Improved Camshift with adaptive searching window

*Aixia Wang, Jingjiao Li,Zhenlin Lu
Northeastern University, Shenyang, China

e-mail: wax526@tom.com

Abstract. **Camshift is widely used real-time algorithm in video target tracking field. The size of searching window (SW) is a key factor of Camshift, and bigger or smaller size of SW will both decrease the real-time feature of Camshift. In this paper, a accelerated Camshift with adaptive searching window (ACASW) was proposed. Firstly the meanshift process and computational cost (CC) were modeled, and the relationship between the size of SW and CC was analyzed quantificationally, then the optimized size of SW was deduced, which was used in the proposed algorithm. From the experimental results it can be seen that, compared to the traditional Camshift with fixed EW, the proposed algorithm can reduce the computing time effectively, which improve the real-time feature of the algorithm.**

Keywords: *real-time target tracking, camshift, adaptive searching window*

1. Introduction

Moving target tracking is the visual surveillance, video compression and automatic applications such as visual analysis of the key technologies. The vast majority of applications require target tracking algorithm in the block, light change and the goal of scaling, rotation, is robust. Meanwhile, the tracking algorithm also need to consume as little as possible time to meet the real-time requirements. Tracking algorithm matches the search category as a typical representative of the mean-shift tracking algorithm [1-2], its calculation of a small amount of the target deformation, rotation, adaptability and other characteristics change widely attention and a wide range of applications. In the field of video-tracking is a continuous adaptive mean shift algorithm is the mean offset algorithm Camshift. According to the colors of targets to be tracked back projection calculations, Camshift the vicinity of the region in the current search and track the target to be the same color component, through the mean shift iterations to find them the same color component of the centroid position, thus completing the tracking process. Compared with other tracking algorithms [3-5], Camshift less affected by background noise and calculate the amount of small, so widely used among real-time target tracking. At present, based on a lot of Camshift target tracking applications. [6] gives a Camshift interesting applications. They will support vector machine SVM and extract the image combination Camshift text. First, the use of training of the SVM based image texture image into the text of regional and non-text area, then use the text area Camshift extracted. They also use the text size of the image pyramid to solve the problem of inconsistency. [7] proposed a dynamic background Camshift algorithm ABCshift. Camshift and basic difference is that, ABCshift the search window at the same time the background image histogram statistics, and in which the statistical weight assigned to the negative weights. In this way, objectives and background if there is a similar composition, then the component will be a corresponding reduction in weight, ABCshift through this strategy to weaken the impact of background noise. [8] Statistical shape features will combine the model and Camshift tracking, respectively, on a number of shape features in order to empower the value of comprehensive consideration. [9] use Adaboost classifier for face detection, followed by Camshift track as a way to form the human face detection and tracking system. [10] First, calculation of target appearance under different light conditions and constitute a database to track the use of Camshift process, the right lighting conditions for the current frame to assess

the use of assessment results as a search criteria in the database to extract the corresponding object model, using This means to overcome the light conditions change. [11] will be applied automatically Camshift aid in the automatic tracking problem among wheelchair. In [12] gives a multi-model switching means of simulation using the proposed method of a better place to track major changes in the appearance of target when the tracking problem.

Although the small amount of Camshift algorithm has the advantages of the calculation, but in some strong real-time areas, such as video capture device IO operations accounted for most of the time frames, how to further reduce the computational Camshift algorithm is still an urgent need resolved. Scholars at home and abroad there is no in-depth discussion on this issue. A fast algorithm for continuous adaptive mean shift FCshift, through Camshift in Camshift frames meanshift iterative process and the calculation of the amount of modeling, quantitative analysis of the expansion of the size of the window Camshift the relationship between the computation and thus derive out of the calculation basis for optimal expansion of the window, thereby constituting a continuous adaptive mean-shift algorithm for fast FCshift.

2. Camshift Algorithm

Camshift algorithm needs the probability distribution of the target area of the color images, so need to create a color histogram of the color model of the target area. At this point, the RGB color model used can not meet the requirements, but rather use the HSV color model. Concrete steps are as follows: Compute the color histogram of target to be tracked. Will be received by the camera RGB space images into HSV space images, HSV color space components, said color information, and generate one-dimensional color histogram calculated. According to obtain the original image color histogram into a probability distribution of color images, this process shall get back projection (Back Project) process. Transformation process, in order to color histogram as a look-up table. For the color represented by the 8 Bit value, the probability distribution of the value range is $0 \sim 255$.

Camshift of the tracking process is inherently iterative process of Meanshift. By calculating the reverse projection of the various regions in the search order moments can be obtained, such as target tracking, the size, direction, mass and other characteristic quantities, the specific process is as follows:

Computing zero-order moment:

$$m_{00} = \sum_{x} \sum_{y} I(x, y) \tag{1}$$

Calculating a first-order moment:

$$m_{10} = \sum_{x} \sum_{y} x I(x, y) \tag{2}$$

$$m_{01} = \sum_{x} \sum_{y} y I(x, y) \tag{3}$$

Calculating the centroid position of the search window:

$$x_c = \frac{m_{10}}{m_{00}} \tag{4}$$

$$y_c = \frac{m_{01}}{m_{00}}$$

(5)

Color map is a discrete probability distribution of the grayscale image, its maximum is 255, thus setting the initial window size and the relationship:

$$s = 2 * \sqrt{\frac{m_{00}}{256}}$$

(6)

Taking into account the symmetry, so that the calculated results for the nearest odd number. By calculating the two-order image moments can be tracked target direction:

$$m_{20} = \sum_x \sum_y x^2 I(x, y)$$

(7)

$$m_{02} = \sum_x \sum_y y^2 I(x, y)$$

(8)

$$m_{11} = \sum_x \sum_y xy I(x, y)$$

(9)

Can get the length and width of search window are as follows:
Window direction:

$$\theta = \frac{\arctan\left(\dfrac{b}{a-c}\right)}{2}$$

(10)

Where：

$$a = \frac{m_{20}}{m_{00}} - x_c^2$$

(11)

$$b = 2\left(\frac{m_{11}}{m_{00}} - x_c y_c\right)$$

(12)

$$c = \frac{m_{02}}{m_{00}} - y_c^2$$

(13)

When using Camshitf algorithm to track a specific target color when the images do not have to calculate all the pixels of each frame the color of the probability distribution, only the calculation window larger than the current search area pixel within the color of the probability distribution, it can reduce the computation and improve Camshitf algorithms in computational speed the process to ensure that in the video tracking process in real time. When the video sequence frame by frame change, Camshift algorithm processed frame by frame so that you can complete the continuous tracking of the moving target.

3. FCShift

In order to clarify the following discussion, the section later in this article, will lock the target location of the window is called the Target box, but will expand the window is called after the search box. Frame set up the first goal box length and width, respectively, and the target centroid location box, the search box length and width, respectively, and, where. The first frame of the goal box, centroid position in the target box and the length and width, respectively. The sake of discussion, first of the three basic assumptions given in this article.

1. target color distribution is uniform.
2. ignoring the background noise interference.
3. between the two adjacent targets the size of the basically unchanged.

These three assumptions are necessary, they can be discussed within the scope of norms in Geometry. In fact, in many tracking scenarios, such as football matches of the soccer tracking, the three basic assumption is usually satisfied.

3.1. The estimation of the number of iterations

First give an algorithm to determine the number of iterations to determine Camshift theorem.
Theorem 1. To indicate the first frame after the first iterations the search box in center of mass frame with the target centroid position of residuals. In the Camshift iterative process, with the increases, when, and when we can see the first frame Camshift iteration number of times.
The proof of this theorem will be given in the following analysis.

The first frame in the target locations with the first frame to the position of search window corresponding to five kinds of relationship shown in Figure 1. Diagram, the shaded rectangle, respectively, as well as the first frame as the first frame of the target location, a large rectangular frame in the search box for the first position. The following will be the relationship of these five kinds of the corresponding Camshift iterations discussed.
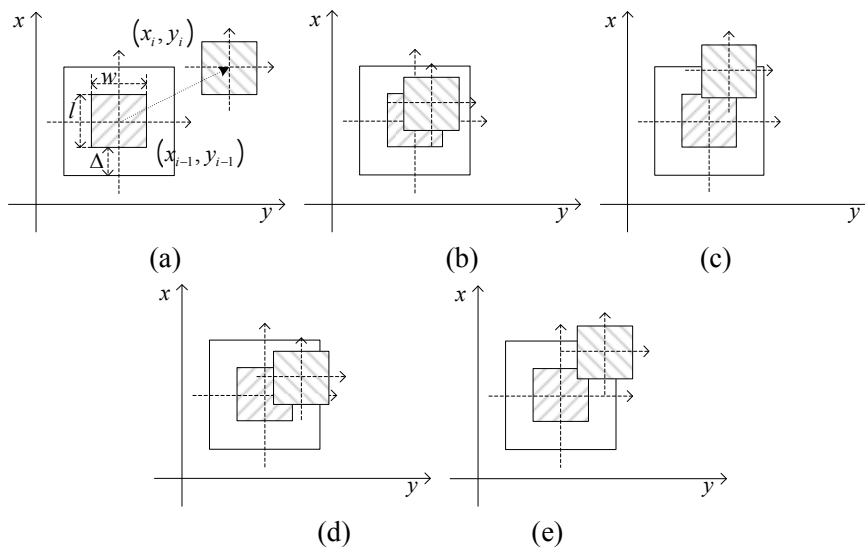

Figure 1. Example of typical situations

(1) When the first frame the target beyond the first frame the scope of the search box in Figure 3.a, that is or, then the tracking fails.

(2) When the first frame the target of at least part of the first frame in the search box, as Figure 3.b-3.e, this time to meet and, first of all, when the first goal of the frame all in the first frame in the search box, as Figure 3.b, this time to meet and, with that the first frame after the first iterations the search box centroid position in the goal after the first iteration centroid position:

$$x_{i,1}^c = x_{i-1} - \frac{w}{2} + \Delta x_{i-1} + \frac{w}{2} = x_{i-1} + \Delta x_{i-1} \qquad (14)$$

$$y_{i,1}^c = y_{i-1} - \frac{w}{2} + \Delta y_{i-1} + \frac{w}{2} = y_{i-1} + \Delta y_{i-1} \qquad (15)$$

After the first iteration the location of the residuals as follows:

$$x_{i,1}^e = x_{i-1} + \Delta x_{i-1} - (x_{i-1} + \Delta x_{i-1}) = 0 \qquad (16)$$

$$y_{i,1}^e = y_{i-1} + \Delta y_{i-1} - (y_{i-1} + \Delta y_{i-1}) = 0 \qquad (17)$$

We can see the first frame in Camshfit iterative process is complete.
Based on the above conclusions, in the Camshfit iterative process, when, and when we can see the first frame Camshift iteration number of times. Theorem 1 may permit. In the following discussion will use this conclusion.

(3) the nature of c, and d are the same, here to c as an example for discussion. In the c, satisfy, and. After the first iteration, the search box centroid:

$$x_{i,1}^c = x_{i-1} - \frac{w}{2} + \Delta x_{i-1} + \frac{w}{2} = x_{i-1} + \Delta x_{i-1} \qquad (18)$$

$$y_{i,1}^c = y_{i-1} - \frac{l}{2} + \Delta y_{i-1} + \frac{l + \Delta - \Delta y_{i-1}}{2} = y_{i-1} + \frac{\Delta y_{i-1} + \Delta}{2} \qquad (19)$$

Position residuals as follows:

$$x_{i,1}^e = x_{i-1} + \Delta x_{i-1} - (x_{i-1} + \Delta x_{i-1}) = 0 \qquad (20)$$

$$y_{i,1}^e = y_{i-1} + \Delta y_{i-1} - (y_{i-1} + \frac{\Delta y_{i-1} + \Delta}{2}) = \frac{\Delta y_{i-1} - \Delta}{2} \qquad (21)$$

Easy to deduce the location of the first residuals after the second iteration as follows:

$$x_{i,1}^e = 0 \qquad (22)$$

$$y_{i,j}^e = \frac{\Delta y_{i-1} - \left(2^j - 1\right)\Delta}{2^j} \qquad (23)$$

According to Theorem 1, if satisfied:

$$\Delta > \frac{\Delta y_{i-1} - \left(2^j - 1\right)\Delta}{2^j} \qquad (24)$$

Camshift second iteration is completed after the first calculation.
By the formula 2 can be, we can see for Figure 3.c of the situation, that is to satisfy and conditions, the first frame in the number of iterations Camshift:

$$N_i = \begin{cases} +\infty, & \Delta = 0; \\ \left\lceil \log_2\left(\frac{\Delta y_{i-1}}{\Delta} + 1\right) \right\rceil, & \Delta > 0; \end{cases} \qquad (25)$$

Which is rounded up.
Similarly, we can see the situation for the diagram, d, that is to satisfy and conditions, the first frame in the number of iterations Camshift:

$$N_i = \begin{cases} +\infty, & \Delta = 0; \\ \left\lceil \log_2\left(\frac{\Delta x_{i-1}}{\Delta} + 1\right) \right\rceil, & \Delta > 0; \end{cases} \qquad (26)$$

Integrated map c and Figure d the conclusions, we can see for the plan e of the situation, that is to satisfy and conditions, the first frame in the number of iterations Camshift:

$$N_i = \begin{cases} +\infty, & \Delta = 0; \\ \max\left(\left\lceil \log_2\left(\frac{\Delta x_{i-1}}{\Delta} + 1\right) \right\rceil, \left\lceil \log_2\left(\frac{\Delta y_{i-1}}{\Delta} + 1\right) \right\rceil\right), & \Delta > 0; \end{cases} \qquad (27)$$

It can be seen from the formula, when the search window size is equal to the size of the tracking window, real-time, in theory, Camshfit algorithm will never converge to the centroid position. At that time, the number of iterations only with the amount of displacement and expansion of the value of the ratio of the relevant. The actual operation in the algorithm process, taking into account the residual item, when the video track is the minimum pixel block and when to meet, when the following formula:

$$\frac{\Delta x_{i-1} - (2^n - 1)\Delta}{2^n} - \Delta < 0.5 * b \qquad (28)$$

$$\frac{\Delta y_{i-1} - (2^n - 1)\Delta}{2^n} - \Delta < 0.5 * b \qquad (29)$$

By the formula 4.a and 4.b can deduce the formula when the number of iterations to meet the 5 pm:

$$N_i = \max\left(\left\lceil \log_2(\frac{\Delta x_{i-1} + \Delta}{0.5b + 2\Delta}) \right\rceil, \left\lceil \log_2(\frac{\Delta y_{i-1} + \Delta}{0.5b + 2\Delta}) \right\rceil\right) + 1 \tag{30}$$

Camshift algorithm converges to the target centroid position.

3.2. Computation load estimation

Camshift with said inter-frame computation, according to the calculation of Camshift process, we can see with the expansion of the box, and the number of iterations, that is, which is a single increasing function, for a fixed value of the situation:

$$k(w, l, \Delta, n) = n(w + 2\Delta)(l + 2\Delta) \tag{31}$$

Through analysis and,, and to analyze the relationship between the size of search window Camshift the relationship between the computation. Without loss of generality, set up, you can get:

$$k(w, l, \Delta, n) = n(w + 2\Delta)(l + 2\Delta) = \left\lceil \log_2(\frac{\Delta y}{\Delta} + 1) \right\rceil (w + 2\Delta)(l + 2\Delta) \tag{32}$$

In order to simplify the calculation, so that:

$$\left\lceil \log_2(\frac{\Delta y}{\Delta} + 1) \right\rceil (w + 2\Delta)(l + 2\Delta) \approx \left(\log_2(\frac{\Delta y}{\Delta} + 1) \right)(w + 2\Delta)(l + 2\Delta) \tag{33}$$

To calculate the convenience, can be used instead:

$$g(w, l, \Delta, n) = \left(\log_2(\frac{\Delta y}{\Delta} + 1) \right)(w + 2\Delta)(l + 2\Delta) \tag{34}$$

where $\Delta y < l + \Delta$。
Rewrite the formula 7:

$$g = w^2 \left(\log_2(\frac{\alpha}{\chi} + 1) \right)(1 + 2\chi)(\beta + 2\chi) \tag{35}$$

Where:

$\alpha = \dfrac{\Delta y}{w}$, is the inter-frame velocity ratio;

$\chi = \dfrac{\Delta}{w}$, is the inter-frame expansion ratio;

$\beta = \dfrac{l}{w}$ is the target aspect ratio.

In the last equation, is determined by objective characteristics of their own decisions; with the adjacent two image target centroid is proportional to the distance. The following diagram shows specific, and situations between and images:
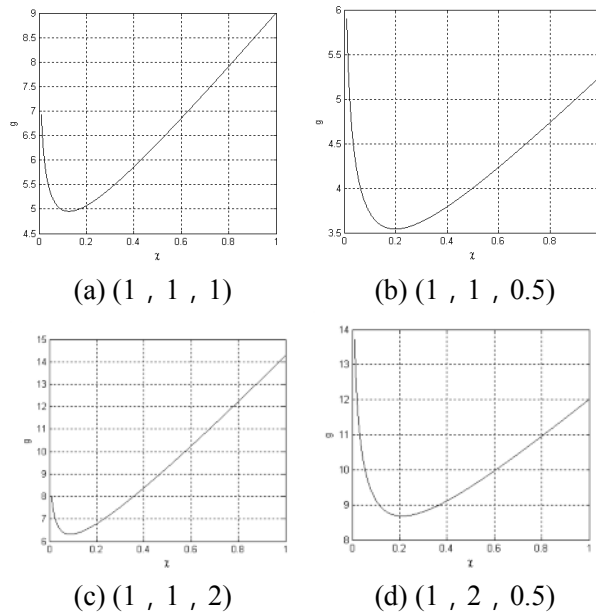

(a) (1 , 1 , 1)   (b) (1 , 1 , 0.5)

(c) (1 , 1 , 2)   (d) (1 , 2 , 0.5)

Figure 2. **Example of different situations**

Figure in brackets represent the triples, and values. Can be seen from Figure 4, specific, and the minimum conditions exist, and different, and the corresponding is different. Therefore, select the appropriate search window to allow the calculation of Camshift reached minimum.

3.3. determining the size of the searching window

In order to obtain the minimum value corresponding to the formula 8 on the partial derivatives of demand and make it equal to zero, to be the formula 9

$$\frac{\alpha(1+2\chi)(\beta+2\chi)}{(\alpha\chi+\chi^2)} - 2(\log_2(\frac{\alpha}{\chi}+1))(1+\beta+4\chi) = 0 \tag{36}$$

Equation 36 on the direct solution and the analytical solution is difficult, you can pre-set and values, and use a numerical method for solving the corresponding. In practice, the calculation method used in look-up table and using interpolation algorithm to obtaining the necessary values. The following table shows part of step 0.4 with the corresponding relations.

Table 1. Relationship between searching window, target speed and target size

β \ α	0.4	0.8	1.2	1.6	2.0
0.4	0.0986	0.0640	0.0522	0.0461	0.0423
0.8	0.1907	0.1215	0.0966	0.0838	0.0759
1.2	0.2660	0.1693	0.1324	0.1134	0.1018
1.6	0.3306	0.2106	0.1626	0.1380	0.1230
2.0	0.3880	0.2473	0.1889	0.1590	0.1409

As can be seen from Table 1, with the increase, that is, between the two target displacement increases, the optimal expansion of the window by expanding content is gradually increasing. In the actual calculation process, we can use formulas to estimate the 10 pairs and:

$$\Delta x_{i-1} = 2\Delta x_{i-2} - \Delta x_{i-3} \tag{37}$$

$$\Delta y_{i-1} = 2\Delta y_{i-2} - \Delta y_{i-3} \tag{38}$$

calculate the current through the formula:

$$\alpha = \frac{\max(\Delta x_{i-1}, \Delta y_{i-1})}{w} \tag{39}$$

In the calculation, depending on the target box, the information look-up table and interpolation calculation of the corresponding search window. It should be noted that, due to space constraints, Table 1 only give part of the step size of 0.4, compared with the corresponding relations. In the actual calculation can be more small step in a long pre-calculation of data in order to enhance the accuracy of the actual use. Due to the need before the two displacement information to determine the current frame, and therefore is calculated starting from the first three, the 1,2-frame displacement Camshift use the basic algorithm to obtain the information.

4. Experiments

The main contents of this experiment is done under different conditions pendulum movement in table tennis, track. Experiment consists of three parts, including the use of different rates of expansion of Camshift algorithm and FCshift in the background noise and background noise, the smaller the larger context of performance comparisons, as well as failure analysis algorithm. The main criteria include comparison test average number of iterations of each frame, the average processing time of each frame, and the algorithm failure rate. Involved in the operation of this algorithm is the hardware platform, P4 3.0GHz, 1G memory, the software platform for Matlab. For your convenience, will use the expansion rate of 0.1 Camshift marked Camshift0.1, the remaining cases and so on.

4.1. Background with less nosie

In Experiment 1, selected 10 video, which features video and background color of table tennis, there is a clear difference between the colors, and the experiment has always been involved in a variety of algorithms are able to lock on to targets. A total of 100 for each video. Figure 5 shows the tracking of FCshift algorithm. Which, 5.a target initial position, 5.b for the corresponding back-projection map, 5.c for the color histogram, 5.d ~ 5.f screenshot for the tracking process.
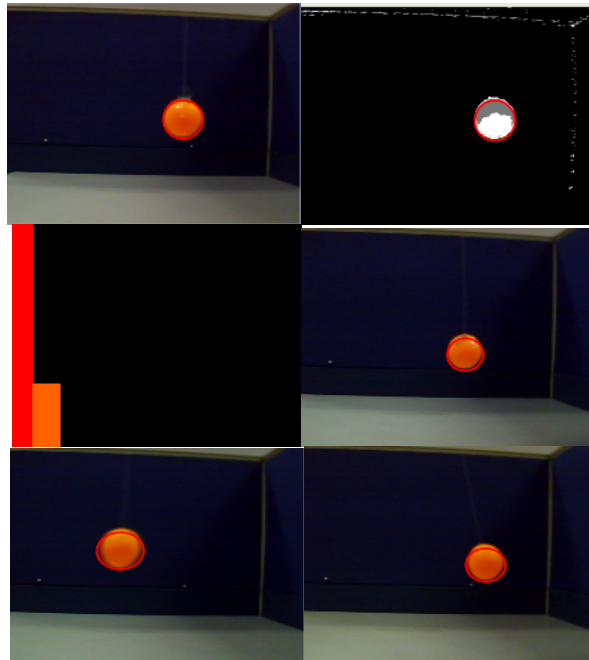
Figure 3. Experimental results of test 1

Table 1. Relationship between searching window, target speed and target size

β \ α	0.4	0.8	1.2	1.6	2.0
0.4	0.0986	0.0640	0.0522	0.0461	0.0423
0.8	0.1907	0.1215	0.0966	0.0838	0.0759
1.2	0.2660	0.1693	0.1324	0.1134	0.1018
1.6	0.3306	0.2106	0.1626	0.1380	0.1230
2.0	0.3880	0.2473	0.1889	0.1590	0.1409

Table 2 shows the different rate of expansion Camshift circumstances, as well as FCshift the average number of iterations and time of each frame ratio

Table 2. Relationship between searching window, target speed and target size

No.	Extend ratio	Iter-time	Time
1	0	3.312	1.579
2	0.05	2.697	1.454
3	0.1	2.105	1.385
4	0.2	1.872	1.231
5	0.5	1.564	1.606
6	1.0	1.178	1.950
7	FCshift	1.713	1

As can be seen from Table 2, using a different rate of expansion of the Camshift corresponds to the average number of iterations of each frame is different. With the rate of expansion increases, the average number of iterations of each frame is reduced. Which, Camshift1.0, the average number of iterations of each frame is only 1.178, that

is, most of the frame in this condition only requires a calculation of iterations have been completed. FCshift the average number of iterations of each frame as 1.713, with similar Camshift 0.2. From the experimental data can know, for the background noise is relatively small, increased rate of expansion will reduce the number of iterations the algorithm. However, the rate of expansion through a variety of Camshift run time and the ratio of the running time of FCshift can be seen, first of all extended Camshift rate was higher than the running time of FCshift, which Camshift 1.0 the running time of almost twice the FCshift; secondly as the rate of expansion increases, the running time of the beginning of steady decline, but then increased rapidly.

4.2. Background with more nosie

In experiment 2, the same choice of 10 video, the video is characterized by the existence of more background noise, but the experiment has always been involved in a variety of algorithms are able to lock on to targets. A total of 100 for each video. Figure 6 shows the tracking of FCshift algorithm. Which, 6.a target initial position, 6.b for the corresponding back-projection map, 6.c for the color histogram, 6.d ~ 6.f screenshot for the tracking process.
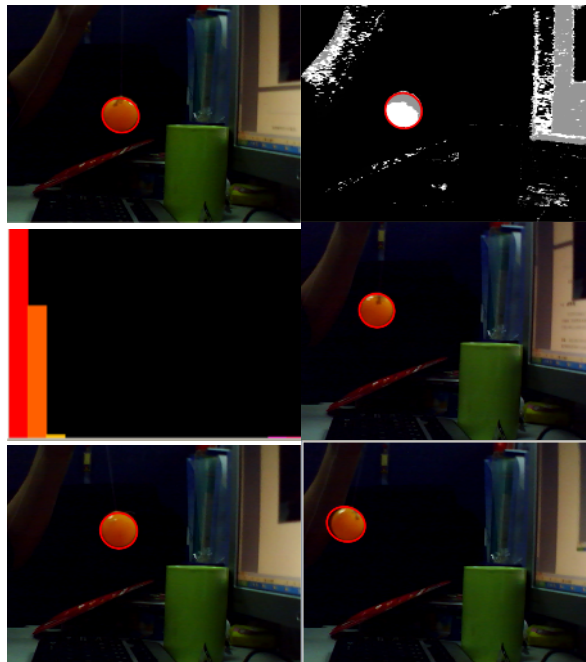

Figure 4. **Experimental results of test 2**

Table 3 shows the different rate of expansion Camshift circumstances, as well as FCshift the average number of iterations and time of each frame ratio

Table 3. Relationship between searching window, target speed and target size

No.	Extend ratio	Iter-time	Time
1	0	3.673	1.513
2	0.05	2.854	1.494
3	0.1	2.350	1.336
4	0.2	1.915	1.262
5	0.5	1.698	1.587
6	1.0	1.231	2.135
7	FCshift	1.814	1

Tables 2 and 3 by comparison of the data can be seen that the expansion of the use of different rates of Camshift and FCshift, when there are more background noise, the average number of iterations of each frame there is less noisy than the background value when the big . This can be interpreted as each iteration the search box will contain the new background noise, resulting in the search box centroid position change. From the ratio between the running of it, using a different rate of expansion of the Camshift running time and the ratio of the running time FCshift have also been enhanced.

5. Conclusion and Future Work

Continuously adaptive mean shift algorithm Camshift is an excellent real-time tracking algorithm. However, real-time tracking for certain occasions, such as real-time demanding, but the video capture device IO occupy a larger proportion of the time, how to further improve the calculation speed Camshift algorithm remains a serious problem. In this paper, the calculation of the tracking process and the amount of data to model their impact on algorithm performance Camshift the key parameters - expansion of the rate selected on the basis, which constitutes a rapid and continuous adaptive mean-shift algorithm FCshift. Experiments proved that FCshift algorithm can reduce the computing time at the same time a certain extent, to overcome the background noise. In the future, we will try to further improve the algorithm performance by combining some optimization method with it.

References

[1] Bradski G R. Real time face and object tracking as a component of a perceptual user interface. In: Proceedings of the 4th Workshop on Application of Computer Vision. Princeton, USA: IEEE, 1998. 214-219.

[2] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-577

[3] Comaniciu D, Meer P. Mean Shift: a robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 603-619

[4] Collins R T, Liu Y X, Leordeanu M. Online selection of discriminative tracking features. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10):1631-1643.

[5] Matthews L, Ishikawa T, Baker S. The template update problem. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(6): 810-815.

[6] Kwang K, Keechul J, Jin HK. Texture-Based Approach for Text Detection in Images Using Support Vector Machines and Continuously Adaptive Mean Shift Algorithm, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2003, 25(12): 1631-1693.

[7] Prahlad V, Peter L, Liyanage C, Liu J. Multimodal Approach to Human-Face Detection and Tracking IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, 2008, 55(3): 1385-1393

2222222222222222222222222222222