A target tracking method based on feature fusion

Li dan[1,2,a], Dang xiangying[1,2]

[1]Xuzhou Institute of Technology, Xuzhou, China

[2]Xuzhou Machine Vision Application Engineering Technology Research Center, Xuzhou, China

[a]E-mail: lidanonline@163.com

Keywords:city traffic;feature fusion;target tracking; Camshift

Abstract. The paper presented a adaptive template update method of multi feature fusion based on Camshift algorithm. When one feature to distinguish the foreground and background is not obvious, other features can be complement each other. The algorithm combines texture, edge and color features, using the different characteristics contribution degree set the different weight calculated in the multi feature space. It can better solve the problems of background color similar, objective morphological changes and the change of illumination.

Introduction

At present, the research of video object tracking technology has become the core issue in the video monitoring field. Color feature is not affected by the shape of the object changes, it has the scale and rotation invariant characteristics, the amount of calculation is small, the existing algorithms mostly use the features to represent the target. But in complex traffic scene with the uneven brightness, high noise, close background color[1,2], it's easy to be effected by the disruptor and cause tracking failure. According to the complex traffic environment, Camshift fused LTP texture model, HSV model and edge features, made feature contribution degree to obtain the feature weight. Experiments show that, the new algorithm search fast, low complexity, there is a better robustness to traffic high noise, nonuniform illumination, color similar complex environment.

Camshift Algorithm

Meanshift algorithm is the core of Camshift, make the target object color distribution is $\mathrm{q_u}$, moving target in the Ith frame color distribution is $\hat{\mathrm{p}}_{\mathrm{u}}(y_i)$. Use Bhattacharrya coefficient to judge the similarity of color distribution. Let y_{i+1} be the next frame motion target center and seek y_{i+1} let a, b most similar.

$$\hat{\rho}(y_{i+1}) = \sum_{u=1}^{m}\sqrt{\hat{\mathrm{p}}_{\mathrm{u}}(y_{i+1})\mathrm{q}_u}$$

$$\approx \frac{1}{2}\sum_{u=1}^{m}\sqrt{\hat{\mathrm{p}}_{\mathrm{u}}(y_i)\mathrm{q}_u} + \frac{1}{2}\sum_{u=1}^{m}\hat{\mathrm{p}}_{\mathrm{u}}(y_{i+1})\sqrt{\frac{\mathrm{q}_u}{\hat{\mathrm{p}}_{\mathrm{u}}(y_i)}} \tag{1}$$

Known $\hat{\mathrm{p}}_{\mathrm{u}}(y_i)$、 q_u, the weight is:

$$w(x_i) = \sum_{u=1}^{m}\delta(b(x_i)-u)\sqrt{\frac{\mathrm{q}_u}{\hat{\mathrm{p}}_{\mathrm{u}}(y_i)}} \tag{2}$$

If maximum similarity, it can maximize the probability density estimation of the search window of position y_{i+1}.

$$\sum_{u=1}^{m}\hat{\mathrm{p}}_{\mathrm{u}}(y_{i+1})w(x_i) = C_h\sum_{i=1}^{n}k\left(\frac{\| y_{i+1}-x_i \|^2}{h^2}\right)w(x_i) \tag{3}$$

Algorithm used kernel function Epanechnikov, in the Meanshift vector formula, y_{i+1} moves to the place of largest probability density, using continuous iteration to calculate the centroid position.

$$M_h(x)=\frac{\sum_{i=1}^{n}G(\frac{x_i-x}{h})w(x_i)x_i}{\sum_{i=1}^{n}G(\frac{x_i-x}{h})w(x_i)}-x \qquad (4)$$

The Camshift algorithm takes the steps as follows:

1) To determine the initial position of the search window (x_c, y_c),set the size of the search window;

2) Around (x_c, y_c),computing color probability distribution of search window within the 1.1 times;

3) To obtain the new search window location, size of Meanshift is as to be the initial position of the next frame;

4)Using the value of step 3) to initialize the search window size, position of the next frame image, then turn to step 2) to continue calculating until convergence.

We can see that Camshift searches window centroid iteratively is adaptive to changes in target position and size to realize the real-time target tracking. But the amount of information provided by the model of single character is limited after all. Appearance characteristics cannot describe all, it's easy to lose the target when change in appearance, so need to rely on multi feature model to describe.

Texture Model Establishment

The time complexity of Camshift algorithm is low, anti occlusion and deformation ability is better, and can better tracking under simple environment. But in a complex environment, the color information is easy to be undisturbed. Because the texture information is relatively stable, the paper used texture information[3,4] to integrate into Camshift. So as to solve the problems of easy to loss goals. At present, most of the texture model such as gray level co-occurrence matrix, wavelet and so on although can get the element distribution of texture, but they do not support the point sample estimation, so it is difficult to fusion of color and other characteristics.

Local Ternary Patterns Texture model was proposed in 2007 by Tan and Bill in the literature [5] and has a good application in face identification. Solved the problem of sensitivity to noise by LBP model. This paper combined LTP in moving target detection to strengthen the ability of feature space classification. LTP threshold function is as the following:

$$s'(u,p_c,t)=\begin{cases}1 & u\geq p_c+t\\ 0 & |u-p_c|<t\\ -1 & u\leq p_c-t\end{cases} \qquad (5)$$

LTP makes the adjacent area pixels quantizatized to 0 within the range of ±t. Pixels which are less than the width quantizatized to -1, pixels which are more than the width quantizatized to 1. Let t be a noise threshold. $s(u)$ convert into a value of three mode $s'(u, p_c, t)$. This transformation uses the symmetry properties of model and noise threshold to enhance the anti noise ability and filter the noise effectively. Code is as the following:

$$LTP_{P,R}(x_c,y_c)=\sum_{p=0}^{P-1}s(u)3^p \qquad (6)$$

In order to divide LTP into two parts, it can simplify the calculation: Upper LBP positive and Lower LBP negative calculation are two parts:

$$s_{upper}(x_c,y_c)=\begin{cases}1, s'(x_c,y_c)=1\\ 0, s'(x_c,y_c)\neq 1\end{cases}$$

$$s_{lower}(x_c,y_c)=\begin{cases}1, s'(x_c,y_c)=-1\\ 0\ , s'(x_c,y_c)\neq 1\end{cases} \qquad (7)$$

Establish the LTP model texture histogram can enhance texture information and solve the problem of LBP has poor anti noise performance. Fuzzy and illumination variation have smaller effects on LTP, it has better anti noise performance.

Establish and Update Multiple Features Template

This paper proposed the algorithm based on multi feature template fusion tracking, it can be described in each feature space using histogram separately. By defining the contribution rules and set feature model, each feature has different weight and update feature template dynamically, so that it greatly improved tracking stability.

A. Establish a multi-feature template

Feature selection should distinguish the background and objectives. Feature model can not be set too much, otherwise it will increase the amount of computation. Let K is Characteristic species, u_j is the feature vector, $\{x_1, x_2, ...x_n\}$ is a set of pixels of the window. Candidate target for j types of characteristic distribution is $\hat{p}_{u_j}(y)$, $j = 1, 2, 3, ...K$. Feature vector is u_j, feature space and the mapping between pixels is $b_j(x_j)$, u_j=1,2,3…m_j, quantization level is m. Candidate target center position y, normalization constant C_j. The probability density distribution of the candidate target template is as the following formulas:

$$\hat{p}_{u_j}(y) = C_j \sum_{i=1}^{n} k(\left\| \frac{y - x_i}{h} \right\|^2) \delta(b_j(x_i) - u_j)$$

$$\hat{q}_{u_j}(y) = C_j \sum_{i=1}^{n} k(\left\| \frac{y - x_i}{h} \right\|^2) \delta(b_j(x_i) - u_j) \tag{8}$$

Multi-feature space needs to integrated various types of features and not only consider certain features. So it is necessary for each feature weighting, it can modify theformula weights as the following:

$$w_i = \sum_{j=1}^{k} a_j \sum_{u_j=1}^{m_j} \delta(b_j(x_i) - u_j) \sqrt{\frac{\hat{q}_{u_j}}{\hat{p}_{u_j}(y_0)}}$$

$$= a_1 w_i(u_1) + ...a_k w_i(u_k) \tag{9}$$

$\sum_{j=1}^{k} a_j = 1$, $a_1, ...a_k$ are weight coefficients. The proposed multi-feature templates can reduce the interference intensity of the background at different spatial. But the target environment changes at any time during moving. Therefore, the weighting factor a_1……a_k should also be different and can change adaptively. By comparing the objectives and background environmental histogram with different characteristics as color, texture, edges, etc. to determine the value of each characteristic coefficient a_j.

B. Update Template

Camshift matches based on nuclear histogram and run faster. However, after selected the initial model, it's not updated with the tracking process, when lighting situations encountered or the target morphology changes, model will bias because of the color, texture and other characteristics to lead tracking failure.

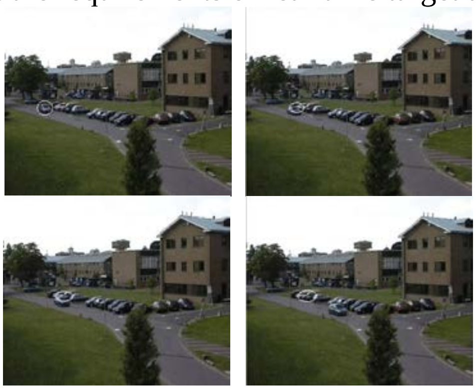
Target background often change during moving. The paper enhanced the applicability of the algorithm through model update method. The current template update usually use target template as a whole,But the impact of each component to the overall features is diffrernt while the update process. Not all vectors will be updated, therefore, the feature vector quantization histogram is m-level, So update the template between two frames as the following:

Let $f(u) = \sqrt{\hat{p}_u \hat{q}_u} \Big/ \sum_{u=1}^{m} \hat{p}_u \hat{q}_u$ be a characteristic component matching function, from small to large order matching each component. If the value of $\sum_{u=1}^{m} \hat{p}_u \hat{q}_u$ is small, indicating there is a greater interference impact on the tracking, does not update the model. If the value of $\sum_{u=1}^{m} \hat{p}_u \hat{q}_u$ is more than 0.8, then the matching degree which is smaller than 0.5* $\sum_{u=1}^{m} \hat{p}_u \hat{q}_u$ the n-th component update according to the weight. As the formula (10), the current and historical results weighted compromise, it can reduce the sensitivity of the model update process for environmental change.

$$\hat{q}_u(t) = C_q(\alpha p_u(t) + (1-\alpha)\hat{q}_u(t-1)) \tag{10}$$

Experimental Results

HSV color model histogram described the image has greater redundancy, and the edge, texture recognition are also strong, so HSV quantization level can be reduced, the color space can be quantized to 16 * 4 * 4 colors. Edge features can test local feature discontinuities and if there is a step gray change of the surrounding pixels. The edge histogram quantized to 12 parts, using 9-class LTP texture pattern to reduce the computational overhead. Although each feature space vector need to update the weights, increase the time overhead, but only to update the smaller n-th sub-feature matching degree components, but also reduces the time consuming. Overall, the computational complexity of the new algorithm is slightly higher than the original algorithm, but with HSV, LTP and edge feature fusion. Tracking moving targets can be more accurate. Experimental results show that the new algorithm can meet the requirements of real-time target tracking.



(a)

(b)

Figure 1. Comparison of substation original and new algorithms

Shown in Figure 1, using the Camshift algorithm and multi-feature template update algorithm tracking video object. It can be seen, in complex environment, Camshift complex environments easily converge at similar HSV color in Figure (b). The results of this algorithm is shown in Figure (b). Due to the integration of the edge, texture and HSV feature space. When a feature interference, other features may increase the weight, adaptively update the template and track better.

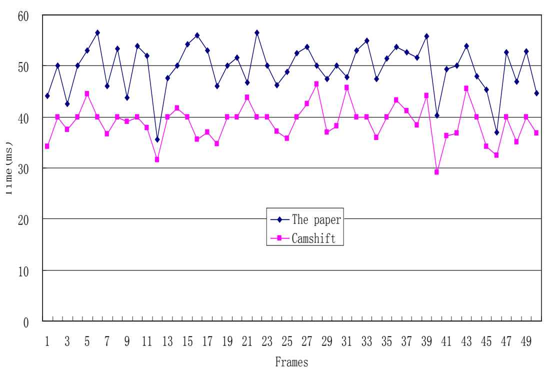


Figure 2. Time-consuming

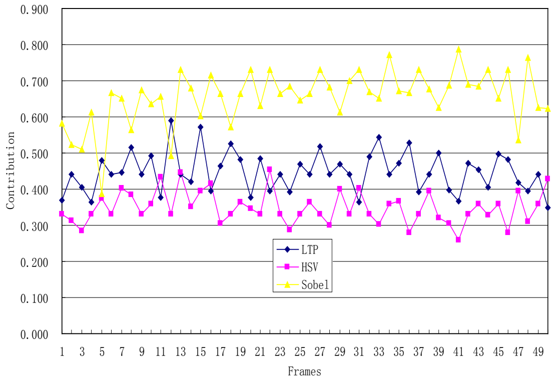


Figure 3. Feature contribution

Figure 2 is 50 consecutive frames of the video in Figure 2 processed statistical results. The results can be seen, although the complexity of the algorithm is slightly higher than the original algorithm, however, it can meet the monitoring requirements of real-time and tracking accurately. The figure 3 is the distribution of feature contribution.

Conclusion

In order to adapt to the complex environment of traffic video surveillance needs, this paper presented a multi-feature fusion method based on Camshift algorithm. The new algorithm fused textures, edges and color features. Set weights using different contribution of different characteristics, calculated in the multi-feature space. In case of foreground and background color approximation, morphological changes, target illumination changes and other issues, it can be a better solution. Since in the matching function, update only poor matching characteristic component, not all features of the component updates. The new algorithm reduced the time complexity. It has good prospects in traffic video surveillance.

Acknowledgment

This work is partly supported by the Ministry of Housing and Urban Construction Science and Technology Plan Project(2014-K5-027) and Xuzhou Machine Vision Application Engineering Technology Research Center.

References

[1] Chen Weimin, Zhou Gang, Wang Gang, et al. Evaluation method of miners' safety behavior based on gray-fuzzy-improving momentum BP algorithm [J]. Journal of China Coal Society, 2010, 35(1): 101-105.

[2] Sun Jiping. Networking technology for safety supervision system in a coal mine[J]. Journal of China Coal Society, 2009, 34(11): 1547–1549.

[3] LIU Guo-Ying, QIN Qian-Qing, WANG Lei-Guang et al. New Algorithm Of Multiscale Texture Image Segmentation Based On Contourlet Transform [J]. Journal of Infrared and Millimeter Waves. 2009, 28(6):450-455.

[4] Zhao Hai Ying; Xu Guang Mei; Sun Feng Yu. A metric algorithm based on three elements of texture visual feature [C]. Conference on Theories and Applications (BIC-TA)， 2010: 1221-1224.

[5] Tan Xiao-Yang,Bill Triggs. Enhanced Local Texture Feature Sets For Face Recognition Under Difficult Lighting Conditions[C].Analysis and Modeling of Faces and Gestures,Rio de Janeiro, Brazil, Oct.20, 2007:168-182.