Research and Implementation of Person Tracking Method Based on Multi-feature Fusion

Fang Fang(✉), Kun Qian, Bo Zhou, and Xudong Ma

Key Laboratory of Measurement and Control of CSE, School of Automation, Southeast University, Nanjing 210096, Jiangsu, People's Republic of China
ffang@seu.edu.cn

Abstract. Aiming at the problem of person tracking for mobile robot in complex and dynamic environment, a multi-feature tracking strategy is proposed in this paper, by which the target can be determined based on the joint similarity. The joint similarity consists of motion model similarity, color histogram similarity and human HOG feature similarity. The tracking of target is realized by the method of joint likelihood data association. The above strategy can solve the problems such as similar color interference, target loss, and target occlusion. In addition, considering the lost target, a fast search strategy is proposed to search the target. Finally, the method is tested with the mobile robot. The experimental results show that the proposed method is robust and effective when the target is moving rapidly, and it can satisfy the real-time requirement of the system.

Keywords: Tracking · Multi-feature · Joint similarity · Joint likelihood data association

1 Introduction

Person tracking is one of the focuses in service robots field [1, 2]. In recent years, the RGB-D camera is successfully applied in mobile robots platform for the advantage of easy access to depth information as well as the good stability and cost-effective feature [3]. The person detecting and tracking based on Kinect camera are researched by Armando Pesenti Gritti, et al. realizing the person tracking by detecting the legs for the height limit of camera [4]. The researches of faces detection and upper body detection based on RGB-D camera are studied by Duc My Vo, realizing the tracking by the Kalman filter algorithm which can deal with the pose change and target occlusion problem [5]. Christian Dondrup used laser and RGB-D camera information as inputs and fused different types of sensor data, which finally realizes human tracking with the Kalman filter algorithm [6]. The face detection and recognition based on RGB-D camera are studied by Wolfgang Rosenstiel and finally the person tracking is achieved [7]. However, the above researches have the common precondition that the motion of target is smooth, which means it is easy to lose the target when the motion is not smooth. Therefore, to improve the performance of tracking with the challenges such as rapid motion and target occlusion, the tracking strategy based on multi-feature fusion is

Y. Huang et al. (Eds.): ICIRA 2017, Part III, LNAI 10464, pp. 141–153, 2017.
DOI: 10.1007/978-3-319-65298-6_14

proposed in the paper, which contains the motion model similarity, color histogram similarity and human HOG feature similarity, realizing the person tracking by the method of joint likelihood data association. In addition, to deal with the target loss problem in tracking process, the quick search strategy is proposed which can deal with the problems such as similar color interference, target loss and occlusion.

2 Overall Design of Tracking Method

The person tracking method based on multi-feature fusion involves three aspects: (1) motion trend estimation of target person; (2) the feature of color distribution statistics for target area; (3) morphological feature of target. The three channels of information are fused and the tracking is realized by the joint likelihood data association, which can improve the stability of tracking with the challenge of dynamic interference. The overall design of tracking method is shown as Fig. 1.

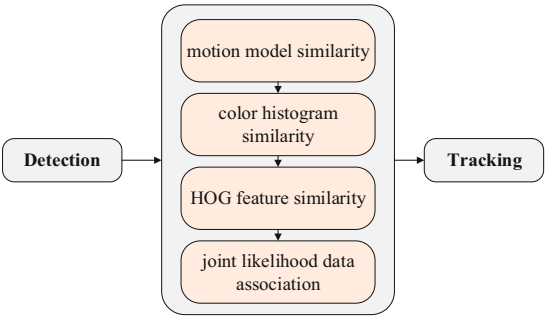

Fig. 1. Framework of target person tracking method

As shown in Fig. 1, The calculated confidence of detected body is the input of the tracking module. Firstly, the motion model is built with the target position forecast by the particle filter algorithm, after which the similarity is calculated based on the predicted position and actual position. Besides, the color histogram similarity is calculated based on color histogram of target body and detected body, based on which the HOG feature similarity is calculated later. Finally, the tracking is realized by the method of joint likelihood data association which synthetically considers the motion model similarity, color histogram similarity and HOG feature similarity. The joint similarity is used to judge whether the detected body is the target person, which is stable and reliable in the dynamic environment.

3 Similarity Calculation

3.1 Motion Model Similarity

Motion Model Establishment

The motion model similarity can be calculated by the Mahalanobis distance between the predicted position and actual detected position. In dynamic and complex environment, the constant velocity model is established considering the interferential objects and occlusion, which can predicate the person position conveniently based on the particle filter algorithm [8]. Taking x and y to describe the position in environment. In motion model, the particle filter need to be updated based on the position of each frame. The state vector and observation vector are as Eqs. (1) and (2).

$$x_k = (x \quad y \quad \dot{x} \quad \dot{y})^T \tag{1}$$

$$z_k = (x \quad y)^T \tag{2}$$

The mean value is zero and the variance value is σ_a^2 with the assumption that the directions of x and y are satisfied with the normal distribution. The motion equation is shown as Eq. (3).

$$x_k = F \cdot x_{k-1} + G \cdot a_k \tag{3}$$

$$F = \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, G = \begin{pmatrix} \frac{\Delta t^2}{2} & 0 \\ 0 & \frac{\Delta t^2}{2} \\ \Delta t & 0 \\ 0 & \Delta t \end{pmatrix}, a_k = \begin{pmatrix} \ddot{x} \\ \ddot{y} \end{pmatrix} \tag{4}$$

The dynamic system model of filter is shown in Eq. (5):

$$\begin{cases} x_k = F \cdot x_{k-1} + w_k \\ z_k = H \cdot x_k + v_k \end{cases} \tag{5}$$

In Eq. (5), the process noise w_k stratifies the facts that mean value is zero and it is multi-parameter normal distribution. The variance is $Q(w_k \sim \mathrm{N}(0, Q))$ where:

$$Q = G^T G \cdot \sum\nolimits_a = \begin{pmatrix} \frac{\Delta t^2}{2} & 0 \\ 0 & \frac{\Delta t^2}{2} \\ \Delta t & 0 \\ 0 & \Delta t \end{pmatrix} \cdot \begin{pmatrix} \frac{\Delta t^2}{2} & 0 \\ 0 & \frac{\Delta t^2}{2} \\ \Delta t & 0 \\ 0 & \Delta t \end{pmatrix}^T \cdot \begin{pmatrix} \sigma_{ax}^2 \\ \sigma_{ay}^2 \end{pmatrix} \tag{6}$$

The parameters of observation model are as followings:

$$\begin{aligned} v_k &\sim \mathrm{N}(0, R_k) \\ H &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \\ R_k &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot (\sigma_v^2 + \sigma_d^2) \end{aligned} \tag{7}$$

The covariance matrix of noise is the sum of σ_v^2 with σ_d^2, which respectively represent the quantization error of voxel grid filter and the measuring error of RGB-D camera.

Target Position Prediction

The target position prediction is accomplished by the particle filter algorithm. The posterior probability density is estimated based on the priori probability density and current observation value. Firstly, the point cloud data is obtained from the RGB-D camera and the points are initialized as particles, after which each particle is assigned a weight and the weights normalization is finished. The target position can be replaced by the weighted average value of all particles [9].

① State probability distribution

The state probability density can be approximated by the empirical probability distribution. $z_{1:k}$ is the measurement set at time of k, $x_k^{(i)}$ represents the particle i, $\delta(\cdot)$ is the Dirac function, and $P(x_{0:k}|z_{1:k})$ is the probability density of x.

$$P(x_{0:k}|z_{1:k}) = \frac{1}{N}\sum_{i=1}^{N} \delta x_k^{(i)}(dx_{0:k}) \tag{8}$$

② Particles generation

The particles $X_0^i \sim P(x_{0:k}|z_{1:k}), i = 1, \ldots, M$ can be sampled from $P(x_{0:k}|z_{1:k})$.

③ Weight calculation

To solve the difficulty in sampling from $P(x_{0:k}|z_{1:k})$, the known probability distribution function $q(x_{0:k}|z_{1:k})$ is introduced and the weight can be calculated as Eq. (9).

$$w_k^i = w_{k-1}^i \frac{P(z_k|x_k^i)P(x_k^i|x_{k-1}^i)}{q(x_k^i|x_{0:k-1}^i, z_{0:k})} \tag{9}$$

④ Weight normalization

Weights normalization can be calculated as Eq. (10).

$$w_k^{(i)} = \frac{w_k^i}{\sum_{i=1}^{M} w_k^i} \tag{10}$$

⑤ Target position prediction

The target position can be replaced by the weighted average value of all particles. At the time of k, the weight of i-th particle is $w_k^{(i)}$, the target position can be predicted based on the following equation.

$$x_k = \sum_{i=1}^{N} x_k^{(i)} w_k^{(i)} \tag{11}$$

Motion Model Similarity Calculation

The motion model similarity can be calculated by the Mahalanobis distance between the target predicted position and the actual detected position. The Mahalanobis distance between target person i and detected person j can be calculated as Eq. (12), where $i = 1, j = 1, 2, 3, 4$.

$$D_M = \tilde{z}_k^T(i,j) \cdot S_k^{-1}(i) \cdot \tilde{z}_k(i,j) \tag{12}$$

$S_k(i)$ is the covariance matrix of tracking i, $\tilde{z}_k(i,j)$ is the residual vector between measurement vector $z_k(i,j)$ and the forecast vector $\hat{z}_{k|k-1}(i)$ as Eq. (13).

$$\tilde{z}_k(i,j) = z_k(i,j) - \hat{z}_{k|k-1}(i) \tag{13}$$

Where $z_k(i,j)$ consists of position information of measurement j and velocity information of tracking i. When $i = 1$, D_M can be represented as:

$$D_{_M} = [D_{1,1}, D_{1,2} \ldots, D_{1,j}]^T \tag{14}$$

The similarity between predicted target position and actual detected position can be represented as Eq. (15), where $D_{\max} = \max\{D_{1,1}, D_{1,2}, \cdots, D_{1,j}\}$. The similarity is lower if the Mahalanobis distance is larger.

$$\rho_M = \begin{bmatrix} 1 - D_{1,1}/D_{\max} \\ 1 - D_{1,2}/D_{\max} \\ \cdots \\ 1 - D_{1,j}/D_{\max} \end{bmatrix} \tag{15}$$

3.2 Color Histogram Similarity

In the process of target tracking, it is necessary to compare the color feature information in the current view. The similarity between the color histogram feature of the regional image and the target is calculated to judge whether the detected object is the target person. Commonly used similarity calculation methods are Euclidean distance, Pasteur distance. The Pasteur coefficient is used in this paper to represent the similarity of the target area. The Bhattacharyya coefficient is a discrete probability density function. Take $\vec{p}_i$ and $\vec{q}_i$ respectively as the color histogram eigenvector of the target template and the region to be observed. The similarity coefficient is defined as Eq. (16).

$$\rho_{color} = \rho(\vec{p}, \vec{q}) = \sum_{i=1}^{m} \sqrt{\vec{p}_i, \vec{q}_i} \tag{16}$$

It can be seen from Eq. (16) that the Bhattacharyya coefficient is proportional to the similarity of the color histogram, which means that the larger the Bhattacharyya coefficient is, the higher the similarity. The distance between the target template and the observation area is calculated as in Eq. (17), where the smaller the Pasteur distance is, the greater the similarity.

$$d(\vec{p}, \vec{q}) = \sqrt{1 - \rho(\vec{p}, \vec{q})} \tag{17}$$

The color histogram feature of the target person is extracted, and whether the detected body is the target person is determined by calculating the similarity between the color histogram feature of the target person and each detected body.

3.3 Human HOG Feature Similarity

Human body detection based in HOG and SVM can only judge whether an object is a body, which may be the target person, the interferential person or human-like object. Therefore, it is necessary to calculate the human feature similarity to exclude the interferences. The HOG feature similarity can be obtained from the HOG feature of the human body. After a large number of experiments, the results can be summarized that the minimum value of confidence threshold min_*confidence* is −1.5 and the maximum confidence max_*confidence* is 2.0.

$$\rho_{HOG} = \frac{person\_confidence - \min\_confidence}{\max\_confidece - \min\_confidence} \tag{18}$$

The HOG feature confidence test is performed on the standing postures of different people, different sexes and similar human body. The similarity of HOG feature is calculated, and the test results are shown in Table 1. From the table it can be seen that the similarity is 0.35 when the human-like body is taken for the real person by mistaken, while the real body similarity is 0.70 or so, which means the similarity of human-like body is lower compared to the real human body.

Table 1. Similarity of human features based on HOG

Situation	HOG feature similarity
Stand upright	0.71
Stand side	0.73
Stand upright (Male)	0.69
Stand upright (Female)	0.71
Human-like body	0.35

4 Joint Likelihood Data Association

4.1 Correlation Probability Calculation

Data association is the key to tracking implementation. There are many methods of data association, such as Nearest Neighbor Data Association (NNDA), Probabilistic Data Association (PDA), Joint Probability Data Association (JPDA, often referred to as Joint Likelihood Data Association). The joint likelihood data association is a good compromise in the performance and computational loss when the human body is not very dense. Therefore the joint likelihood probability is used in this paper to deal with data. The correlation probability between each detection object j and the tracking target i is calculated as Eq. (19), which involves three likelihood probabilities which respectively based on motion model, color information and HOG feature.

$$L_{TOT}^{i,j} = L_{motion}^{i,j} \cdot L_{color}^{i,j} \cdot L_{HOG}^{i,j} \tag{19}$$

If the three likelihood probabilities are large, the combined likelihood probability obtained by multiplication is greater than a certain threshold, which means the detected body can be considered as the tracking target. The greater the value is, the greater the likelihood that the human body will be the target, that is the likelihood of detecting the human body as a tracking target is proportional to the joint likelihood. In order to simplify the calculation, Eq. (19) is changed to Eq. (20). After the conversion, the monotonic relationship has changed. The smaller $l_{TOT}^{i,j}$ is, the possibility is larger.

$$l_{TOT}^{i,j} = -\log(L_{TOT}^{i,j}) = \gamma \cdot D_M^{i,j} + \alpha \cdot c_{color}^{i,j} + \beta \cdot c_{HOG}^{j} \tag{20}$$

The $D_M^{i,j}$ is the Mahalanobis distance between the tracking i and detection j. The $c_{color}^{i,j}$ is the color histogram confidence between tracking i and detection j. The c_{HOG}^{j} is the HOG feature confidence of detection j. Coefficients γ, α and β are determined by the experience, which are respectively 0.9, 0.9 and 0.7, so the minimum value of joint confidence is 0.567.

4.2 Tracking Process Realization

Only the target person is in the robot view at the beginning. The detected body is initialized as the tracking target.

① Weight update
When the human body appears occlusion, the visibility of the human body color feature decreases, and the weight of the color feature need to be reduced. When the target is lost and then returned to the view of robot, the similarity of the motion model is unreliable, and the weight of the motion model based on the motion model needs to be reduced.

② Target distinction
If the interferential body has the same color with target person, it cannot be distinguished by the feature information based on the color histogram. In this situation, the target person position can be predicted by the particle filter algorithm. Since the position of the tracking target person is constant in a short time. Therefore, using the particle filter algorithm to predict the location of the human body can deal with the interference of same color feature, and the robot can distinguish the target to achieve tracking.

③ Tracking recovery
When the target is lost, the robot need to search the target (the search strategy will be introduced in next part). In the search process, firstly the color histogram similarity is calculated if a possible target is found and then the HOG feature similarity is calculated to judge whether the searched body is the human body.

5 Search Strategy for Target Loss

It was found that when the target person quickly disappeared form the robot vision, the robot would not detect the target person. In this situation, the target is lost and the robot need to retrieve the target again. There are many ways to search the target, such as blob detection, autonomous rotation detection. These methods are global search strategy, and the search takes a long time [11]. In order to improve the reliability, the quick search strategy need to be designed.

Assume that before the target is lost, the target center is defined as C, the angle of the movement as $\phi(\phi \in [-180^{\circ}, +180^{\circ}])$, and the outer rectangle of the target is determined by the height and width. According to the principle of motion continuity, the most likely position is the position near the center and at the range of ϕ.

The N is the search layer, the d is the search radius. On each layer, there are $8 * N$ search directions. If search starts at the direction n ($n = 0, 1, \ldots, 8 * N$), the following directions are: n, $(n+1) \bmod (8 * N)$, $(n-1) \bmod (8 * N)$, $(n+2) \bmod (8 * N) \cdots$. The search radius d is defined as $d \in ((N+1) * width, (N+1) * height)$, the n can be obtained by $n = floor(\frac{angle}{360/(N*8)}, n \subset \{0, 1, \ldots, 7\})$, where $angle = \begin{cases} \varphi, & if(\varphi \geq 0) \\ -\varphi, & if(\varphi < 0) \end{cases}$. The $floor(x)$ means to get the maximum value that is not larger than x. For example, if $N = 1$, $\phi = 45^{\circ}$. The search sequence is $(1, 0, 2, 7, 3, 6, 4, 5)$. Actually, if the first layer search is finished and the target is still not be found, the robot will stop searching next layer, alert and wait the target to back to view, avoiding useless search.

6 Experimental Results and Analysis

Considering the fact that the camera is moving, the joint likelihood tracking strategy involving the motion model, color features and HOG feature is proposed in this paper. In order to verify the effectiveness of the tracking strategy, the proposed method is validated by the mobile service robot experiment platform (Fig. 2). The tracking program is developed on the ROS platform. The tracking algorithm is implemented with OpenNI and point cloud library PCL. The tracking algorithm is evaluated by a series of experiments. For the particle filter algorithm, the number of particles is initialized to 100.

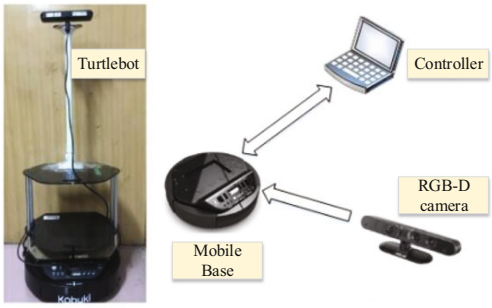

Fig. 2. The hardware platform

6.1 Target Tracking Experiment

The initial distance between the target person and the robot is 1.0 m, and a constant distance of 0.8 m is maintained between the robot and the target during the tracking process. In the case of single target without interference, the robot can reliably track the target person. During the tracking process, the robot and the target are moving. The speed of target person is about 0.3 m/s. The routes are shown in Fig. 3.

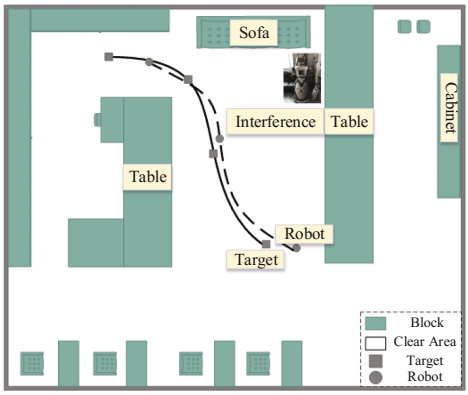

Fig. 3. Trajectory of tracking in case of single person

The video sequences of target tracking for single-person without interference are shown in Fig. 4, which are respectively the first frame, the 67th frame, the 189th frame, and the 265th frame. The first frame is at the beginning time. In the 67th frame, the target is moving forward. In the 189th frame, the target is turning left and the 265th frame shows that the target is going to stop. In this case, the similarity of the motion model is about 0.92, the similarity of the color feature is about 0.91, the similarity of the HOG feature is about 0.7, and the joint similarity is about 0.6. The experiment shows that the tracking is stable and effective.

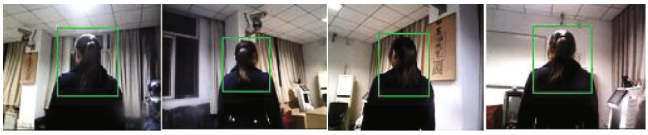

Fig. 4. Video tracking sequence of single person

The tracking sequence of confidence diagram is shown in Fig. 5. It can be seen that there is an adjustment process at the beginning of the tracking process and the tracking similarity is not stable, After 60 frames, the joint similarity tends to be stable (about 0.6), at which time the robot stably tracks the target person.

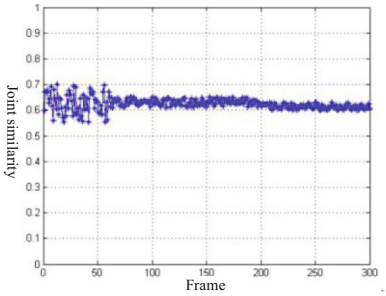

Fig. 5. Confidence diagram of single target

6.2 Target Loss Experiment

In order to verify the validity of the fast search strategy in the case of target loss, the target person quickly moves out of the robot vision and the robot will lose the target. The robot firstly searches in front, turns right and then turns left. If the target is searched, the target will be tracked again. The video sequences of tracking for target loss are shown in Fig. 6. The first row of images shows that the target quickly lost. The second row of images shows that the target be searched.

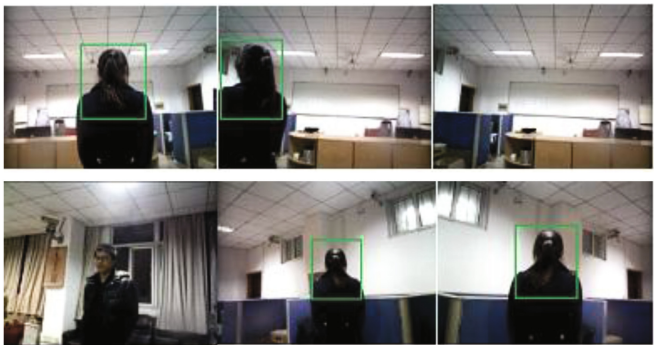

Fig. 6. Video tracking sequence of target loss

The tracking sequence of confidence diagram is shown in Fig. 7. It can be seen that when the target is tracked, the joint similarity is high at the beginning, when the target is quickly lost and the target is not searched in the right direction, the joint similarity is zero. When the target is lost, the motion feature will be unreliable and its coefficient will be zero, and the coefficients based on the color information and the HOG feature will be correspondingly larger.

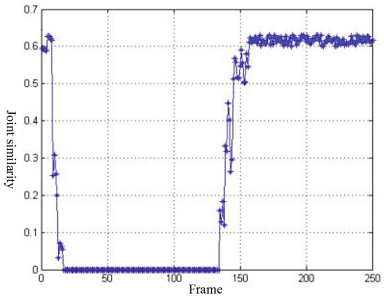

Fig. 7. Confidence diagram of tracking sequence for target loss

6.3 Target Occlusion Experiment

The video sequences of tracking for target occlusion are shown in Fig. 8. The third image shows that the target is occluded by the interferential body and the robot will search forward until the block disappears. The fourth image shows that the block disappears and the robot continues tracking. When the target is blocked, the coefficient based on the color feature will be unreliable and its coefficient will be zero, and the coefficients based on the motion model information and the HOG feature will be adjusted to one-half, respectively.

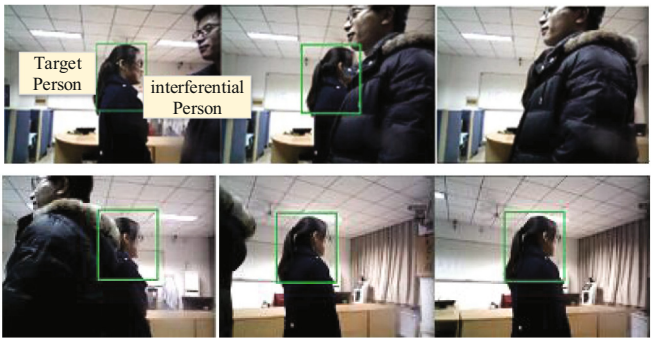

Fig. 8. Video tracking sequence in case of target occlusion

The results of the experiment are shown in Table 2 by using the joint likelihood data association algorithm, the particle filter tracking algorithm and the Cam-Shift tracking algorithm respectively. The experimental results show that joint likelihood data association algorithm has better tracking performance in scenes of lighting change, target occlusion, target loss and interference, compared with the other two algorithm.

Table 2. Comparison of experimental results of three algorithms

Method	Method in this paper	Particle filter	Cam-Shift
Normal	30	25	22
Lighting change	30	11	9
Target occlusion	28	20	5
Interference	30	21	23
Target loss	26	19	2

7 Conclusion

Aiming at the target tracking of mobile service robot in complex and dynamic environment, a multi-feature target tracking strategy is proposed in this paper. The key of the tracking strategy is the joint likelihood data association method which includes three similarities respectively based on the motion model, the color histogram feature and the HOG feature. The strategy can solve the problems such as same color interference, target loss and recovery, target occlusion, etc. In addition, considering the target loss, a quick search strategy is designed. Finally, the strategy is verified by experiments.

Acknowledgment. The authors would like to acknowledge the valuable support of Natural Sciences Foundation (NNSF) of China (No. 61573100, No. 61573101).

References

1. Takashi, Y., Nishiyama, M., Sonoura, T., et al.: Development of a person following robot with vision based target detection. In: Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, pp. 5286–5291. IEEE (2006)
2. Mekonnen, A., Lerasle, F., Herbulot, A.: Cooperative passersby tracking with a mobile robot and external cameras. Comput. Vis. Image Underst. **117**(10), 1229–1244 (2013)
3. Huazhu, F., Dong, X., Stephen, L.: Object-based RGBD image co-segmentation with mutex constraint, pp. 4428–4436. School of Computer Engineering, Nanyang Technological University (2015)
4. Gritti, A., Tarabini, O., Guzzi, J., et al.: Kinect-based people detection and tracking from small-footprint ground robots. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014), pp. 4096–4103. IEEE (2014)
5. Vo, D., Jiang, L., Zell, A.: Real time person detection and tracking by mobile robots using RGB-D images. In: IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 689–694. IEEE (2014)
6. Tasaki, T., Ozaki, F., Matsuhira, N., et al.: People detection based on spatial mapping of friendliness and floor boundary points for a mobile navigation robot. J. Robot. **1**, 1–10 (2011)
7. Choi, W., Pantofaru, C., Savarese, S.: Detecting and tracking people using an RGB-D camera via multiple detector fusion. In: Computer Vision Workshops (ICCV Workshops), pp. 1076–1083 (2011)
8. Munaro, M., Menegatti, E.: Fast RGB-D people tracking for service robots. Auton. Robots **37**(3), 227–242 (2014)
9. Bergen, J.R., Anandan, P., Hanna, K.J., Hingorani, R.: Hierarchical model-based motion estimation. In: Sandini, G. (ed.) ECCV 1992. LNCS, vol. 588, pp. 237–252. Springer, Heidelberg (1992). doi:10.1007/3-540-55426-2_27
10. Nummiaro, K., Koller-Meier, E., Van Gool, L.: An adaptive color-based particle filter. Image Vis. Comput. **21**(1), 99–110 (2003)
11. Kailath, T.: The divergence and Bhattacharyya distance measures in signal selection. IEEE Trans. Commun. Technol. **15**(1), 52–60 (1967)