PAPER

Multiple-Object Tracking in Large-Scale Scene

Wenbo YUAN[†], **Zhiqiang CAO**[†a)], **Min TAN**[†], ***Nonmembers***, ***and*** **Hongkai CHEN**[†], ***Student Member***

SUMMARY In this paper, a multiple-object tracking approach in large-scale scene is proposed based on visual sensor network. Firstly, the object detection is carried out by extracting the HOG features. Then, object tracking is performed based on an improved particle filter method. On the one hand, a kind of temporal and spatial dynamic model is designed to improve the tracking precision. On the other hand, the cumulative error generated from evaluating particles is eliminated through an appearance model. In addition, losses of the tracking will be incurred for several reasons, such as occlusion, scene switching and leaving. When the object is in the scene under monitoring by visual sensor network again, object tracking will continue through object re-identification. Finally, continuous multiple-object tracking in large-scale scene is implemented. A database is established by collecting data through the visual sensor network. Then the performances of object tracking and object re-identification are tested. The effectiveness of the proposed multiple-object tracking approach is verified.

key words: *visual sensor network, HOG, improved particle filter, re-identification, object tracking*

1. Introduction

In recent years, with the development of communication technology, signal processing technology and computer technology, the visual sensor network is developing rapidly. Compared with a single camera, the visual sensor network has advantages in solving tracking problems in large-scale and complex environments. Compared with the traditional sensor network, the visual sensor network is more focused on the video and image acquisition and processing, so it can more accurately and comprehensively capture and process object information. Visual sensor networks have attracted wide attentions in the fields of security monitoring and control, smart home, etc. Object detection and tracking technology based on visual sensor network can predict object location, automatically detect object and switch scene. It is the core technology of visual sensor network information processing, and it plays an important role in implementing and expanding the function of visual sensor network.

Aiming at fulfilling the wide-area video surveillance, Wang *et al*. proposed an adaptive Gaussian mixture model and an unscented Kalman filter to solve target tracking for each camera node. The correspondence between the targets in different camera views is established by homography transformation of target positions. The optimal node selection is achieved and the accurate tracking is implemented in real scenes [1]. Based on distributed PTZ (pan, tilt, zoom) camera network, Choudhary *et al*. used particle filter tracking to track the objects in each camera view and multi-layered belief propagation for seamlessly tracking objects across cameras [2]. Fang *et al*. presented a collaborative multi-tier nodes strategy for visual target tracking in wireless multimedia sensor networks. A novel function measuring the value of utility of tracking as well as energy cost is used to select the optimal camera node. Therefore, a better tradeoff between performance and energy consumption is achieved in two-tier network structure [3]. Dai *et al*. addressed the problem of object association and consistent labeling through exploring geometrical correspondences of objects, not only in sequential frames from a single camera view but also across multiple camera views [4]. A standard single shot re-identification approach based on offline training is combined with a Dynamic Time Warping (DTW) distance. Then a novel multiple shots re-identification approach is proposed to recognize humans from different views [5]. To limit the computational load, Cabrera *et al*. proposed to reuse the same set of Haar-features based tracking-by-identification for detection and identification. An Adaboost cascade is used for tracking. The method yields good results on standard datasets, without the need to update the model online [6].

Different from the works in [1]–[4], we more concern the key regions, and the visual sensors are non-calibrated with non-overlapping views, which increases the difficulty of continuous tracking in different scenes. Furthermore, our research adopts object features instead of historical information [5], [6] to achieve re-identification and tracking, which improves the ability to cope with the situation where the objects appear with a short-term period. In this paper, a multiple-object tracking approach in large-scale scene is proposed based on the wireless visual sensor network. Object detection is first carried out by extracting HOG features. Then, object tracking is performed based on an improved particle filter method. Furthermore, continuous object tracking in visual sensor network is implemented through object re-identification based on object features.

The remainder of this paper is organized as follows. Section 2 presents the implementation of multiple-object tracking in large-scale scene. The experiment results are given in Sect. 3 and Sect. 4 concludes the paper.

Manuscript received November 27, 2015.
Manuscript revised March 19, 2016.
Manuscript publicized April 21, 2016.

[†]The authors are with The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, China.
a) E-mail: zhiqiang.cao@ia.ac.cn (Corresponding author)

DOI: 10.1587/transinf.2015EDP7481

2. Multiple-Object Tracking Approach in Large-Scale Scene

2.1 Problem Statement

A visual sensor network is adopted to monitor the objects in large-scale scene, where large-scale refers to that the coverage of monitoring area can be enlarged by multiple cameras with non-overlapping views. These cameras are installed on the top of the room with a downward installation angle, and they are arranged according to the expected monitoring areas. In this paper, we mainly concern the object tracking issue after these cameras have already arranged. Figure 1 gives the schematic diagram of the monitoring scene. We consider two categories of objects: pedestrian and robot. We label s as the scale of object, and Ω is denoted with the evaluation of HOG features of a particle. $\mathfrak{R}^k$ and V^k represent the weighted HSV colour histogram and the shape information vector of k^{th} region, respectively, where the regions are obtained through region segmentation.

To achieve this goal, objects need to be detected firstly as the initial state of tracking. Then object tracking will be conducted by solving the problems such as nonlinear/none-Gaussian condition, the change of object's feature and accumulated error. It shall be noted that the object tracking should be continuously tracked with high precision and real-time capability. Besides, considering that the objects is occluded or move among different scenes, continuous multiple-object tracking can be realized by using object re-identification technique with high real-time capability.

The objective of this paper is stated as follows. Given the video streams from the visual sensor network, design an effective multiple-object tracking approach to achieve continuous multiple-object tracking in large-scale scene.

2.2 Object Detection Based on HOG

At present, object detection in visual sensor networks are normally based on feature extraction methods, such as HOG, SIFT and Haar. HOG feature [7] describes the distribution of gradient intensity and gradient direction of local area in the image, and it can characterize the local appearance and shape of object well. Moreover, HOG feature is not sensitive to direction and illumination variance. Compared with other features, HOG feature gets better performance, and it is widely used in object detection.

In this paper, the image is collected from camera whose size is 1280*720, and sample sizes of pedestrian and robot are set to 64*128 and 64*64, respectively, so that the computation quantity can be accepted. Figure 2 is a schematic diagram of the extraction of HOG feature for 1 block. Each sample is composed of multiple blocks.

The positive and negative sample databases are constructed to train classifiers. The purpose of detection is finding out whether the object is in the image. It is a binary classification problem actually. As a supervised learning method, the SVM method can be used to find a super plane in high dimensional space as the segmentation of two classifications, while guaranteeing the minimum classification error rate. So linear SVM method is selected to train classifiers.

2.3 Object Tracking Based on an Improved Particle Filter

If the object size and shape is unknown, detection based HOG descriptor needs to detect all scales of the image sequentially. It is high computational complexity and takes a long time. To achieve real-time tracking of multiple objects in the visual sensor network, a simple method with low complexity is needed. Object tracking is essentially a problem of state estimation. Particle filter is widely used to solve the problem of state estimation for nonlinear/none-Gaussian system.

The traditional particle filter generally uses a series of weighted particles to characterize the feature of object. Positioning and iterative tracking of moving object are implemented based on the cycle process of "prediction and update" [8], [9]. In practical use, the tracking precision is decreased because the feature of object is changing over time. Considering the actual motion situation of object, a temporal and spatial dynamic model is designed to solve the problem. After state transition of particles, the state of each particle needs to be evaluated for selecting the best particle. In the appearance model of traditional particle filter, the particle's evaluation is obtained by comparing HOG feature of current time with that of initial time. This will cause the ac-

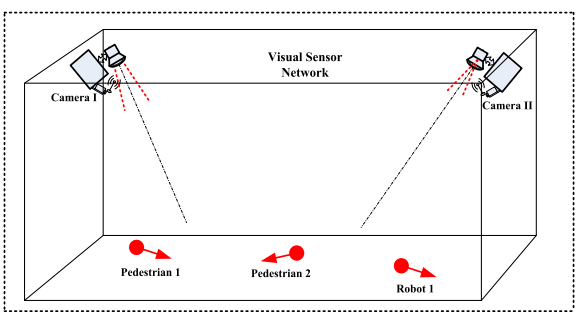


Fig. 1 The schematic diagram of the monitoring scene.

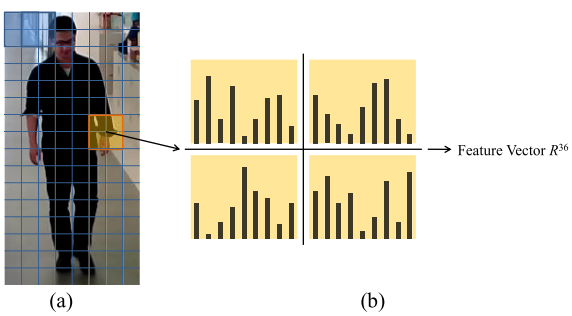


Fig. 2 A schematic diagram of the extraction of HOG feature for 1 block. (a) Blocks displayed on the original image. (b) Histograms of the 4 cells of a block.

cumulated error, which leads to tracking failure. So a new appearance model is given to evaluate particles.

One important problem of the traditional particle filter method is the degradation, i.e., after a few iterations, the weights of majority particles are small, which leads to the waste of computing resources. Resampling is normally used to solve the degradation problem, i.e., the particles with small weights are abandoned, while particles with large weights are propagated to progenies with different quantities. The total number of particles is kept constant. In addition, for state update, all of ω_t^j should be compared with a given threshold ε. If all of them do not exceed ε, it means that tracking process is failure and re-detection is inevitable. Otherwise, the optimal particle of all particles is selected as the final state at current time.

The temporal dynamic model is designed using n order Bessel curve with Gaussian noise:

$$P_{t+1} = \sum_{i=0}^{n} B_{i,n}(\zeta) \cdot P_{t-n} + \upsilon \tag{1}$$

where $X_t = [P_t, s_t]^T$ describes the state vector of a particle at time t, $P_t = [x_t, y_t]^T$ represents the coordinate of a particle at time t in the image, υ is zero-mean Gaussian noise, $B_{i,n}(\zeta) = \frac{n!}{i!(n-i)!}\zeta^i(1-\zeta)^{n-i}$, and ζ represents the proportional parameter.

The spatial dynamic model is given by:

$$s_{t+1} = s_t + v_{t+1} + \varpi \tag{2}$$

where s_t is the scale of object for current detection, v_{t+1} is average variation of scale in the last several time, and ϖ is zero-mean Gaussian noise.

The appearance model is shown as follows:

$$\omega_t^j = e^{\theta \times \Omega} \tag{3}$$

where ω_t^j is the weight of particle j at time t, θ is the coefficient.

2.4 Object Re-Identification

Losses of the tracking will be incurred for several reasons, such as occlusion, scene switching and leaving. When the object is in the scene again, object tracking in visual sensor network requires that the object can be re-identified. Then object tracking will be continuous in visual sensor network by associating with the tracking history.

As for the re-identification, an object re-identification method based on fusion of local features is proposed. First of all, region segmentation for object is carried out by the combination of colour and shape feature. Feature descriptors are further designed to extract colour and shape information of key regions. Finally, object re-identification is implemented by evaluating similarities among the objects.

2.4.1 Region Segmentation

For human visual characteristics, the recognition process will be carried out with the neglect of some unnecessary small details. Then local information will be summarized and sorted such as colour, texture, shape and other information. The re-identification is accomplished by matching with relevant regional characteristics of candidates. Reasonable segmentation for objects is a meaningful preliminary work for extracting feature information. According to the feature characteristics, it is necessary to complete the region segmentation to obtain relatively stable local region. Compared with the global feature, the feature vector extracted from the local region has better descriptive ability and is more robust to characteristic changes owing to occlusion and deformation, etc.

Fig. 3 Demonstration of region segmentation for pedestrians.

The size and quantity of the region segmentation should be related to the category and characteristics of object. For pedestrian as an example, the head information is similar and few, while the colour feature of upper body and lower body is relatively stable. Besides, the shape feature of upper body is also very stable. So the object of pedestrian is divided into three parts including the head, upper body and lower body.

The selection of segmentation line takes into account the colour feature and shape feature. Operators are respectively designed to characterize the changes of colour and shape in a certain region. Segmentation line based on the colour feature or shape feature is the place where has most dramatic change. The final segmentation line is determined by weighting the colour segmentation line and shape segmentation line. The weights of two kinds of segmentation lines are different in different regions.

Take pedestrians as examples, the results of region segmentation are given in Fig. 3. The pedestrians are divided into 3 parts by 2 segmentation lines, and the upper body and lower body are considered as key regions with relatively stable features.

2.4.2 Feature Descriptors Design

After region segmentation, feature descriptors need to be designed to extract the feature information of object's key regions. Feature descriptors include weighted HSV colour histogram and shape information descriptor.

Usually, object re-identification assumes that appearance feature of object will not change in a certain time,

so colour is the most commonly used and effective feature in re-identification. But traditional colour histogram does not contain spatial position information, which means descriptive ability of features is reduced to some extent. The weighted HSV colour histogram (Using two parameters of hue and saturation) is designed by importing the weight transformed from the spatial position.

Each dimension of $\mathfrak{R}^k$ can be expressed as follows:

$$\mathfrak{R}^k_{h,s} = \sum_{j=f_k}^{f_{k+1}} \sum_i \delta^{h,s}_{i,j} \eta_1 Q_{i,j} \qquad (4)$$
$$h \in 1, 2, \cdots, H;\ s \in 1, 2, \cdots, S$$

where H is the quantized series of hue, and S is the quantized series of saturation. If the hue value of pixel $p_{i,j}$ is within h, while the saturation value is within s, then the value of $\delta^{h,s}_{i,j}$ is 1, otherwise it is 0. $Q_{i,j}$ represents the weight of pixel $p_{i,j}$, obeying two-dimension normal distribution with regional barycenter G as expectation, and η_1 is the normalization constant of $Q_{i,j}$. The pixels near the barycenter have bigger weights in histogram statistics than others. A better robustness is achieved based on combination of spatial position information and colour information. The dimension of $\mathfrak{R}^k$ are fixed and determined by quantized series of S, H.

Colour information has some invariance to the pose changes of object. But it is hard to re-identify accurately by using colour information only when complex situations appear. For examples, the colours between object and background are similar; the colours of different objects are close or just same. Considering the differences of shape features among objects in most situations, so a kind of shape information descriptor is designed. The detailed shape information extraction based on shape information descriptor is shown in Algorithm 1.

Algorithm 1: Shape information extraction

Input: the region segmentation result f_k, f_{k+1} of k^{th} region, the foreground image.

Output: shape information vector V^k.

1 According to the region segmentation result f_k, f_{k+1} of k^{th} region, the foreground image M of this region is obtained;
2 The regional barycenter G of M is calculated;
3 The tangent L_l and R_l of the left and right boundary of M are calculated, respectively;
4 M is divided into four sub parts M_1-M_4 according to the regional barycenter G, as shown in Fig. 4;
5 The origins of M_1-M_4 are intersections between L_l, R_l, f_k, and f_{k+1}, respectively. Then the shape information of each sub part is calculated. The specific method is shown in the Eq. (5) and (6);
6 V^k is obtained by cascading the shape information of four sub parts.

$$V^k_{m,a} = \sum_{p_{x,y} \in M_m} \delta^a_{x,y} \left(1 - \eta_2 T_{x,y}\right) \qquad (5)$$
$$m \in 1, 2, 3, 4;\ a \in 1, 2, \cdots, A$$

$$\delta^a_{x,y} = \begin{cases} 1, & \dfrac{2A \arctan\left(\dfrac{y}{x}\right)}{\pi} \in [a-1, a) \\ 0, & \text{else} \end{cases} \qquad (6)$$

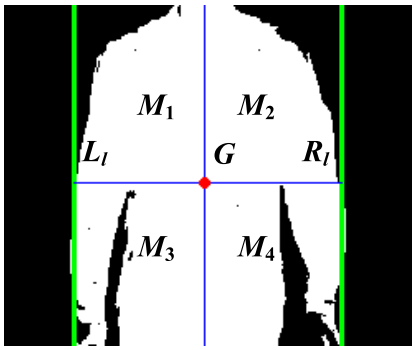


Fig. 4 The schematic diagram of partition of foreground image M.

where A is the dimension of shape information of each sub part, $p_{x,y}$ is the pixel within M, x, y are the coordinates relative to the origin of corresponding sub part, and $\delta^a_{x,y}$ is calculated according to the Eq. (6). $T_{i,j}$ represents the weight of pixel $p_{x,y}$, obeying two-dimension normal distribution with regional barycenter G as expectation, and η_2 is the normalization constant of $T_{i,j}$. The pixels that are far away from the barycenter contain more shape information of object, and thus they are attached bigger weights in shape information statistics than others.

2.4.3 Fusion and Usage of Features

After extracting features, object re-identification is accomplished by measuring the similarities of features. Because of using different methods for feature extraction, the use of measurement is not same as well.

We assume that I_1, I_2 are two objects to be compared. The measurement of colour histogram should be able to characterize the similar degree of colour feature in two regions. So the histogram intersection is selected to compare colour histograms of the corresponding regions and the similarity $r_c(I_1, I_2)$ of colour features is calculated. Shape information vector is composed of four cascaded parts of shape information. The measurement of shape information vector needs to focus on the variation trend of the corresponding parts, so the Pearson correlation coefficient is selected to calculate the similarity $r_s(I_1, I_2)$ of shape features. The final similarity $r(I_1, I_2)$ is obtained by weighting the similarities of two kinds of descriptors.

In the tracking process, the feature information of objects which lose tracking will be recorded. When an object is detected in the scene, the local features are extracted to calculate the similarities with all records. If the max similarity is higher than the threshold ξ, the new object is further associated to the corresponding object which has max similarity. Otherwise, a new tracking sequence will be build.

3. Experimental Results

The hardware platform adopts 4 wireless camera nodes with the view angle of 60° to form a visual sensor network with non-overlapping views. All cameras are placed down-

(a) (b)

Fig. 5 The snapshots of the monitoring video with two pedestrians.

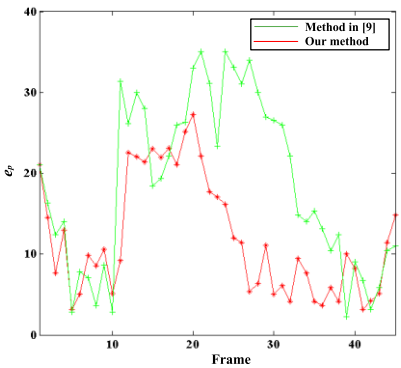


Fig. 6 The deviation result of object tracking for pedestrian 1 in pixel unit.

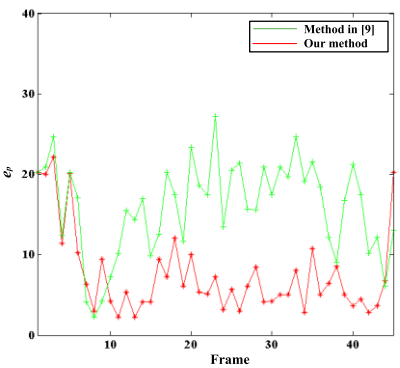
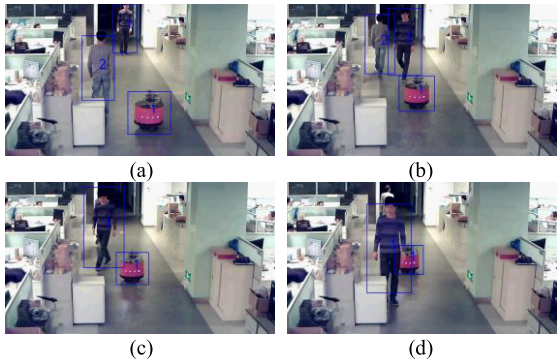


Fig. 7 The deviation result of object tracking for pedestrian 2 in pixel unit.

ward. The data is provided as 720P video stream with 15 frames/second. This section is used to testify the proposed approach. After the performances of tracking and re-identification are verified, respectively, the performance of the proposed multiple-object tracking approach in large-scale scene is testified.

3.1 Object Tracking Performance

The performance of the tracking method based on an improved particle filter is evaluated in this section. Experiment is carried out aiming at the monitoring video with multiple objects. The relevant parameters of the experiment are as follows: $n = 2$, $\zeta = 1.5$, $\theta = 4$, $\varepsilon = 1$. The deviation value e_p between tracking result and the ground truth calibrated manually is used as the tracking precision under current conditions, and the initial state of tracking is obtained through object detection.



(a) (b) (c) (d)

Fig. 8 The snapshots of two-category objects tracking experiment.

3.1.1 Tracking Precision Test

A monitoring video with two pedestrians is used for tracking precision test. The snapshots of the video are shown in Fig. 5, where the right person is pedestrian 1.

The results including the comparison with the method in [9] for pedestrian 1 and pedestrian 2 are demonstrated in Fig. 6 and Fig. 7, respectively. Due to the fixed aspect ratio of particle and lanky stature, some deviation values in Fig. 6 are a bit big. On the whole, the deviation result of our approach is acceptable.

3.1.2 Two-Category Objects Tracking Test

Figure 8 shows the scenes of tracking two pedestrians and a robot in the visual sensor network. It is seen from Fig. 8 that the tracking performs well, even if the pedestrian is occluded by a cabinet in the environment.

3.2 Object Re-Identification Performance

In order to verify the effectiveness of the object re-identification method, pedestrian database is constructed based on the data collected from visual sensor network. There are 237 samples of 19 pedestrians from 4 scenes.

The relevant parameters of the experiment are as follows: $A = 30$, $\xi = 0.5$. For weighted HSV colour histogram, $H = 12$, $S = 8$. Hue is quantized to 12 series uniformly, and saturation is quantized to 8 series non-uniformly, which are {0, 0.075, 0.15, 0.275, 0.4, 0.575, 0.75, 0.875, 1.0}.

In this paper, the performance of our method is evaluated by using the cumulative matching characteristic (CMC) curve [10]. This is the mainstream evaluation criterion in the field of object re-identification. Our method is compared with two methods including using colour feature only and using shape feature only. Experimental results are shown in Fig. 9. In each round of the test, the rank score represents that the former results with highest similarity to goal sample will be chosen. If the correct sample is included in these results, it is considered that the experimental result of current round is correct. It can be seen from Fig. 9 that the recognition rate of using shape feature only is lower than that of

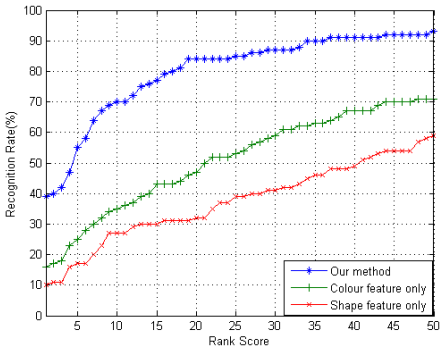


Fig. 9 Experimental results of object re-identification.

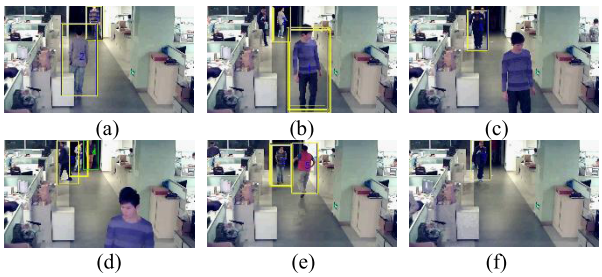


Fig. 10 The snapshots of multiple-pedestrian tracking experiment.

using colour feature only because of the mutability of shape feature. The local features are extracted and fused in our method, which leads to stronger robustness. The recognition rate of our method is nearly 20% higher than that of using colour feature only.

Fig. 10 describes a case where there are 5 pedestrians all together in the monitoring scene. These pedestrians move autonomously according to their demands. From the experimental result, our approach has achieved the tracking for all pedestrians. Especially, for pedestrian 2, he is first detected in the scene shown in Fig. 10 (a). Then, he is occluded by pedestrian 3 (see Fig. 10 (c)). When he appears again in the scene (see Figs. 10 (d)-(e)), continuous tracking for pedestrian 2 is implemented by our re-identification algorithm.

3.3 The Performance of Proposed Multiple-Object Tracking Approach in Large-Scale Scene

In this section, we first consider the experiment of object tracking under multiple scenes with three cameras. The experimental result is given in Fig. 11, where the green and red curves are corresponding to the method in [9] and our proposed method, respectively. The results shows that our method has lower error than the method in [9]. Figure 12 depicts the result where a pedestrian moves through 3 different scenes. One can see that the switching of different scenes has no influence for the tracking performance.

To verify the generality of our method, we choose two sequences of pedestrian motion images from the PETS 2009 dataset as our test environment, where these two sequences share a common intersection. We emphasis on two pedestrians that are labelled by blue rectangles in Fig. 13. From the result, our method performs well even under different views where the illumination conditions have greatly changed. Due to the fact that the objects to be tracked is usually small in PETS 2009 dataset, the extracted features is not always stable, which will lead to a reduction of tracking accuracy to some extent. Overall, our approach is considered as an effective one.

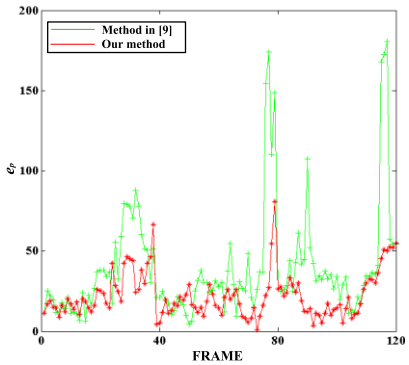


Fig. 11 The deviation result of object tracking with re-identification in pixel unit.

Fig. 12 The snapshots of multiple-scene tracking experiment.

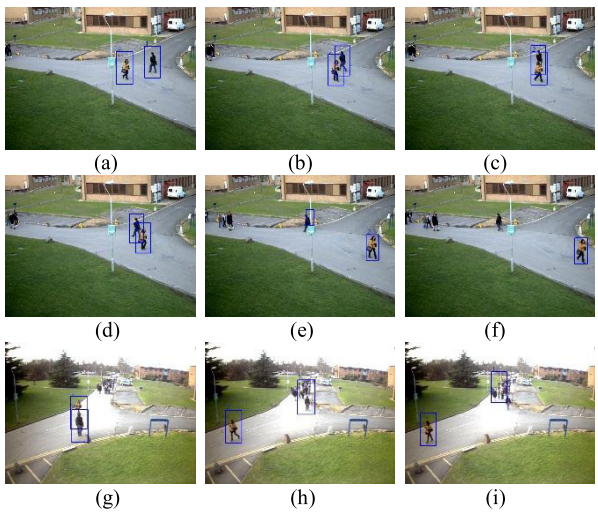


Fig. 13 The snapshots of tracking experiment using PETS 2009 dataset.

4. Conclusion

In this paper, a multiple-object tracking approach in large-scale scene is proposed based on visual sensor network. The approach of this paper is composed of three parts: firstly, object detection is carried out by extracting HOG features. Then, object tracking is performed based on an improved particle filter method. In addition, continuous object tracking in visual sensor network is implemented through object re-identification. A database is established by collecting data through the visual sensor network. Then the performances of object tracking and object re-identification are tested. The effectiveness of the proposed multiple-object tracking approach is verified. In the near future, we will improve the performance of classifiers so that pedestrians with different body types can be detected and tracked accurately. Besides, more stable features will be considered and used and our approach can apply to rich scenes.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant 61273352, and by Beijing Natural Science Foundation under Grant 4161002, and by the 863 Program of China under Grant 2015AA042307, and by the Open Foundation of the State Key Laboratory of Management and Control for Complex Systems, CASIA under Grants 20130101 and 20140107.

References

[1] Y. Wang, D. Wang, and W. Fang, "Automatic node selection and target tracking in wireless camera sensor networks," Comput. Electr. Eng., vol.40, no.2, pp.484–493, 2014.
[2] A. Choudhary, S. Sharma, I. Sreedevi, and S. Chaudhury, "Real-time distributed multi-object tracking in a PTZ camera network," Proc. International Conference on Pattern Recognition and Machine Intelligence, Warsaw, Poland, vol.9124, pp.183–192, 2015.
[3] F. Wu, D.H. Wang, and Y. Wang, "Collaborative nodes strategy for target tracking in two-tier wireless multimedia sensor networks," J. Netw., vol.9, no.7, pp.1803–1810, 2014.
[4] X. Dai and S. Payandeh, "Tracked object association in multi-camera surveillance network," Proc. IEEE Int. Conf. Syst., Man, Cybern., Manchester, United Kingdom, pp.4248–4253, 2013.
[5] D. Simonnet, M. Lewandowski, S.A. Velastin, J. Orwell, and E. Turkbeyler, "Re-identification of pedestrians in crowds using dynamic time warping," Proc. European Conference on Computer Vision, Florence, Italy, vol.7583, pp.423–432, 2012.
[6] R.R. Cabrera, T. Tuytelaars, and L. Van Gool, "Efficient multi–camera detection, tracking, and identification using a shared set of haar-features," Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit., Providence, RI, United states, pp.65–71, 2011.
[7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit., San Diego, United states, pp.886–893, 2005.
[8] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," IEEE Trans. Signal Process, vol.50, no.2, pp.174–188, 2002.
[9] R. Hess, and A. Fern, "Discriminatively trained particle filters for complex multi-object tracking," Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit., Miami, United states, pp.240–247, 2009.
[10] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," Proc. IEEE Int. Workshop Perform. Eval. Track. Surveill., Rio de Janeiro, Brazil, 2007.

Wenbo Yuan received the B.S. degree in Measure and Control Technologies and Instruments from University of Science and Technology Beijing, Beijing, China, in 2011. He is currently pursuing the Ph.D. degree from the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His current research interests include networked robot system, computer vision, robot navigation, and service robot.

Zhiqiang Cao received the B.S. and M.S. degrees from Shandong University of Technology, China, in 1996 and 1999, respectively. In 2002, he received the Ph.D. degree in Control Theory and Control Engineering from Institute of Automation, Chinese Academy of Sciences, Beijing, China. He is currently an associate professor in the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His research interests include multi-robot systems, embedded vision and visual scene cognition, networked intelligent robotic system, etc.

Min Tan received the B.S. degree in Control Engineering from Tsinghua University, Beijing, China, in 1986, and the Ph.D. degree in Control Theory and Control Engineering at the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1990. He is a Professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include advanced robot control, multirobot systems, biomimetic robots, and manufacturing systems.

Hongkai Chen received the B.S. degree in Control Science and Engineering from Xiamen University, Xiamen, China, in 2011. He is currently pursuing the Ph.D. degree from the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His current research interests include robot vision, especially for visual tracking and target detection.