

# A Research of Targets Tracking and Positioning Algorithm Based on Multi-feature Fusion

Pengfeng Chen and Long Zhao

**Abstract** In order to improve the accuracy and robustness of targets tracking and positioning, this paper proposes a particle filter algorithm based on multi-feature fusion. According to the diversity of character information, a multi-feature fusion strategy based on color, texture, and edge character has been developed, which can realize the comprehensive utilization of various visual features information. A feature criterion function has been addressed to the weighted strategy of multi-feature fusion, which can adjust adaptively the weight of feature and enhance the reliability of target tracking. Combined with binocular stereo vision technology, this algorithm can locate the target by calculating the geometrical relationship between the correspondence pixel points and spatial points. The test results show that the algorithm can realize the target tracking and positioning more accurately.

**Keywords** Multi-feature fusion · Adaptive weights · Particle filter · Target tracking · Binocular stereo vision · Target positioning

## 1 Introduction

Target tracking and locating is a prosperous research filed in computer vision which can be applied to many fields such as the pedestrian tracking [1], security surveillance, intelligent transportation, military, etc. However, restricted by the complex environment, positioning accuracy and device factor, the target tracking and locating methods still remain to be further improved.

---

P. Chen  
Digital Navigation Center, Beihang University, Beijing, China

L. Zhao (✉)  
Science and Technology on Aircraft Control Laboratory, Beihang University,  
Beijing, China  
e-mail: buaa\_dnc@buaa.edu.cn

Tracking location research can be divided into two parts: target tracking and target positioning. In target tracking, there is no single visual feature that can describe the target information completely. Tracking algorithms based on single feature, therefore, are often limited by the practical application environment. According to the intrinsic differences between different features, a feasible solution can be obtained by multi-feature fusion [2] to improve the accuracy and robustness of tracking algorithm. In the research of target positioning, the commonly used methods include global positioning system (GPS), Bluetooth, radio frequency identification (RFID), infrared, computer vision, etc. GPS signal is easily weakened in indoor or high dense-constructed area; Infrared positioning technology needs to launch modulation infrared ray, it can only support the line-of-sight transmission because the ray cannot through the obstacles. In addition, most position technologies need expensive reception terminal devices and relevant communication sensors for targets, so that they are limited to be widely used. Traditional positioning technology [3] based on computer vision can be divided into three sub-processes, i.e., stereo matching, three dimensional reconstruction, and calculating spatial coordinates. Because the stereo vision matching needs to spend much time to extract the feature points and has mismatching feature points, real-time performance and robustness of the algorithm are reduced.

In order to solve this problem, an algorithm combined the adaptive weighting particle filter algorithm and multi-feature fusion is presented according to the observation probability density fusion of the particle filter algorithm. In tracking model, color histograms [4], local binary pattern (LBP) and edge character were chosen to describe the target area. Color histogram, LBP, and edge character describe the color distribution of target area, texture information, and the association between the surrounding pixels and target edge information, respectively. In the fusion strategy, the adaptive weighting can be obtained through calculating the probability density distribution of corresponding characteristics for the target area and background area. The adaptive weighting particle filter algorithm based on multi-feature fusion, not only focuses on the similarity between the candidate area and the target area, but also considers distinction between target region and background region. The proposed tracking algorithms can enhance the target tracking precision, and estimate the centroid position of motion target in the image, which can avoid the complex stereo vision matching process in visual positioning. For this reason, the spatial location of target can be calculated by using binocular vision technology and the performance of the algorithm presented in this paper is verified by a practical experimental test.

The paper is organized as follows. The feature extraction is presented in Sect. 2. Section 3 focuses on the target tracking with multi-feature fusion. Section 4 states how to calculate the spatial location. Section 5 gives the experiment analysis. Finally, we draw some conclusions and shed light on future work in Sect. 6.

## 2 Target Feature Extraction

In the image sequences, there are many typical characteristics that can be used for target tracking. Because each character has its own special properties and different sensitivity to the environment, the choice of characteristics is very significant problem. Taking the distinguishability, stability, and independence of the features into account, the color histogram, LBP, and edge character are chosen as target visual features in this paper.

### 2.1 Color Histogram

Color histogram describes the color distribution in the form of histogram by computing the value of each pixel in the target area. In order to increase the reliability of the distribution, a kernel function is used to compute the weights for every pixel of the target. Then the target model are described as follows:

$$q_{\text{color}}(u) = f \frac{1}{N} \sum_{i=1}^N k\left(\left\|\frac{x_i - x_0}{h}\right\|\right) \delta[b(x_i) - u], \quad u = 1, 2, \dots, m \quad (2.1)$$

$$k(x) = \begin{cases} 1 - x^2 & x \leq 1 \\ 0 & \text{others} \end{cases} \quad (2.2)$$

$$f = 1 / \sum_{i=1}^N k\left(\left\|\frac{x_i - x_0}{h}\right\|\right) \quad (2.3)$$

where  $k(x)$  is the weighted kernel function,  $\delta$  is the Kronecker delta function,  $f$  is the normalization factor, and  $h$  is defined by  $h = \sqrt{h_x^2 + h_y^2}$ , in which  $h_x$  and  $h_y$  is the half width, and half height of the target area, respectively.

### 2.2 Local Binary Pattern

The LBP operator [5], shown as a powerful measure of image, can extract the texture feature of target easily. The LBP operator labels the pixel in an image by comparing its neighborhood with the center value and considering the results as a binary number (binary pattern). The LBP operator is described by

$$\text{LBP}(x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(i_p - i_c) \quad (2.4)$$

where  $i_c$  denotes the gray value of the center pixel  $(x_c, y_c)$ ,  $i_p$  denotes the gray value of the neighborhood, and  $P$  is the number of its neighborhood. The function  $s(x)$  is defined as follow

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (2.5)$$

Calculate LBP value of each pixel according to Eq. (2.5), then the texture histogram of LBP character is normalized as

$$q_{\text{color}}(u) = \frac{1}{N} \sum_{i=1}^N \delta[b(x_i - u)] \quad (2.6)$$

where  $b(x_i)$  is the index function, returning the histogram serial number of pixel  $x_i$ ,  $N$  is the total number of pixels in target area.

### 2.3 Edge Feature

The edge feature [6] can be described by histogram of oriented gradient. In gray image, the horizontal edge  $G_x(x_i)$  and vertical edge  $G_y(x_i)$  can be acquired by using edge detection operators to calculate the edge information in  $X$  and  $Y$  direction, respectively. The gradient  $G(x_i)$  and orientation  $\theta$  of each pixel can be defined as

$$\begin{aligned} G(x_i) &= \sqrt{G_x^2(x_i) + G_y^2(x_i)} \\ \theta &= \arctan(G_y(x_i)/G_x(x_i)) \end{aligned} \quad (2.7)$$

The weighted oriented gradient histogram is calculated by dividing the orientation into  $m$  bins. The histogram of oriented gradient is given by

$$q_{\text{edge}}(u) = C \sum_{i=1}^N k\left(\left\|\frac{x_i - x_0}{h}\right\|\right) G(x_i) \delta[b(x_i) - u] \quad (2.8)$$

where kernel function  $k(x)$  is shown as Eq. (2.2), the normalization factor  $C$  is defined as

$$C = 1 / \sum_{i=1}^N G(x_i) \quad (2.9)$$

### 3 Target Tracking with Adaptively Multi-feature Fusion

#### 3.1 Multi-feature Fusion

After extracting the target features, the multi-feature fusion is realized by the Gaussian weighted strategy. Given the target state  $x_k$  and time  $k$ , the entire observation likelihood can be calculated by

$$p(z_k|x_k) = \exp\left\{-\frac{1}{2}\left(\alpha\frac{d_{\text{color}}^2}{\sigma_c^2} + \beta\frac{d_{\text{texture}}^2}{\sigma_t^2} + \gamma\frac{d_{\text{edge}}^2}{\sigma_e^2}\right)\right\} \quad (3.10)$$

where  $\sigma_c^2$ ,  $\sigma_t^2$  and  $\sigma_e^2$  denote the noise variance of the color histogram, LBP character, edge feature, respectively;  $d_{\text{color}}$ ,  $d_{\text{texture}}$  and  $d_{\text{edge}}$  denote the similarity distance of the color histogram, LBP character, edge feature between target area, and candidate area;  $\alpha$ ,  $\beta$  and  $\gamma$  are the weights of the color, texture, and edge features, respectively.

The similarity distance  $d_f$  can be calculated by Bhattacharyya coefficient. During the current frame, assuming that the likelihood of candidate area is  $p_f$  while the target area is  $q_f$ , the similarity between candidate area and target area is described as follows

$$\rho_f(q_f, p_f) = \sum_{u=1}^m \sqrt{q_f(u)p_f(u)}, \quad f \in \{\text{color}, \text{texture}, \text{edge}\} \quad (3.11)$$

where

$$d_f = \sqrt{1 - \rho_f(q_f, p_f)}, \quad f \in \{\text{color}, \text{texture}, \text{edge}\} \quad (3.12)$$

The smaller value of  $d_f$  indicates the higher similarity of character between candidate area and target area, then particle gets bigger weighting in Eq. (3.10).

#### 3.2 Adaptive Weights

If multiple features are characterized by equal or fixed weighting without discrimination in multi-feature fusion, the target tracking would be seriously disturbed by some features, which is sensitive on the occasion of varying environment. In this paper, an evaluating function, used to reflect the ability of identification, has been addressed to calculate the weights of features [7].

Based on the sum of intraclass and interclass variance, the evaluation function for the effectiveness of characteristic is defined as follow:

$$\text{var}(q_f; \mu_f^t, \mu_f^b) = \frac{w_f^t w_f^b (\mu_f^t - \mu_f^b)^2}{w_f^t \sigma_{ft}^2 + w_f^b \sigma_{fb}^2}, \quad f \in \{ \text{color}, \text{ texture}, \text{ edge} \} \quad (3.13)$$

where  $\mu_f^t$  and  $\mu_f^b$  denote the average of target and background area's correspondence characteristic value;  $w_f^t$  and  $w_f^b$  denote the weights of target and background area's correspondence characteristic value;  $\sigma_{ft}^2$  and  $\sigma_{fb}^2$  denote the variance of target and background area's correspondence characteristic value. In Eq. (3.13), the molecules are the sum of interclass, the denominators is the sum of intraclass variance. According to pattern recognition theory, the larger variance in interclass, while the smaller variance in intraclass, the discernment for target and background is more robust.

As mentioned above, the weights can be calculated by

$$\begin{cases} \alpha = \frac{\text{var}(q_{\text{color}}; \mu_{\text{color}}^t, \mu_{\text{color}}^b)}{\sum \text{var}(q_f; \mu_f^t, \mu_f^b)} \\ \beta = \frac{\text{var}(q_{\text{texture}}; \mu_{\text{texture}}^t, \mu_{\text{texture}}^b)}{\sum \text{var}(q_f; \mu_f^t, \mu_f^b)} \\ \gamma = \frac{\text{var}(q_{\text{edge}}; \mu_{\text{edge}}^t, \mu_{\text{edge}}^b)}{\sum \text{var}(q_f; \mu_f^t, \mu_f^b)} \end{cases}, \quad f = \{ \text{color}, \text{ texture}, \text{ edge} \} \quad (3.14)$$

### 3.3 Tracking Algorithm Based on Particle Filter

This paper uses the color histogram, LBP texture, and histogram of oriented gradient to describe the target information and embeds it into the particle filter tracking.

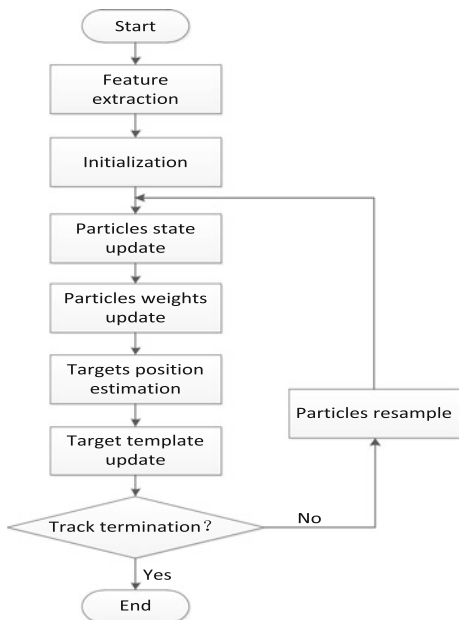
The diagram of the tracking algorithm is shown in Fig. 1.

## 4 Target Positioning Based on Binocular Vision

After robust tracking, centroid coordinates of targets in binocular video sequences can be acquired. On the basis, the spatial location of target can be calculated with the use of binocular vision technology [8].

In stereo vision system, an arbitrary point  $P(X_w, Y_w, Z_w)$  is fallen the left and right image on each  $P_L(u_l, v_l)$  and  $P_R(u_r, v_r)$ . Assume that projection matrix  $M_L$  and

**Fig. 1** The structure of tracking algorithm based on particle filter



$M_R$  denote the geometric relationship from  $P$  to  $P_L$ , from  $P$  to  $P_R$ , respectively, then model is shown as follows

$$z_{cl} \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^l & m_{12}^l & m_{13}^l & m_{14}^l \\ m_{21}^l & m_{22}^l & m_{23}^l & m_{24}^l \\ m_{31}^l & m_{32}^l & m_{33}^l & m_{34}^l \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M_1 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (4.15)$$

$$z_{cr} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11}^r & m_{12}^r & m_{13}^r & m_{14}^r \\ m_{21}^r & m_{22}^r & m_{23}^r & m_{24}^r \\ m_{31}^r & m_{32}^r & m_{33}^r & m_{34}^r \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M_2 \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (4.16)$$

where  $M_i, i = L, R$  is a  $3 \times 4$  matrix,  $Z_{cl}$  and  $Z_{cr}$  denote the coordinates of the  $Z$  axis of  $P$  in the correspondence camera coordinate system.

To further simplify the formula, the model be calculated as follows

$$\begin{cases} u_l = \frac{m_{11}^l X_w + m_{12}^l Y_w + m_{13}^l Z_w + m_{14}^l}{m_{31}^l X_w + m_{32}^l Y_w + m_{33}^l Z_w + m_{34}^l} \\ v_l = \frac{m_{21}^l X_w + m_{22}^l Y_w + m_{23}^l Z_w + m_{24}^l}{m_{31}^l X_w + m_{32}^l Y_w + m_{33}^l Z_w + m_{34}^l} \end{cases} \quad (4.17)$$

$$\begin{cases} u_r = \frac{m_{11}^r X_w + m_{12}^r Y_w + m_{13}^r Z_w + m_{14}^r}{m_{31}^r X_w + m_{32}^r Y_w + m_{33}^r Z_w + m_{34}^r} \\ v_r = \frac{m_{21}^r X_w + m_{22}^r Y_w + m_{23}^r Z_w + m_{24}^r}{m_{31}^r X_w + m_{32}^r Y_w + m_{33}^r Z_w + m_{34}^r} \end{cases} \quad (4.18)$$

The model can be expressed in matrix equation as

$$A_i P = b_i \quad (4.19)$$

where  $A_i = \begin{bmatrix} m_{31}^i u_i - m_{11}^i & m_{32}^i u_i - m_{12}^i & m_{33}^i u_i - m_{13}^i \\ m_{31}^i v_i - m_{21}^i & m_{32}^i v_i - m_{22}^i & m_{33}^i v_i - m_{23}^i \end{bmatrix}$ ,  $b_i = \begin{bmatrix} m_{14}^i - m_{34}^i u_i \\ m_{24}^i - m_{34}^i v_i \end{bmatrix}$ ,  
 $i = 1, r$ .

Let  $A = (A_1, A_r)^T$ ,  $b = (b_1, b_r)^T$ , then

$$AP = b \quad (4.20)$$

Finally, based on least square method, the coordinate of  $P$  can be obtained by  $P = (A^T A)^{-1} A^T b$ .

## 5 Experiment

To evaluate the performance of the proposed algorithm, a series of experiments are implemented. The experiments are divided into two parts. The former aims to prove that the accuracy and robustness of particle filter tracking, and the latter is to verify the effectiveness of the positioning.

The test video 1 which has complicated environment is from the Visual Tracker Benchmark (video is available at <http://www.visual-tracking.net>), and its resolution is  $320 \times 240$  pixel. The test result is shown in Fig. 2.

As can be seen from Fig. 2, the particle filter based on color histogram is easily disturbed when the target's color is similar to background, and the algorithm presented in this paper can track the target accurately.

The test video 2 which is used to test the location accuracy is from Digital Navigation Center of Beihang University, and its resolution is  $1280 \times 960$  pixel. The test result is shown in Fig. 3.

Figure 3 gives the results of human tracking indoor. After tracking the target accurately, we can estimate the geometric center coordinate. Then the target location information can be calculated by Eq. (4.20).

In this experiment, we chose front-left corner of room as the origin of coordinate in the binocular positioning in general. The related parameters were measured as follows: camera installation height is 2 m, it's lens optical axis is inclined downward and the angle with vertical plane is  $\pi/3$ , two camera's axes are parallel with the horizontal space 45 cm, and the calibrated focus length is 3.6 mm. The calculated result of target location is shown in Table 1.



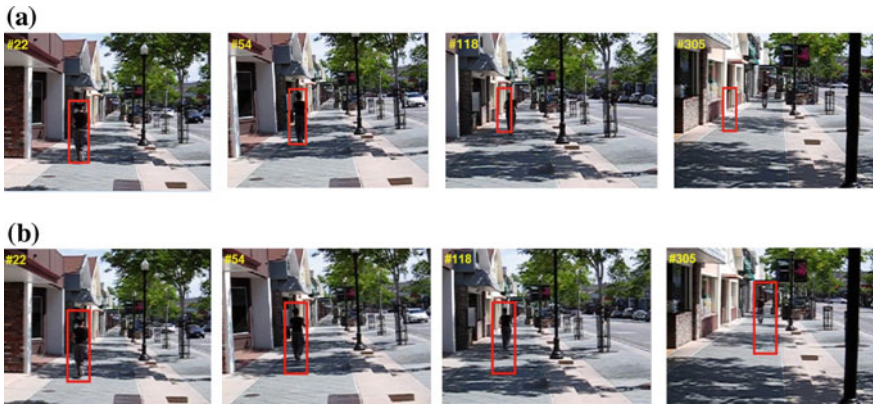


Fig. 2 The results of the targets tracking, a particle filter based on color histogram in literature [3], b particle filter based on multi-feature fusion

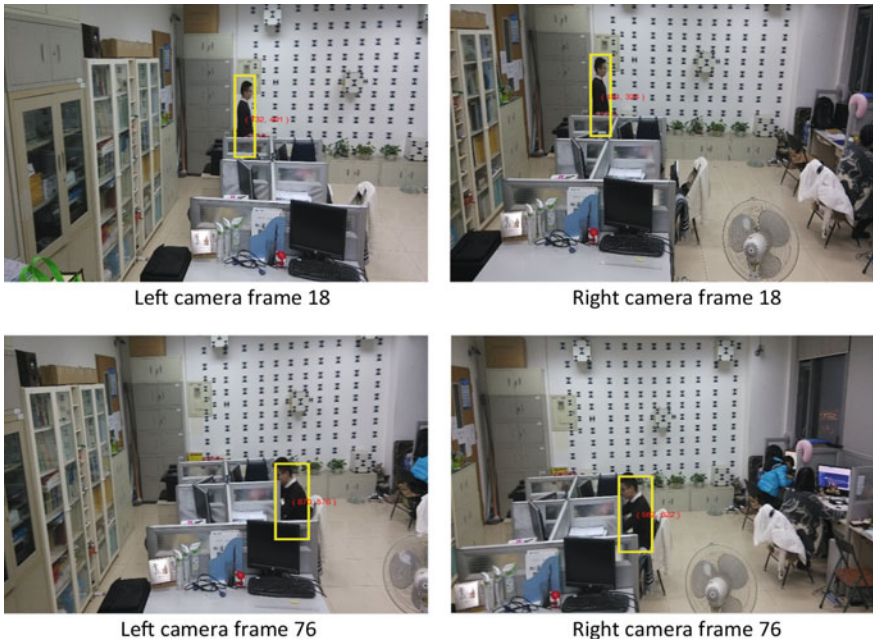


Fig. 3 The tracking results in left and right camera

Some samples of final positioning results are demonstrated in Table 1. The frame indexes are 18, 25, 34, 53, 58, 76, 83, and 97. The calculated mean absolute error of the top 100 frame is (13.32, 12.76, 8.15). The reason is that the estimation of targets' geometric center coordinate may not completely match in left and right

**Table 1** The tracking results in left and right camera

Frame number	World coordinate (cm)	Calculating coordinate (cm)
Frame 18	(137, 219, 110)	(131, 212, 108)
Frame 25	(142, 245, 110)	(134, 237, 106)
Frame 34	(148, 250, 110)	(156, 258, 117)
Frame 53	(275, 330, 110)	(269, 337, 103)
Frame 58	(278, 362, 65)	(287, 373, 73)
Frame 76	(405, 382, 65)	(398, 371, 58)
Frame 83	(527, 416, 65)	(543, 407, 70)
Frame 97	(599, 471, 110)	(608, 479, 102)
Mean absolute error of the top 100 frame (13.32, 12.76, 8.15)		

video sequences. Then this will lead to calculating error during binocular vision positioning. According to Table 1, we can see that the proposed positioning method can realize target positioning in the range of permitted error. The error mainly comes from the tracking error between left and right camera while parts from measurement error.

## 6 Conclusion

In this paper, a target tracking and locating algorithm was reported, including feature extraction, multi-feature fusion, particle filter tracking, and target positioning. The video test results have proved that the proposed algorithm can realize the target tracking and positioning. However, there are certain limitations of proposed algorithm. First, the tracking error may cause estimated error of targets' pixel coordinate. Second, the target real-time positioning is still challenging. All the mentioned contents need further improvement and are under investigation.

**Acknowledgements** This project is supported by the National Natural Science Foundation of China (Grant No. 41274038, 41574024), the Aeronautical Science Foundation of China (Gratis No. 2013ZC51027), and the Fundamental Research Funds for the Central Universities.

## References

1. Pai CJ, Tyan HR, Liang YM et al (2004) Pedestrian detection and tracking at crossroads. *Pattern Recogn* 37(5):1025–1034
2. Perez P, Vermak J, Blake A (2004) Data fusion for visual tracking with particles. *Proc IEEE* 92(3):495–513
3. Martins HA, Birk JR, Kelley RB (1981) Camera models based on data from two calibration planes. *Comput Graphics Imaging Process* 17:173–180

4. Nummiaro K, Koller-Meier E, Gool LV (2003) An adaptive color-based particle filter. *Image Vis Comput* 21(1):99–110
5. Ojala P, Harwood DM (1996) A comparative study of texture measures with classification based on feature distributions. *Pattern Recogn* 29(1):51–59
6. Guo S (2012) Approach of image edge detection based on wavelet scale multiplication. *Appl Mech Mater* 130–134:4282–4285
7. Shoushtarian B, Bez HE (2005) A practical adaptive approach for dynamic background subtraction using an invariant color model and object tracking. *Pattern Recogn Lett* 26(1):5–26
8. Zhang ZY (1997) Motion and structure from two perspective views: from essential parameters to euclidean motion via fundamental matrix. *J Opt Soc Am A* 14(11):2938–2950