Wenyi and Decun *EURASIP Journal on Embedded Systems* (2016) 2016:16
DOI 10.1186/s13639-016-0053-7

EURASIP Journal on Embedded Systems

RESEARCH Open Access

Embedded tracking algorithm based on multi-feature crowd fusion and visual object compression

Zheng Wenyi[1,2*] and Dong Decun[1]

Abstract

The accuracy and poor real-time performance of moving objects in a dynamic range complex environment become the bottleneck problem of the target location and tracking. In order to improve the positioning accuracy and the quality of tracking service, we propose an embedded tracking algorithm based on multi-feature fusion and visual object compression. On the hand, according to the feature of the target, the optimal feature matching method is selected, and the multi-feature crowd fusion location model is proposed. On the other hand, to reduce the dimension of the multidimensional space composed of the moving object visual frame and the compression of the visual object, the embedded tracking algorithm is established. Experimental results show that the proposed tracking algorithm has high precision, low energy consumption, and low delay.

Keywords: Visual object compression, Embedded tracking, Crowd fusion, Multi-feature systems

1 Introduction

Moving target tracking is one of the most active areas in the development of science and technology. The target tracking algorithms have been widely valued by all countries in the world [1]. With the performance of continuous improvement and expansion, location and tracking algorithm for successful application in industry, agriculture, health care, and service industry [2], in the urban security, defense and space exploration have dangerous situations [3] is to show their talents.

The low-complexity and high-accuracy algorithm was presented in [4], for reducing the computational load of the traditional data-fusion algorithm with heterogeneous observations for location tracking. Trogh et al. [5] presented a radio frequency-based location tracking system, which could improve the performance by eliminating the shadowing. In [6], Liang and Krause proposed the proof-of-concept system based on a sensor fusion approach, which was built with considerations for lower cost, and higher mobility, deplorability, and portability, by combining the drift velocities of anchor nodes. The scheme of [7] could estimate the drift velocity of the tracked node by using spatial correlation of ocean current. The distributed multi-human location algorithm was researched by Yang et al. [8] for a binary piezoelectric infrared sensor tracking system.

The model-based approach was presented in [9], which is used to predict the geometric structure of an object using its visual hull. The task-dependent codebook compression framework was proposed to learn a compression function and adapt the codebook compression [10]. Ji et.al [11] proposed a novel compact bag-of-patterns descriptor with an application to low bit rate mobile landmark search. The blocks were flagged by lying in the object regions flagging compression blocks and an object tree would be added in each coding tree unit to describe the object's shape in its additional object tree [12].

However, how to provide guarantee for the high precision tracking of moving objects in the dynamic range and complex environment and the complexity of the optimization algorithm is one of the most difficult problems. Based on the results of the above researches, the embedded tracking algorithm based on multi-feature crowd fusion and visual object compression was proposed for mobile object tracking.

The rest of the paper is organized as follows. Section 2 describes the location model based on multi-feature

* Correspondence: weyzheng@sina.com
[1]The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China
[2]School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, Jiangsu, China

crowd fusion. In Section 3, we show the embedded tracking algorithm for visual object compression. We analyzed and evaluated the proposed scheme in Section 4. Finally, we conclude the paper in Section 5.

2 Location model based on multi-feature crowd fusion

Generally, the target location is divided into two stages.

In the first phase, the feature of moving target is extracted. Feature extraction would be completed with the following steps.

(1) Capture moving target image frame sequence.
(2) The characteristics of real-time target image frames would be extracted.
(3) From the current image frame to the still image frame between the target, search and the extracted features of the image frame the most similar to the target motion characteristics.

The second stage involves the characteristics of the moving target matching.

Choosing different features, according to the characteristics of the target, is selecting the best feature matching scheme.

The above localization scheme has the following defects:

(1) The extracted features are single. Such feature extraction is difficult to locate for the complex moving objects with multiple states.
(2) The change features of the moving object in complex scenes such C_{def} as various deformation, C_{lgf} (light), C_{siz} (size), and C_{col} (color), making the single feature matching success rate $\mathrm{SR}_{\mathrm{FM}}$ very low, as shown in formula (1).

$$\begin{cases} f(\mathrm{IF}_i = \mathrm{if}) = \sum_{i=1}^{N}\left(G_i, h\left(\mathrm{if}, h_{i-1}, \sum_{j=1}^{i} h_j\right)\right) \\ \mathrm{if}\left(C_{\mathrm{def}}, C_{\mathrm{lgf}}\right) h\left(\mathrm{if}, h_{i-1}, \sum_{j=1}^{i} h_j\right) = |\mathrm{if}_i(C_{\mathrm{siz}}, C_{\mathrm{col}}) - \alpha| h_{i-1}\left(\alpha, \sum_{j=1}^{i} h_j\right) \\ \mathrm{SR}_{\mathrm{FM}} = \sum_{i=1}^{M} f(\mathrm{if}_i, h_i) \frac{1}{\sum_{j=1}^{N} f(\mathrm{if}\ j)} \leq \frac{f(\mathrm{if}_M, h_M)}{N} \end{cases} \tag{1}$$

Here, if represents the image frames. IF is the representation of image frame sequence. G is the vector representation of image frame matrix. H is the function used to solve the image frame characteristics and frame similarity. N represents the captured motion target image frame sequence length. M represents the frames of image feature matching. From formula (1), it is found that the upper bound of the matching success rate is $\frac{f(\mathrm{if}_M, h_M)}{N}$. But the success rate of image frames captured is inversely proportional to the number of captured image frames. The conclusion shows that the captured image frames will restrict the single feature matching characteristic.

(3) The accuracy and robustness of target motion in real time are poor, as shown in formula (2). In order to improve the accuracy and robustness, a single feature set is tracking the feature series, but the complexity of the transition algorithm is too high, as shown in formula (3).

$$\begin{cases} A_{\mathrm{TR}} = \sum_{i=1}^{M}\left(\sin\beta - \alpha^i\right)\frac{1}{N} + \sum_{i=1}^{M} h_i\sqrt{\alpha} \\ \mathrm{RUS}_{\mathrm{TR}} = \rho f(\mathrm{IF}_M)\mathrm{SR}_{\mathrm{FM}} \end{cases} \tag{2}$$

Here, A_{TR} indicates the positioning accuracy. $\mathrm{RUS}_{\mathrm{TR}}$ indicates the location robustness. β is the included angle between adjacent image frames. ρ denotes the expressed error vector.

$$\begin{cases} \mathrm{CLE}_{\mathrm{TSA}} = |\mathrm{IF}_i - \rho\sin\beta| \leq M * g(h_i, \alpha) \\ g(h_i, \alpha) \approx \sum_{i=1}^{M}\left(h_i \mathrm{if}\left(C_{\mathrm{def}}, C_{\mathrm{lgf}}, C_{\mathrm{siz}}, C_{\mathrm{col}}\right)\right) \\ \qquad \geq M * \mathrm{if}_M\left(C_{\mathrm{def}}, C_{\mathrm{lgf}}, C_{\mathrm{siz}}, C_{\mathrm{col}}\right) \end{cases} \tag{3}$$

Here, $\mathrm{CLE}_{\mathrm{TSA}}$ represents the complexity of the transition algorithm. The function $g(h_i, \alpha)$ represents transition algorithm. It can be found that the complex image frames are proportional to the degree and feature matching. This shows that more space, time, and computation must be paid in order to get more features to match the image frames.

In order to solve the above problems, we propose a multi-feature crowd fusion location model. The model analyzes the dynamic motion of the target, the moving track, and the structure parameters of the image frame. The state characteristics of different targets are captured; the composition of multiple feature vectors such as formula (4) is presented. This vector integrates the characteristics of motion state and deformation, light, size, and color and can effectively improve the low matching success rate of single feature extraction, such as formula (5).

$$\begin{cases} \mathrm{ML}_F = \begin{bmatrix} v_{\mathrm{mot}}\,\mathrm{if}_{11} & \cdots & v_{\mathrm{mot}} f_{1L} \\ \vdots & \ddots & \vdots \\ v_{\mathrm{mot}}{}^K f_{K1} & \cdots & v_{\mathrm{mot}}{}^K f_{KL} \end{bmatrix} \\ v_{\mathrm{mot}} = \frac{M}{N}\sum_{i=1}^{N-M} f(\mathrm{if}_i, h_i) \end{cases} \tag{4}$$

Here, v_{mot} is the target motion trajectory fitting function. K is the representation of time series features. L is the representation of spatial sequence features.

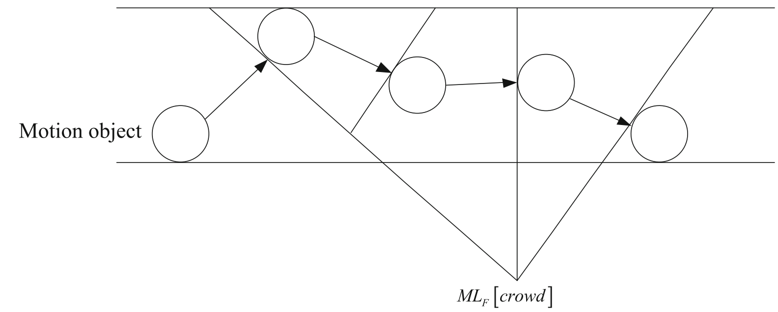

Fig. 1 Crowd fusion between multiple features

$$\mathrm{SR_{FM}} = \begin{cases} \alpha\cdot\mathrm{rank}\{\mathrm{ML_F}\}\tan\beta, \beta < \alpha || \beta > \rho \\ 1, \alpha \le \beta \le \rho \end{cases} \tag{5}$$

Here, rank $\{\mathrm{ML_F}\}$ is the rank of multi-feature vector. From formula (5), we can see that the high matching success rate can be guaranteed as long as the multiple feature vectors are solved correctly.

In order to further reduce the complexity and improve the accuracy and reliability of multi-feature matching, this model combines the multi-feature fusion mechanism based on crowd feature analysis. Multi-feature vector and the target motion curve of the knowledge combination are shown in Fig. 1. In the crowd analysis characteristics, the curve and the multiple features are relatively independent, the relative independence between the arc and the multiple features. By crowd analysis, multi-feature vectors are optimized. Multi-features in this vector are not mutually exclusive. Multi features in this vector are not mutually exclusive for improving the performance of multi feature fusion. This can reduce the amount of fusion operations, as shown in formula (6).

$$\begin{cases} \mathrm{fu_{comp}} = \dfrac{1}{M^2}\displaystyle\sum_{i=1}^{K}\sum_{j=1}^{L} G(i,j)\mathrm{ML}_{F(i,j)} \\ G(i,j) = \dfrac{f(\mathrm{IF}_i)f(\mathrm{IF}_j)}{h(f, \mathrm{ML}_{Fi}, \mathrm{ML}_{Fj})} \end{cases} \tag{6}$$

In summary,the multi feature crowd fusion algorithm is shown in Fig. 2.

3 Embedded tracking algorithm for visual object compression

The visual frame of the moving object constitutes the multidimensional space of Q dimension. The visual frame in this space is denoted as x^{Q}. The visual frame of

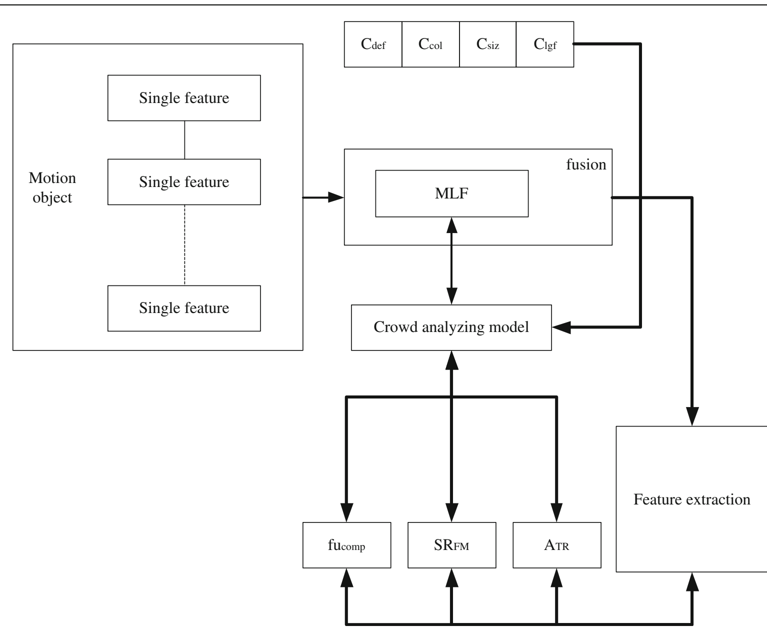

Fig. 2 Multi-feature crowd fusion location algorithm and workflow

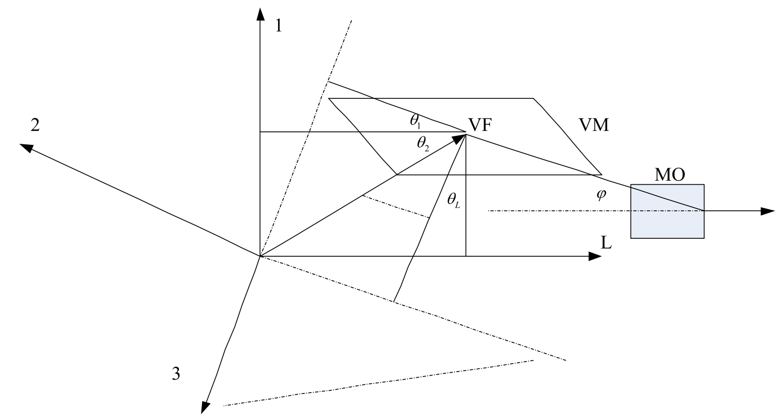

Fig. 3 L-dimensional space tracking system for visual frame

the internal elements is used to form a visual matrix VM. The element of the visual matrix receives the interference of the multidimensional space, and the information is easy to be distorted. In order to solve this problem, the visual matrix VM can be compressed. The visual frame of the L-dimensional space tracking system is shown in Fig. 3. Here, the VM is selected as the object of the center visual frame VF, perpendicular to the coordinate axes of the L-dimensional space. The angle between the vertical line is denoted as θ_1, ..., θ_Q. MO represents a moving target. In the moving process of objects, ϕ is the angle between the VF point and the motion direction. The VF point can collect the visual targets with different degrees. These targets are used to update the elements of the visual matrix VM. The fusion results of VF and VM are mapped into the MO plane. The compression matrix must satisfy the omni-directional low-dimensional characteristics, as shown in formula (7). After the matrix is compressed, the characteristics of the distribution of the visual frame must be satisfied, as shown in formula (8).

$$\begin{cases} F(\theta,\phi) = |\mathrm{VM}|^H \dfrac{\sin\theta\cos\phi}{2d} \\ D_{\mathrm{eff}} = \left\lceil \dfrac{2\pi}{Q}\displaystyle\sum_{i=1}^{Q}\sin\theta_i\left\{\dfrac{\cos\phi}{\theta_i}\right\}_{\min} \right\rceil \end{cases} \tag{7}$$

Here, $F(\theta,\phi)$ represents the direction function of visual frames in L-dimensional space. D represents the distance between the VF point and the origin of the L-dimensional space. D_{eff} represents the effective dimension of visual tracking space. The parameter value is obviously less than L, which can effectively reduce the dimension and improve the compression efficiency of the visual frame.

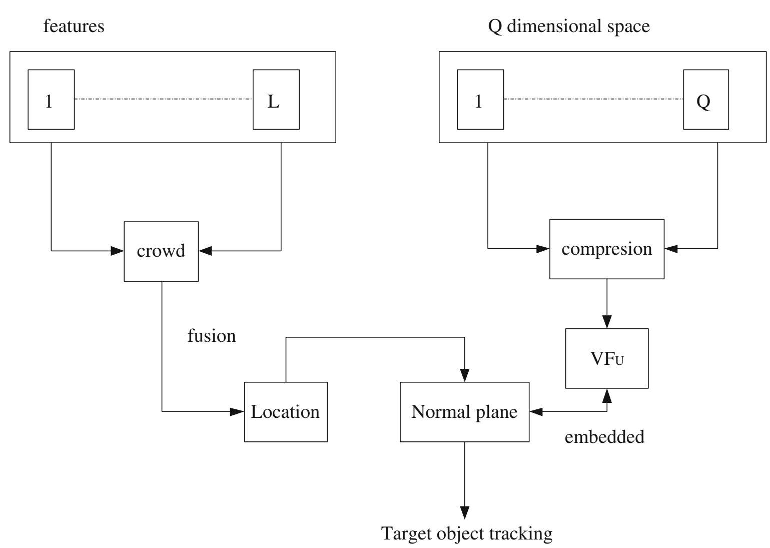

Fig. 4 Embedded tracking algorithm architecture with visual frame target compression

Table 1 Delay with 1000 visual frame samples

Normal plane radian	Location delay with single feature	Location delay with multiple features and fusion
30	25.7 ms	1.9 ms
50	89.4 ms	2.0 ms
110	345.2 ms	1.8 ms

$$f(\mathrm{VF}) = \begin{cases} \dfrac{|\mathrm{VM}|^H}{D_{\mathrm{eff}}}, \theta_1 < \phi < \theta_{D_{\mathrm{eff}}} \\ \dfrac{|\mathrm{VM}|^H}{Q}, \phi = \theta_1, \cdots \theta_Q \\ \{D_{\mathrm{eff}}\}_{\max}^{i=1,\ldots,|VM|^H}, \phi > \theta_{D_{\mathrm{eff}}} \| \phi < \theta_1 \end{cases} \tag{8}$$

Here, f(VF) is the visual frame distribution density function. $\{D_{\mathrm{eff}}\}_{\max}^{i=1,\ldots,|\mathrm{VM}|^H}$ represents the largest spatial dimension in the VM rank of the visual matrix.

After the visual matrix VM is compressed and reconstructed by the visual frame, the tracking signal is shown in formula (9).

$$\mathrm{vf}^{D_{\mathrm{eff}}} = \mathrm{VM}^{\sin\theta \cos\phi} \sum_{i=1}^{|\mathrm{VM}|^H} x_i \tag{9}$$

The visual object compression method can obtain the mapping relationship between the visual frame and the moving object from the L dimension or D_{eff} dimensional space by choosing the $\{D_{\mathrm{eff}}\}_{\min}^{i=1,\ldots,|\mathrm{VM}|^H}$ and realize the target motion prediction. The specific steps of the visual target compression algorithm are as follows:

(1) The visual matrix of the core visual frame-oriented migration: The VFC state of the visual frame is $\{\theta_C, \phi_C, D_{\mathrm{eff}\_C}\}$, which is captured by the current moving target. The visual frames propagate along the direction of the $F(\theta_C, \phi_C)$. The new state $\{\theta_U, \phi_U, D_{\mathrm{eff}\_U}\}$ of the visual frame is obtained after spreading on the dimensional space D_{eff_C}.

(2) The moving object, the current state of the visual frame, and the diffusion of the visual frame form a compressed plane PCV: The compressed point set PT is formed in the plane. Arbitrary two points PT_j and PT_i in the plane into a visual line: PT_i normal vector is $\mathrm{NV}_i = \sin\theta_C \cos\phi_U \parallel \mathrm{PT}_j \parallel$. The normal vector of any point PT_j on the plane is $\mathrm{NV}_j = \sin\theta_C \cos\phi_U \parallel \mathrm{PT}_i \parallel$.

(3) The included angle between the normal plane of the method vector NV_i and the normal plane of the normal vector NV_j: The relation between the plane angle and the direction arc is $\sin\gamma = \mathrm{NV}_i \cdot \mathrm{NV}_j \|\mathrm{PT}\| \arctan\left(\frac{|\theta_C - \theta_U|}{|\phi_C - \phi_U|}\right)$. The vector mapping relation between the plane angle and the direction field of the moving object is shown in formula (10).

$$\begin{cases} M_\gamma = \begin{bmatrix} M_{\mathrm{PT}_i} & M_{\mathrm{PT}_j} \\ M_{\mathrm{VF}_C} & M_{\mathrm{VF}_U} \end{bmatrix} \\ M_{(i,j)} = F(\sin\theta_C, \cos\phi_U) \|\mathrm{PT}_i\| \|\mathrm{PT}_j\| \sum_{i=1}^{|\mathrm{VM}|^H} x_i \end{cases} \tag{10}$$

The mapping matrix M_γ is divided into 4 submatrices. Each submatrix $M_{(i,j)}$ is obtained by solving the direction function, compressing the visual frame point and the signal strength.

(4) The 4 submatrices of the moving object multi-feature fusion mapping matrix are obtained by the visual frame analysis. Tracking matrix MT is obtained through the target compression.

$$M_T = \begin{bmatrix} \mathrm{ML}_{F(i,j)} \mathrm{fu}_{\mathrm{comp}} & \mathrm{SR}_{\mathrm{FM}} \cdot \mathrm{rank}\{\mathrm{ML}_F\} \tan\gamma \\ M * \mathrm{if}_M(C_{\mathrm{def}}, C_{\mathrm{lgf}}, C_{\mathrm{siz}}, C_{\mathrm{col}}) & \sum_{i=1}^{M} (\sin\gamma - \alpha^i) \frac{1}{N} \end{bmatrix}$$

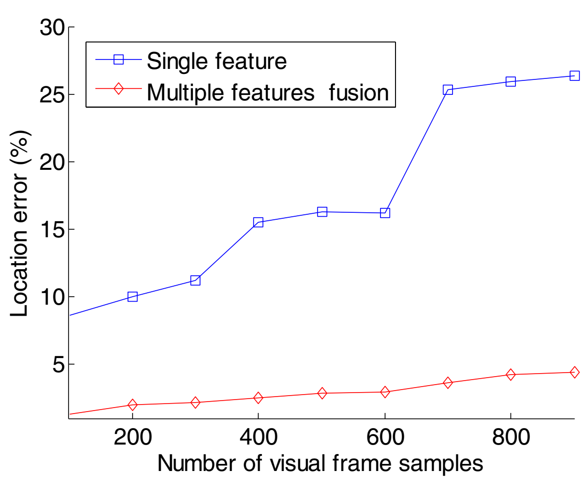

Fig. 5 Location error

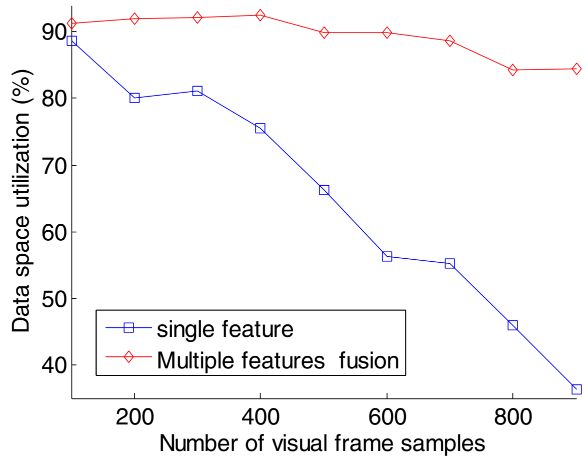

Fig. 6 Data space utilization

(5)M_γ and M_T through the operational matrix of integration to predict the moving object space of the latest state.

To sum up, the embedded tracking algorithm based on multi-feature fusion and localization of visual target compression is shown in Fig. 4.

4 Performance analysis of embedded tracking algorithm

We focus on the library environment and the playground environment to test the proposed algorithm in this paper. The server of the experimental platform is core i5 Intel, physical memory is 8G, and virtual memory is 4G (algorithm using C language programming). In the library environment, the digital cameras were used to obtain the actual value of the pedestrian trajectory; the measurement range is 120 m^2. During the experiment, the pedestrians once every move, capture a visual frame image, and upload to the server through the Wi-Fi. In the playground, the scope of the experiment scene is larger. The maximum circumference of the playground is 700 m. By using an outdoor electric car combined with HD industrial cameras, the pedestrian trajectory is captured.

In order to analyze the improvement of the positioning accuracy of the multi-feature fusion and location algorithm for moving objects, we test three different methods. When the number of visual frames is 100200300, the single feature and multi-feature fusion and location algorithm are the work delay. The number of random sampling is set to 1000 times. Table 1 shows the statistical positioning delay with different curvature method pinging, feature extraction and location analysis, and the execution time required. It was found that the greater the plane curve, the greater the positioning delay. Visual object compression algorithm can effectively reduce the spatial dimension and reduce the plane curve. Single feature location delay is several times of the multi-feature fusion and localization. In the worst case, the single feature location delay is about 150 times of the multi-feature fusion. This is because the multi-feature fusion location algorithm using multi-feature extraction acceleration and different dimensions of feature fusion estimation moving trajectory of the target object not only reduces the processing time but also the computational complexity and the positioning method, so that the

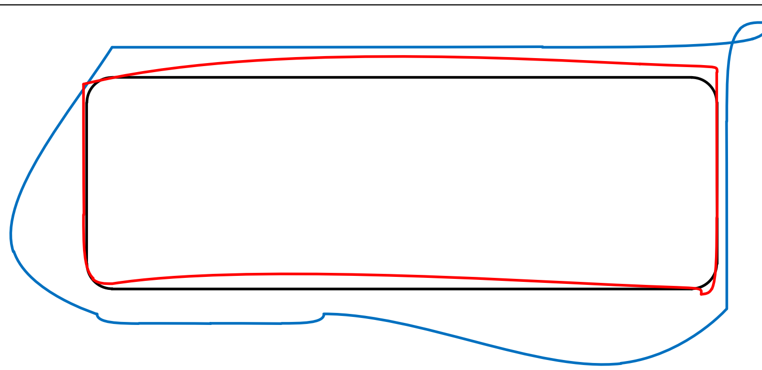

Fig. 7 Reconstruction track

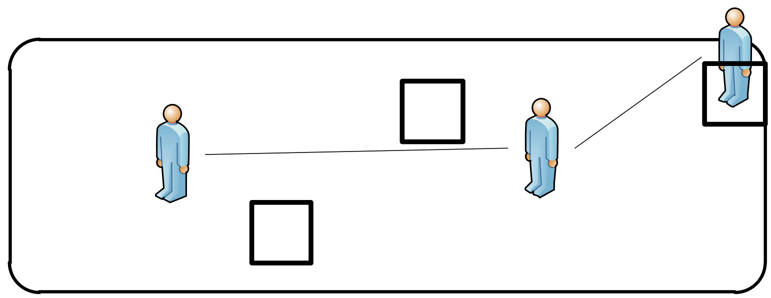

Fig. 8 Tracking results of moving object with T-OTF algorithm

positioning acceleration effect is very obvious. At the same time, Fig. 5 shows the positioning accuracy of single and multi-feature fusion and location algorithm under different samples. It is found that, with the increase of the visual frame samples, the error of the single feature location algorithm increases obviously. When the sample number is two times the speed of the moving object, the position error is 25 %. The single feature location has been unable to capture the visual frame and the normal tracking object. In contrast, multi-feature fusion algorithm always maintains high precision, which benefits from the feature fusion.

Figure 6 shows the storage space of the server with the tracking algorithm. It is found that the proposed algorithm has a better space utilization. The proposed algorithm can achieve accurate target localization using a small amount of data.

In the playground, the pedestrians walk at a speed of 10 m/s. The experiment time is 9 a.m. The weather is cloudy, the sun is not enough, the embedded tracking algorithm based on multi-feature fusion and visual target compression is recorded as ET-MVC, and the target object feature tracking algorithm is denoted as T-OTF. Pedestrian trajectory reconstruction is shown in Fig. 7. Blue track represents the T-OTF tracking results. Black locus represents the actual value. The red trace represents the ET-MVC's location tracking results. The blue trace is the most distant from the actual value of the black locus. The main source of error is that the space dimension is too large. High spatial dimension leads to the excessive discretization of the visual frame, and the reconstruction error will be relatively large. ET-MVC can eliminate these errors significantly. The reconstruction accuracy of the motion trajectory is effectively restrained by the error. As seen from the graph, the red trajectory generation optimization results are significantly improved, which is obviously better than the blue trajectory.

Figures 8 and 9 show the effect of the moving target tracking on the actual scene of the experiment. As shown in Figs. 8 and 9, the proposed algorithm is almost always able to accurately track the target in the entire tracking process. The proposed algorithm can eliminate the effect of occlusion and interference in the scene. However, the T-OTF algorithm quickly lost the tracking target.

5 Conclusions

Positioning accuracy, real-time performance, and robustness are important performance indexes of moving target location and tracking. In order to improve the positioning accuracy and improve the quality of tracking service, we propose an embedded tracking algorithm based on multi-feature crowd fusion and visual object compression. First of all, it analyzes the dynamic motion of the target, the moving track, and the structure parameters of the image frame. By capturing the state characteristics of different targets, the multiple feature vectors are formed. Secondly, the visual matrix of the core visual frame-oriented migration is obtained. Moving object, the current state of the visual frame, and the diffusion of the visual frame form a compression plane. Embedded tracking is implemented.

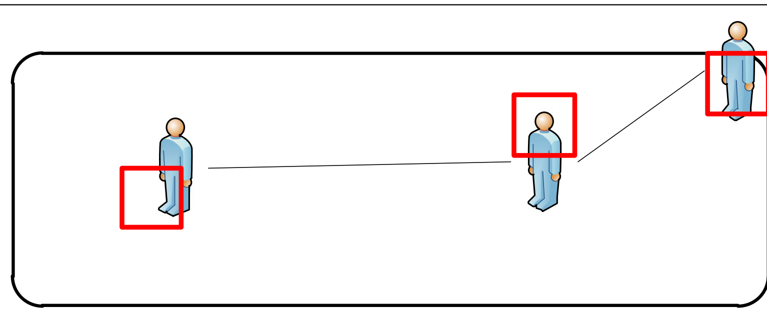

Fig. 9 Tracking results of moving object with ET-MVC algorithm

The experimental results show that the multi-feature fusion can effectively reduce the positioning error and shorten the delay. Compared with the target object tracking algorithm, the embedded tracking algorithm based on visual object compression has a significant advantage in the reconstruction of the moving target.

Acknowledgements
This work is supported in part by the National Youth Fund Project No. 61300228.

Competing interests
The authors declare that they have no competing interests.

Received: 9 June 2016 Accepted: 7 September 2016
Published online: 21 September 2016

References

1. PH Tseng, KT Feng, YC Lin et al., Wireless location tracking algorithms for environments with insufficient signal sources. IEEE Transactions on Mobile Computing 8(12), 1676–1689 (2009)
2. CT Chiang, PH Tseng, KT Feng, Hybrid unified Kalman tracking algorithms for heterogeneous wireless location systems. IEEE Transactions on Vehicular Technology 61(61), 702–715 (2012)
3. YC Lai, JW Lin, YH Yeh et al., A tracking system using location prediction and dynamic threshold for minimizing SMS delivery. Journal of Communications & Networks 15(15), 54–60 (2013)
4. YS Chiou, F Tsai, A reduced-complexity data-fusion algorithm using belief propagation for location tracking in heterogeneous observations. IEEE Transactions on Cybernetics 44(6), 922–935 (2014)
5. J Trogh, D Plets, A Thielens et al., Enhanced indoor location tracking through body shadowing compensation. IEEE Sensors Journal 16(7), 2105–2114 (2016)
6. P-C Liang, P Krause, Smartphone-based real-time indoor location tracking with 1-m precision. IEEE Journal of Biomedical and Health Informatics 20(3), 756–762 (2016)
7. R Diamant, LM Wolff, L Lampe, Location tracking of ocean-current-related underwater drifting nodes using Doppler shift measurements. IEEE Journal of Oceanic Engineering 40(4), 887–902 (2015)
8. B Yang, Y Lei, B Yan, Distributed multi-human location algorithm using naive Bayes classifier for a binary pyroelectric infrared sensor tracking system. IEEE Sensors Journal 16(1), 1–1 (2015)
9. SS Hwang, WJ Kim, J Yoo et al., Visual hull-based geometric data compression of a 3-D object. IEEE Transactions on Circuits & Systems for Video Technology 25(7), 1151–1160 (2015)
10. R Ji, H Yao, W Liu et al., Task-dependent visual-codebook compression. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society 21(4), 2282–2293 (2012)
11. R Ji, LY Duan, J Chen et al., Mining compact bag-of-patterns for low bit rate mobile visual search. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society 23(7), 3099–113 (2014)
12. T Huang, S Dong, Y Tian, Representing visual objects in HEVC coding loop. IEEE Journal on Emerging & Selected Topics in Circuits & Systems 4(1), 5–16 (2014)