2011 Sixth International Conference on Image and Graphics

Global Trajectory Construction across Multi-Cameras via Graph Matching

Xiaobin Zhu, Jing Liu, Jinqiao Wang, Wei Fu, Hanqing Lu
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
Beijing, China
{xbzhu, jliu, jqwang, wfu, luhq}@nlpr.ia.ac.cn

Yikai Fang
System Research Center
Nokia Research Center
Beijing, China
ykfang@gmail.com

***Abstract*—Behavior analysis across multi-cameras becomes more and more popular with the rapid development of camera network in video surveillance. In this paper, we propose a novel unsupervised graph matching framework to associate trajectories across partially overlapping cameras. Firstly, trajectory extraction is based on object extraction and tracking and is followed by a homographic projection to a mosaic-plane. And we extract appearance and spatio-temporal features for trajectory description. Then a robust graph matching algorithm based on reweighted random walk is adopted for trajectory association. The association is formulated as node ranking and selection on an association graph whose nodes represent candidate correspondences of trajectories. Finally, the pairs of corresponding trajectories in overlapping regions are fused by an adaptive averaging scheme, in which trajectories with more observations and longer length is given higher weight. Experiments and comparison on real scenarios demonstrate the effectiveness of the proposed approach.**

Keywords-trajectory association; fusion; trajectory projection; graph matching; camera-network; reweighted random walk

I. INTRODUCTION

Due to limited vision scope but low cost of single camera, the cooperative camera-network system for video surveillance in a broad area, e.g., airport, coast, and road, has attracted increasing attention recently. In the camera-network system, how to construct the global trajectory of moving objects across multiple cameras becomes an essential research topic. However, there are various possible disturbances in the task, such as false moving objects, object re-entrances and occlusion. Thus, how to perform trajectory association robustly and efficiently under such complex scenarios is challenging and difficult.

Previous methods [5, 7, 8, 12, 18] often model trajectory association as a point-wise matching problem. They ignore the structural or context information among trajectories and cannot obtain an optimal association sets from a global view.

To overcome this shortcoming, graph matching had been adopted in performing trajectory association for multiple cameras [4, 13]. They constructed an undirected bipartite graph for two cameras. The vertex in bipartite graph denotes the trajectory in camera, and the weight of edge denotes the affinity of two trajectories. The maximum matching of the bipartite graph is found by Hopcroft and Karp [15]. However the graph matching algorithms, which are adopted in trajectory association, focused on exploiting relatively weak unary and pair-wise attribute and didn't aim at optimizing a well defined objective function [10]. Instead, based on Integer Quadratic Programming (IQP), graph matching formulation can take into consideration both unary and pair-wise terms. Since IQP is known to be NP-hard, approximate solutions are required. And graph matching by random walk in an association graph is a good approximate solution.

In this paper, to independent on the calibration knowledge and inter-topology of cameras, we consider the problem of inferring the correct object association across partially over-lapping cameras using local trajectories, which are extracted from individual cameras. Then we perform the association using the projected trajectories on the mosaic-plane, which is computed by the background images of the neighboring two cameras. Graph matching algorithm based on reweighted random walk is adopted to recover the optimal association sets. The reweighted step can strengthen the effect of reliable nodes in random walk by adopting a jump or teleport. Consequently, it can be more robust to noise and outliers. For the similarity estimation in graph construction, we attempt to combine appearance correlations and spatio-temporal correlations between any pair of trajectories. Then candidate associations are established using reweighted random walk on the affinity graph. After association, the pairs of corresponding trajectories in overlapping regions are fused by adaptive averaging, in which trajectory with more observations and longer length is given higher weight. Figure 1 illustrates the framework of the proposed approach.

The rest of the paper is organized as follows. The related work about object association across multiple cameras is overviewed in Section 2. The pre-processing for trajectory association is given in Section 3. The details of our proposed algorithm are discussed in Section 4. Section 5 presents the experimental evaluations. The conclusions are given in Section 6.

978-0-7695-4541-7/11 $26.00 © 2011 IEEE
DOI 10.1109/ICIG.2011.101

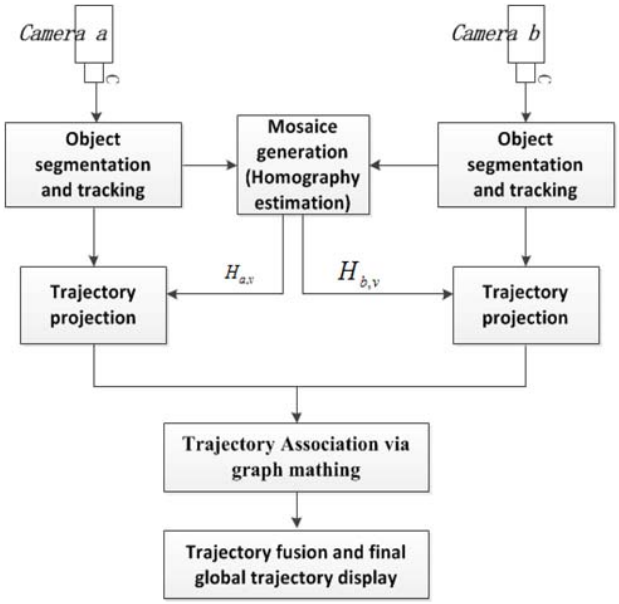

Figure 1. Framework of our proposed approach

II. Related Work

Trajectory association approaches can be categorized into three main classes, namely appearance based, geometry based and hybrid approaches.

Appearance based approaches [1 - 4, 13] adopt color information to match trajectories across cameras. Kim [3] adopted a tracking method based on template matching and on color histogram back-projection to solve occlusion problem. Kuo [4] proposed an on-line learned discriminative appearance affinity model for associating tracks across multiple non-overlapping cameras. Appearance based methods generally suffer from illumination variation.

Geometry based approaches [2, 5 - 8, 14, 18] established association between objects generally by exploiting epipolar geometry, homography correlation and camera calibration. Junejo [5] proposed a linear approach to auto-calibrate cameras through an unsupervised learning. After auto-calibrating a camera and metric rectifying the input trajectories, path models are constructed from these trajectories and similarity measure is adopted to compare the input trajectory with the path model for abnormal activity detection. Ariel [6] proposed an approach using spatial information for associating trajectories from multiple views, which are represented as consecutive points of a joint ground plane in the world coordinate system. Methods based on pure geometric constraints heavily rely on the accuracy during the correspondence process.

Hybrid methods [8, 9, 12, 13] combined appearance and geometry approached in association process. Sheikh [8] associated trajectories across multiple views from airborne cameras with a statistical approach. They make two basic assumptions: 1) cameras are significantly high with respect to the ground and 2) at least one object is covered by two cameras simultaneously at a minimum duration. They posed the problem of maximizing the likelihood function as a k-dimensional matching and used the matching result as an approximate association. Wei [9] proposed a framework that combines particle filtering and belief propagation to track athletes in team sports using multiple cameras. The former algorithms based on hybrid methods always neglect the robustness to noise and outlier in association.

Our work is highly inspired by the approaches of Anjum [12], and Javed [13]. Anjum [12] presented a trajectory association for partially overlapping cameras. Multiple features, including appearance, motion, spatial distance, etc., are adopted in their algorithm. But they neglected the temporal information between trajectories from two cameras. Because two objects across the overlapping region between two adjacent cameras exceed a maximum transition time can't be corresponding object. We adopted temporal information in our algorithm. Javed [13] established object correspondence across non-overlapping cameras by using motion trends and appearance of objects. Then he adopted a bipartite graph matching method to find corresponding trajectories between two non-overlapping cameras. However, this method is not so robust to noise and outliers. There are various possible disturbances in the association, such as false moving objects, object re-entrances and occlusion. So a more robust graph matching algorithm based on reweighted random walk is adopted in our algorithm to promote robustness and accuracy in association.

III. Trajectory Extraction and Projection

In this paper, we perform association on mosaic-plane with trajectories from individual cameras. The transformation matrix is achieved in the process of mosaic-plane construction. Also we will present the problem arise from the transformation of trajectories in overlapping region.

Let $\mathrm{C} = \{\mathrm{C}^1, \mathrm{C}^2, \ldots, \mathrm{C}^N\}$ be a set of N partially semi-overlapping cameras in a typical camera-network for large area surveillance shown in Figure 2. For clarity, in this paper, we only focus on the trajectory association between two semi-overlapping cameras, whose solution is easy to extend to the case of multiple cameras.

A. Background Subtraction and Trajectory Extraction

Let C^a and C^b be two overlapping cameras from camera-network. To extract object trajectories from individual camera, we adopt background subtraction followed by a graph-based tracking [17]. Let $\mathrm{T_a} = \{\mathrm{T_a^1}, \mathrm{T_a^2}, \ldots, \mathrm{T_a^n}\}$ represents the resulting n trajectories in local camera $\mathrm{C^a}$, and $\mathrm{T_b} = \{\mathrm{T_b^1}, \mathrm{T_b^2}, \ldots, \mathrm{T_b^m}\}$ represents the resulting m trajectories in local camera $\mathrm{C^b}$. Every trajectory is a set of observations, e.g., $\mathrm{T_a^i} = \{\mathrm{T_a^i}(\mathrm{x_1}, \mathrm{y_1}, \mathrm{t_1}), \mathrm{T_a^i}(\mathrm{x_2}, \mathrm{y_2}, \mathrm{t_2}), \ldots, \mathrm{T_a^i}(\mathrm{x_N}, \mathrm{y_N}, \mathrm{t_N})$, where t represents the time of the observation and N represents the total observation number of the trajectory.

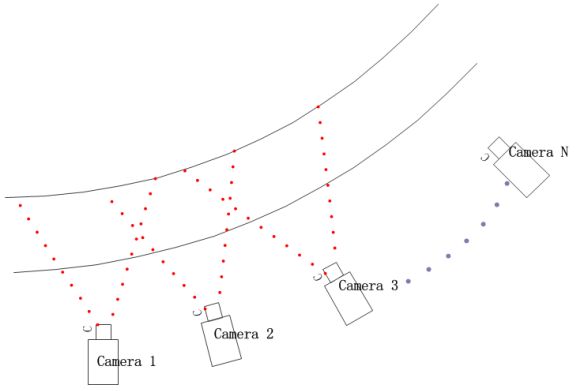

Figure 2. A road with N cameras and their fields of view differences on the mosaic-plane in overlapping region.

B. *Trajectory Projection*

For each camera, we obtain a representative background image by computing a temporal median over the related background images. Based on the representative background images from two cameras, we construct a panorama image [11] as the virtual ground-plane (mosaic-plane) to prepare for trajectory association. During constructing the panorama image, the two homography matrices (H_b^v and H_a^v) are estimated by performing bundle adjustment. The projection from image-plane to mosaic-plane is estimated by the homography matrix, e.g.,

$$\overline{T}_b^i(\overline{x},\overline{y},t) = H_b^v T_b^i(x,y,t), \quad (1)$$

where $\overline{T}_b^i(\overline{x},\overline{y},t)$ is the projection of on mosaic-plane.

However, the projections are prone to cause the corresponding trajectories unaligned in the overlapping region ($V_{a,b}$) on the mosaic-plane (Figure 3(d)). Additionally, because of noise and errors in the trajectory extraction, existing trajectory association algorithms often fail to correctly associate. In next section, we will introduce our robust and efficient graph matching based on reweighted random walk to solve the trajectories association in partially overlapping cameras.

IV. Trajectory Association and Fusion

A. *Trajectory Graph Model*

In the case of partially overlapping cameras, we need to establish the correspondence between the transformed trajectories ($\overline{T}_a^i, \overline{T}_b^j$) on the mosaic-plane. Let G^a and G^b be two graphs standing for camera a and b, respectively. The nodes in each graph are corresponding to trajectories in each camera (Figure 4(a)). Then a bipartite graph based on two graphs was constructed, and the association problem is to find the optimal matching set in the bipartite graph (Figure 4(b)). However, the existing graph matching algorithms, such as adopted in [13], focused on exploiting relatively weak unary and pair-wise attributes and didn't aim at optimizing a well defined objective function.

Different from the graph matching algorithms mentioned above, we construct an association graph G^{rw} from G^a and

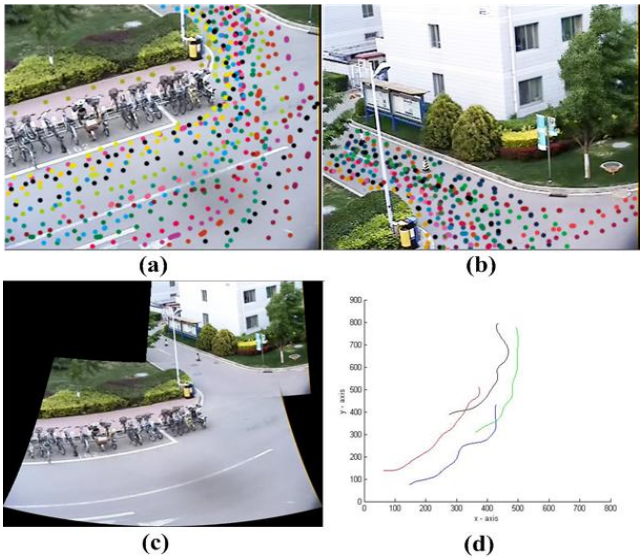

Figure 3. Two partially overlapping cameras. (a) and (b) are accumulated trajectories in each view; (c) mosaic-plane of the cameras; (d) two pairs of trajectories which are different on the mosaic-plane in overlapping region.

G^b as illustrated in Figure 5. Hence, the original graph matching problem between G^a and G^b is equivalent to node ranking or selection on G^{rw}. To select the nodes in G^{rw}, we adopt the statistics of Markov random walks. However, some nodes in G^{rw} correspond to false candidate correspondence. In this case, the normalization of affinity matrix can strengthen the adverse effect of outliers and prevent random walkers. To avoid this case, we adopt a jump during the random walk, i.e., the random walker moves by traversing an

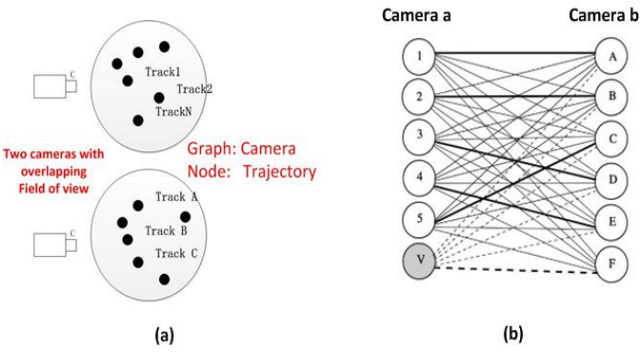

Figure 4. Bipartite graph matching for find corresponding trajectories.

edge with probability α. The probability α represents the bias between the two possible actions, e.g., following edge or jumping. This method can strengthen the effects of reliable nodes in random walks, thus more robust to noise and outliers. The detailed algorithm can be referenced to the algorithm in [10].

In our reweighted random walk based graph matching algorithm for trajectory association, the critical issue is how

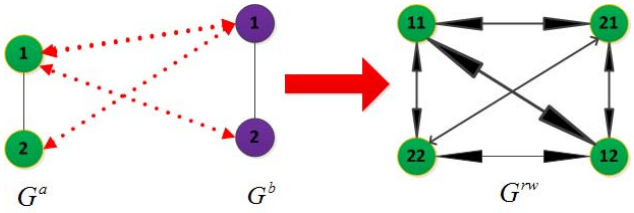

Figure 5. Association Graph G^{rw} from G^a and G^b

to construct the pair-wise affinity matrix W. We combine spatio-temporal information with appearance feature and velocity feature to construct the affinity matrix. The appearance features includes: 1) RGB color histograms are used to model the appearance of a trajectory. In our implementation, we use 8 bins for each channel to form a 24-dimensional histogram; 2) To describe the image texture, we use a descriptor based on covariance matrices of image features proposed in [19]; 3) Shape feature: A 32-dimentional HOG feature f_{HOG_k} is extracted from the observation region. In summary, the appearance descriptor of one observation in the trajectory can be written as $A_a^k=(\{f^a_{RGB_k}\},\{C^a_k\},\{f^a_{HOG_k}\})$ in camera a, where $f^a_{RGB_k}$ is the color feature, C^a_k is the texture feature, and $f^a_{HOG_k}$ is the 32D HOG feature, where k is the subscript of observation. In our design, we average appearance feature on all observations belong to one trajectory. So the final appearance feature is:

$$\beta_a^i = \frac{1}{N}\sum_{k=0}^{N-1} A_a^k, \quad (2)$$

where N is the total observation number in trajectory T_a^i. Additionally, we expand the feature set by including the average object velocity on mosaic-plane for helping in describing the rate of change of the object and calculated as:

$$V_a^i = \frac{1}{N}\sum_{k=0}^{N-1}(\bar{x}_{k+1}-\bar{x}_k, \bar{y}_{k+1}-\bar{y}_k). \quad (3)$$

The combined distance on appearance feature and velocity feature of two trajectories is calculated as:

$$DF_{ai,bj} = \frac{1}{C}\left(\left\|\beta_a^i - \beta_b^j\right\| + \lambda_1\left\|V_a^i - V_b^j\right\|\right), \quad (4)$$

where C is a normalization term, λ_1 is empirically determined coefficient. The spatial distance of trajectory i and trajectory j is obtained by taking the absolute difference between the spatial coordinates of observation in overlapping region on mosaic-plane. Resampling is done in order to have equal length trajectories. The spatial distance is computed as:

$$DS_{ai,bj} = \sum_{\forall k \text{ in } V_{a,b}}\left|\bar{T}_a^i(\bar{x}_k,\bar{y}_k,t_k) - \bar{T}_b^j(\bar{x}_k,\bar{y}_k,t_k)\right|, \quad (5)$$

Combining the two distances computed in Equation (4) and Equation (5), the final distance between two trajectories is:

$$D_{i,j} = \lambda_2 DF_{ai,bj} + \lambda_3 DS_{ai,bj}, \quad (6)$$

where λ_2 and λ_3 are empirically determined coefficients. In our experiments, λ_2 and λ_3 are set to 0.5 to show the best performance.

Let i and p be two trajectories in camera a, j and q be two trajectories in camera b. The node in pair-wise affinity matrix W is computed with above distance as follows:

$$W_{ip,jq} = e^{-\left|D^a_{i,j}-D^b_{p,q}\right|^2/\sigma_s^2}. \quad (7)$$

Let t_{si} and t_{ei} be the starting and ending time of trajectory i. We adopt a temporal threshold T, which is a maximum transition time of objects across the overlapping region between two adjacent cameras. Empirically, we set $T = 25$ in our implementation. If trajectories i and j are close in temporal domain, i.e.,

$$(t_{si} \le t_{sj} \le t_{ei} + T) \text{ or } (t_{sj} \le t_{si} \le t_{ej} + T), \quad (8)$$

this means that trajectory i and trajectory j may be the same object since they are observed by cameras around the same time. If trajectory i and trajectory j, or trajectory p and q trajectory, are not close in temporal domain, then

$$W_{ip,jq} = 0. \quad (9)$$

After constructing the pair-wise affinity matrix, the quasi-stationary distribution of this reweighted random walk is efficiently computed using the power iteration method. In the final discretization step, the Hungarian algorithm is adopted to cover the optimal association sets.

B. *Global Trajectory Fusion*

Once association is finished, the next procedure is to fuse pairs of corresponding trajectories in overlapping regions, and construct a global trajectory across the whole field on mosaic-plane. To fuse two corresponding trajectories $\bar{T}_a^i$ and $\bar{T}_b^j$, which are captured by two cameras ,we adopt an adaptive weighting algorithm:

$$T'_{new} = \begin{cases} \omega_1 \bar{T}_a^i + \omega_2 \bar{T}_b^j, & \text{overlapping region} \\ \bar{T}_a^i, & \text{region in camera a} \\ \bar{T}_b^j, & \text{region in camera b} \end{cases}. \quad (10)$$

The trajectory that has more observations and longer length will be given higher weight than other. The weights are calculated as function of number of observation and length for each trajectory:

$$\omega_1 = \frac{|T_a^i| + L_{T_a^i}}{|T_a^i| + |T_b^i| + L_{T_a^i} + L_{T_b^i}}, \quad \omega_2 = \frac{|T_b^i| + L_{T_b^i}}{|T_a^i| + |T_b^i| + L_{T_a^i} + L_{T_b^i}}, \quad (11)$$

where $|.|$ is the number of observations in the trajectory and L_T is the length of the trajectory. The example of fusion result can be seen in Figure 6. The two pairs of trajectories have been described in Figure 3(d), and the dotted points are the fusion results in overlapping region.

V. EXPERIMENTAL RESULTS

A. *Experimental Setting*

We evaluate the performance of the proposed algorithm on two real world datasets. The first dataset (*D1*) is outdoor road surveillance video sequence, which consists of 12000 frames (images at 25 frames/sec and 704x576 resolution), describing a scene simultaneously recorded by two cameras located at different viewpoints [12]. We split it into two segments, namely Seg_1a and Seg_1b, with equal length. The second dataset (*D2*) is a more complex outdoor pedestrian activity video sequence, which consists of 9000 frames (images at 25 frames/sec and 704x576 resolution), describing a scene simultaneously recorded by two cameras located at different viewpoints. Also, we split it into two segments, namely Seg_2a and Seg_2b, with equal length. In both datasets, the closeness of objects' movement and similarity in color make the association task challenging.

Projecting trajectories onto mosaic-plane with the homography computed will result in discrepancy between two trajectories belong to one object in overlapping region. We expect the robustness and accuracy of our algorithm, when suffering from discrepancy of the trajectories in overlapping region, similarity of object's appearance, and errors in trajectories extraction.

To evaluate the association and fusion results, we adopt two measures, Recall (R) and Precision (P) are calculated as:

$$\begin{cases} \mathrm{R} = \frac{|\mathrm{G} \cap \mathrm{E}|}{|\mathrm{G}|} \\ \mathrm{P} = \frac{|\mathrm{G} \cap \mathrm{E}|}{|\mathrm{E}|} \end{cases}, \tag{12}$$

where G is the ground truth pairs of trajectories, and is E the estimated results.

B. *Experiment Results of Trajectory Matching*

We compare the proposed method with standard Dynamic Time Wrapping (DTW) [16] (*M1*), and two of state-of-the-art approaches presented in [18] (*M2*) and [12] (*M3*), in term of R and P on two datasets. Table 1 shows the detailed performance. And Figure 7 and Figure 8 show the selected complete global trajectories of two videos on individual mosaic-plane.

The results show the proposed approach is better by 22% and 17% for R and P, respectively, compared to *M1*. It reveals that in complex datasets such *D1* and *D2*, where objects are close in time and space, trajectory statistics can promote the association performance tremendously. Compared to *M2*, the proposed approach outperforms 12% and 10% for R and P, respectively. *M2* adopts geometry based approach. It lacks in a procedure to resolve conflict situations. This results in a low R and P scores, especially in a more complex segment in *D2*. Compared to *M3*, the proposed approach outperforms 6% and 5% for R and P respectively. Although *M3* adopts multiple features, it fails to distinguish objects with similar appearance and close spatial-temporal feature in crowed scenario, such as Seg_2a in *D2*. The results on the two real complex datasets show the high performance of our proposed approach.

VI. CONCLUSION

In this paper we presented a framework for global trajectory construction for partially-overlapping camera-network and applied it to large area video surveillance. The proposed framework uses trajectories generated by individual cameras and performs trajectory association on the mosaic-plane by graph matching. The experimental results demonstrate that our method is robust to noise and outliers, and high associating accuracy can be achieved in complex environments.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No. 60833006, 60905008 and 60903146).

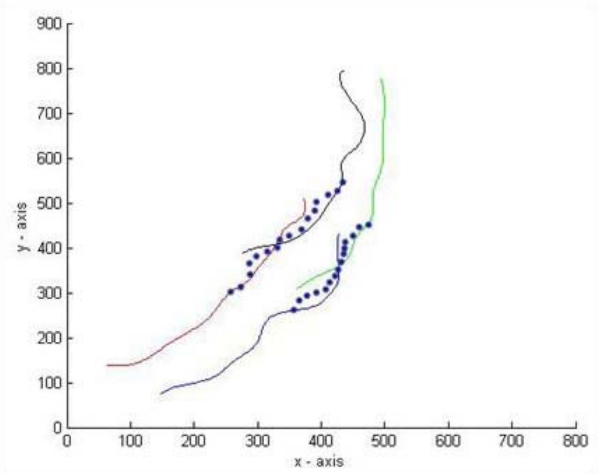

Figure 6. An example of trajectory fusion: dotted line is the fusion part in overlapping region.

Figure 7. Selected trajectory fusion results on dataset 1. Each trajectory is shown in different color.

Figure 8. Selected trajectory fusion results on dataset 2. Each trajectory is shown in different color.

Table 1. Comparison of trajectory association Precision (*P*) and Recall (*R*) on Dataset 1 and Dataset 2.

Methods	*D1*				*D2*			
	Seg_1a		*Seg_1b*		*Seg_2a*		*Seg_2b*	
	R	P	R	P	R	P	R	P
M1	0.78	0.85	0.69	0.79	0.56	0.66	0.70	0.79
M2	0.88	0.92	0.82	0.88	0.78	0.83	0.85	0.90
M3	0.93	0.96	0.87	0.93	0.84	0.89	0.91	0.94
Ours	**1.00**	**1.00**	**0.96**	**0.99**	**0.91**	**0.95**	**0.94**	**0.96**

REFERENCES

[1] K. Nummiaro , E. Koller-Meier, T. Svoboda, D. Roth, and J-L. Van Gool, "Color-based object tracking in multi-camera environments", in DAGM-Symposium, pp. 591-599, 2003.

[2] J. Kang, L Cohen, and G. Medioni, "Tracking people in crowded scenes across multiple cameras", in Proceeding Of Asian Conference on Computer Vision, 2004.

[3] H. Kim, Y. Seo, S. Choi, and K.S. Hong, "Where are the ball and players? soccer game analysis with color-based tracking and image mosaic", in Lecture Notes in Computer Science, Volume 1311, pp. 976-983, 1997.

[4] C.-H. Kuo, C. Huang, and R. Nevatia, "Inter-camera Association of Multi-target Tracks by On-Line Learned Appearance Affinity Models", In European Conference on Computer Vision, Greece, pp. 383-396, 2010.

[5] L.N. Junejo and H. Foroosh, "Trajectory rectification and path modeling for video surveillance", in IEEE 11th International Conference on Computer Vision,Rio de Janeiro, pp. 1-7, 2007.

[6] Y.L. de Meneses, P. Roduit, F. Luisier, and 1. Jacot, "Trajectory analysis for sport and video surveillance", in Electronic Letters on Computer Vision and Image Analysis, vol. 5, no. 3, pp. 148-156, 2005.

[7] A. Ariel, A. H. Murad, "Trajectory fusion for multiple camera tracking", in Computer Recognition Systems, pp.19-26 , 2007.

[8] Y.A. Sheikh and M. Shah, "Trajectory association across multiple airborne cameras", in IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.30, no.2, pp.361-367, 2008.

[9] W. Du, B. Hayet, "Collaborative multi-camera tracking of athletes in team sports", in Workshop on Computer Vision Based Analysis in Sport Environments, pp. 2-13, 2006.

[10] M. Cho, J. Lee, K. M. Lee, "Reweighted Random Walks for Graph Matching", in European Conference on Computer Vision, Greece , pp. 383-396, 2010.

[11] M. Brown, D. G. Lowe, "Recognising Panoramas", in IEEE 9th International Conference on Computer Vision, Nice, France, pp.1218-1225, 2003.

[12] A. Nadeem, C. Andrea, "Trajectory association and fusion across partially overlapping cameras", in Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance(AVSS), Genoa, Italy, 2009.

[13] J. Omar, R. Zeeshan, "Tracking across multiple cameras with disjoint views", IEEE 9th International Conference on Computer Vision, Nice, France, Vol2, pp. 952-957, 2003.

[14] R. Pflugfelder, H. Bischof, "Tracking across non-overlapping views via geometry", in International Conference on Pattern Recognition, Tampa, FL, pp. 1-4, 2008.

[15] J. Hopcroft and R. Karp. "An n2.5 algorithm for maximum match ings in bipartite graphs". SIAM J. Computing, 1973.

[16] E. Keogh, "Exact indexing of dynamic time warping", In Proceedings Of International Conference on Very Large Data Bases, Hong Kong(china), 2002.

[17] M. Taj, E. Maggio, and A. Cavallaro, "Multi-feature graph-based object tracking", in Proceedings of the 1st international evaluation conference on Classification of events, activities and relationships, Southampton, UK, pp. 190-199, 2006.

[18] K. Gabin, A. Nadeem, "Global trajectory reconstruction from distributed visual sensors", in Second ACM/IEEE International Conference on Distributed Smart Cameras (ISDSC), Stanford, pp. 1-8, 2008.

[19] O. Tuzel, F.Porikli, "Region covariance: A fast descriptor for detection and classification", In European Conference on Computer Vision, Heidelberg, pp. 589–600 , 2006.