Multimed Tools Appl (2017) 76:1941–1957
DOI 10.1007/s11042-015-3163-7

Multiple human tracking based on distributed collaborative cameras

Zhaoquan Cai[1] **· Shiyi Hu**[2] **· Yukai Shi**[2] **· Qing Wang**[2] **·**
Dongyu Zhang[2]

Received: 28 February 2015 / Revised: 1 November 2015 / Accepted: 15 December 2015 /
Published online: 5 January 2016

Abstract Due to the horizon limitation of single camera, it is difficult for single camera based multi-object tracking system to track multiple objects accurately. In addition, the possible object occlusion and ambiguous appearances often degrade the performance of single camera based tracking system. In this paper, we propose a new method of multi-object tracking by using multi-camera network. This method can handle many problems in the existing tracking systems, such as partial and total occlusion, ambiguity among objects, time consuming and etc. Experimental results of the prototype of our system on three pedestrian tracking benchmarks demonstrate the effectiveness and practical utility of the proposed method.

Keywords Multi-object tracking · Collaborative cameras · Video surveillance

1 Introduction

Video based multi-object tracking is a challenging and attractive research area in computer vision. Multi-object tracking technology can be widely used in video surveillance, intelligent transportation and intelligent robots. However, due to the horizon limitation of single camera, it is very difficult for single camera-based multi-object tracking system to track the objects all the time. Besides, because of the occlusions and ambiguous appearances among different objects, the performance of single camera-based multi-object tracking system is often degraded.

With the development of multi-object tracking technology, multi-camera network has been proposed to address the issues of single camera based multi-objecting. Multi-camera network can provide different camera views on objects, so that these objects can be tracked well even when they are occluded in some cameras. Besides, multi-camera network can also alleviate the

✉ Dongyu Zhang
cszhangdy@163.com

[1] Huizhou University, Huizhou, China

[2] Sun Yat-sen University, Guangzhou, China

ambiguity among similar objects with different object appearances captured by multiple views. Though much progress has been made in multi-camera based tracking, the current multi-object tracking system usually suffers from two main issues: unreliable detection results and time consuming [2, 4, 13, 17, 18, 20, 21, 36, 40]. Therefore, how to take advantage of multi-camera network to track objects robustly and efficiently has become a hot topic.

In this paper, we propose and implement a tracking-by-detection based system for multiple human tracking in multi-camera network. As shown in Fig. 1, there are mainly two stages in our method, i.e., single camera tracking, and multi-camera collaborative tracking. In the single camera tracking stage, we adopt an efficient pedestrian detection method, in which the foreground connected regions are extracted through codebook background modeling algorithm and then more accurate detection on the foreground connected regions is achieved by using the histogram of oriented gradients (HOG) pedestrian detection algorithm. With the results of pedestrian detection, tracking is firstly performed on each single camera, in which a particle filter is used for multi-object tracking and Hungarian algorithm is employed for data-association. In the multi-camera collaborative tracking stage, by using the tracking results of each single camera, a multi-camera view object correspondence scheme is proposed to track objects in multiple camera views. In order to improve the system efficiency, we adopt a multi-thread strategy in the implementation of the proposed method. Because of the complexity of multi-camera calibration process, traditional camera calibration based multi-camera tracking approach is hard to use in practice. In this paper, we employ a pure image approach for multi-camera fusion and collaborative tracking, which is easy to be implemented and effective in practice.

The contributions of this paper are mainly three-fold: First, we propose a new method and implement a prototype system of multi-object tracking by using multi-camera network. The proposed system can handle partial and total occlusion issues, as well as alleviate ambiguity among objects in current multi-object tracking system; Second, we present an efficient pedestrian detection method to improve the practicality of tracking system; Third, we create a new dataset for facilitating the research of multi-object tracking in multi-camera network.

The rest of this paper is organized as follows. In Section 2, we briefly present the related works about multi-object tracking. In Section 3, we describe the proposed multi-camera collaborative multi-object tracking system. The experimental results are provided in Section 4. Finally, Section 5 concludes this paper.

2 Related work

With the progress of detection algorithms [10, 12, 23–25, 27, 28, 31–33], many detection-based tracking methods have been proposed for multi-object tracking [1, 5, 9, 38]. The

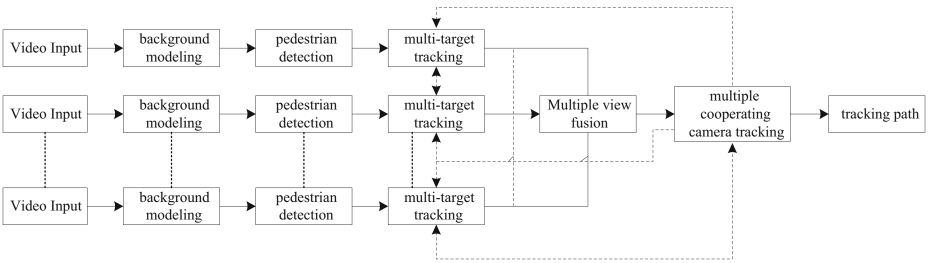


Fig. 1 The Framework of proposed multi-camera collaborative tracking method

detection-based tracking method adopts detection algorithm to locate the interested objects in the video sequences, and associates the detection results of successive frames to form the trajectory [29, 30]. However, the detection results of detector are sparse and unreliable as the detector only provides a set of discrete detection reaction. Therefore, the false positives and miss detection may often occur, especially when objects suffering from long-term occlusion. In order to solve these issues, different approaches have been proposed. For example, Michael D. Breitenstein et al. proposed an algorithm for robust multi-object tracking by combining the particle filter and the confidence of the detector [6]. C. Huang et al. proposed an approach to track multi-object by using the hierarchical association of detection responses [15]. W. Brendel et al. treated the data association problem as the max weight independent set problem, and linked the tracklets into trajectory iteratively [8]. Besides, several global optimal approaches using network flow [3, 41] or continuous energy minimization [34] have also been proposed to solve the multi-object tracking problem [26, 39].

Recently, multi-camera based object tracking has been proposed to overcome the disadvantage of single camera based object tracking, such as the horizon limitation of camera, occlusion and ambiguous appearances among different object. Orwell et al. proposed a color based approach for multi-camera based tracking [36]. F. Fleuret and J. Berclaz adopted probabilistic occupancy map model and K-shortest path algorithm to track multi-objects in multi-view environment [2, 13]. Recently, S.M. Khan and M. Shah developed a pure image based approach which combined multi-layer planar homography and Overlap Field of View constraint for multi-view tracking [20, 21]. Yun et al. proposed to [40] combine planar homography and epipolar line constraint for multi-view tracking. In [4, 17], the authors proposed to track multi-object with single camera using particle filter and then use the epipolar geometry for consistent and collaborative tracking.

3 Multi-camera collaborative tracking method

In this section, we describe the proposed multi-camera collaborative tracking method for human tracking. As shown in Fig. 1, there are mainly two stages in the proposed method, i.e., single camera tracking, and multi-camera collaborative tracking. In the following, we will describe the two stages respectively.

3.1 Single camera tracking

In the single camera tracking stage, our system will initialize a new tracker if an object has subsequent detections with overlapping bounding box which is neither occluded nor associated to any existing tracker in all the camera views. If a tracker survives for a limited number of frames without being associated with detection in any views, it will be terminated automatically. The single camera tracking can be roughly divide into two steps, i.e., detection and data association.

(a) Detection. In the proposed system, objects are initialized by the automatic detection method. In order to overcome the speed bottleneck in the detection step, we adopt the Codebook + HOG approach for pedestrian detection. For each frame of the video, we first employ the Codebook background subtraction algorithm [22] to locate the foreground regions, and then adopt HOG algorithm [10] to detect pedestrians in the foreground regions.

(b) Data Association. In the data association phase, we use the Hungarian algorithm for optimal data association. The progress is done as follows: For detections at time t of camera i, we first use particle filters to predict the states of the tracked objects. Then by defining the gate function and likelihood function, the distance matrix of tracked object regions and the corresponding detection bounding box could be calculated, which are used as the input of Hungarian algorithm for the optimal data association assignment. In the following, we will describe this progress in detail.

In our system, the distribution of each object's state is estimated by a particle filter [7]. Given a state $S=(x,y,u,v)$, where (x, y) denotes the position in the 2D image, and (u, v) denotes the velocity along x and y direction, we use the following motion model to propagate the particles.

$$(x,y)_t = (x,y)_{t-1} + (u,v)_{t-1} + \varepsilon_{(x,y)}, \tag{1}$$

$$(u,v)_t = \alpha*(u,v)_{t-1} + (1-\alpha)*(u_T, v_T) + \varepsilon_{(u,v)}, \tag{2}$$

where $\varepsilon_{(x,y)}$ and $\varepsilon_{(u,v)}$ are Gaussian noise, and u_T, v_T are the object's velocity. Equation (2) is used to handle the case that the object accelerates suddenly. We predict an object's state by the particles ranked top 25 % by weights:

$$S_{predict} = S_i * w_i, \tag{3}$$

where $i = 1, \ldots, \frac{N}{4}$. Besides, in order to prevent particles' degradation, a resample process also must be done.

For the tracked object region T and detection bounding box d, the distance of them can be defined by the matching score as:

$$D(T,d) = 1 - Gate(T,d) * L_{single}, \tag{4}$$

where $Gate$(T, d) and L_{single} denotes the gate function [17] and the likelihood function [35] of single camera i that are defined in Eqs. (5) and (6) respectively.

$$Gate(T,d) = \begin{cases} 1, & if\,overlap(T,d) \geq \sigma\,and\,M(T,d) \leq \delta(T) \\ 0, & otherwise \end{cases}, \tag{5}$$

where $overlap(T,d) = \frac{T \cap d}{|T| \cup |d|}$, $\delta(T) = \gamma*[h(T) + w(T)]$, and $h(T)$, $w(T)$ indicate the height and width of region T respectively. For matched regions T and d, in order to calculate their distance, we define the likelihood function by using the Bhattacharyya distance of HSV histogram [35] and a normal distribution evaluation of the distance between position of detection d and a particle p [7] as,

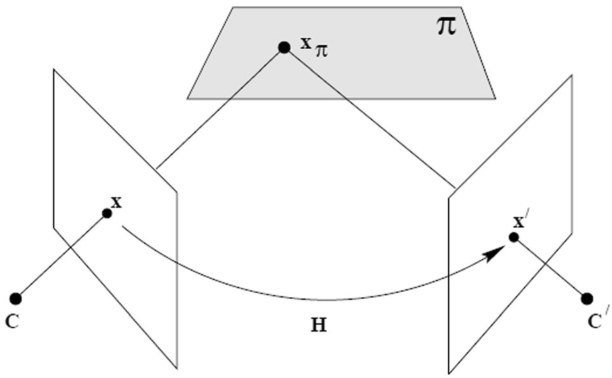
$$L_{single} = \alpha * L_{hsv} + (1-\alpha) * \gamma * \sum\nolimits_{p \in T}^{N} p_N(d-p), \tag{6}$$

$$L_{hsv}(d,T) = e^{-\lambda * BhattaDist(H_d, H_T)}, \tag{7}$$

$$p_N(d-p) \sim N\left(pos_d - pos_p, 0, \sigma^2\right). \tag{8}$$

In, Eq. (5), σ, γ ($\gamma<1$) are two parameters. Generally, small σ will lead to big ambiguous when two or more objects coupled together, while big σ cannot handle large motions of the

Fig. 2 Illustration of planar homography. One can see that the points between two image plane can be mapped one to one by a homography matrix H



tracked objects. *Gate*(*T*,*d*)=1 denotes that *T* and *d* are match, and vice versa. In this paper, we empirically set σ as 0.7.

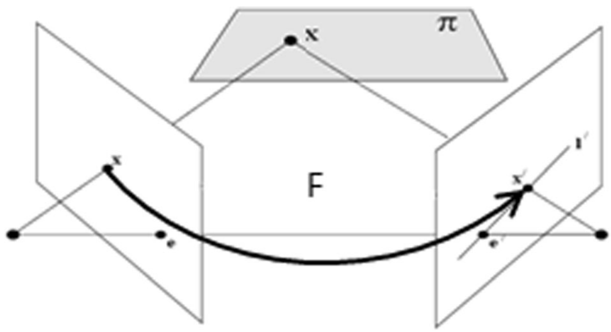
Algorithm1. Single Camera Tracking

Input: Detections of camera c at time t $D_c^t=\{d_1^t,\ldots,d_N^t\}$,
Trajectories of time t-1 $\mathbf{T}^{t-1}=\{T_1^{t-1},\ldots,T_L^{t-1}\}$

1: for each $T^{t-1}\in\mathbf{T}^{t-1}$ do
2: Predict the trajectories' state by particle filter
3: end for
4: $\mathbf{T}^t$ = Optimal_Assign_by_Hungarian_Algorithm ($\mathbf{T}^{t-1}$, D_c^t)
5: for $T^t\in\mathbf{T}^t$ do
6: if detection d and trajectory T can associate
7: Update object's state
8: else
9: Multi-camera collaborative tracking
10: end if
11: end for

Output: $\mathbf{T}^t$

Fig. 3 Illustration of epipolar constraint. One can see that an arbitrary point in one image plane can be constrained on an epipolar line in another image plane by a fundamental matrix F



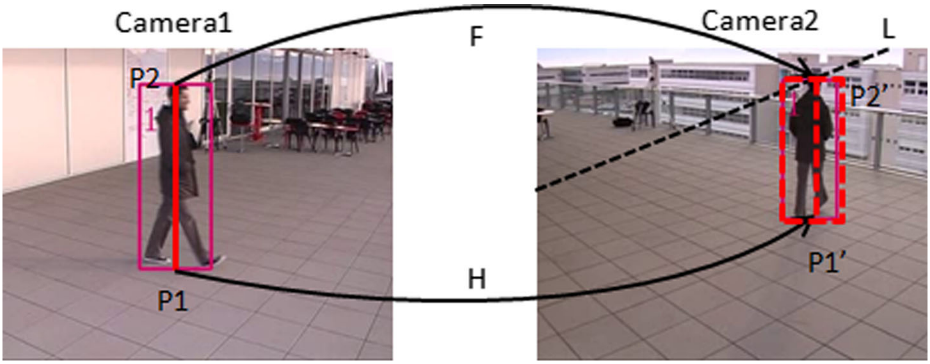


Fig. 4 Illustration of mapping objects across different views

Finally, the distance matrix of tracked object regions T and the corresponding detection d bounding box can be calculated as:

$$dist = \begin{bmatrix} dist_{1i} & \dots & dist_{1N} \\ \dots & \dots & \dots \\ dist_{Mi} & \dots & dist_{MN} \end{bmatrix}, \tag{9}$$

where M and N are the object numbers in adjacent frames. The distance matrix will be employed as the input of the Hungarian algorithm for optimal assignment.

We also define an observed model for guiding the particle propagation

$$w_{T,p} = \alpha{*}I(T){*}p_N(p{-}d) + \beta{*}L_{hsv}(p,T), \tag{10}$$

where $I(T)$ is an indicate function. If $I(T)=0$, the detection d and the tracker T cannot be associated in the data association phase, and we compute the weight for a particle just by the

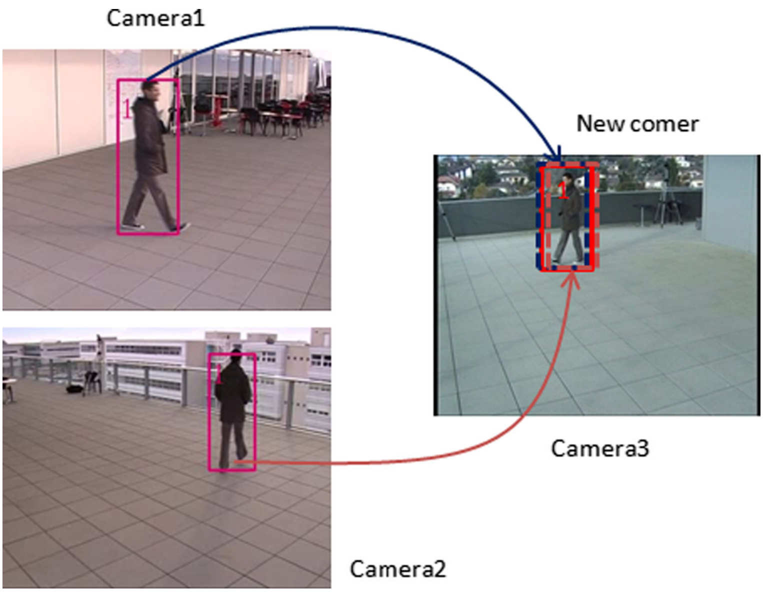


Fig. 5 Illustration of object correspondence across different views

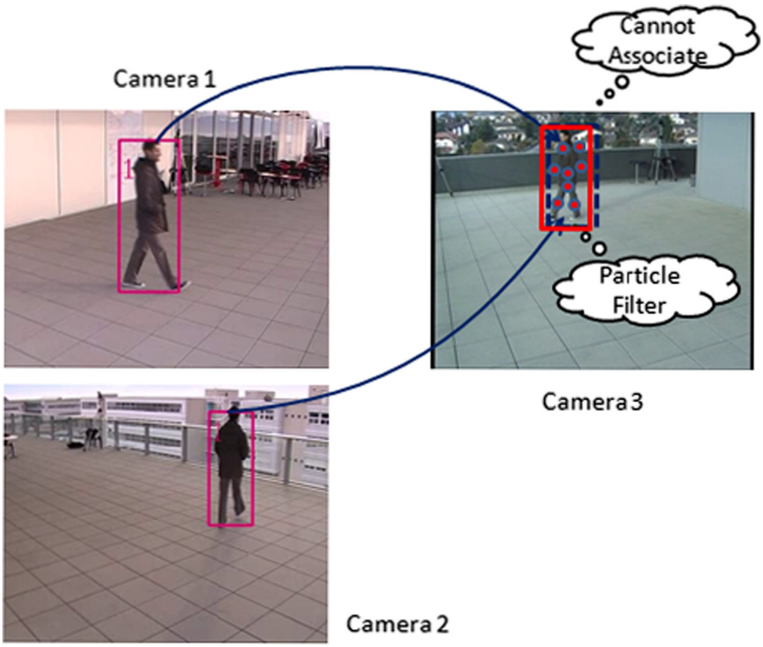


Fig. 6 One typical sample of collaborative tracking. The tracked objects in camera 1 and 2 can be employed to track objects that are failed to associate detections in camera 3

Bhattacharyya distance of HSV histogram. If $I(T)=1$, a normal distribution evaluation for the distance between the position of detection d and a particle p is added to guide the particles to move to the most possible area. The single camera

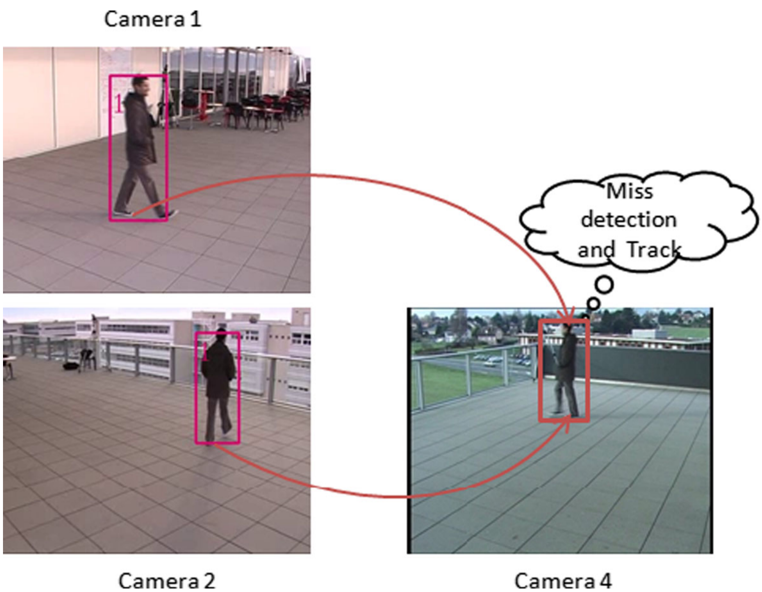


Fig. 7 Another typical sample of collaborative tracking. The tracked objects in camera 1 and 2 can be employed to track objects that are miss detections and tracks in camera 4

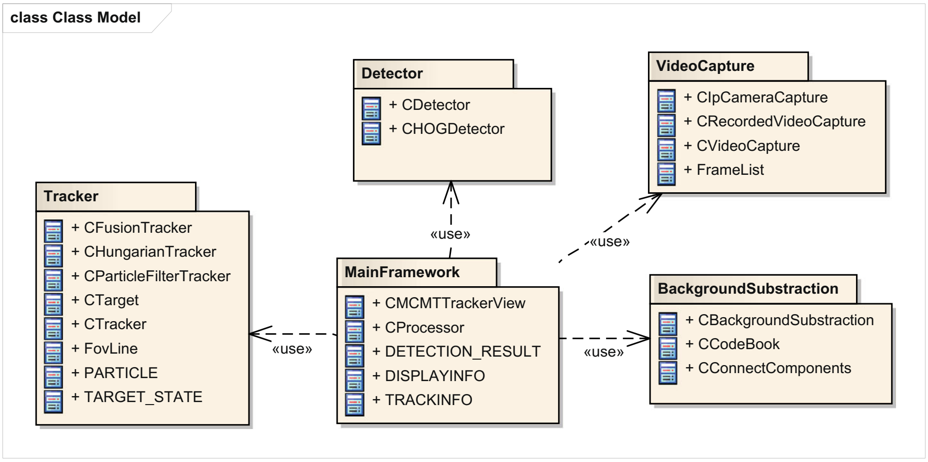


Fig. 8 Fundamental modules of our prototype system

3.2 Multi-camera collaborative tracking

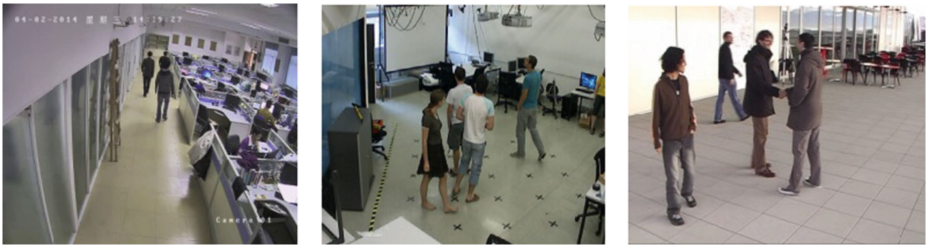
Multi-camera collaborative tracking is the second stage of the proposed method. The input of this stage is from the tracking results of each single camera in the former stage. By fusion the different camera views and tracking results of single camera, multi-camera collaborative tracking can handle the camera view limitation problems, as well as the object occlusion and ambiguity among similar objects problems in single camera tracking. In multi-camera collaborative tracking, the key issues are the object mapping and correspondence in different camera views. In the following, we first present some basic knowledge in multi-view geometry, and then provide the object mapping, object correspondence method and the multi-camera collaborative tracking.

(a) *Multi-view geometry.*

Figure 2 shows an illustration of point projection in multi-view geometry [37]. The projections of point x_π on camera C and camera C' are x and x' respectively. Point x and x' satisfy the planar homography, which means that $x' = \mathrm{H}_{2\pi}\mathrm{H}_{1\pi}^{-1}x = \mathrm{H}x$, and H is a 3×3 matrix which can be calculated by at least 4 correspondence points.

Table 1 The fundamental modules and corresponding function of our prototype system

Module name	Module function
Main framework module	Schedule module
Video capture module	Obtain the input video from video files or IP camera video streams
Background subtraction module	Builds the background model of the scene and extracts the foreground regions
Detector module	Distinguish pedestrians and locate their positions and sizes
Tracker module	For single-camera tracking and multi-camera collaborative tracking,



(a) (b) (c)

Fig. 9 Sample frames of **a** our group dataset, **b** ICG dataset, **c** EPFL dataset

Furthermore, point x and x' also satisfy the epipolar constraint. As Fig. 3 shows, the point x in camera C can be mapped to a line in C' by the 3×3 fundamental matrix F [37], and the correspondence point x' must lie in the line l'

$$l' = [e']_x \mathrm{H}_\pi = \mathrm{F}x. \tag{11}$$

where the fundamental matrix F can be calculate by at least 7 correspondence points.

(b) *Object Mapping*

We combine the ground plane homography and epipolar constraint for object mapping in different cameras. Figure 4 illustrates the mapping progress. We use the ground plane homography matrix H to project the point P_1 in Camera 1 to P_1' in Camera 2, and use the fundamental matrix F to project the point P_2 in Camera 1 to L in Camera 2. Then we can obtain the point P_2' by the intersection of the line L and the vertical line pass point P_1'. Finally we can estimate the object's location and size in different views. Furthermore, in order to make sure the mapping is valid, a constraint of overlapping view [20] is also used in our system.

(c) *Object Correspondence across different views*

Object correspondence across different views is a key issue in multi-camera collaborative tracking. Figure 5 shows the object correspondence process in our system. Once there is a new comer in one camera view, we first check whether it has been tracked in other views by performing the object mapping, and calculating the overlapping ratio of the new comer and the other tracking objects. By a certain threshold, we could decide whether give the newer comer a new ID or assign it with the ID of the most similar existing object according to their overlapping ratio.

(d) *Collaborative Tracking*

Collaborative tracking can be used to tackle the situation with no associative detections. As shown in Fig. 6, there is no associated detections with a tracked object in camera 3. We can map the correspondent object from other views into view 3 and estimate the object's position and size. After that, we can add particles and propagate them in this area, and finally use particle filter to predict object's state.

Table 2 Comparison result of detection evaluation on our Lab dataset

Approach	DR	FPR	MDR	FPS
HOG	0.63	0.01	0.37	4.35
HOG + Codebook	0.60	0.01	0.40	10.70

Table 3 The comparison of Ours-Multiple with Ours-Single in Lab sequence camera 1

Approach	MOTA	MT	FP	ID-Switch
Ours-Single	0.34	952	23	19
Ours-Multiple	0.64	439	75	27

Collaborative tracking can also solve the situation that the detector has no response in some regions in a view for a long time. As illustrated in Fig. 7, the object cannot be detected and tracked the man in camera 4 for a long time. But the man can be correctly detected and tracked in Camera 1 and Camera 2 respectively. In this situation, we can directly map and estimate position and size of the object in camera 4 by combining the data from camera 1 and camera 2. After that, we can add a new object to the single camera tracking phase of camera 4.

4 Experimental result

Utilizing the OpenCV library for related image and video processing, we implement a prototype system of proposed multi-camera collaborative tracking method on the platform of C++. Figure 8 shows the fundamental modules of the prototype, and their functions are listed in Table 1.

To compressively evaluate the proposed multi-camera collaborative tracking method, we run the prototype system on three different video datasets, including ICG dataset [16], EPFL dataset [11], and our Lab dataset. Figure 9 shows same sample frames from the videos of the three datasets respectively. All the experiments were carried on a desktop with an Intel i3 3.0 GHz CPU and 4 GB RAM. The computational speed of our system achieves about 7 FPS.

4.1 Detection performance evaluation

In the single camera tracking stage of the proposed method, we adopt the HOG + codebook method to speed up the detection progress. Thus, we first evaluate the performance of this method by comparing its detection results and the detection results of HOG on our Lab dataset (including 944 frames). The comparison results including detection rate (DR), false positive rate (FPR), miss detection rate (MDR) and frames per second (FPS), are listed in Table 2. From Table 2, we can see that, the FPRs of the two method are same. Although the DR and MDR of HOG + codebook are slightly worse than that of HOG, but the FPS of HOG + codebook is 10.47, which is much better than that of HOG. This means that compared with HOG algorithm, our method, i.e., HOG + codebook, can speed up the detection progress more than two times

Table 4 The comparison of Ours-Multiple with Ours-Single and Two-Granularity in ICG sequence camera 1

Approach	MOTA	MT	FP	ID-Switch
Two-Granularity [14]	0.79	576	10	15
Ours-Single	0.39	145	7	12
Ours-Multiple	0.94	67	62	41

Table 5 The comparison of Ours-Multiple with Ours-Single and Two-Granularity in EPFL terrace sequence camera 2

Approach	MOTA	MT	FP	ID-Switch
Two-Granularity [14]	0.67	749	2	14
Ours-Single	0.35	224	6	12
Ours-Multiple	0.77	467	64	15

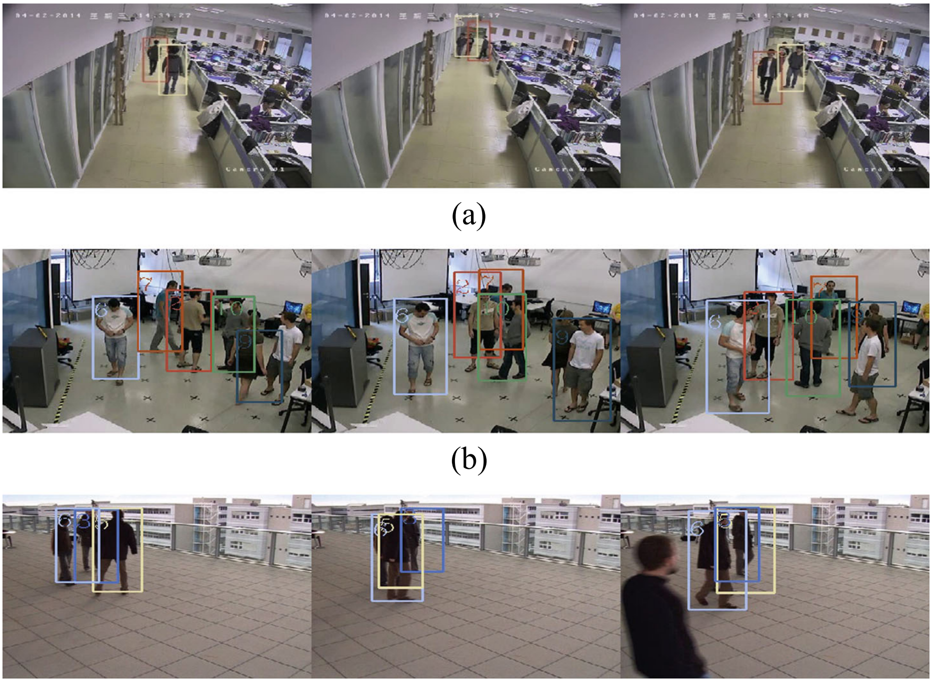
4.2 Tracking performance evaluation

In the tracking performance evaluation, we adopt the metric MOTA [19] to evaluate the tracking results on our Lab sequence (944 frames), ICG hard sequence (first 1500 frames), EPFL Terrace sequence (first 1140 frames). The MOTA metric is defined as,

$$MOTA = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t}, \qquad (12)$$

where m_t, fp_t, and mme_t are the number of misses, false positives, and mismatches at time t, respectively.

We first compare the tracking results of the proposed multi-camera collaborative tracking method (Ours-Multiple) with single camera based multi-object tracking method, including the single tracker of our method (denoted as Ours-Single) described in single camera tracking stage, and Two-Granularity [14]. The comparison results are listed in Tables 3, 4, and 5



(a) (b) (c)

Fig. 10 Sample tracking results of Ours-Multiple on: **a** our Lab sequence, **b** ICG hard sequence, **c** EPFL Terrace sequence

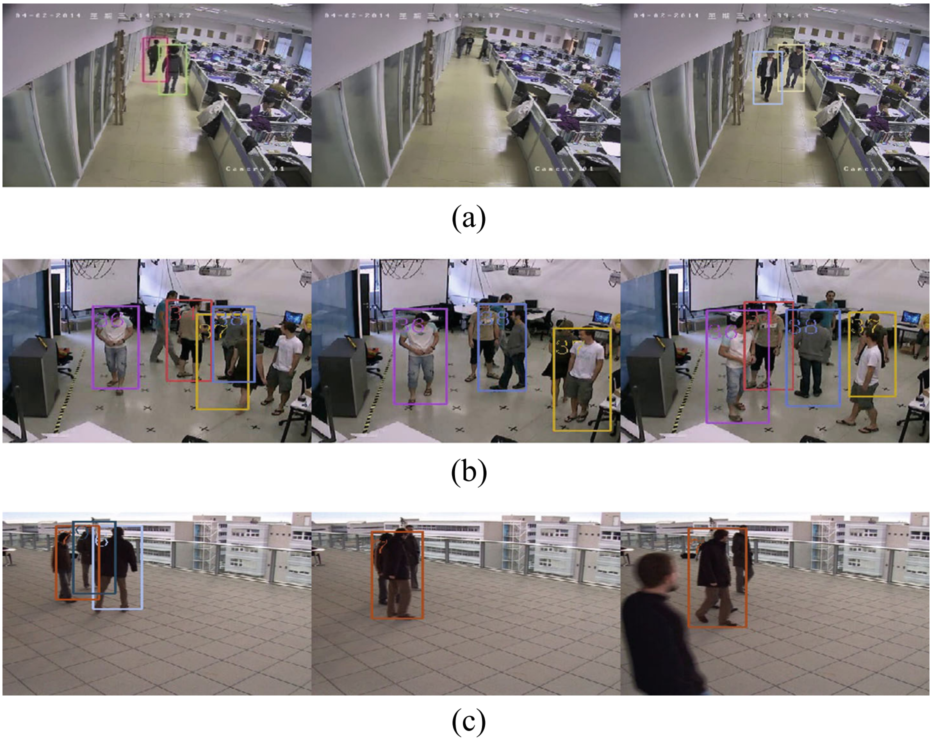


Fig. 11 Sample tracking results of Ours-single on: **a** our Lab sequence, **b** ICG hard sequence, **c** EPFL Terrace sequence

respectively. Besides, Figs. 10, 11, and 12 show same samples of tracking result of comparison methods on the three dataset respectively.

From Tables 3, 4, and 5, we can find out that, the tracking results of Ours-multiple are better than Ours-single in term of MOTA in all three dataset. Besides, the tracking results of Ours-

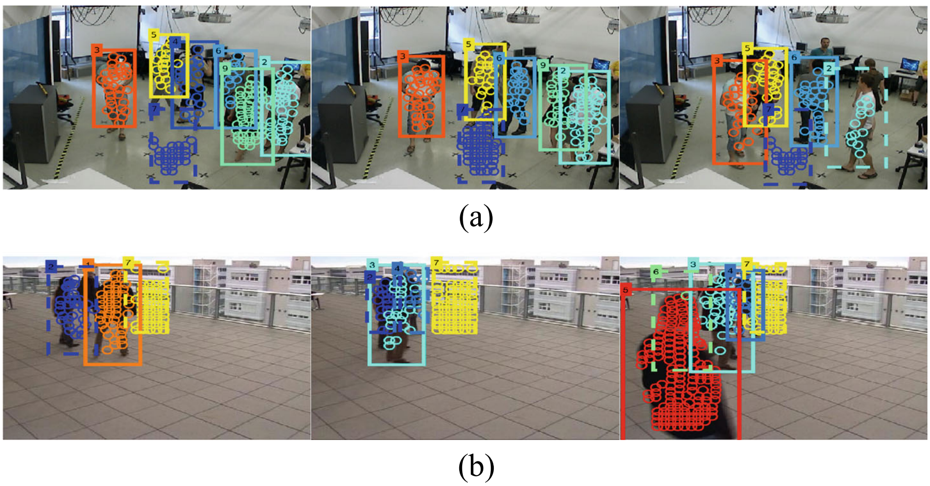


Fig. 12 Sample tracking results of Two-Granularity on: **a** ICG hard sequence, **b** EPFL Terrace sequence

Table 6 The comparison of Ours-Multiple with other three multi-camera multi-object methods in EPFL passageway sequence

Approach	MOTA	FN(rate)	FP(rate)
POM [13]	0.45	0.32	0.05
POM + KSP [2]	0.38	0.58	0.05
CNOV [18]	0.40	0.54	0.03
Ours	0.48	0.50	0.01

multiple are also better than that of Two-Granularity. These results demonstrate that compared with single camera tracking, the collaboration of multi-camera in the proposed method could indeed promote the performance of multi-object tracking.

To further evaluate the performance proposed method, we compare our method with other multi-camera-based multi-object tracking method. However, to the best of our knowledge, none of multi-camera multi-object methods release the complete code online. Thus, we compare our method with three multi-camera multi-object tracking methods, i.e., POM, POM + KSP, and CNOV, by using the evaluation results on EPFL passageway sequence presented in [18]. We also present our variant tracker (denoted as Ours-Single), which leverages single camera to track multiple objects, to further demonstrate the effectiveness of the proposed multi-camera collaborative tracking method. The comparison results on our Lab dataset, ICG and EPFL are presented in Table 6. We can see that in terms of MOTA and FP, Ours-Multiple outperforms POM, POM + KSP, and CNOV, which demonstrates the effectiveness of the proposed method in multi-object tracking.

5 Conclusion

Due to the horizon limitation of a single camera, it is difficult to track multiple objects for single camera based multi-object tracking system. In addition, the possible object occlusion and ambiguous appearances often degrade the performance of tracking system. In this paper, we propose a multi-camera collaborative based multi-object tracking method. The proposed method adopt the HOG + Codebook strategy for object detection, which can greatly speed up the detection progress in the single camera tracking stage. Besides, by fusing the tracking results of each single camera with the object correspondence model, the proposed method can solve the view limitation, occlusion and ambiguity appearance problems in the single camera-based tracking. We implemented a prototype system of the proposed method on the C++ platform. The experimental results of the prototype system on three datasets demonstrated that the proposed method is very effective for the pedestrian tracking.

Acknowledgments This work is partially supported by National High Technology Research and Development Program of China (863 program) under Contract No. 2013AA013801, the NSFC funds of China under Contract No. 61370185, 61170193, 61401125, the Natural Science Foundation of Guangdong Province, China, under Contract No. S2013010013432, S2012020011081, S2013040012570, and the Science and Technology Plane of Guangdong Province, under Contract No. 2013B010406005.

References

1. Andriluka M, Roth S, Schiele B (2008) People-tracking-by-detection and people-detection-by-tracking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Anchorage, USA, pp 1-10

2. Berclaz J, Fleuret F, Turetken E (2011) Multiple object tracking using k-shortest paths optimization. IEEE Trans Pattern Anal Mach Intell 33(9):1806–1819
3. Pascal Fua (2013) Multi-commodity network flow for tracking multiple people. IEEE Trans Pattern Anal Mach Intell 36(8):1614–1627
4. Bredereck M, Jiang X, Korner M, Denzler J (2012) Data association for multi-object tracking-by-detection in multi-camera networks. In: Proceedings of the 6th ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC), Hong Kong, pp 1–6
5. Breitenstein MD, Reichlin F, Leibe B, Koller-Meier E, Van Gool L (2009) Markovian tracking-by-detection from a single, uncalibrated camera. In: IEEE CVPR Workshop on Performance Evaluation of Tracking and Surveillance (CVPR Workshop), pp 1–10
6. Breitenstein MD, Reichlin F, Leibe B, Koller-Meier E, Van Gool L (2009) Robust tracking-by-detection using a detector confidence particle filter. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), Kyoto, pp 1515–1522
7. Breitenstein MD, Reichlin F, Leibe B (2011) Online multiperson tracking-by-detection from a single, uncalibrated camera. IEEE Trans Pattern Anal Mach Intell 33(9):1820–1833
8. Brendel W, Amer M, Todorovic S (2011) Multiobject tracking as maximum weight independent set. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Kyoto, pp 1515–1522
9. Choi W, Savarese S (2010) Multiple target tracking in world coordinate with single, minimally calibrated camera. In: Proceedings of the European Conference on Computer Vision (ECCV), pp 553–567
10. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, pp 886–893
11. "EPFL" data set: Multi-camera Pedestrian Videos http://cvlab.epfl.ch/data/pom/
12. Felzenszwalb PF, Girshick RB, McAllester D (2010) Object detection with discriminatively trained part-based models. IEEE Trans Pattern Anal Mach Intell 32(9):1627–1645
13. Fleuret F, Berclaz J, Lengagne R (2008) Multicamera people tracking with a probabilistic occupancy map. IEEE Trans Pattern Anal Mach Intell 30(2):267–282
14. Fragkiadaki K, Zhang W, Zhang G, Shi J (2012) Two-granularity tracking: mediating trajectory and detection graphs for tracking under occlusions. In: Proceedings of the European Conference on Computer Vision (ECCV), Firenze, Italy, pp 552–565
15. Huang C, Wu B, Nevatia R (2008) Robust object tracking by hierarchical association of detection responses. In: Proceedings of the European Conference on Computer Vision (ECCV), Marseille-France, pp 788–801
16. ICG Lab 6 (Multi-Camera Multi-Object Tracking) http://lrs.icg.tugraz.at/download.php
17. Jiang X, Rodner E, Denzler J (2012) Multi-Person tracking-by-detection based on calibrated multi-camera systems. In: Proceedings of the International Conference on Computer Vision and Graphics (ICCVG), Warsaw, Poland, pp 743–751
18. Jiuqing W, Achuan L (2015) Multiple people tracking using camera networks with overlapping views. Int J Distrib Sens Netw. doi:10.1155/2015/591067
19. Keni B, Rainer S (2008) Evaluating multiple object tracking performance: the CLEAR MOT metrics. EURASIP J Image Video Process. doi:10.1155/2008/246309
20. Khan SM, Shah M (2006) A multi-view approach to tracking people in crowded scenes using a planar homography constraint. In: Proceedings of the European Conference on Computer Vision (ECCV), pp 133-146
21. Khan SM, Shah M (2009) Tracking multiple occluding people by localizing on multiple scene planes. IEEE Trans Pattern Anal Mach Intell 31(3):505–519
22. Kim K, Chalidabhongse TH, Harwood D (2005) Real-time foreground–background segmentation using codebook model. Real-Time Imaging 11(3):172–185
23. Kuhn HW (1955) The Hungarian method for the assignment problem. Nav Res Logist Q 2(1–2):83–97
24. Leibe B, Leonardis A, Schiele B (2008) Robust object detection with interleaved categorization and segmentation. Int J Comput Vis 77(1–3):259–289
25. Lin L, Zhu S-C, Wang Y (2007) Layered graph match with graph editing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 1-8
26. Lin L, Wu T, Porway J, Xu Z (2009) A stochastic graph grammar for compositional object representation and recognition. Pattern Recogn 42(7):1297–1307
27. Lin L, Lu Y, Pan Y, Chen X (2012) Integrating graph partitioning and matching for trajectory analysis in video surveillance. IEEE Trans Image Process 21(12):4844–4857
28. Lin L, Zeng K, Wang Y, Xu Y-Q, Zhu S-C (2013) Video stylization: painterly rendering and optimization with content extraction. IEEE Trans Circuits Syst Video Technol 23(4):577–590
29. Lin L, Xu Y, Liang X, Lai JH (2014) Complex background subtraction by pursuing dynamic spatio-temporal models. IEEE Trans Image Process 23(7):3191–3202
30. Lin L, Wang X, Yang W, Lai JH (2015) Discriminatively trained and-or graph models for object shape detection. IEEE Trans Pattern Anal Mach Intell 37(5):959–972

31. Lin L, Cu Y, Li C, Cheng H, Zuo W (2015) Detection-free Multi-object Tracking by Reconfigurable Inference with Bundle Representations. IEEE Trans Cybern. doi:10.1109/TCYB.2015.2478515
32. Liu X, Lin L, Jin H, Yan S, Tao W (2011) Integrating spatio-temporal context with multiview representation for object recognition in visual surveillance. IEEE Trans Circuits Syst Video Technol 21(4):393–407
33. Liu N, Wu H, Lin L (2015) Hierarchical ensemble of background models for PTZ-based video surveillance. IEEE Trans Cybern 45(1):89–102
34. Milan A, Roth S, Schindler K (2014) Continuous energy minimization for multi-target tracking. IEEE Trans Pattern Anal Mach Intell 36(1):58–72
35. Okuma K, Taleghani A, De Freitas N (2004) A boosted particle filter: multitarget detection and tracking. In: Proceedings of the European Conference on Computer Vision (ECCV), pp 28–38
36. Orwell J, Massey S, Remagnino P (1999) A multi-agent framework for visual surveillance. In: Proceedings of the IEEE Conference on Image Analysis and Processing, Venice, pp 1104–1107
37. Richard H, Zisserman A (2004) Multiple view geometry in computer vision, 2nd edn. Cambridge University Press, Cambridge
38. Schindler K, Van Gool L, Leibe B (2007) Coupled detection and trajectory estimation for multi-object tracking. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), Rio de Janeiro, pp 1–8
39. Wang K, Lin L, Lu J, Li C, Shi K (2015) PISA: pixelwise image saliency by aggregating complementary appearance contrast measures with edge-preserving coherence. IEEE Trans Image Process 24(10):3019–3033
40. Yun Y, Gu IYH, Aghajan H (2012) Maximum-likelihood object tracking from multi-view video by combining homography and epipolar constraints. In: Proceedings of the Sixth International Conference on Distributed Smart Cameras (ICDSC), Hong Kong, pp 1–6
41. Zhang L, Li Y, Nevatia R (2008) Global data association for multi-object tracking using network flows. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Anchorage, AK, pp 1–8

Zhaoquan Cai is a professor of computer science at Huizhou University, China. His research interests include computer networks, intelligent computing, and database systems. Cai received his PhD in computer science from Huazhong University of Science and Technology.

Shiyi Hu is a master student majoring in computer science at the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, P.R. China. His research interests include computer vision and machine learning. He received his BE degrees from Heilongjiang University.

Yukai Shi is a master student majoring in software engineering at the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, P.R. China. His research interests include computer vision and machine learning. He received his BE degrees from Heilongjiang University.

Qing Wang Ph.D. & Associate professor of School of Software, Sun Yat-sen University, experienced researcher on Human Computer Interaction, User Experience, Collaborative Software, and Interoperability of Enterprise Software and Application, and especially interested in the Human Information Interaction.

Dongyu Zhang received his B.S. degree and Ph.D. in Harbin Institute of Technology University, Heilongjiang, China. His research interests include computer vision and machine learning.