An Improved Camshift Algorithm Based on Grabcut with a LBP Model of Correction Tracking Centroid

Xianggong Hong*, Xiying Zheng, Huimei Xiao, Zhiyi Xue

Nanchang University Information and Engineering College, Jiangxi, 330031, China.
393472615@qq.com

According to the analysis of advantages and disadvantages of traditional Camshift algorithm, it can be known that background noise and interference of similar color objects will have a greater impact on Camshift algorithm, which may cause tracking errors easily. This paper presents an improved Camshift algorithm based on Grabcut and LBP centroid tracking correction model. Current frame image enhanced by enhanced coefficient, the Camshift algorithm will apply Grabcut object segmentation to achieve pure histogram for target object and then move the centroid which has been distracted due to similar color objects' obstacles to uncovered parts of the target object by using LBP centroid tracking correction model. As a result, the problems caused by background noise and similar color obstacles distraction can be solved effectively. Tests presented in the paper can prove that the algorithm can be used to operate the real-time tracking much more steadily and accurately.

1. Introduction

Visual tracking, one of the hottest research directions, plays a significant rolein image processing and computer vision at present. Moving object tracking is widely used in many military and civilian fields, such as vision-guiding, unmanned aerial vehicles tracking, security monitoring, public scene monitoring, ITS and so on. Currently, common used methods for tracking moving targets are particle filter, compressive sensing algorithm, background subtraction, neighbor frame difference method, optical flow method, Camshift and so forth. Nevertheless, all of these methods, instead of being perfect, are defective. For example, although optical flow method has adopted subtraction method, it is poor in real-time algorithm which would easily lead to tracking fail under the complex environment. Particle filter algorithm is strong in anti-jamming capability, but it presences a phenomenon of particle degeneracy and it is weak in stability. Background subtraction and adjacent frame difference method cannot be used for background transforming situation. According to the above analysis, this paper has adopted Camshift algorithm to be the main body of tracking method.

This paper introduces an improved Camshift algorithm. Taking into account the selected initial histogram would be easily mixed with background noise, the image is enhanced as a whole, which increases contrast between object and background. The next our algorithm adopts Grabcut to separate the target object and then to gain the pure color histogram. For similar color obstacles, we get LBP histogram by processing the search window firstly and then get the discriminant coefficient through modeling. With image analysis from the search window, the S-Grabcut algorithm is used to get uncovered target area and works out its centroid. If this centroid replaces its original counterpart, the effect of object tracking will not be affected when similar color obstacles moving in. As to the fast occlusion problem, the paper has applied the Kalman filter which were confirmed (Wang and Li (2010)) to improve the real-time tracking accuracy.

2. The Constitutes of the Model of Correction Tracking Centroid

2.1 Fundamental of Grabcut

The Grabcut, N. Otsu (1979) reported, expands Graphcut from monochromatic space to color space and we use GMM (Gauss mixture model) instead of statistical histogram to model for foreground and background color modeling. Grabcut expresses images as vectors Z and definesthe opacity arrayof those images as values α. We assume that if α is equal to 1 , it indicates the corresponding image is foreground. If α is equal to 0, the images are background. The parameter θ is the eigenvalue of histogram that describes the statistical

properties of the color of foreground region and background region. The value Z is expressed as the pixel gray values.

Image Gibbs's energy function model formula which were confirmed (C. Rother et al (2004)):

$$E(\alpha, k, \theta, z) = U(\alpha, k, \theta, z) + V(\alpha, z) \tag{1}$$

Histogram gauss modelθformula:

$$\theta = \{\pi(\alpha, k), u(\alpha, k), \sum(\alpha, k), \alpha = 0,1, k = 1 \dots k\} \tag{2}$$

2.2 LBP algorithm

Texture is one of the inherent characteristics of the surface that can be considered to be a pattern in gray space in the form of a certain change. The basic idea of LBP (Local binary patterns) algorithm is that the value of the center pixel gray scale image is used as a threshold value, which is compared with its neighboring pixels so as to obtain a binary code which can be used to express the local texture features and reflect the texture information in the region. In order to adapt to the texture features of different scales, T. Ojala et al proposed an "equivalent model" method in 1996 which is used to reduce the dimension of LBP operator and named it as$LBP_{p,R}^{riu2}$.

With such improvements, the type of binary mode could be greatly reduced without losing the edge information.

2.3 S-Grabcut algorithm

Surf which were confirmed (Bay et al (2006)) applies the same method as what evaluates the approximate value of Hessa in matrix determinant to extract key points, and adds detailed information to these points, describing the main direction and building subvectors. Then it compares the features of the two images so as to find numbers of matching feature points. Finally it establishes a corresponding relationship between scenes. When Surf carried out non-overlapping regions' detection and segmentation in similar color interferences, which first obtained feature point between template and current marquee images, Surf made match points connect into a curve. Getting together with the current marquee division, they are viewed as Grabcut initialization parameters. At last, it extracts the non-overlapping target area by using Grabcut to divide. By this means Surf works out the centroid. The breaking through, mixture of Surf and Grabcut, develops Grabcut into automatic dividing and no longer needs any manual operation. Therefore, this kind of combination is called S-Grabcut algorithm. Experiment results prove that it can be automatically segmented effectively.

3. Principle and Implementation of Improved Algorithm

3.1 Purify histograms of the target object

Because background noise will be mixed into the color histogram when the original Camshift algorithm artificially selects the search window under complex backgrounds, the search window will be unable to achieve the best convergence and it will bring unnecessary deviation. In order to solve this problem, we first extract the target's three channels mean of RGB and then enhance current frame image by the enhanced coefficient.

$$Z = \left\lfloor {255}/{\max(r_1, g_1, b_1)} \right\rfloor \tag{3}$$

$$R_{(i,j)} = Z * r_{(i,j)}, G_{(i,j)} = Z * g_{(i,j)}, B_{(i,j)} = Z * b_{(i,j)} (i \leq W, j \leq H) \tag{4}$$

Z is the enhanced coefficient. r1, g1 and b1 represent three channels mean of RGB respectively, and W,H represent the overall image of width and height respectively. r(i,j), g(i,j) and b(i,j) stand for the gray value of corresponding points in the original image respectively. Then we conduct foreground segmentation on the selected target object by Grabcut.

A color histogram is established in the light of separating target object and treated the target as the initialized tracking templates of Camshift through Grabcut. Then basing on the histogram, Camshift and Kalman will make it erative calculation directly in the enhanced video. Next, the position of search window will be worked out. At last, map the size and position of the search window into the original video.

Figure 1: Grabcut flow chart

3.2 The model of track centroid correction

First of all,we should detect the size of the search window when it tends to be stable in the enhanced video stream. If the continuously three frame search window sizes vary in the range of error (The variation range of current frame's length and width is within the 0.2 times first frame's length and width), we should extract the third frame as a LBP template of target tracking. When doing above to work, we can get the corresponding LBP histogram. According to the gray-scale value, each feature statistics are stored in different bins. Next, we have the real-time mark for the current frame of the target object frame as the same process used to template. At last, we get LBP histograms of each frame image and the values of different bins.

$$B_{max} = Max(B_{0i}) , \Delta b_i = |B_{0i} - B_{ni}|, i \leq N \tag{5}$$

$$a = \begin{cases} 1, (\Delta b_i \leq 0.3 B_{max}) \\ 0, (\Delta b_i > 0.3 B_{max}) \end{cases} \tag{6}$$

N stands for the number of bin, i represents the position of the bin, B on behalf the value of bin, ais the judgment coefficient, B_0 represents the value of the bin of the template, B_n represents the bin value of LBP's histogram in the selected Nth frame.

In the test, it is clear that when a non-similar color obstacle enters, the search window becomes smaller. In contrast when similar color obstacles enter, the search window becomes inclusive and bigger. So it can be assumed that when a is 0, the search window becomes smaller, the model has no feedback and the tracking follows the same process as Camshift; When a is 0, the search window becomes bigger (The current frame's length and width is greater than the 1.2 times template's length and width), and then the S-Grabcut Segmentation begins to conduct At this time, S-Grabcut extracts uncovered the similar-color parts and works out their centroid position (x_1, y_1). Next, it replaces the center position of the current search window (x_0, y_0). Take this result as a starting point for next frame iteration. In a short,we adjust the centroid to the target object rather than obstacle so that the similar color interference could be eliminated, which reduce the image storage consumption that caused by numerous pictures matching divisions. At last, the algorithm is improved in the timeliness and accuracy.
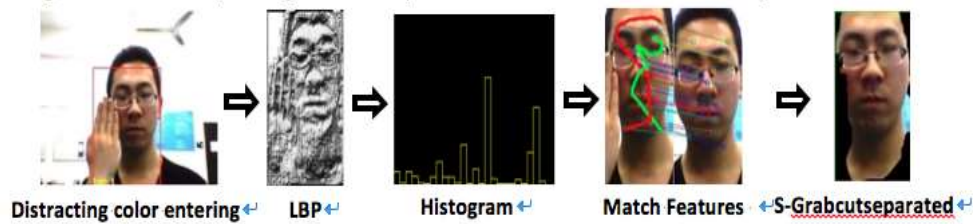

Distracting color entering   **LBP**   **Histogram**   **Match Features**   **S-Grabcutseparated**

Figure 2: Real-time extraction process of corection model

4. Experimental Result and Analysis

In the scene of test 1, we chose human face as the target object. On the premise of no obstacle, apply the tradition Camshift Algorithm and the improved Algorithm to test the matching accuracy of the search window. In this test the man moves his head from side to side, so that we could compare the degree of match the search window and object (head) when the object (head) moves or changes its size.
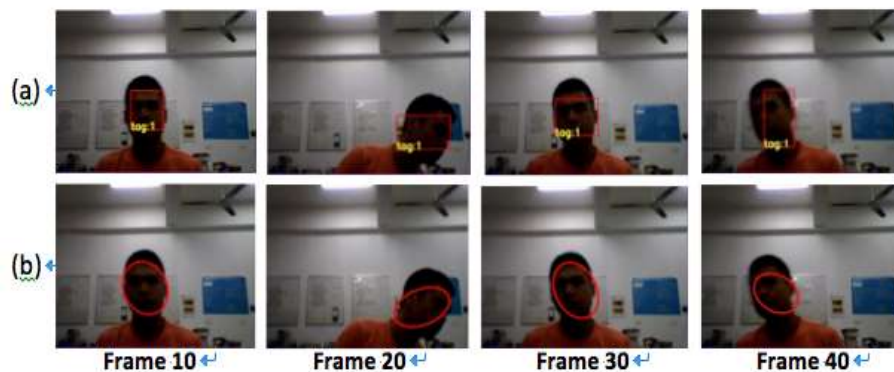

(a)

(b)

Frame 10   **Frame 20**   **Frame 30**   **Frame 40**

Figure 3: Pictures of test 1: (a) experimental pictures of this paper (b) experimental pictures of tradition Camshift

Table 1: The processing of experimental data from test 1

	Tradition Camshift		Histogram after purification		The ratio of two relative distances
	Actual center a	Search window center a1	Actual center a	Search window center a2	Δ
10	(103,354)	(121,355)	(103,354)	(115,356)	0.6753
20	(290,300)	(285,306)	(290,300)	(291,295)	0.6530
30	(440,337)	(437,338)	(440,337)	(437,337)	0.9494
40	(510,373)	(492,359)	(510,373)	(498,358)	0.8425
50	(284,312)	(295,310)	(284,312)	(289,310)	0.4821
60	(239,318)	(231,320)	(239,318)	(243,320)	0.5418

Note:$\Delta= \sqrt{(x_{a_2} - x_a)^2 + (y_{a_2} - y_a)^2}/\sqrt{(x_{a_1} - x_a)^2 + (y_{a_1} - y_a)^2}$

From the analysis result of the test 1, enhancing and purifying the target histogram is useful to reduce the interference which is caused by edge background noise. Compare the data of search window centroids with actual centroids from frame10 to frame 60, all the relative distance rate between centroids are less than 1. It indicates that after improvement the distance between the search window centroids and the actual centroids of objects is smaller than before. Therefore it can show that the fitting degree between the Search window and the object increases as well as the tracking effect is improved.

The test 2 aims to test and compare the tracking effects of those three tracking algorithms when objects are not all under similar color obstacles. As before, the human face is the object being tracking and the hand is regarded as the similar color distraction obstacle. In the test, the hand is placed at the 2/3 of the face and then moved from one side to the other. We test the effects of tradition Camshift, the fusion algorithm of Camshift and Kalman as well as the algorithm which mentioned in this paper under some same frames and compare them. The results are as follows:
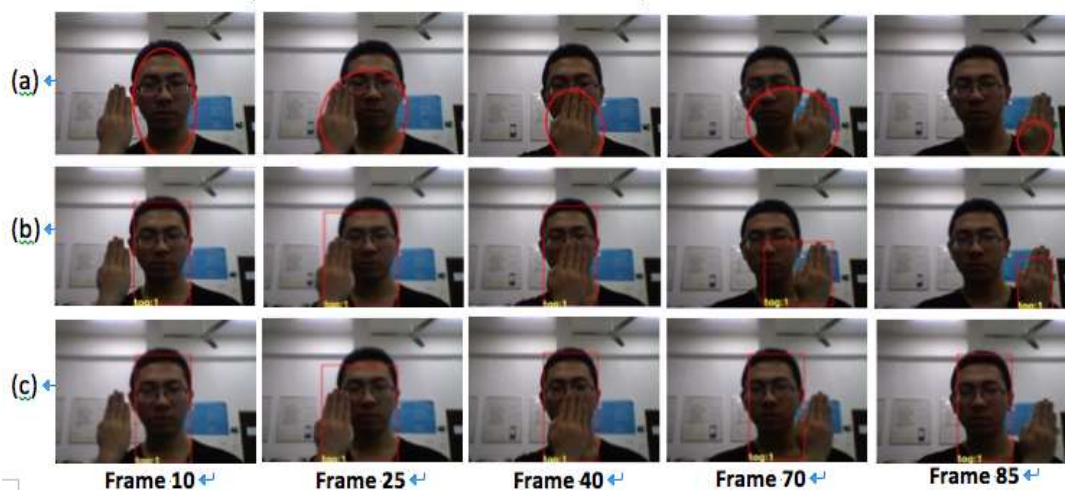

Figure 4: Pictures of test 2: (a) experimental pictures of tradition Camshift (b) experimental pictures of the fusion algorithm of Camshift and Kalman (c) experimental pictures of this paper

Table 2: The processing of experimental data from test 2

	Original		After image matting		Relative moving distance	
	Actual center a	Search window center a'	Actual center b	Search window center b'	Δ	
10	(344,299)	(89,181)	(345,300)	(90,182)	1.414	
25	(311,319)	(122,160)	(276,320)	(87,161)	35.0143	
40	(326,312)	(106,166)	(280,289)	(60,143)	51.4296	
55	(329,302)	(99,174)	(317,195)	(87,67)	107.6708	
70	(340,298)	(91,180)	(334,200)	(85,82)	98.1835	
85	(341,295)	(91,174)	(321,229)	(63,108)	71.6938	
100	(359,297)	(97,176)	(340,230)	(78,109)	69.6419	
115	(349,296)	(88,177)	(343,289)	(82,170)	9.2195	
130	(344,295)	(87,178)	(341,292)	(84,175)	4.2426	

Note:$\Delta = \sqrt{(x_{a'} - x_{b'})^2 + (y_{a'} - y_{b'})^2}$

According to the test 2, when the target (the face) is covered by similar color obstacle(the hand), the tradition Camshift algorithm is under great influences. When the hand enters in, the search window is deformed irregularly and the centroid is totally moved. In the fusion algorithm of Camshift and Kalman, when the hand enters, the search window becomes smaller while the window is still connected with the head. However, when hand moves out, the centroid is moved together with the hand and its place is changed obviously. In the 85th frame, the centroid leaves from the target. The results of the improved algorithm show that when the hand enters, the search window keeps the same effect with that under the previous algorithms——their containments are the same (the frame expands). When the hand moves away from the face, after the frame 55, the window does not change correspondingly. Therefore, the corrected centroid still stays at the face. The search window stays even when the hand moves out of the face picture.

According to the forms and the relative movement distances which derived from the centroid correction model, which shows that when similar color obstacles enter into video stream, this model does move the original centroid of the current frame, so the centroid would be kept within the target object but not be distracted or moved by similar color obstacles and lead to tracking mistakes. The results above show that when targets are distracted by similar color obstacles, by using the improved algorithm in this paper, the method of centroid correction which is based on LBP and S-Grabcut can decrease the error-tracking probabilityand increase the tracking robustness.

5. Conclusions

This paper, after analysis of the processes of the tradition Camshift algorithm, has presented a Camshift algorithm that is based on Grabcut with a LBP model of correction tracking centroid, which can be applied into tracking moving target objects. Setting a fixed number to enhance the tracking video and with the help of Grabcut draw out a colorful histogram of pure color of the object which could help to decrease the background noise and distraction. Build a LBP-centroid tracking correction model, discern the similar color obstacles distraction and use the S-Grabcut algorithm to move the original centroid to the un-covered part of target object. What's more, the result of automatic segmentation is good and the movement of the centroid is high-efficiency. According to the tests, this model has a good effect on against the distraction of similar color and it can satisfy the requirement of real-time and stability. At last, the Kalman filter is used to predict the location of the target object and track effectively when most parts of the object are covered or the object moves quickly. As a result, the robustness of the whole system could be enhanced.

References

Bay H., Tuytelaars T., and Gool L.V. 2006. SURF: Speeded up robust features. Computer vision–ECCV 2006. Springer Berlin Heidelberg. Vol. 17(8) (pp. 404-417). DOI: 10.1007/11744023_32.

378

Cheng Y.Z. 1995. Mean shift, mode seeking, and clustering. Pattern Analysis and Machine Intelligence, IEEE Transactions on. Vol. 17(8) (pp. 790-799). DOI: 10.1109/34.400568.

Exner D., Bruns E., Kurz D., Grundhoefer A., Bimber O. 2010. Fast and Robust CAMshift Tracking. Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. (pp. 9-16). DOI: 10.1109/CVPRW. 2010.5543787.

Fukunaga K., Hostetler L.D. 1975. The estimation of the gradient of a density function with applications in pattern recognition. Information Theory, IEEE Transactions on. Vol. 21(1) (pp. 32-40). DOI: 10.1109/TIT.1975.1055330.

Hou Y.C., Peng W.C. 2014. Distance Between Uncertain Random Variables. Mathematical Modelling and Engineering Problems. Vol. 1(1). (pp. 17-24).

Jianhui W., Guoyun Z., Longyuan G. 2013. Study The improved CAMSHIFT algorithm to detect the moving object in fisheye image. Mechatronic Sciences, Electric Engineering and Computer (MEC), Proceedings 2013 International Conference on. (pp: 1017-1020). DOI: 10.1109/MEC. 2013.6885210.

Li Z.Y. 2011. Modified local entropy-based transition region extraction and thresholding. Applied Soft Computing. Vol. 11(8) (pp. 5630-5638). DOI: 10.1016/j.asoc. 2011.04.001.

Ojala T., Pietikainen M., Harwood D. 1996. A comparative study for texture measures with classification based on feature distributions. Pattern Recognition. Vol. 29(1) (pp. 51–59). DOI: 10.1016/0031-3203(95)00067-4.

Ojala T., Pietikäinen M., Mäenpää T. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol: 24(7) (pp. 971-987). DOI: 10.1109/TPAMI. 2002.1017623.

Otsu N. 1979. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man & Cybernetics. Vol. 9(1). (pp. 62-66). DOI: 10.1109/TSMC.1979.4310076.

Rother C., Kolmogorov V., Blake A. 2004. "Grabcut"-Interactive foreground extraction using iterated graph cuts. ACM Transactions on Graphics (TOG). Vol. 23(3) (pp. 309-314). DOI: 10.1145/1015706.1015720.

Tu J.H. 2014. A Novel Building Boundary Extraction Method for High-Resolution Aerial Image. Review of Computer Engineer Studies. Vol. 1(2). (pp. 19-22).

Wang C., Zhu S.M. 2015. A Design of FPGA-Based System for Image Processing. Review of Computer Engineer Studies. Vol. 2(1). (pp. 23-28).

Wang X.Y., Li X.J. 2010. The study of Moving Target tracking based on Kalman-CamShift in the video. Information Science and Engineering (ICISE), 2010 2nd International Conference on. (pp. 1-4) DOI: 10.1109/ICISE. 2010.5690826.

Zhao S., Gao Y. 2008. Establishing Point Correspondence Using Multidirectional Binary Pattern for Face Recognition. Pattern Recognition, 2008.ICPR 2008.19th International Conference on. (pp: 1–4). DOI: 10.1109/ICPR. 2008.4761248.