

## A Method for Tracking Multiple Targets based on Multi-feature Fusion

Yanhai WU<sup>1</sup>, Rongrong ZHANG<sup>1</sup>, Nan Wu<sup>2</sup>, Jing Wang<sup>1</sup>

<sup>1</sup>Department of Telecommunication and Information Engineering, Xi'an University of Science and Technology, Xi'an, 710054, China

<sup>2</sup>Department of Computer Science and Technology, Shaanxi Xueqian Normal University, Xi'an, 710061, China

**Keywords:** multi-target tracking, feature fusion, particle Swarm optimization, Meanshift tracking algorithm, adjustable window.

**Abstract:** Traditional target tracking methods are easy to lose the target in complex background, especially tracking multiple targets. In this paper PSO (Particle Swarm Optimization) algorithm is presented, it is used to calculate the weights of features. and weight the features to get optimal feature, then apply the optimal feature to Meanshift. And, combine the adjustable window algorithm with Meanshift to adapt to the change of the target scale. The experiments' results show that the proposed algorithm has improved the robustness and accuracy than the traditional tracking algorithm in complex background and with multi-target.

### 1. Introduction

Multi-target tracking is actually calculate 2-D coordinate of moving targets in images<sup>[1]</sup>. In literature [2], track target by fusing R, G, B three color features, calculating the separability of each feature between target and background and weighting them. Considering the track box the Meanshift selects is fixed, it always lead to a failure when the target's scale is changeful. In literature [3], the method selects target by multiple tracking boxes for target template matching, and chose size of the best matching rate tracking box to update tracking box size, this method is easy to realize. In this paper, PSO is introduced to fuse of color, texture and edge, the fused feature is used as tracking feature of Meanshift, and a method for adjusting window is also mentioned.

### 2. A method of target tracking based on multi-feature fusion

#### 2.1 Target feature selection

Color feature<sup>[8]</sup> is robust to partial cover, changing in scale and target rotation, etc. Considering the image obtained directly is RGB format, we choose RGB color space for feature extraction and color histogram is used to describe target and background.

Texture<sup>[4]</sup> is robust to rotation and light changing. An operator of LBP is used to describe texture. The calculation for LBP is as follows: for an image, select a pixel as center point, set it's gray value as threshold. If the gray value of surrounding pixel is smaller than the threshold value, set the pixel's value to 0, else set to 1. Then obtain a binary sequence and weight it, this value is center's LBP value. In this paper, template for calculating LBP value is selected by 3 \* 3, the calculation for LBP can be described as:

$$\text{LBP}(x_k, y_k) = \sum_{n=0}^7 2^n S(i_n - i_k) \quad (1)$$

Where  $S(u) = \begin{cases} 0, & u < T \\ 1, & u \geq T \end{cases}$ ,  $i_n$  and  $i_k$  are gray value of pixel n and k, T is threshold value.

Edge features are insensitive to light or color changing. In this paper, Canny<sup>[5]</sup> edge detector is selected. Firstly, a Gaussian Filter is used to process the image and a gradient map is obtained by calculating the image's gradient. Then restrain non-maximum value of gradient map and reserve the maximum local gradient point. Last, use dual threshold to detect edge connection, that is, retain

edge pixel by a high threshold and remove noise by a low threshold. The edge obtained is close to the real.

## 2.2 Feature fusion criterion

In this paper, RGB color, LBP texture and Canny edge are selected. contrast of target and background under each feature was calculated firstly, the weight was calculated by the discrimination next, and last, weighting the three features.

The description of target and background is shown in Fig. 1, The region of the small rectangle means the target, the portion between rectangles is the background.



Fig. 1 description of target and background

Giving a feature  $g$ , the target histogram under this feature is  $H_{f_E}(i)$ , the background's is  $H_{b_E}(i)$ ,  $q(i)$  and  $p(i)$  respectively are empirical distribution probability density functions of background and target. The specific mathematical expression is:

$$\begin{cases} p(i) = H_{f_E}(i)/N_{f_E} \\ q(i) = H_{b_E}(i)/N_{b_E} \end{cases} \quad (2)$$

Where  $i$  is quantified series,  $N_{f_E}$  and  $N_{b_E}$  represent the number of pixels in target and background region. The distinction degree under this feature is represented by log-likelihood ratio function  $L(i)$ , which is defined as:

$$L(i) = \max \left\{ -1, \min \left[ 1, \log \left( \frac{\max(p(i), \epsilon)}{\max(q(i), \epsilon)} \right) \right] \right\} \quad (3)$$

Where  $\epsilon$  is a very small value to guarantee the denominator is not 0, set to 0.001 .

Use the variance of  $L(i)$  to describe discrimination between target and background, as follow:

$$v(L;p) = E[L^2(i)] - E^2[L(i)] \quad (4)$$

## 2.3 Multi-feature fusion based on PSO

In this paper, PSO<sup>[6]</sup> algorithm was used to compute features' weights, record the candidate features of color, texture and edge as C,L,E, denote the weights as  $\omega_1, \omega_2$  and  $\omega_3$ , the fusion feature is:

$$F = \omega_1 * C + \omega_2 * L + \omega_3 * E \quad (5)$$

The weight can be expressed in vector form:

$$W = [\omega_1, \omega_2, \omega_3] \quad (6)$$

The optimal solution is denoted as  $W'$ , and the calculation of  $W'$  is setting  $m$  particles in the  $D$ -dimensional space, the  $D$ -dimensional position vector of the  $i$ -th particle is expressed as  $z_i = (z_{i1}, z_{i2}, \dots, z_{id}, \dots, z_{iD})$ , move to the next position from the current position at the speed  $v_i = (v_{i1}, v_{i2}, \dots, v_{id}, \dots, v_{iD})$ . And the particle updated itself by two extremums: one is individual extremum  $p_b = (p_{b1}, p_{b2}, \dots, p_{bd}, \dots, p_{bD})$ , the other is global extremum  $p_g = (p_{g1}, p_{g2}, \dots, p_{gd}, \dots, p_{gD})$ . The expressions are as follows:

$$v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (p_{bd} - z_{id}^k) + c_2 r_2 (p_{gd} - z_{id}^k) \quad (7)$$

$$z_{id}^{k+1} = z_{id}^k + v_{id}^{k+1} \quad (8)$$

Where  $i = 1, 2, \dots, m$ ;  $d = 1, 2, \dots, D$ ;  $k$  are iterative times,  $r_1$  and  $r_2$  are random numbers between  $[0, 1]$ ,  $c_1$  and  $c_2$  are learning factors.

The fitness function value is calculated by the formulas (3) and (4). When all particles finish the iterations or the position invariant, get the  $p_g$  which is  $W'$ .

### 3. Meanshift algorithm

$\{x_i^*\}, i=1, \dots, n$  is a pixel set in target region. Kernel function  $k(x)$  is a function of isotropy and monotone decreasing<sup>[7]</sup>. Description of target's histogram is as follow:

$$\hat{q}_u = C \sum_{i=1}^n k(\|x_i^*\|^2) \delta[b(x_i^*) - u] \quad (9)$$

Where  $\delta(x)$  is Delta function,  $\delta(b(x_i^*) - u)$  is to judge whether the feature value of the pixels in target region is  $u$ , if true, set to 1, else set to 0.  $C$  is a standardized constant coefficient, to ensure  $\sum_{u=1}^m \hat{q}_u = 1$ . So there is  $C = [\sum_{i=1}^n k(\|x_i^*\|^2)]^{-1}$ .

Similarly, the Kernel histogram of candidate target model is described as follow:

$$\hat{p}_u(y) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \delta[b(x_i) - u] \quad (10)$$

Where  $C_h = [\sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right)]^{-1}$ .

The similarity of  $\hat{p}_u(y)$  and  $\hat{q}_u$  is expressed by Bhattacharyya (BH) coefficient :

$$\rho(y) \equiv \rho[\hat{p}(y), \hat{q}] = \sum_{u=1}^m \sqrt{\hat{p}_u(y) \hat{q}_u} \quad (11)$$

Taylor expansion of formula (12) in  $\hat{y}0$ , then we get:

$$\rho[\hat{p}(y), \hat{q}] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{\hat{p}(\hat{y}0) \hat{q}_u} + \frac{C_h}{2} \sum_{i=1}^n \omega_i k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \quad (12)$$

Where  $\omega_i = \sum_{u=1}^m \delta[b(x_i) - u] \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}0)}}$ .

When  $\rho(y)$  get maximum value, calculating for new position  $\hat{y}1$  is describes as:

$$\hat{y}1 = \frac{\sum_{i=1}^{n_h} x_i \omega_i k\left(\left\|\frac{\hat{y}0 - x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} \omega_i k\left(\left\|\frac{\hat{y}0 - x_i}{h}\right\|^2\right)} \quad (13)$$

Where  $g(x) = -k(x)$ .

### 4. Scale adaptive strategy

On the basis of the Meanshift algorithm, SPMSA algorithm<sup>[4]</sup> is proposed, which means to regard the traced box of previous frame as a standard, get three size of traced boxs, size are the original's 110%, 100%, 90%. Matching the selected area by them with the area selected by the original trace box. Selecting the optimal size as the current frame size which has the biggest matching coefficient. The algorithm is as follows:

1) Calculate the similarity of the template at the current position of  $y$ , represented by BH coefficient, expressed as BH1;

2) Use the Meanshift algorithm, calculate the new position  $y1$  of the target;

3) Calculate the BH coefficient at the position of  $y1$ , expressed as BH2, and compare BH1 with BH2, for each iteration, adjust the width three times, size are original's 110%, 100%, 90%, iterate until the BH2 is maximum. Similarly, adjust the height three times, until the BH2 is maximum;

4) Compare BH1 with BH2, if  $BH2 > BH1$  and  $> \epsilon$ , replace the original coefficient by BH2, and return to step 2, until  $BH2 < BH1$  or  $< \epsilon$  is satisfied.

## 5. Experiment Result and Analysis

Capture the 1,50,100,150,200,250,300 frames from the same video sequence respectively, and compare the result of the algorithm of traditional Meanshift added the scale adaptive transform method with the proposed algorithm.

It can be seen from Fig. 2 that the middle vehicle is lost in the 100th frame because of background interference. In the 150th frame, the tracking of the left vehicle is inaccurate, It's all dependent on single color feature.

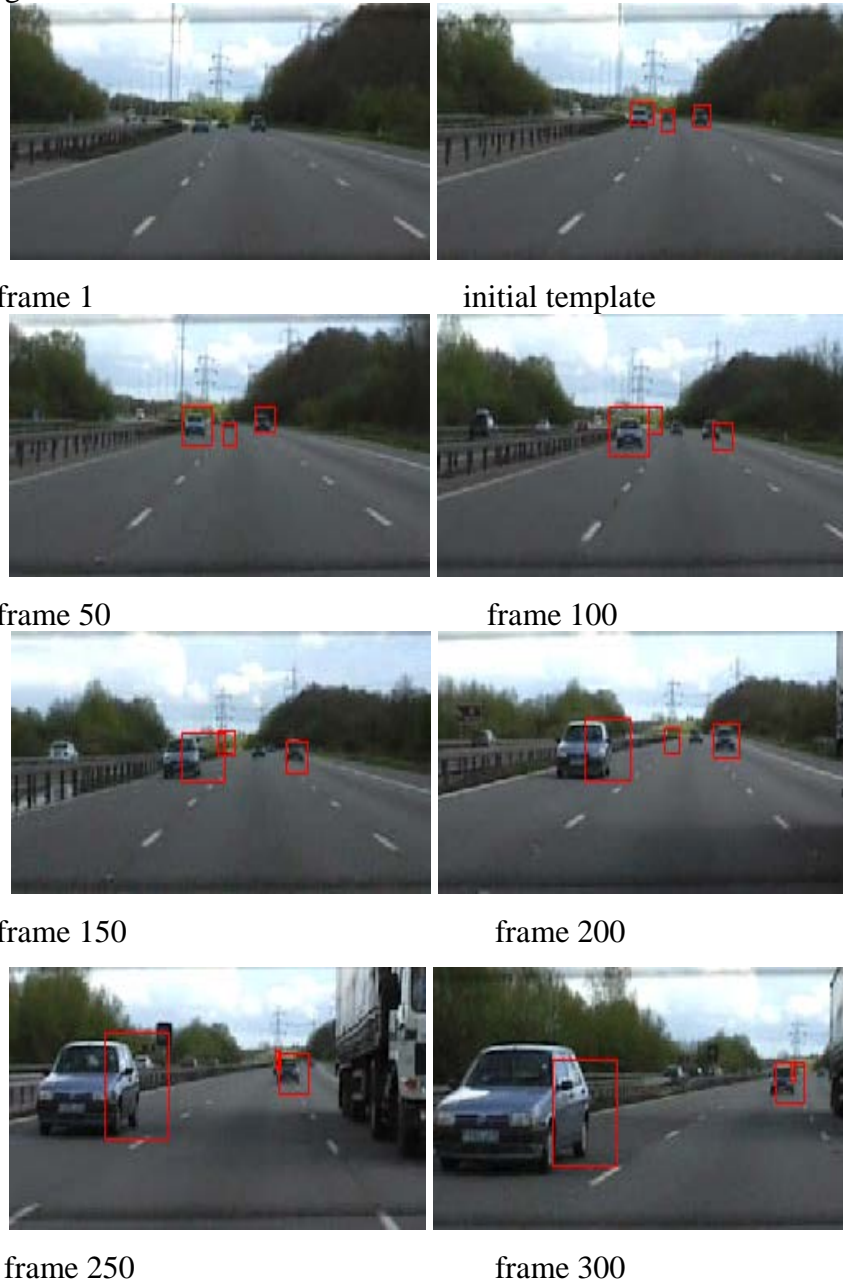


Fig. 2 traditional Meanshift combine with adjustable window algorithm

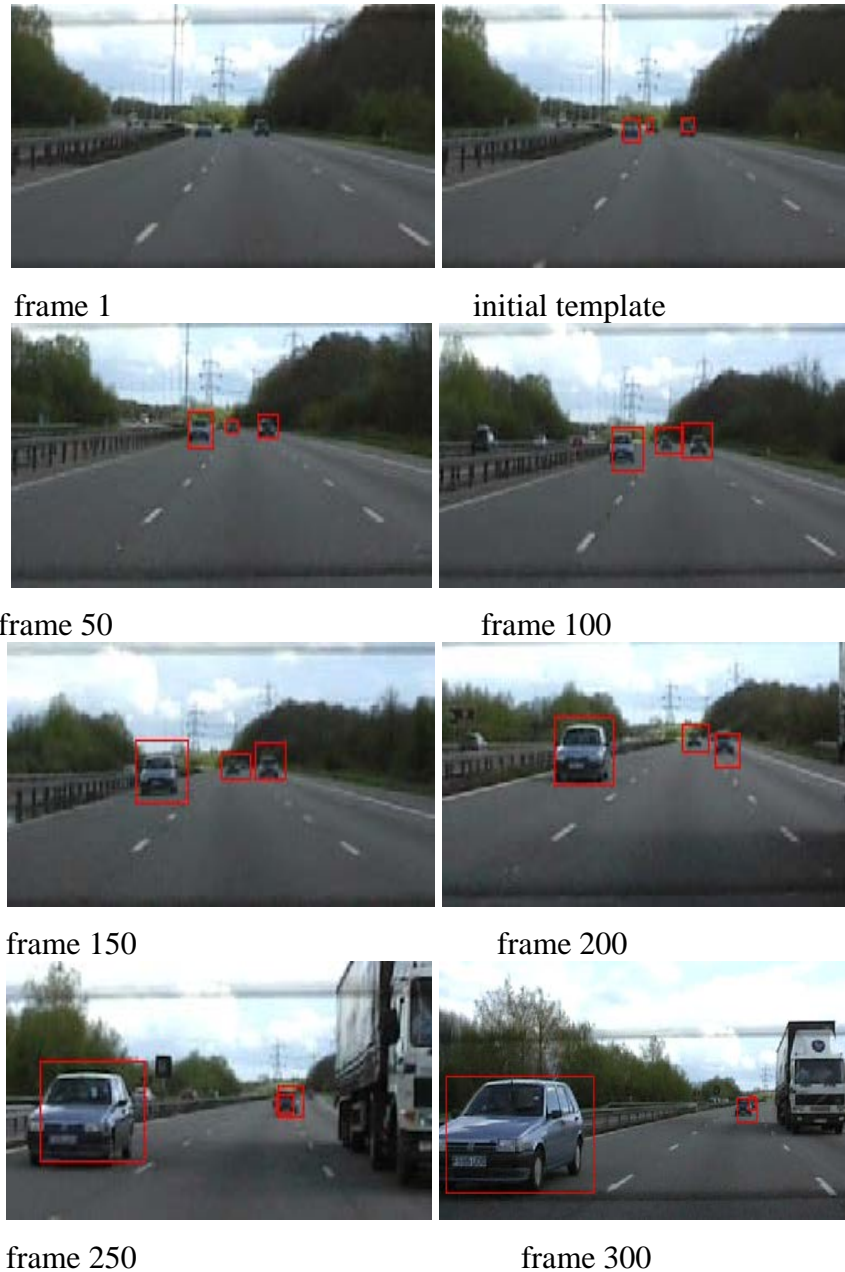


Fig. 3 herein algorithm

Fig. 3 shows the results by the proposed algorithm. Because of multi-feature fusion, in the 50th and 100th frames of the video sequences and after that, even though the interference of background is strong, tracking can also accurate because of the other two features. Compared with traditional Meanshift, it has better accuracy and robustness.

## 6. Conclusions

In this paper, the PSO algorithm is combined with Meanshift, and an operation of adjustable frame size is added. This method makes the results better when the background is complex or the size of object is changeful. The proposed algorithm calculate the contrast of objectives and background about different features, get the optimal weights by PSO, and weight these feature, then, match several different sizes template with the target template, and find the best size. Comparing the results of the traditional Meanshift with proposed algorithm by experiment, the algorithm in this paper shows better accuracy and robustness when background is complex or target's scale is changeful.

## Acknowledgements

The work is supported by the National Natural Science Foundation of China (Grant No. 61302133), Scientific and Technological Research Project of Science and Technology Department of Shaanxi Province (Grant Nos. 2011K09-36 and 2012K06-16) and Special project of Education Department of Shaanxi Province (Grant No. 12JK0528).

## References

- [1] KANG Li, XIE Weixin, HUANG Jingxiong. New method for multi-target tracking using ant algorithm. *Systems Engineering and Electronics*, 30(9), 1782-1784, 2008.
- [2] Collins R T, Yanxi L, Marius L. Online selection of discriminative tracking features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10), 1631-1643, 2005.
- [3] MA Wei, CHEN Jianguo, ZHANG Maolei. Video target tracking based on scale adaptation and tracking box rotations. *Tsinghua Univ (Sci & Tech)*, 52 (1), 92-95, 2012.
- [4] T. Ojala, M. Pietikäinen, D. Harwood. A Comparative Study of Texture Measures with Classification Based on Feature Distributions. *Pattern Recognition*, 29, 51-59, 1996.
- [5] NIU Fafa, CEHN Li, ZHANG Yongxin, LI Qing. Canny edge detection algorithm based on robust principal component analysis. *Journal of Computer Applications*, 34 (6), 1727-1730, 2014.
- [6] Yang SY, Ma Q, Huang WJ. Particle Swarm Optimized Unscented Particle Filter for Target Tracking, *International Congress on Image and Signal Processing, Tianjin, China*, 1-5, 2009.
- [7] Comaniciu D, Ramesh V, Meer P. Kernel-Based Object Tracking. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 25(5), 564-575, 2003.
- [8] Sun Jun, He Fazhi, Chen Xiaoa, ang Chen Yilin. Multi-feature Extraction and Stepwise Refinement Based High-speed Moving Target Tracking Algorithm. *Journal of Computer-Aided Design & Computer Graphics*, 26(10), 1747-1752, 2014.