International Journal of Multiphysics
Volume 18, No. 2, 2024
ISSN: 1750-9548

Multi-Feature Adaptive Target Tracking Algorithm Based on Rotational Inertia

Yebing Ding[1,*], Guilin Tang[2]
[1]School of electronic and Information Engineering, Anhui Post and Telecommunication College, China
[2]School of computer and network, Anhui Post and Telecommunication College, China
*Corresponding Author.

Abstract

In motion target tracking algorithms based on video, traditional mean shift target tracking algorithms use color as the target feature, which is easily affected by homochromatic interference, and the adaptive adjustment of kernel function bandwidth is insufficient. A multi feature video object tracking algorithm based on rotational inertia is proposed in this paper, which uses target color and texture features to jointly create the target sample model. The target sample is projected to generate a probability density distribution map. In the density distribution area of the target, the method of rotational inertia is used to calculate length, width, and angle of the target, thereby adaptively adjusting kernel function bandwidth, and locking the size and angle of video target. The experimental results show that the algorithm can adaptively adjust video tracking window according to the size and angle of the target, and can resist interference from similar colors, with good robustness.

Keywords: target tracking, Mean Shift, Kernel function bandwidth, texture, probability density, rotational inertia.

1. Introduction

Computer vision is a crucial research domain within artificial intelligence, with video target tracking serving as its foundational research. It finds applications in areas such as video surveillance, intelligent transportation, and human-computer interaction, etc [1-4]. The Mean Shift algorithm [5], which is simple, easy to implement, and quick, has been successfully integrated into video object tracking algorithms. Typical examples include the Continuously Adaptive Mean Shift (CAM Shift) algorithm [6] and the classic Mean Shift target tracking algorithm [7].

CAM Shift algorithm uses H component in the HSV color space as the target feature, which is susceptible to interference from the same color. Stolkin R proposed an adaptive background color tracking algorithm [8]. Zivkovic Z combined the maximum expected algorithm with the Mean Shift algorithm [9]. Peihua Li used clustering methods to adaptively divide the target color space to establish a target model [10]. The above improved algorithms are similar to traditional Mean Shift tracking algorithms, using color to represent target features.

The target texture features are color and rotation invariant, and the calculation is simple, which can resist the interference of the same color. Local Binary Pattern (LBP) [11-13] is an operator that describes the local texture features of an image. Combining LBP with color will have better results in video object tracking research.

The traditional kernel function Mean Shift target tracking algorithm has a problem in that its similarity metric reaches a local maximum in a small window, and its bandwidth cannot be adjusted adaptively [14]. Yilmaz A proposed an asymmetric kernel function to obtain the size and direction of the target [15], but this method is only applicable to targets with unchanged contours. Peng Ningsong used affine transformation parameters to modify the kernel bandwidth, but this method is not applicable to non-rigid objects [16].

In classical mechanics, the rotational moment is a measure of the inertia of an object when it rotates around an axis. The magnitude of the rotational inertia depends on the distribution of the object's mass and the distance between the mass elements and the rotation axis. For solid objects, the moment of inertia is calculated by dividing the object into multiple particles and integrating the moment of inertia for each particle [17-20].

This article combines target color and texture to form target features, creates a probability density distribution map, represents color using H and S components of HSV color space that are less susceptible to variations in light intensity, and uses patterns related to target boundaries and corners for texture. The target position is found through iterative seeking by Mean Shift algorithm. Length, width and angle of the target are calculated using the method of rotational inertia at its optimal position, thus adaptively adjusting bandwidth of kernel function.

2. Classic Target Tracking Algorithm based on Mean Shift

The classic tracking algorithm based on Mean Shift seeks target's position through an iterative process, wherein it establishes the optimal match between the histogram of candidate targets and that of the target model. Initially, a feature space is defined, wherein the feature values of each point in this space are denoted by u, with u=1,2,…,m. x_i (i=1,2,…,N) represents the location of each point within the target model, with the model's center serving as the origin. The kernel profile function k(x) is isotropic and monotonically decreasing, and the weight of a pixel diminishes as it moves away from the center. Function $b(x_i)$ denotes the feature value at position x_i, and thus the probability calculation for each feature value within the target sample model as follows:

$$q_u = \frac{\sum_{i=1}^{N} k\left(\left\|x_i\right\|^2\right)\delta\left[b(x_i)-u\right]}{\sum_{i=1}^{N} k\left(\left\|x_i\right\|^2\right)}, \tag{1}$$

Among them, the function $\delta[b(x_i)-u]$ is 1 when $b(x_i)=u$, otherwise it is 0.

Let y_i (i=1,2,…,N_h) denote the positions of points in the candidate target centered at y, with the bandwidth of the profile function being h. The probability of each eigenvalue of the candidate target is:

$$p_u(y) = C_h \sum_{i=1}^{N_h} k\left(\left\|\frac{y-x_i}{h}\right\|^2\right)\delta\left[b(x_i)-u\right], \tag{2}$$

Normalization constant C_h is:

$$C_h = \frac{1}{\sum_{i=1}^{N_h} k\left(\left\|\frac{y-x_i}{h}\right\|^2\right)}. \tag{3}$$

The similarity function assesses the level of matching between the target model and the candidate, that is

$$\rho(y) = \sum_{u=1}^{m} \sqrt{q_u p_u(y)}. \tag{4}$$

Target tracking starts from the previous frame video target position, and iteratively searches for the optimal target position.

Assuming that the previous frame video target position is y_0, expanding (4) using Taylor's formula at y_0 yields,

$$\rho(y) \approx \frac{1}{2}\sum_{u=1}^{m}\sqrt{q_u p_u(y_0)} + \frac{1}{2}\sum_{u=1}^{m} p_u(y)\sqrt{\frac{q_u}{p_u(y_0)}}, \tag{5}$$

Substitute the expression (2) into the above expression, and simplify it to

$$\rho(y) \approx \frac{1}{2}\sum_{u=1}^{m}\sqrt{q_u p_u(y_0)} + \frac{C_h}{2}\sum_{u=1}^{m}\omega_i k\left(\left\|\frac{y-x_i}{h}\right\|^2\right), \tag{6}$$

Wherein the weight ω_i is

$$\omega_i = \sum_{u=1}^{m}\sqrt{\frac{q_u}{p_u(y_0)}}\delta[b(x_i-u)]. \tag{7}$$

From equation (6), maximizing the similarity function is equivalent to maximizing the probability density function estimated by the kernel profile function $k(x)$ centered at y. Taking its first derivative, we have

$$\begin{aligned}
&\left[\frac{C_h}{2}\sum_{u=1}^{m}\omega_i k\left(\left\|\frac{y-x_i}{h}\right\|^2\right)\right]' \\
&= C_h\sum_{u=1}^{m}(y-x_i)\omega_i\left[k\left(\left\|\frac{y-x_i}{h}\right\|^2\right)\right]' \\
&= C_h\sum_{u=1}^{m}(x_i-y)\omega_i g\left(\left\|\frac{y-x_i}{h}\right\|^2\right) \\
&= C_h\sum_{u=1}^{m}g\left(\left\|\frac{y-x_i}{h}\right\|^2\right)\left[\frac{\sum_{u=1}^{m}x_i\omega_i g\left(\left\|\frac{y-x_i}{h}\right\|^2\right)}{\sum_{u=1}^{m}g\left(\left\|\frac{y-x_i}{h}\right\|^2\right)}-x\right]
\end{aligned}, \tag{8}$$

In the above formula, $g(x)=-k'(x)$, then along direction of projected probability density gradient, target location center point y_1 is obtained as

$$y_i = \frac{\sum_{i=1}^{N_h}x_i\omega_i g\left(\left\|\frac{y_0-x_i}{h}\right\|\right)}{\sum_{i=1}^{N_h}\omega_i g\left(\left\|\frac{y_0-x_i}{h}\right\|^2\right)}. \tag{9}$$

By continuously iterating along the direction of the projected probability density gradient, the optimal target position is obtained once the extreme condition is met.

3. Texture Model

LBP is an operator that describes local texture features of an image by comparing the pixel values of the center and its neighboring areas. The specific calculation formula is

$$LBP_{N,R}(x_c,y_c) = \sum_{p=0}^{N-1}s(g_p-g_c)2^p, s(x) = \begin{cases}1, x\ge 0\\ 0, x<0\end{cases}, \tag{10}$$

In the above formula, g_c represents the pixel value at the center position (x_c, y_c), N denotes the number of pixels within a radius R, and g_p is the pixel value of the neighboring position p.

Texture model with grayscale and rotation invariance is

$$LBP^{r}_{N,R}(x_c, y_c) = \begin{cases} \sum_{p=0}^{N-1} s(g_p - g_c - a), U \le 2 \\ p+1, U > 2 \end{cases}, \tag{11}$$

Where, *a* is the threshold of gray fluctuation, and *U* is

$$\begin{aligned} U = &\left| s(g_{p-1} - g_c - a) - s(g_0 - g_c - a) \right| + \\ &\sum_{p=1}^{N-1} \left| s(g_{p-1} - g_c - a) - s(g_0 - g_c - a) \right| \end{aligned}. \tag{12}$$

The $LBP^{r}_{8,1}$ texture operator has nine standard mode results. Mode 0 and mode 1 are bright spots, indicating noise. Mode 7 and mode 8 indicate dark spots or smooth areas. mode 2 and mode 6 indicate line ends. mode 3 and mode 5 indicate corners. mode 4 indicates boundaries. Mode 2-mode 6 are related to boundaries and corners, and the texture features are more obvious. Therefore, the texture model with five modes is obtained as follows

$$LBP^{r}_{8,1}(x_c, y_c) = \begin{cases} \sum_{p=0}^{7} s(g_p - g_c - a), \{2,3,4,5,6\} \\ 0, other \end{cases}. \tag{13}$$

4. Rotational Inertia

The rotational inertia is a measure of the inertia of an object when it rotates around an axis in classical mechanics. For a particle, its moment of inertia as follows

$$I = mr^2, \tag{14}$$

Where, m represents quality of the particle, and r denotes perpendicular distance between the particle and the axis of rotation.

The planar graph can be regarded as a homogeneous object with thickness *t*.Its rotational inertia and product of inertia about the x and y axes are respectively

$$\begin{cases} I_x = t\int_A y^2 dA \\ I_y = t\int_A x^2 dA \\ I_{xy} = t\int_A xy dA \end{cases}. \tag{15}$$

The principal axis passing through the centroid of an image is called the centroid principal axis, and its inertia product is zero. The rotational inertia corresponding to the centroidal axis has a maximum and a minimum value, respectively, as

$$\begin{cases} I_{\max} = \dfrac{I_x + I_y}{2} + \sqrt{\left(\dfrac{I_x - I_y}{2}\right)^2 + I_{xy}{}^2} \\ I_{\min} = \dfrac{I_x + I_y}{2} - \sqrt{\left(\dfrac{I_x - I_y}{2}\right)^2 + I_{xy}{}^2} \end{cases}. \tag{16}$$

The angle between the centroid principal axis and the x-axis is φ, and its relationship with the x-axis and y-axis rotational inertia is

$$tg2\varphi = -\frac{2I_{xy}}{I_x - I_y}. \tag{17}$$

The relationship between the moment of inertia of the centroid principal axis of an elliptical planar object with mass M and the fitted ellipse's major and minor axes, l and w, respectively, is

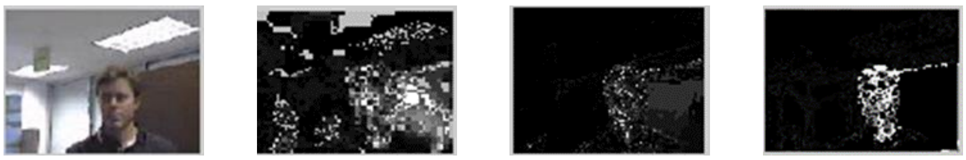
$$\begin{cases} I_{\max} = \dfrac{Ml^2}{4} \\ I_{\min} = \dfrac{Mw^2}{4} \end{cases}. \tag{18}$$

5. The Improved Tracks Algorithms

The target model uses the H and S components (in the range of 0-255) in the HSV color space, and the texture information is constructed using 5 patterns related to the target boundary and corners.

In cases of interference from identical background colors, the target model demonstrates enhanced discrimination.

As depicted in Figure 1, various algorithms are employed to illustrate the probability density distributions of face targets. The figure reveals that the CAM Shift algorithm, which relies on hue, results in a face being obscured within the same background color, while the traditional Mean Shift target tracking algorithm can identify the target, it is susceptible to the influence of identical background colors, the algorithm integrated with texture and color in this study provides a clearer depiction of the target.



(a)Original image (b) CAM Shift (c) Traditional Mean Shift (d)The improved algorithm

Figure 1 Face target probability density distribution diagram

In the probability density distribution diagram projected using the target model, the centroid coordinates (x_c, y_c) of the target can be calculated using the following formula

$$\begin{cases} x_c = \dfrac{\int_A x dm}{M} = \dfrac{\sum xf(x,y)}{\sum f(x,y)} \\ y_c = \dfrac{\int_A y dm}{M} = \dfrac{\sum yf(x,y)}{\sum f(x,y)} \end{cases}, \tag{19}$$

Where, M represents the cumulative value of all pixel values at the target location, and $f(x, y)$ denotes the pixel value at coordinates position(x, y), which,similar to the particle mass.

For a homogeneous object passing through the centroid, the rotational inertia and product of inertia about the x and y axes are respectively as

$$\begin{cases} I_x = \int (y-y_c)^2 dm = \sum (y-y_c)^2 f(x,y) \\ I_y = \int (x-x_c)^2 dm = \sum (x-x_c)^2 f(x,y) \\ I_{xy} = \int (x-x_c)(y-x_c) dm = \sum (x-x_c)(y-y_c) f(x,y) \end{cases}. \tag{20}$$

The rotation angle of the centroid principal axis is given by equation (17). The target is fitted as an ellipse,and the size of the ellipse's long radius l and short radius w as follows

$$\begin{cases} l = 2\sqrt{\dfrac{I_{\max}}{M}} \\ w = 2\sqrt{\dfrac{I_{\min}}{M}} \end{cases} . \tag{21}$$

The maximum and minimum values of the centroid principal axis rotational inertia are shown in formula (16).

The kernel function bandwidth h is determined by the long and short radii of the target approximation ellipse, and has a size of

$$h = \sqrt{\left(\frac{l}{2}\right)^2 + \left(\frac{w}{2}\right)^2} . \tag{22}$$

The size of the target in the video may change at any time, and the probability density distribution of each image projected by the target model will be adjusted accordingly. Its approximate ellipse will also adjust its size in a timely manner, thereby adaptively adjusting the kernel function bandwidth..

The improved algorithm flow is:

(1) Select the tracked target area by frame, obtain the target size and position information, and set the center position as y_0.

(2) Create a target model that integrates color and texture features.

(3) Calculate the candidate target feature model the previous frame centered on y_0.

(4) Calculate the weight ω_i (i=1,2,…, N_h) and obtain the new center position y_1 of the target from the Mean Shift vector.

(5) Compare the distance between y_1 and y_0, and if the distance difference reaches the threshold, stop iterative, Otherwise, $y_0 \leftarrow y_1$ and return to step (3).

(6) Create a projected probability density distribution map in the current frame image based on the target feature model.

(7) Calculate the length, width, and angle of the target probability density distribution based on the rotational inertia in the optimal target location area, and lock the target with an ellipse.

(8) Adaptively adjust the kernel bandwidth h and search window size based on the ellipse parameters, and return to step (3) until the video target tracking is complete.

6. Experiment Results and Analysis

6.1 Car tracking

In the car tracking experiment, there are cases of similar color interference and occlusion. Figure 2 shows the tracking results of the classic Mean Shift algorithm, which, can be seen from the figure that tracking box is offset, even when passing through obstacles, the target is lost and the direction of the tracking box cannot be adjusted. Figure 3 shows some typical tracking results of the improved algorithm in this article, which can be seen from Figure 3 that in the process of tracking black car, when there is a large amount of similar background color interference and almost occlusion, the improved algorithm can still accurately track the car and adjust the size and angle of the tracking box as the car moves.

Figure 2 Classic Mean Shift algorithm tracking results

Figure 3 The tracking results of improved algorithm

6.2 Cup tracking

In the water cup tracking experiment, the size of the water cup changes with distance, and it frequently changes its posture in the hand.

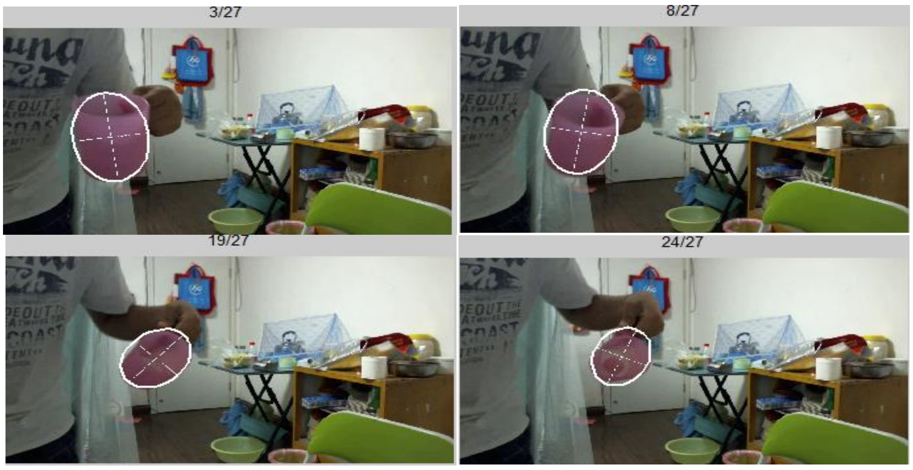


Figure 4 The cup tracking results of improved algorithm

As shown in Figure 4, the improved algorithm in the article can adaptively adjust the tracking window and accurately locate the cup under changes in target size and direction.

The pixel size of the cup video is 720×1280. After moving the origin of the coordinate axis to the target centroid position, rotational inertia relationship and statistical data of various target parameters are presented in Table 1.

It can be seen from the data in Table 1 that rotational inertia of the probability density of the water cup to the X axis in the figure is greater than that to the Y axis, because the water cup is always in a vertical state, although it is slightly skewed.

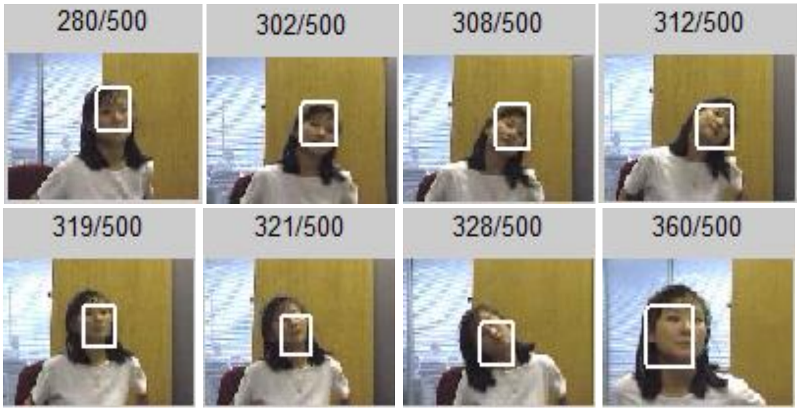
Table 1 Cup tracking parameters analyses

Frame number	Center coordinates	Long and short radius	Angle(°)	Rotational inertia	
				I_x and I_y	I_{xy}
3	(308,381)	156,108	-7.8861	$I_x>I_y$	$I_{xy}>0$
8	(345,368)	148,102	9.3155	$I_x>I_y$	$I_{xy}<0$
19	(430,355)	114,86	43.3941	$I_x>I_y$	$I_{xy}<0$
24	(462,345)	105,79	23.9083	$I_x>I_y$	$I_{xy}<0$

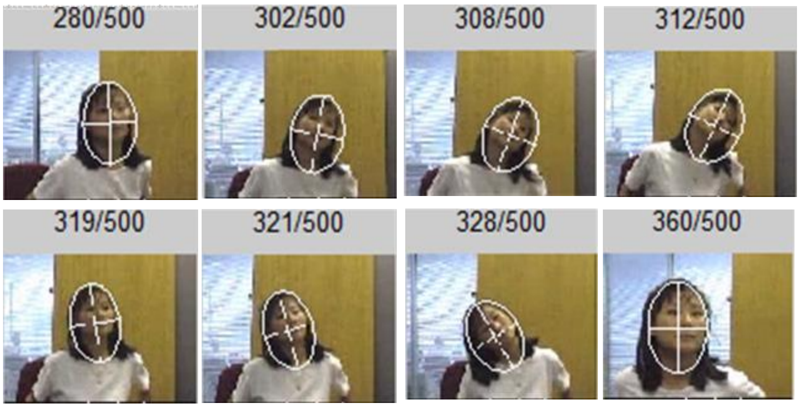
The deflection angle of the cup can be calculated by formula (17). In the case of left deviation, the product of rotational inertia is greater than zero, and the deflection angle is negative. In the case of right deviation, the product of rotational inertia is less than zero, and the deflection angle is positive. Comparing the scale changes caused by the distance of the water cup from the lens, it can also be clearly seen from the values of long and short radii in Table 1.

6.3 Face tracking

The video size of the face tracking experiment is 96×128, with a frame rate of 500. There are influences from changes in lighting, as well as interference from similar background colors, and the face is also rotated left and right and constantly moving.



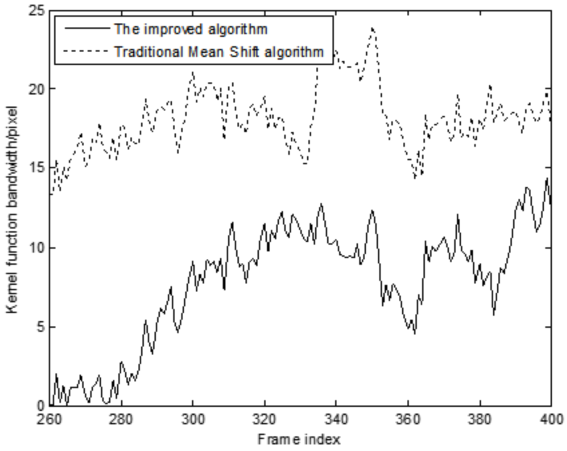
(a) Classic Mean Shift tracking algorithm



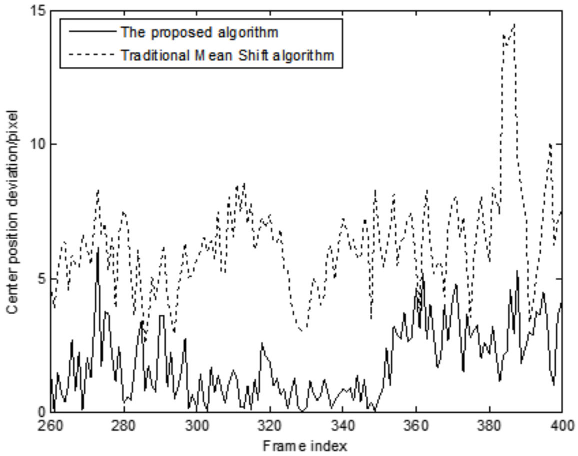
(b)The improved tracking algorithm

Figure 5 The face tracking result

Figure 5 shows the face tracking results of Classic Mean Shift tracking algorithm and the improved tracking algorithm based on rotational inertia. As shown in Figure 5, the improved algorithm is significantly better than the traditional algorithm. When the face moving to left and right, the tracking box can also be deflected accordingly to accurately lock the face. When the distance between the face and the camera is different, the tracking box can adaptively adjust kernel bandwidth according to the size of the face. When there are changes in illumination and background color, the tracking box can still accurately locate the target face.



(a)Kernel bandwidth error curve



(b) Center point position error curve

Figure 6 Face tracking error curve

Figure 6 describe the graph of tracking result deviation data between the traditional Mean Shift tracking algorithm and the improved algorithm, mainly selecting video tracking data from frames 260 to 400. Figure 6(a) is a graph of the kernel bandwidth error comparison, and Figure 6(b) is a graph of the center point location error comparison. As shown in Figure 6, the algorithm error in this article is relatively small.

Table 2 lists the error data of the algorithm in this article and the traditional (classic) Mean Shift tracking algorithm. Using mean and standard deviation statistics, the data results in Table 2 show that the algorithm error in this article is relatively smaller.

Table 2 Face tracking error analyses

Algorithm	Center position（pixel）	Kernel bandwidth（pixel）
The traditional Mean Shift algorithm	3.7125±2.1340	7.6944±3.7900
The improved algorithm	6.2386±1.9861	18.1381±2.0867

From frame 260 to frame 297, the face is basically left-biased. From frame 298 to frame 313, the face is right-biased, reaching a maximum angle of 29.785º. From frame 317 to frame 328, the face is left-biased, reaching a maximum angle of 30.752º, as shown in Figure 7 by the blue curve. The left vertical axis is the angle.Right deviation, positive angle,Left deviation, negative angle.

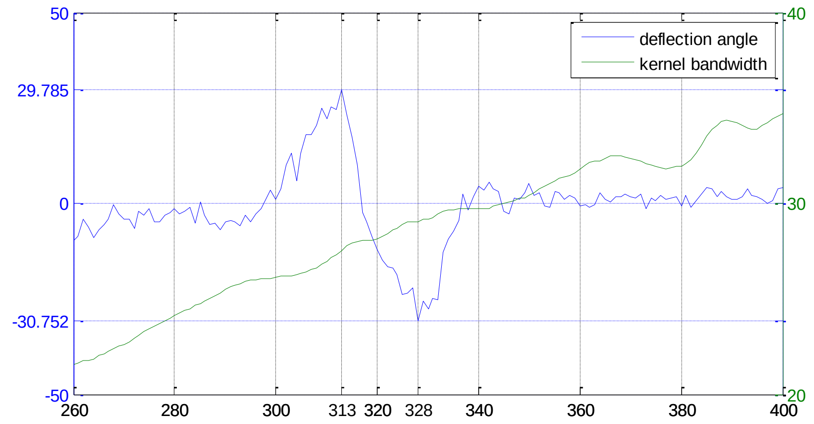


Figure7 Face deflection angle and kernel bandwidth

From frame 260 to frame 400, the distance between the face and the camera from far to near, and the face gradually increases in size, and the long and short axes of the tracking box also increase accordingly. The kernel bandwidth is calculated from the long and short axes and then adaptively adjusted, as shown in the green curve in Figure 7. The right vertical axis shows the pixel value of kernel bandwidth.

7. Conclusions

This paper proposes an improved multi-feature bandwidth adaptive video object tracking algorithm based on rotational inertia. The algorithm integrates target color and texture information to construct a target model, enhancing the ability to resist interference. By calculating the rotational inertia of the target probability density distribution, the length, width, and angle information of the target are obtained, and the kernel function bandwidth is adaptively adjusted. From the results of the above experiments, it can be seen that the algorithm can accurately locate the target and adaptively adjust the tracking window with changes in target size and angle. The algorithm constructs a target model based on two-dimensional video, and is susceptible to background interference with the same color and texture features. Future research can consider introducing three-dimensional features of the target.

Acknowledgments

This paper is supported by Anhui Province Natural Science Research Key Project: Research on Motion Object Tracking Algorithm Based on Mean Shift (2022AH052958).

Refrences

[1] Mavropoulos T, Symeonidis S, Tsanousa A, Giannakeris P, Rousi M, et al, Smart integration of sensors, computer vision and knowledge representation for intelligent monitoring and verbal human-computer interaction. Journal of Intelligent Information Systems, 2021, 57(2): 1-25.

[2] Vishva P, Ayush G, Anupama B, Suryanarayan S, Kumar A D, Object detection and activity recognition in video surveillance using neural networks. International Journal of Web Information Systems, 2023, 19(3-4): 123-138.

[3] Li Y, Chu L, Zhang Y, Guo C, Fu Z, et al, Intelligent transportation video tracking technology based on computer and image processing technology. Journal of Intelligent and Fuzzy Systems, 2019, 37(3): 3347-3356.

[4] Ioannis D, Dimitrios M, Niki P, Vaios L, Vassilios K, et al, Camera-Based Local and Global Target Detection, Tracking, and Localization Techniques for UAVs. Machines, 2023, 11(2): 315-315.

[5] Fukunaga K, Hostetler L D. The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition. IEEE Transactions on Information Theory, 1975, 21(1): 32-40.

[6] Bradski G. R, Computer Vision Face Tracking for Use in a Perceptual User Interface. Intel Technology Journal, 1998, 2(2): 1-15.

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

[8] Stolkin R, Florescu I, Baron M, Harrier C, Kocherov B, Efficient visual servoing with the ABCshift tracking algorithm //IEEE International Conference on Robotics ＆ Automation. USA: IEEE, 2008: 3219－3224.

[9] Zivkovic Z, Kr B, An EM-like algorithm for color-histogram-based object tracking //IEEE Computer Society Conference on Computer Vision ＆ Pattern Recognition. USA: IEEE, 2004.

[10] Peihua Li, An Improved Mean Shift Algorithm for Object Tracking. ACTA AUTOMATICA SINICA, 2007: 33(4): 347-354.

[11] Zitouni A, Nini B, Local Binary Pattern Regrouping for Rotation Invariant Texture Classification. Journal of Information Technology Research (JITR), 2022, 15(1): 1-15.

[12] Devi S S, Texture classification with modified rotation invariant local binary pattern and gradient boosting. International Journal of Knowledge-based and Intelligent Engineering Systems, 2022, 26(2): 125-136.

[13] Pratap R S, Ratnakar D, Kumar R M, LBP and CNN feature fusion for face anti-spoofing. Pattern Analysis and Applications, 2023, 26(2): 773-782.

[14] Junyi Z, Liang Y, Chunhui Z, Quan P, Hongcai Z, Researches on Scale Adaption Strategy in Mean Shift Tracking Algorithm. Journal of Image and Graphics, 2008:13(9):1750-1757.

[15] A. Y, Object tracking by asymmetric kernel mean shift with automatic scale and orientation selection. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007.

[16] Ningsong P, Jie Y, Zhi L, FengChao Z, Automatic Selection of Kernel-Bandwidth for Mean-Shift Object Tracking. Journal of Software, 2005, 16(9): 1542-1550.

[17] Nan L, An approach on characteristic lines extracting of gray images using moment of inertia // IEEE International Conference on Advanced Computer Control. China: Shenyang, 2010: 379-382.

[18] Licini C. J, White O. R, Awad G, Choi J. Y, Enhanced Rolling Moment of Inertia Demonstration. The Physics Teacher, 2024, 62(2): 135-138.

[19] Ahmadreza R, Gilbert M. S, Georg S, Koen L. V, Johan N, et al, Rigid motion tracking using moments of inertia in TOF-PET brain studies. Physics in medicine and biology, 2021, 66(18).

[20] Melia F, Inertia, gravity and the meaning of mass. Physica Scripta, 2024, 99(1).