,k}\{\mathrm{L}(L_s^i, L_k^j)\}, \quad s \in [1,M], k \in [1,N] \tag{1}$$

where $\mathrm{L}(L_s^i, L_k^j)$ is the correspondence likelihood for a pair of principal axes $\{L_s^i, L_k^j\}$.

2.1. Maximum correspondence likelihood

We use the correspondence likelihood to evaluate the similarity of principal axis pairs across views. In Fig.1, let L_s^i denote the detected principal axis of person s in camera i. g_s^i is the projection of L_s^i on ground plane Π. For person k in camera j, L_k^j and g_k^j are similarly defined. The ground plane homography from the ith image plane to the jth one is denoted as H_i^j, which can be recovered by hand using the method in [9]. Let L_s^{ij} be the line obtained by applying H_i^j to L_s^i. L_s^{ij} and L_k^j will have one intersection, denoted as Q_{sk}^{ij}. According to the property of H_i^j, if person s in camera i and person k in camera j correspond to the same person in the world coordinate system, Q_{sk}^{ij} will correspond to the "land-point" (Q_{sk}^{π}), the intersection of the principal axis and ground plane Π. Thus, the distance between the observation of "land-point" and the intersection Q_{sk}^{ij} can be used to calculate the correspondence likelihood for principal axis pairs.

Let X_s^i denote the observation of the "land-point" of person s in the ith camera view, and X_k^j denote that of person k in the jth camera view. Q_{ks}^{ji} is the point corresponding to Q_{sk}^{ij} on the ith image plane. We assume that X_s^i and X_k^j are independent of each other, for X_s^i and X_k^j are independently detected in two camera views. The correspondence likelihood for the pair of persons $\{s, k\}$ is defined as

$$\mathrm{L}(L_s^i, L_k^j) = p(X_s^i, X_k^j | Q_{sk}^{\pi}) = p(X_s^i | Q_{ks}^{ji}) p(X_k^j | Q_{sk}^{ij}) \tag{2}$$

To compute the correspondence likelihood in Equ.2, we need to specify the probability densities $p(X_s^i|Q_{ks}^{ji})$ and $p(X_k^j|Q_{sk}^{ij})$. Without loss of generality, Gaussian distributions are assumed.

$$p(X_k^j|Q_{sk}^{ij}) = (2\pi|\Sigma_k^j|)^{-2/2}\exp\{-\frac{1}{2}(X_k^j - Q_{sk}^{ij}){\Sigma_k^j}^{-1}(X_k^j - Q_{sk}^{ij})^T\}$$

where the covariance Σ_k^j is a diagonal matrix with two components of $(\sigma_{x_k}^j)^2$ and $(\sigma_{y_k}^j)^2$.

$$p(X_s^i|Q_{ks}^{ji}) = (2\pi|\Sigma_s^i|)^{-2/2}\exp\{-\frac{1}{2}(X_s^i - Q_{ks}^{ji}){\Sigma_s^i}^{-1}(X_s^i - Q_{ks}^{ji})^T\}$$

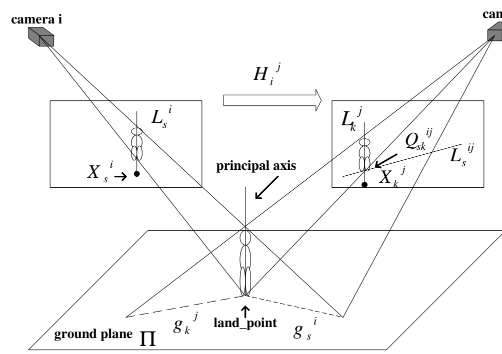

Fig. 1. Projection of principal axes.

where the covariance Σ_s^i is a diagonal matrix with two components of $(\sigma_{x_s}^i)^2$ and $(\sigma_{y_s}^i)^2$.

If we assume that Σ^i is independence of image positions, the correspondence problem in Equ.1 is solved as

$$arg\max_{s,k}\mathrm{L}(L_s^i, L_k^j) \Longleftrightarrow arg\min_{s,k} D_{sk}^{ij} \tag{3}$$

where D_{sk}^{ij} is the correspondence distance for a pair of principal axes, defined as

$$D_{sk}^{ij} = (X_s^i - Q_{ks}^{ji}){\Sigma^i}^{-1}(X_s^i - Q_{ks}^{ji})^T + (X_k^j - Q_{sk}^{ij}){\Sigma^j}^{-1}(X_k^j - Q_{sk}^{ij})^T \tag{4}$$

2.2. Correspondence algorithm in multiple cameras

The major steps of the correspondence algorithm are summarized as follows:

(1) Create a list of all possible correspondence pairs of people for each camera view.

(2) Search for the best match among the list of pairs. The best match of the person (m) in the ith camera view and the person (n) in the jth camera view must satisfy:

$$\{m,n\} = arg\min_{s,k} D_{sk}^{ij} < T_d \tag{5}$$

where T_d is the threshold pre-defined.

(3) Delete the pair of $\{m, n\}$ from the list and return to step 2. If no best match found, go to step 4.

(4) Update the human positions with the intersection and label the corresponded people.

3. DETECTION OF PRINCIPAL AXES

3.1. Detecting principal axes of isolated people

Least Median of Squares method [11] is applied to determine the principal axis based on the global shape constraint

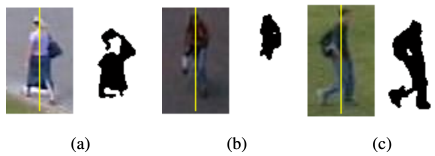

Fig. 2. Principal axes and foreground regions of isolated people.

that human body is typically close to symmetric around its principal axis while in upright fashion. The best fitting principal axis is computed by minimizing median of squared perpendicular distances between the foreground pixels and the axis.

$$L = \min_{l} median_i\{D(X_i, l)^2\} \tag{6}$$

where X_i is the coordinates of the ith foreground pixel, l is the principal axis to be determined, $D(X_i, l)$ is the perpendicular distance between the ith foreground pixel and the principal axis to be determined.

Fig.2 demonstrates the detected principal axes of isolated people. In Fig.2-(b), though the bottom part of the person does not be detected for some reasons, the principal axis of it is accurately detected and less affected by the poor motion detection result that influences the centroid very much.

3.2. Detecting principal axes of multiple people in a group

When people move in a group, it first needs to segment individuals from the group. Then, principal axes of the individuals are determined as Section 3.1.

We segment people by introducing the peak region, defined as a region between two valleys in the vertical projection histogram [10]. Not all peak regions correspond to individuals; only significant peak regions are selected as the candidate individuals. A peak region is significant if it is satisfies two conditions: maximum value of a peak region is above a peak threshold (T_p) and the values of two valleys must be lower than a valley threshold (T_c). In this paper, the valley threshold value (T_c) is selected as the mean value of the entire histogram. The peak threshold value (T_p) is selected as eighty percent of human image height.

Fig.3 shows an example of segmented individuals in a group. The foreground region is given in Fig.3-b, and its vertical histogram is shown in Fig.3-d. According to the three significant peak regions detected in the histogram, three individuals are segmented and their principal axes are accurately determined (Fig.3-c).

3.3. Detecting principal axes of occluded people

In case of occlusion, we use probabilistic appearance model [12] to segment people and then determine the principal axes using the method in Section 3.1. The appearance model is an rgb color model, which records the rgb color of each pixel of an object, with an associated probability mask, which records the likelihood of the object being observed at that pixel. Fig.4 shows an example of segmenting a person from the occlusion of a car. Based on the segmented visible part of the person (Fig.4-c), the principal axis is accurately detected in Fig.4-d.

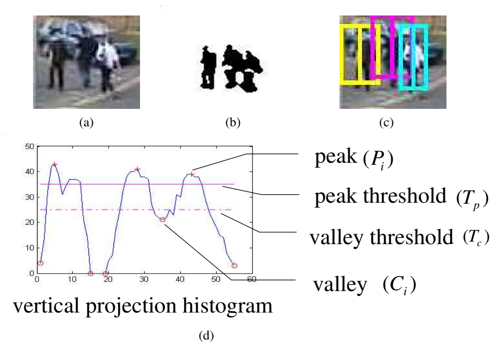

Fig. 3. Segmenting multiple people in a group: (a) input image, (b) detected foreground region, (c) segmented individuals and their principal axes, (d) vertical projection histogram of the foreground region.

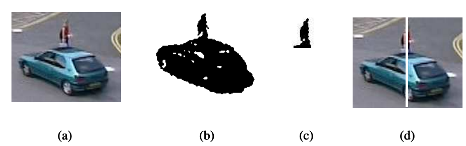

Fig. 4. Detecting the principal axis of a occluded person: (a) input image, (b) detected foreground region, (c) segmented foreground pixels of the occluded person, (d) detected principal axis.

4. EXPERIMENTS

We have tested the proposed method on the testing part of the PETS 2001 dataset1. This dataset includes two-viewpoint video sequences, each of which containing 2688 frames with 768*576 resolution. Fig.5 shows an example of tracking a group of people through occlusion. In this example, a group of three people and a car are moving towards each other. As we can see, two persons in the group have black clothes. In this case, it's difficult to identify them in two cameras with color information. During tracking, people are patially occluded by the car in both cameras, which introduces extra difficulty for segmenting and tracking. In this case, traditional feature points are very difficult to extract and not reliable for matching. Using our method, the individuals in the

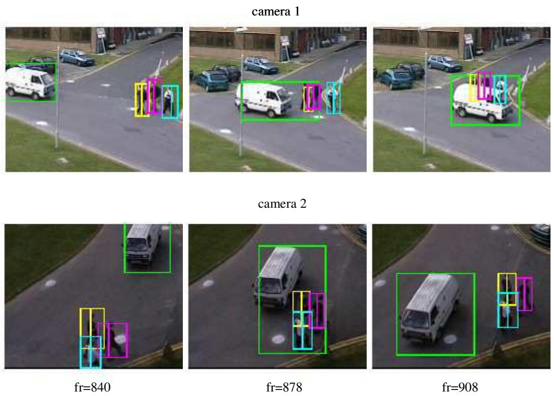

Fig. 5. Tracking a group of people through occlusion.

group are correctly segmented and matched.

To evaluate the algorithms we have also compared the tracking results of the proposed method with those of the point-based method on the same portion of the PETS 2001 dataset1 from frame 2127 to frame 2150 in camera 1. The comparison is done between the trajectory obtained by each method and the ground truth data that we obtained by manual tracking. We measure the trajectory error which is defined as the mean of the distance between estimated human position and the truly human position at each frame. The first result, shown in Fig.6-(a), is a comparison between the trajectory obtained by the proposed principal axis-based method and the ground truth data. The trajectory error for principal axis-based method is $\sigma = 3.2$ pixel. The second result, shown in Fig.6-(b), is a comparison between the trajectory obtained by the point-based method and the ground truth data. Without loss of generality, point-based method use centroid to estimate human position. The trajectory error for centroid trajectory is $\sigma = 5.8$ pixel. From the trajectory error, we can see that the trajectory of the proposed principal axis-based method is more accurate than that of the point-based method.

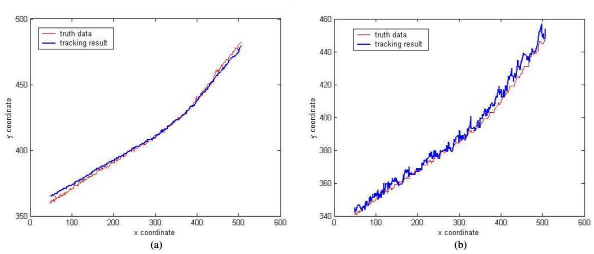

Fig. 6. Comparison: (a) trajectory of the principal-axis based method and truth data, (b) centroid trajectory of the point-based method and truth data .

5. CONCLUSION

We have proposed a novel principal axis based method to solve multi-camera correspondence problem. Contrast to traditional methods, it does not need camera calibration and is less dependent on exact motion detection results. Furthermore, it can accurately localize human positions in each view. The proposed method is tested on real video sequences of the PETS 2001 dataset. The experimental results have demonstrated the effectiveness and efficiency of the method.

Acknowledgements

This work is partly supported by NSFC (Grant No. 60105002, 60373046, 60335010, 60121302), Natural Science Foundation of Beijing (Grant No. 4031004 and 4041004), the National 863 Program of China (Grant No. 2002AA117010-11).

6. REFERENCES

[1] J. Orwell, P. Remagnino, G. A. Jones, *Multi-Camera Color Tracking*, in Proc. of The 2ed IEEE Workshop on Visible Surveillance, 1999, pp. 14-24.

[2] J. Krumm, S. Harris, B. Meyers, B. Brumitt, M. Hale and S. Shafer, *Multi-Camera Multi-Person Tracking for EasyLiving*, in Proc. of The 3rd IEEE Intl. Workshop on Visual Surveillance, 2000, pp. 3-10.

[3] A. Mittal and L. S. Davis, *M2Tracker: A Multi-View Approach to Segmenting and Tracking People in a Cluttered Scene Using Region-Based Stereo*, in Proc.of The 7th ECCV, May-June, 2002.

[4] H. Tsutsui, J. Miura, and Y. Shirai, *Optical Flow-Based Person Tracking by Multiple Cameras*, in Proc. of IEEE Conf. on Multisensor Fusion and Integration in Intelligent Systems, 2001.

[5] A. Utsumi, H. Mori, J. Ohya and M. Yachida, *Multiple-Human Tracking Using Multiple Cameras*, in 3rd IEEE Int. Conf. on FG, 1998, pp. 498-503.

[6] Q. Cai and J. K. Aggarwal, *Tracking Human Motion in Structured Environments Using a Distributed-Camera System*, in IEEE trans. PAMI, November, 1999, Vol.2, No.11, pp. 1241-1247.

[7] S. Khah and M. Shah, *Consistent Labeling of Tracked Objects in Multiple Cameras with Overlapping Fields of View*, in IEEE Trans. On Pattern Analysis and Machine Intelligence, vol.25, No.10, October 2003, pp.1355-1360.

[8] J. Black and T. Ellis, *Multi Camera Image Tracking*, in Proc. 2ed IEEE Int. Workshop on PETS, 2001.

[9] K. J. Bradshaw, L. D. Reid and D. W. Murray, *The Active Recovery of 3D Motion Trajectories and Their Use in Prediction*, in IEEE Trans. on PAMI, March, 1997, Vol.19, No.3, pp. 219-234.

[10] S. Stillman, R. Tanawongsuwan and I. Essa, *A system for Tracking and Recognizing Multiple People with Multiple Cameras*, in Proc. of The 2ed Intl. Conf. on Audio and Video-Based Biometric Person Authentication, 1999, pp. 96-101.

[11] Y. Yang and M. Levine, *The Background Primal Sketch: an Approach for Tracking Moving Objects*, Machine Vision and Applications, vol. 5, pp.17-34, 1992.

[12] A. Senior, *Tracking People with Probabilistic Apperarance Models*, in the Proc. of 3rd IEEE Int. Workshop on PETS, 2003.