Camshift Algorithm Combined with Contour Coordinate Point Target Tracking

Chao Han¹, Fucheng You¹, Shuai Zhao¹ and Shaomei Wang¹

¹Beijing Institute of Graphic Communication. No. 1, Xinghua Street (two sections), Daxing District, Beijing
Email: m17865675907@163.com

Abstract Aiming at the deficiency of traditional CamShift algorithm in losing targets easily in video sequence of background with color similarity, this paper explores the reasons of losing targets, gives the numerical requirements to ensure that targets are followed precisely and proposes a target tracking algorithm combining contour coordinate points. The algorithm modifies the tracking window according to the coordinate points of the target contour, so as to obtain the size and position of the target in the current image and improve the tracking stability. The experimental results show that the target can be tracked effectively under the interference of color similarity background.

1. Introduction
Based on the color features of the target to be tracked by Camshift algorithm, the subsequent video sequence adopts the method of reverse projection to process the pixels of the image frame, and then the image frame is transformed into a reverse projection graph with the gray value as the standard to represent the probability density of the target to be tracked. The brighter the pixel in the grayscale projection, the more likely it is to be the target pixel. Due to the use of a single color feature, the color information of the target is relatively stable in the process of motion, so the algorithm is less affected by the deformation, scaling and rotation of the target, and can achieve better tracking effect in a simple background environment.

For similar color backgrounds and targets with interference, the color feature histogram model has poor discrimination effect, and the traditional Camshift algorithm is prone to lose targets. Firstly, this paper analyzes the relation and difference between the color histogram and the reverse projection diagram calculated by the video sequence in the target tracking of the traditional Camshift algorithm, vividly displays the reason of target tracking loss, and gives the numerical requirements for effective target tracking. Then, an improved algorithm combined with tracking target contour coordinate points is proposed. When the algorithm determines that the target is lost, it corrects the tracking box and updates the color model by calculating the coordinate values of the upper left and lower right corner of the contour, and then enters the tracking work of the next frame.

2. Camshift Algorithm
Camshift is a continuous adaptive Meanshift algorithm and Meanshift is a mean-shift tracking algorithm based on gradient calculation. In a two-dimensional space, the mean shift vector is searched step by step along the direction in which the probability density increases, so as to obtain the local maximum value of the spatial probability density. The expression is shown in formula (1).

\[ M = \frac{1}{n} \sum_{k=1}^{n} (a_k - \bar{a}) \]
Where, $\overrightarrow{a_k}$ is the set of all points contained in a circle with $\overrightarrow{a}$ as the center and $h$ as the radius, as shown in figure 1.

![Figure 1. Mean drift vector](image)

In the Camshift tracking algorithm, the probability density is the pixel value of the reverse projection image represented by the gray value. The mean shift vector is the zero order distance and the first order distance of the pixel value of the search window. Meanshift algorithm is to calculate the optimal iterative results on a specific image, while Camshift is a continuous video-oriented sequence, and Meanshift is carried out on each image frame one by one to obtain the optimal search results [1]. The specific flow chart is shown in figure 2.

![Figure 2 flow chart of the camstreamline algorithm](image)

Firstly, the initial search window is set, and then the H channel value of each pixel in the window is sampled, so as to obtain the color histogram of all pixels in the window and take it as the target color model. In the tracking process, the gray value of each pixel in the video image processing area is compared with the color model of the target, and the probability that the pixel belongs to the target pixel is obtained, while other areas outside the image processing area are regarded as the area with a probability of 0. After the above processing, the video image is transformed into a probability projection diagram, also known as the target color projection diagram [2]. To facilitate display, the projection is converted into an 8-bit grayscale projection. The pixel value with probability 1 is set to 255, and the pixel value with probability 0 is 0. Other pixels are also converted into corresponding grayscale values. Therefore, the brighter the pixel in the grayscale projection, the more likely it is to be the target pixel.

The part marked with dotted lines in figure 2 is the core of CamShift algorithm. Let $(x, y)$ be the pixel position in the search window, and $I(x, y)$ be the pixel value at $(x, y)$ in the projection. The
zero-order moment $M_{00}$ and first-order moment $M_{01}$ and $M_{10}$ of the search window are defined as follows:

$$M_{00} = \sum_x \sum_y I(x, y)$$  \hspace{1cm} (2)  

$$M_{01} = \sum_x \sum_y yI(x, y)$$  \hspace{1cm} (3)  

$$M_{10} = \sum_x \sum_y xI(x, y)$$  \hspace{1cm} (4)  

The centroid position in the search window can be obtained as:

$$(x_c, y_c) = \left( \frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}} \right)$$  \hspace{1cm} (5)  

Then resize the search window according to $M_{00}$ and move the center of the search window to the center of mass. If the moving distance is greater than the preset fixed threshold, the adjusted window center of mass will be recalculated for a new round of window position and size adjustment. Until the moving distance between the center of the window and the center of mass is less than the preset fixed threshold, or the number of cyclic operations reaches a certain maximum, the convergence condition is considered to be satisfied, and the next frame of image is entered for a new target search [3]. In the new image, the position and size of the new search window are set by using the final centroid position and zero order moment $M_{00}$ of the last frame image [4]. The CamShift tracking algorithm adaptively adjusts the size of the search window within and between the image frames according to the $M_{00}$ obtained in the previous step, so it can adapt to the dynamic deformation of the tracking target [5-6].

3. Algorithm Improvement

Before algorithm improvement, the reason of target missing is analyzed first. Because the tracking algorithm adopts a single color feature, it is easy to lose the target when the interference target with similar color appears in the video frame. In the experiment, the target color model and the video sequence histogram were compared to obtain the numerical requirements for effective tracking. In figure 3, a shows the histogram of the target color model; in figure 3, b shows the histogram of the selected video sequence to be tracked; in figure 3, the tonal group distance is set as 68. Assuming that the histogram of the target color model has a tonal value of $m$ and an area of $a$, the area of the video sequence histogram is $b$ within the interval of $(m-5, m+5)$. According to the tracking effect, the video frame satisfying formula (6) can achieve effective tracking.

$$\frac{(b-a)}{b} \leq 20\%$$  \hspace{1cm} (6)  

In Figure 4, the projection diagram of the aircraft and the background can be visually observed.
Due to the color similarity, the aircraft and the background are blurred and inseparable, and the target is lost in the tracking process.

The cause of target loss is clearly visible. Therefore, in the improved algorithm, the three-frame difference method is used to obtain the contour map of the tracking target, calculate the coordinate points of the contour map, and correct the tracking frame. Three frame difference method, first of all, the sequence of gray level, then the current frame and the previous frame difference algorithm, the next frame and current frame difference algorithm, the result of the two difference by selecting appropriate threshold binarization processing, and finally to logic "and" operation [7], corrosion expansion, the output, outline of moving targets is obtained.

Since there are a lot of holes and noises in the contour after the three-frame difference method, multiple contour coordinate points can be obtained. In the correction process, first of all, all the coordinate points are traversed and the horizontal and vertical coordinates are compared to obtain the coordinate values of the upper left corner w1(x,y) and the lower right corner w2(x,y). Then, the distance between w1(x,y) and the rect coordinate points obtained by Camshift algorithm is calculated. If the distance is greater than the set threshold value n, as shown in formula (7), it is considered to be at risk of losing the target. The tracking box is modified as a rectangle with diagonal lines of w1(x,y) and w2(x,y). Finally, the template is updated with this new area to track the next frame. If the distance between w1(x,y) and rect is less than the set threshold value n, the target tracking under the traditional camshift algorithm will be continued.

\[
\sqrt{(rect(x) - x)^2 + (rect(y) - y)^2} > n
\]

(7)

4. Experimental Results and Analysis
In the experimental verification, video sequence of airshow aircraft in zhuhai was taken as an example and the traditional camshift algorithm was used to track the target. The results are shown in figure 5. When the background color is similar to that of the plane, the tracking frame will deviate from the target. It can also be seen from that figure that the size of the trace box also deviate from the size of the target. The size of the search window is adjusted according to M00, that is, the grayscale image pixel value converted from similar colors of the target tone value after reverse projection. It can be seen that similar color also affects that size of the trace box.
Figure 5. Effect of traditional camshift tracking algorithm for random four frames

In the improved algorithm, the contour map is obtained according to the three-frame difference method, then the coordinate points are calculated, and finally the target is tracked. The function in OpenCV extracts the contour and calculates the display coordinate points. The contour coordinate points are shown in Figure 6, and the tracking effect after the improved algorithm is shown in Figure 7.

Figure 6. Random four frame contour coordinate points

(a) Random frame 1 coordinate point  (b) Random frame 2 coordinate point
(c) Random frame 3 coordinate point  (d) Random frame 4 coordinate point

Figure 7. Effect diagram of random four-frame improved camshift algorithm tracking

It can be seen from Figure 6 that although the three-frame difference method has carried out two
operations on the three-frame images, corrosion and expansion operations, there are still cavitation and noise phenomena, and the number of coordinate points obtained is multiple. In order not to affect the tracking effect, the improved algorithm traverses the contour coordinate points, calculates the coordinates of the upper left corner and the lower right corner of the overall contour, and obtains the rectangular box that matches the target. It can be seen from figure 7 that the improved algorithm basically achieves accurate tracking, and the position and size of the tracking frame are more suitable for target tracking requirements.

In the figure, most of the tracking frame lines are pressed to the tracking target. This is because most of the contour coordinate points are located inside the target, which causes the tracking frame corrected by the coordinate points to closely fit the target. In figure 7, the tracking frame of picture c is smaller than the target, which is caused by the adaptive window size of traditional Camshift algorithm. The position of the rect coordinate point and w1(x,y) of the traditional algorithm is less than the set threshold value n, and the algorithm determines that there is no danger of losing the tracking target. Therefore, the traditional tracking algorithm is adopted, and the tracking window is smaller than the size of the tracking target due to the influence of the pixel value of similar color reverse projection diagram. In figure 7, the track left by the plane across the sky is delimited by three frames, and then the outline is displayed, as shown in figure 6. The coordinate point of the contour affects the determination of the tracking frame position. Therefore, under the influence of non-target contour, the tracking effect of the improved algorithm will be deviated.

5. Conclusion
Aiming at the deficiency of traditional CamShift algorithm in tracking effect of color similarity background [8], this paper explores the causes of missing targets and proposes a target tracking algorithm combining contour coordinate points. In the case that the algorithm determines that there is a missing target, it first carries out a three-frame difference operation to obtain the contour coordinate points, and then executes the improved algorithm on this basis to obtain the location, size and other parameters of the target, obtaining more accurate and more stable tracking results than the traditional CamShift algorithm. Of course, this method also has shortcomings, such as large deviation in the tracking effect of video sequence with non-target contour interference. The next step of the work is mainly focused on how to eliminate the interference of non-target contour.

6. Acknowledgment
This work was partially supported by Joint Funding Project of Beijing Municipal Commission of Education and Beijing Natural Science Fund Committee (KZ201710015010), Project of National Scientific Found (No.61370188), Project of Beijing Municipal College Improvement Plan(PXM2017_014223_000063) and New Project of Green Printing and Publishing Technology by Cooperative Creating Center (PXM_014223_000025) and BIGC Project (Ec201803 Ed201802Ea201806)

7. References
[1] Wenjing Zhou,Wei Chen. Target tracking algorithm based on improved inter-frame difference and local CamShift[J]. Software Tribune,2018,(17):67-70
[2] Xiaoxiao Xu,Hao Yang,Zonghai Chen. An improved target tracking algorithm is proposed [J]. System simulation technology and its application , 2014,(11):581-584
[3] Hongzhi Zhang , Jinhuan Zhang , Hui Yue , Shilin Huang. Target tracking algorithm based on Camshift[J]. Computer engineering and design,2006,(27):2012-2014
[4] Shirong Liu,Kai Sun,Bo Tao Zhang,Fan Yang. Moving target tracking of mobile robot based on improved Camshift algorithm[J]. Journal of huazhong university of science and technology , 2011,(39):223-226
[5] Mingzhu Liu,Xiaoxiang Ren,Guangming Zhang. Camshift target tracking method based on Euclidean distance prediction[J]. Papers on Chinese science and technology ,
2017,(12):2289-2294

[6] Chunlin Zhang, Xiaojian Li. Camshift target tracking algorithm based on the combination of inter-frame difference and background difference[J]. Industrial control computer, 2014,(27):118-120

[7] Dapeng Wu, Weiping Cheng, Shenglin Yu. Camshift target tracking algorithm based on inter-frame difference and motion estimation[J]. Photoelectric engineering, 2010,(37):55-60

[8] Lei Yang, Meizhi Liu, Hai Gao, Chanjuan Dang. Research on Camshift target tracking algorithm combining texture features [J]. Electronic design engineering, 2018,(26):136-147