Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 2, Issue 3: 92-103, DOI: 10.25236/AJCIS.020311

Object Tracking Based on Multi-Feature Fusion and Kalman Filter

Yibo Min, Jianwei Ma* and Shaofei Zang

School of Information Engineering, Henan University of Science and Technology Luoyang, China
**Corresponding author e-mail: yibo_min@163.com*

ABSTRACT. *Aiming at the problem of tracking drift caused by the variation of illumination variation, deformation and occlusion problems in the target tracking process, a traditional Mean Shift algorithm is proposed. Firstly, LBP and LPQ are used to perform texture feature extraction on the target image, and the extracted color features and texture features are weighted and fused. Secondly, a target tracking method combining kalman filter and mean shift tracking is proposed. The scale adaptive mechanism is used to effectively improve the tracking stability of mean shift when the target scale changes. Experiments show that the improved algorithm is more robust to target deformation, illumination, and occlusion.*

KEYWORDS: *Texture feature, Object Tracking, Kalman filter, Mean shift*

1. Introduction

Object tracking belongs to the field of computer vision and is widely used in intelligent transportation, security monitoring, human-computer interaction and so on. The so-called target tracking is to detect, extract, identify and track a moving target, obtain its state parameters and motion behavior, which is the focus of research in recent years [1-2]. The traditional target tracking algorithm, such as the Mean shift algorithm, has strong real-time performance and is suitable for the description of object motion information. It belongs to the generation tracking, that is, the target region is modeled in the first frame, and is derived from the pixel feature point probability density function. The similarity matching is used to find the candidate target region, the computational complexity is low, and the deformation and rotation of the target are robust, making it a mainstream tracking method in recent times. However, the traditional mean shift only uses the color features of the image, ignoring the spatial features, and is susceptible to illumination changes, occlusion, deformation and other factors. In the long-term tracking process, the target model of single color feature makes the error in the matching process large, and the error accumulation is easy to cause tracking drift [3-4].

ISSN 2616-5775 Vol. 2, Issue 3: 92-103, DOI: 10.25236/AJCIS.020311

Aiming at the problem that the single feature modeling is not strong, a multi-feature fusion tracking strategy is proposed. The literature [5] used the background information and the method of extracting the contour features of the target shape to track, in order to improve the tracking accuracy when the target mutation and the target color feature change, there is still a good tracking effect in the target mutation, but when the system memory When the shape is similar to the target, it is difficult to accurately locate the target; the literature [6] combines the histogram based on the histogram and the pixel-based target template to improve the robustness and accuracy of the tracking algorithm; the literature [7] uses the local sensitive histogram The graph (LSH) describes the target template and proposes a multi-region target tracking algorithm, which can stably track the target under illumination and target rotation deformation conditions. However, the algorithmlacks effective occlusion processing mechanism, and the target template is easily caused under occlusion conditions. The error update causes tracking drift or even failure; the literature [8] selects two features with high target and background discrimination from the seven feature components such as color and shape texture, and uses the mean shift algorithm for tracking. The algorithm is robust. Higher, but computationally intensive. In the framework of MS algorithm, the color feature and SIFT feature are combined to perform target tracking, which improves the robustness, but the algorithm is complex and cannot meet the real-time requirements. Wang Yongzhong et al. [10] proposed a weighted adaptive update mechanism for separability metrics. Combined with color and LBP features, a kernel tracking framework based on multi-feature adaptive fusion was proposed to implement adaptive fusion tracking algorithm. Huang et al. [11] decomposed into small segments and comprehensive length scale for target tracking. Liu [12] proposed a multi-feature adaptive fusion mean shift target tracking algorithm. In the process of tracking the dynamic change of the scene, the multi-feature fusion target model is established by selecting the feature description target with strong target and background distinguishing ability, and the importance weight is set.

Based on the color features, this paper proposes the texture features of the joint target to make up for the inherent defects of the traditional mean shift algorithm. LBP (Local Binary Pattern) [14] can extract local texture features of images with rotation invariance and gray invariance. The texture features extracted by the LPQ (Local Phase Quantization) operation have fuzzy invariance. Therefore, LBP and LPQ are used to extract the texture features of the image, and the color histogram features of the joint image are used to feature weighted fusion of the target region. Another problem with the traditional mean shift algorithm is that there are many iterations and the location update mechanism is lacking. The literature [15] introduces a quadratic polynomial as a fitting function to represent the position of the moving target, and uses the first ten frames of the moving target as the previous fitting value to predict the motion trajectory of the object.

ISSN 2616-5775 Vol. 2, Issue 3: 92-103, DOI: 10.25236/AJCIS.020311

2. Mean Shift

The Mean Shift tracking algorithm belongs to the kernel density estimation method and does not require any prior knowledge and completely relies on the sample points in the feature space to calculate its density function value. The principle of the kernel density estimation method is similar to the histogram method. The given set of sampled data is divided into several equal intervals, and the data is divided into groups according to the interval. The ratio of the number of data of each group to the total number of parameters is each unit. The probability value, while the kernel function is used to smooth the data. The kernel function estimation method can gradually converge to an arbitrary density function when the sampling is sufficient, that is, the density estimation can be performed on the data obeying any distribution. Based on the color shift-based mean shift tracking algorithm, by calculating the weighted normalized color histogram similarity between the target region and the candidate region, the candidate model with the largest similarity function is selected, and the final position of the target is obtained by multiple iterations.

In the first frame of the video sequence, manually select the tracking target area, the width of the rectangular frame or the circular area can be selected as the kernel function window width . Divide the color space of the target area , assuming $\left\{z_i^*\right\}_{i=1\cdots n}$ is normalization, the pixel position with the center of the target area as the origin, and $b(z_i^*)$ is the bin value of the quantized feature space z_i^* . Characteristics of the target probability model $u=1,2,\cdots,m$.

$$\hat{q}_u = C\sum_{i=1}^{n} k(\left\|z_i^*\right\|^2)\delta\left[b(z_i^*)-u\right] \tag{1}$$

$$C = 1\Big/\sum_{i=1}^{n}\left\|z_i^*\right\|^2 \tag{2}$$

$$z_i^* = \frac{\left(x_i - x_0\right)^2 + \left(y_i - y_0\right)^2}{{x_0}^2 + {y_0}^2} \tag{3}$$

Where, $\left(x_0, y_0\right)$ is center position of target, C is the normalization constant to guarantee $\sum_{n=1}^{m}\hat{q}=1$. K is the kernel function, and the Eepanechnikov kernel function is used in this paper. $\delta[b\left(x_i\right)-u]$ is the Kronecker delta function to judge z_i whether it is within the corresponding histogram interval, if it is 1, then 0.

Similarly,to calculate the candidate region histogram features and $\left\{z_i^*\right\}_{i=1\cdots n}$ represent pixels, the probability distribution of the candidate target model is:

ISSN 2616-5775 Vol. 2, Issue 3: 92-103, DOI: 10.25236/AJCIS.020311

$$\hat{p}_u(y) = C_h \sum_{i=1}^{n} k(\left\| \frac{y - z_i}{h} \right\|^2) \delta[b(z_i) - u] \tag{4}$$

Where, (x_0, y_0) is center position of target, C is the normalization constant to guarantee $\sum_{n=1}^{m} \hat{q} = 1$. K is the kernel function, and the Eepanechnikov kernel function is used in this paper. $\delta[b(x_i) - u]$ is the Kronecker delta function to judge z_i whether it is within the corresponding histogram interval, if it is 1, then 0.

Similarly, to calculate the candidate region histogram features and $\{z_i^*\}_{i=1\cdots n}$ represent pixels, the probability distribution of the candidate target model is:

$$\rho[\hat{p}(y), \hat{q}] = \sum_{u=1}^{m} \sqrt{\hat{p}(y)\hat{q}_u} \tag{5}$$

Taylor expansion of the above formula:

$$p(p,q) \approx \frac{1}{2} \sum_{u=1}^{m} \sqrt{p_u(y_0) q_u} + \frac{C}{2} \sum_{i=1}^{n} w_i K\left(\left\| \frac{y - z_i}{h} \right\| \right) \tag{6}$$

$$w_i = \sum_{u=1}^{m} \sqrt{\frac{\hat{q}_u}{\hat{p}_u(y)}} \delta[b(z_i) - u] \tag{7}$$

The second half of the above formula is affected by the value change, so the iterative process can be organized as:

$$y_{k+1} - y_k = \frac{\sum_{i=1}^{n} w_i (y_k - z_i) g\left(\left\| \frac{y_k - z_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n} w_i g\left(\left\| \frac{y_k - z_i}{h} \right\|^2 \right)} \tag{8}$$

Where $g(x) = -k'(x)$; When $y_{k+1} - y_k$ is smaller than the set threshold, the obtained point is the current frame target position, and the position is used as the next frame search center to continue the similarity calculation.

$$M = \frac{\sum_{i=1}^{n} x_i g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n} g\left(\left\| \frac{x - x_i}{h} \right\|^2 \right)} - x \tag{9}$$

ISSN 2616-5775 Vol. 2, Issue 3: 92-103, DOI: 10.25236/AJCIS.020311

3. Feature fusion

3.1 LBP Feature

The traditional mean shift algorithm uses color histogram features to model the target, ignoring the semantic information of the image, resulting in imperfect target information, and tracking drift is easy to occur during the tracking process. The texture information of the image is not affected by factors such as illumination and deformation, and can better compensate for the lost target information. Therefore, this paper uses LBP and LPQ algorithm to extract the texture information of the image and convert it into histogram information as the feature of the image. As a significant visual feature, texture features not only do not depend on color or brightness, but also include the arrangement and organization order of the surface structure of things, showing the connection of contextual content, reflecting the repetitive visual features of homomorphism in the image, so the texture is based on a very important feature for image description and classification in image retrieval methods for content.

The LBP algorithm describes the relationship between the pixels in the image and their neighborhood values. It can extract local texture features of the image, and has rotation invariance and gray invariance. In a grayscale image, the grayscale of a pixel represents its value. The basic LBP algorithm works on a large and fixed matrix block, as shown in Figure 1.

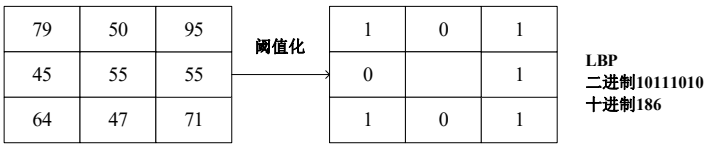

Figure. 1 Basic LBP algorithm

The value of the pixel in the image is obtained through the following steps. First, each pixel in the image and its neighborhood are regarded as a rectangle, and then the pixel value is compared with its neighboring pixels. , Set it to 1, otherwise set to 0, so you can get the binary value corresponding to the original pixel, and finally convert the obtained binary to its decimal equivalent, that is, LBP code, a total of 256 kinds. Then, the value of each pixel in the entire image is calculated in the form of a histogram, which is used as a description of the texture characteristics of the target area image.

$$LBP_{P,R}\left(x_c - y_c\right) = \sum_{n=0}^{7} \delta\left(g_n - g_c\right) 2^n \tag{10}$$

$\delta\left(x\right)$ is Symbolic function, and we define:

ISSN 2616-5775 Vol. 2, Issue 3: 92-103, DOI: 10.25236/AJCIS.020311

$$\delta(x)=\begin{cases}1, & if x\geq 0\\ 0, & otherwise\end{cases} \tag{11}$$

The three-point localization experiment was carried out in 10 groups, and the record data were shown in Table 2. It can be concluded from the table, the estimated value of X axis coordinate system and the Y axis and the target actual value of X axis coordinates and the Y axis deviation was less than 2cm.

The texture feature extraction process is shown in Figure 2.

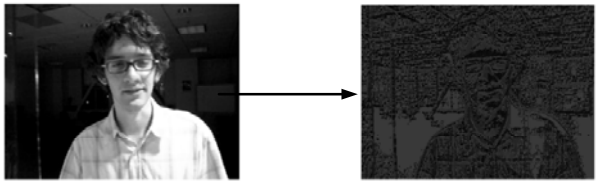

Figure. 2 LBP texture extraction process

3.2 LBP Feature

The texture features extracted by LPQ have fuzzy invariance. For image $f(x)$, $M\times M$ domain N_x uses discrete short-time Fourier transform, as shown in the formula:

$$F(u,x)=\sum_{y\in N_x} f(x-y)e^{-j2\pi u^T y} \tag{12}$$

Where u is the frequency.

The local Fourier coefficients are calculated by 4 frequency points $u_1=[a,0]^T$, $u_2=[0,a]^T$, $u_3=[a,a]^T$, $u_1=[a,-a]^T$, where a is sufficiently small value, $a=1/M$. For each pixel position, represented by a vector

$$F(x)=\left[F(u_1,x),F(u_2,x),F(u_3,x),F(u_4,x)\right] \tag{13}$$

The Fourier coefficient phase can be expressed by the sign of the real and imaginary numbers of each part:

$$q_j=\begin{cases}1, & g_j\geq 0,\\ 0, & g_j<0.\end{cases} \tag{14}$$

ISSN 2616-5775 Vol. 2, Issue 3: 92-103, DOI: 10.25236/AJCIS.020311

g_j is the part of the vector $G(x)=\left[\mathrm{Re}\{F(x)\},\mathrm{Im}\{F(x)\}\right]$. Then q_j binary encode it, as shown in the formula:

$$f_{LPQ}(x)=\sum_{j=1}^{8}q_j 2^{j-1} \tag{15}$$

LPQ_M represents the algorithm whose window size is , and the window shown in the figure below is 5×5 LPQ algorithm.

LPQ operator provides image texture features:

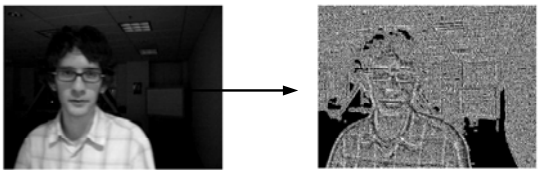

Figure. 3 LPQ texture extraction process

3.3 Feature Fusion

For image signals, the object of spatial domain processing is the image pixel itself, which reflects the change in image gray level; the object of frequency domain processing is the coefficient of change of the image in the transform domain, which reflects the distribution of the image gradient. Both the spatial and frequency domains ignore some information when performing texture analysis, so combining the two can complement each other. Different observation features have different discrimination between target and background in different scenarios. In the tracking process, different weights should be given to the features to make the tracking effect more reliable. Based on this, in the MS object tracking, a multi-feature set $F=\left\{f \mid f\in(c,lbp,lpq)\right\}$ including color and texture is defined, and the weight of the target model with different features f is given. Using the color feature algorithm to get the color-based target center f_c, and using the LBP algorithm to get the texture feature f_{lbp}, and using the LPQ algorithm to get the texture feature f_{lpq}.We have fused different features to get the target model:

$$\hat{q}=\sigma f_c+\beta f_{lbp}+(1-\sigma-\beta) f_{lpq} \tag{16}$$

ISSN 2616-5775 Vol. 2, Issue 3: 92-103, DOI: 10.25236/AJCIS.020311

4. Kalman Filter

Particle filtering and Kalman filtering are combined to realize non-linear filtering, which mainly solves the problem that the system state equation is linear and the observation equation is non-linear in moving target tracking. The problem of nonlinear estimation in video image sequences can be solved by particle filtering. For moving objects, dynamic systems can be described by state equations and observation equations:

$$\begin{cases} x_{t+1} = \mathrm{A}x_t + u_t \\ y_t = h\left(x_t\right) + v_t \end{cases} \tag{17}$$

Where $x_t \in R^m$, x_{t+1} represents the predicted state vector of the current moment dimension, $A \in R^{m \times n}$ represents the parallel or rotating state transition matrix of the moving target at the moment t, y_t represents the observed value of the moment t, $h(\bullet)$ represents x_t the linear or non-linear relation matrix; u_t and v_i represents satisfaction $E\left\{u_n {u_m}^T\right\} = Q\delta_{nm}$, $E\left\{v_n v_m\right\} = \sigma^2\delta_{nm}$ $E\left\{v_n v_m\right\} = \sigma^2\delta_{nm}$ Gaussian noise and observation noise with a mean value of 0, where Q and σ^2 are the covariance matrix of the state noise and the observation noise variance, respectively.

During the tracking process, the state vector $x = \left(x, y, v_x, v_y\right)$ is taken, and the observation vector is $z = \left[x, y\right]^T$, where, x and y are the center coordinate of the target, and v_x and v_y are the partial velocity of the target in the x direction and y direction respectively. Assuming that the interval Δt between consecutive video frames is small, the target can be considered to make a linear motion at a uniform speed in a short period of time, and the state transition matrix can be defined as:

$$A = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The relationship matrix of the observation equation is:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

The covariance matrix is:

ISSN 2616-5775 Vol. 2, Issue 3: 92-103, DOI: 10.25236/AJCIS.020311

$$Q = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

5. Algorithm Details

The steps of the tracking algorithm combined with multi-feature combination and Kalman filtering are as follows:

Step 1: Initialize the target position in the first frame to determine the center of the target position z_0 .

Step 2: A color-texture feature histogram model $\hat{q}_u$ is calculated for the target center z_0 .

Step 3: The mean shift vector is used to iteratively obtain the target position center z of the next frame.

Step 4: Kalman filtering predicts the target position z_k of the same frame, and fuses the two target centers to obtain the target center position of the current frame as $z_1 : z_1 = \lambda z_m + (1-\lambda) z_k$.

Step 5: If $\|z_1 - z_0\| < \varepsilon$ holds, to calculate the next frame of the videos, otherwise to repeat steps 3 and 4.

6. Experimental and analysis

In this paper, experiments are performed in MatlabR2014b, i5 processor and Windows 10. In order to verify the accuracy of the algorithm tracking, four data sets Car, David, Football, and Freeman were selected to verify the accuracy of the algorithm in occlusion, deformation, and lighting changes. The first video sequence tested was a video of car movement under varying lighting conditions, with a total of 659 frames. During the tracking process, the target has a certain degree of lighting change. Figure (a) shows a comparative analysis of the video sequence using the traditional MS algorithm and the feature fusion motion prediction algorithm proposed in this paper. It can be seen that at 16 frames, a similar color background appears to the left of the target, and the traditional algorithm (red box) shows a certain drift, but the improved algorithm can track the target well. At the 189th and 258th frames, it can be seen that the target has changed its lighting. The traditional MS algorithm, which relies solely on color features, drifts due to the cumulative

ISSN 2616-5775 Vol. 2, Issue 3: 92-103, DOI: 10.25236/AJCIS.020311

error, and eventually loses the target completely. The improved algorithm can still track the target.

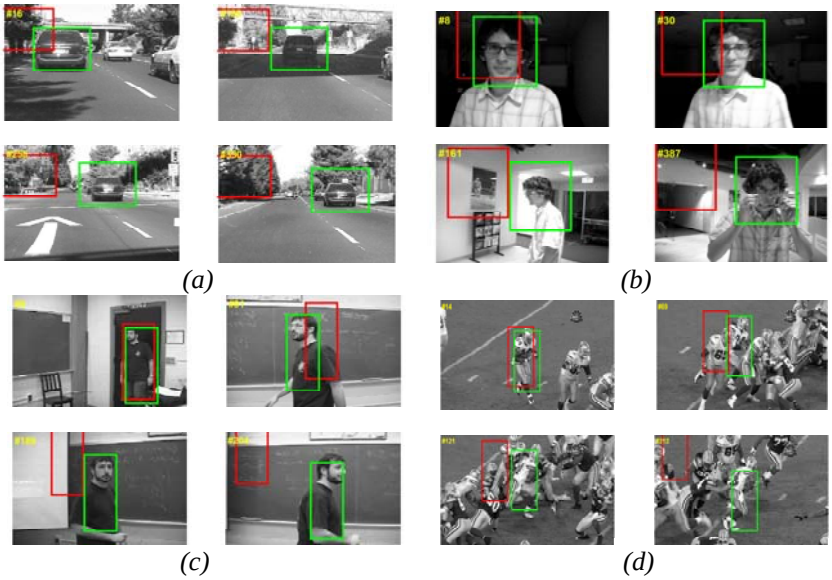

Figure. 4 Comparisions of four video sequences

The second video sequence is face tracking in a complex environment, with a total of 462 frames. During the tracking process, the target has a large rotation deformation and occlusion. In the experiment, the video sequence is tracked using the MS algorithm and the algorithm in this paper, and the tracking results are shown in Figure (b). In the initial frame, the above methods can track the target stably. But by the 30th frame, motion blur appeared on the target, and the MS tracking algorithm began to drift. Since then, the video sequence has undergone large lighting changes and deformations, and the long-term accumulated errors of traditional algorithms have caused the target to be lost. The algorithm proposed in this paper can well adapt to the current scene by fusing the texture characteristics of the target and motion prediction of the target, and continue to track the target stably. Until 387 frames, the method in this paper still has higher robustness and tracking accuracy.

The fourth video sequence is pedestrian tracking in a complex background. The video sequence has a total of 362 frames. During the tracking process, there are severe occlusions, background chaos and deformation. In the experiments, the MS algorithm and the algorithm in this paper are used for tracking, and the results are shown in Figure (5). In the initial frame of the video sequence, the above methods can track the target steadily, but at the 69th frame, due to the effect of background chaos and occlusion, the MS algorithm starts to drift, and the improved method can still track the target. As the number of frames increases, the error of the MS algorithm continues to accumulate, and the lack of an update mechanism to detect

ISSN 2616-5775 Vol. 2, Issue 3: 92-103, DOI: 10.25236/AJCIS.020311

whether the tracking is lost, resulting in the subsequent sequence being in the state of target loss. This article makes corresponding improvements to the MS algorithm, adjusts the tracking frame in time, adapts to the current scene, and continues to track the target steadily.

Table 1 Comparing

Evaluation	Our	KCF	CN
Mean FPS	60	150	45
Mean Precision	85.4	56.3	49.5
Mean Success	83.5	51.3	65.5

7. Conclusion

Aiming at the problem that the single-color feature tracking target of the MS algorithm has weak robustness, a multi-feature fusion modeling method is proposed, and Kalman motion prediction is added to implement a target tracking algorithm in a complex environment. Aiming at the characteristics of different characteristics of the target and background during the dynamic change of the tracking scene, adjust the feature fusion weights to improve the tracking accuracy. In each frame of image, the mean-shift algorithm based on color-texture is used to calculate the target model, and the position of the target is estimated in conjunction with Kalman filtering, and the final target position is obtained through the threshold specification. Through the dynamic update and position prediction of the target model, the tracking robustness is further improved. However, the traditional MS algorithm still has many shortcomings. The next step is to study the tracking target's scale change, reduce the redundant information of the target frame, choose a more descriptive feature model and adaptive adjustment window function, and effectively improve Tracking accuracy and speed.

References

[1] Guo L., Ge PS., Zhang MH, et al. Pedestrian detection for intelligent transportation systems combining AdaBoost algorithm and support vector machine [J]. Expert Systems with Applications, 2012, 39 (4): 4274-4286.

[2] Comaniciu D, Ramesh V, Meer P. Real-time tracking of nonrigid objects using Mean Shift [C] // Pro of International Conference on Information Fusion. 2012: 142-149.

[3] Ning Ji-feng, Zhang Lei, Zhang D, et al. Robust Mean Shift tracking with corrected background-weighted histogram [J]. IET Computer Vision, 2012, 6 (1): 62-69.

[4] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark [C] // Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2411-2418.

ISSN 2616-5775 Vol. 2, Issue 3: 92-103, DOI: 10.25236/AJCIS.020311

[5] SHEN Ding-cheng, XUE Yan-bing, ZHANG Hua, et al. Research on robust object tracking algorithm based on online boosting [J]. Journal of Optotelectronics Laser, 2013, 24 (1): 170-175.
[6] Ulker Y, Gunsel B. Multiple model target tracking with variable rate particle filters [J]. Digital Processing, 2015, (22): 417-429.
[7] Das S, Kale A, Vaswani N. Particle filter with a mode tracker for visual tracking across illumination changes [J]. IEEE Trans. Image Process, 2012, 21 (4): 2340-2346.
[8] Leichter I. Mean shift trackers with cross-bin metrics [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2012, 34 (4): 695-706.
[9] Li S, Wu O, Zhu C, et al. Visual object tracking using spatial Context Information and Global tracking skills [J]. Computer Vision and Image Understanding, 2014, (125): 1-15.
[10] Yao A, Lin X, Wang G, et al. A compact association of particle filtering and kernel based object tracking [J]. Pattern Recognition, 2012, (45): 2584-2597.
[11] Zulfiqar K H, Gu Irene Y H, Andrew G B. Robust visual object tracking using multi-mode anisotropic mean shift and particle filters [J]. IEEE Transactions on Circuits and Systems for Video Technology (CSVT), 2011, 21 (1): 74-87.
[12] Choi H, Kim I, Choi J. Combining histogram-wise and pixel-wise matching for kernel tracking through constrained optimization [J]. Computer Vision and Image Understanding, 2014, (118): 61-70.
[13] SU Yan-zhao, LI Ai-hua, JIN Guang-zhi, et al. Visual tracking of moving object based on double layer features optimization [J]. Journal of Optoelectronics Laser, 2015, 26 (1): 162-169.
[14] He S F,Yang Q X, Rynson W H, et al. Visual tracking via locality sensitive histograms [C]. IEEE Conference on Computer Vision and Pattern Recognition, Oregon, Porland, 2013, 2427-2434.
[15] Wang J Q, Yagi Y. Integrating color and shape-texture features for adaptive real-time object tracking [J]. IEEE Transactions on Image Processing, 2012, 17 (2): 235-240.