High Speed Moving Target Tracking Algorithm based on Mean Shift for Video Human Motion

Zhanwei Feng*
Department of Information Engineering, Heilongjiang International University, Harbin 150025, China
*Corresponding author e-mail: fengzhanwei1981@hiu.edu.cn

Abstract. As an important research field of computer vision, there are a lot of achievements of human action recognition in recent Most of the researches are confined to the daily gentle action of human, such as walking jogging, sitting down. In sports video there were many specific characteristics about human actions, such as the upside pose of human, high speed, moving viewpoint. For human motion characteristics of sports video, a highspeed moving target tracking algorithm was proposed. Mean shift was an effective tracking algorithm, which was employed to track moving object in video, but it loses object frequently when the tracked object in the video moves rapidly. We advanced an approach to tracking object moving rapidly, and it can work effectively when the object was in cluttered environment. Experimental shows that the proposed algorithm can improve the tracking effect of mean shift algorithm and effectively track high-speed moving target.

Keywords: Human Action Recognition, Object Tracking, Mean Shift, Human Motion

1. Introduction
Moving object tracking in the video sequence is a process of determining the position of moving objects in sequence images to set the positional relationship between the objects in the current frame image and in the subsequent frame image. Therefore, the moving object tracking in the video sequence can be regarded as a pattern matching process [1, 2]; that is to search corresponding feature pattern in subsequent frame image according to the feature of the current frame image [3]. In the moving process of moving objects, factors such as background, the transformation of external lighting, mutual occlusion of objects in the motion scene, etc. make the feature of objects change along with time, which makes it difficult to achieve accurate tracking of moving objects [4-6]. Features used for object tracking mainly include shape, texture, color, optical flow, corner point, etc. When adopting shape feature tracking, the object shape in the image would change along with the angle. As to the non-rigid objects, due to their own deformation, the shape would change dramatically with time. In addition, with the blocking by other objects or themselves, the shape of objects would also change. The color feature is likely to be affected by external lighting. In the camera motion scene, tracking method adopting optical flow feature can hardly achieve the tracking of objects [7, 8]. Some methods adopt the local features of objects to track objects [9, 10].
2. Analysis of Mean Shift Tracking Algorithm

Assuming that \( n \) independent identically distributed sampling points \( x_1, x_2, \ldots, x_n \) in \( n \)-dimensional space, the probability density function of \( f(x) \) is estimated by the kernel function. It is shown in Equation (1).

\[
\tilde{f}(x) = \frac{1}{h^d} \sum_{i=1}^{n} K \left( \frac{x - x_i}{h} \right) w(x_i)
\]  
(1)

In Equation (1), to distinguish the importance of different gathering points, give each sampling point a weight \( w(x_i) \) greater than zero. \( K(x) \) is a kernel function that represents the influence of sampling points on the density estimation function of different positions. In order for the probability to be equal to 1, make \( \int K(x) \, dx = 1 \).

\( X \) stands for \( d \)-dimensional Euclidean space \( \mathbb{R}^d \), \( x_i \) represents the \( i \)th component of vector \( x \in X \). \( R \) represents the real domain, and if a function \( K : X \rightarrow \mathbb{R} \) has a section function \( K : [0, \infty) \rightarrow \mathbb{R} \). It is shown in Equation (2).

\[
K(x) = k(||x||^2)
\]  
(2)

And meet the following conditions: \( K \) is non-negative; \( K \) is monotonic; \( K \) is piecewise continuous. So, the function \( K(x) \) is called a kernel function. Commonly used kernel functions include homogeneous kernel function and Gauss kernel function. Unit uniform kernel function, It is shown in Equation (3).

\[
F(x) = \begin{cases} 1 & \text{if } ||x|| < 1 \\ 0 & \text{if } ||x|| \geq 1 \end{cases}
\]  
(3)

Unit Gauss kernel function. It is shown in Equation (4).

\[
N(x) = e^{-||x||^2}
\]  
(4)

\[
g(x) = -k'(x), \text{ it is the negative derivative of } k(x).\text{Define the shadow kernel function of } K(x).\text{It is shown in Equation (5).}
\]

\[
G(x) = g(||x||^2)
\]  
(5)

3. High Speed Moving Body Target Tracking Algorithm

According to the mean shift tracking principle analysis, the current state of the system is estimated according to the observed value, which is a Bayesian optimization filtering problem. When estimates status based on Bayesian theory, if the object model is linear and the noise complies with gauss distribution, then Kalman filtering algorithm can be adopted to obtain a set of optimal analytical solutions. As for the non-gauss nonlinear system, particle filter algorithm can be adopted to achieve filtering estimation. Some traditional difficulties such as object detection, occlusion, intersection, track-losing and other problems can be better solved by particle filter algorithm. To effectively track high-speed moving object, particle filter and mean shift algorithm are combined. First, move the object position \( y_0 \) of the last frame, and use object moving model to estimate the current frame location \( y' \), then use mean shift to converge \( y \) to the current frame location \( y' \). Update the particle weights by observing the Bhattacharyya coefficient between \( y' \) and \( y \).

3.1. Particle Filter Tracking Principle

The moving of objects is a variation process of object status (position, speed, etc.) over time, and the tracking of objects can be regarded to estimate object status at different times through observations. To
analyze and deduce the status of the system at different times, two models are needed generally. One is system model which is used to describe the variation relationship of system status over time; the other is measurement system which is used to build the relationship between system status and observations of band noise. The status estimation based on Bayesian inference is to establish the posterior probability density function based on all the known information. The Bayesian posterior probability density estimation includes two stages: prediction and update. In the prediction stage, the system model is used to estimate the probability density function of system status. There exist biases in the status distribution of prediction due to the system is subject to various types of interference. Update stage uses the latest observations to modify the predicted model.

Assuming that a target is tracked, the state sequence \( \{ Z_k, k \in N \} \) of the target has the following relations. It is shown in Equation (6).

\[
Z_k = f_k(Z_{k-1}, y_{k-1})
\]  

(6)

So the task of tracking is to recursively estimate state \( Z_k \) from the observed value of \( X_k \). It is shown in Equation (7).

\[
X_k = h_k(Z_i, n_i)
\]  

(7)

The target tracking based on bayesian estimation is the probability distribution function \( p(Z_k | X_{1:k}) \) of the target state at k time from the observation sequence \( X_{1:k} \). Assuming the initial state probability density function \( p(Z_0) \) is known, the probability density distribution function \( p(Z_k | X_{1:k}) \) can be recursively used to predict and update the two processes.

3.2. High Speed Algorithm Description

In the two-dimensional video image frame, the horizontal and vertical coordinates of the center point \( P \) of the moving object are represented as \( P_x \) and \( P_y \), and the horizontal and vertical direction speeds of the moving object are represented as \( V_x \), \( V_y \). Represent the status of the moving object as \( Z = [P_x, P_y, V_x, V_y] \). The maneuver of the object is considered as a random interference, the interference size of which is represented by covariance \( w \) of the process noise. Because the variation of movement velocity of objects between adjacent frames in the videos is not large, first-order constant velocity model is used to describe the motion law of objects. The dynamic model can be expressed by the Equation (8).

\[
\begin{bmatrix}
1 & 0 & dt & 0 \\
0 & 1 & 0 & dt \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]  

(8)

After the prediction of system status by the motion model, status estimation value \( Z'_t \) at the time \( t \) can be obtained, then use the observation at that time to modify. Observations can be objects' location, speed, and direction angle, etc.. The process of modifying system status by observations is essentially a process of similarity metric to measure the similarity degree between the possible status and current status of objects. Since each particle represents a possible prediction of the object status, the aim of the system observation is to make particles that are close to the practical situation obtain larger weights, while particles that differ greatly from the practical situation obtain smaller weights. According to this principle, we define the probability density function of observation (likelihood function) as Equation (9).

\[
p(X_t | Z'_t) = \frac{1}{\sqrt{2\pi \sigma}} e^{-\frac{D^2}{2\sigma^2}}
\]  

(9)
The larger the value in Equation (9), the smaller the distance between \( y \) and \( y' \), which means the more reliable of the requirements for adopting mean shift and the bigger probability of the candidate object to be the real object. By selecting the transition probability density function \( p(X_t | Z'_i) \) as the signification density function, the weight of particle is updated to Equation (10).

\[
w'_i = w'_i_{-1} p(X_t | Z'_i) \tag{10}
\]

After describing the particle filter algorithm's tracking motion model and observation model, we present the algorithm of high-speed moving target tracking based on mean shift and particle filter. The algorithm is as follows:

1. According to the selected tracking target position and speed in the initial frame, \( Z_0(i = 1, \ldots, N) \) particles are initialized based on the assumption of united gauss. Particle weight \( w'_i = \frac{1}{N} \).

2. For \( i-1: N \), according to the motion model of Equation (8), we get the sample \( Z'_i \) from the importance function \( p(Z'_i | Z'_{i-1}) \). Taking the location \((x'_i^0, y'_i^0)\) of the target state of particle \( Z'_i \) as the starting point, it converges to \((x'_i, y'_i)\) through the mean shift algorithm. The Euclidean distance of \((x'_i^0, y'_i^0)\) and \((x'_i, y'_i)\) are calculated, and the substitution Equation (9) is used to get \( p(z_i | x'_i) \). The importance weight \( w'_i \) is estimated by Equation (10).

3. Calculate the total weight \( t = \sum_{i=1}^{N} w'_i \).

4. For each particle, the weight is normalized:

\[
w''_i = \frac{w'_i}{t}
\]

5. Calculate the effective sampling scale \( N_{\text{eff}} \). If \( N_{\text{eff}} < N_{\text{bw}} \), simple random resampling is applied to solve the problem of particle degradation. After this step, the obtained particles represent the posterior distribution of the target state at \( t \) time.

6. The target state value is estimated according to the particle \( w''_i \).

4. Experimental Results and Analysis

On the basis of proposing the tracking algorithm for high-speed moving object based on mean shift, VC and OpenCV are combined to achieve the above algorithm. Several sets of videos containing high-speed moving objects are validated using this algorithm and are compared with the camshift algorithm provided in OpenCV. In the experiments, tracking objects are determined manually, different amounts of particle tracking are adopted for each object and the results are compared. To verify the tracking effect of algorithm on high-speed moving objects, high-speed moving footballs are selected as the object to be tracked in the videos. Figure 1 is a result of tracking footballs in the football match videos. The video frame size is 480x360 pixels. The along direction is three consecutive frames in the video. Figure 1 (a) is the tracking effect of camshift algorithm provided by OpenCV algorithm library. Figure 1 (b) is the tracking result by this paper algorithm library.

See from Figure 1 (a). In the two frames after the first frame, the object center tracked by camshift algorithm deviates greatly from the object to be tracked (football) and has a significant feature: the object center obtained from camshift algorithm deviates to the opposite direction from the object and the size of it is to tend to grow bigger. Through the principle analysis of Mean shift algorithm, it is easy to understand the deviation of object center of mean shift algorithm from the practical center and why it moves in the opposite direction to the object. In tracking sequence, the size of the candidate objects obtained by camshift algorithm is larger than the practical one because camshift algorithm adopts the automatic adjustment of the search window size to adapt to changes in the object size. As what have mentioned above, due to the limitations of the mean shift algorithm, assuming that the variation of the location of objects between the adjacent frames is small, in the camshift algorithm, a feature space model closer to the template can be obtained through adjusting the candidate object size.

In the three
frames image in Figure1, by generally increasing the size of candidate objects, the area concluded by the ellipse has a similar color histogram.

(a)Camshift tracking results (b) the tracking results of our method.

Figure 1. Football match video tracking effect comparison

Figure1 (b) is the tracking result based on the tracking algorithm of this paper. Under the circumstances of footballs being in high-speed moving, objects can still be accurately tracked, which means the algorithm described in this paper has better tracking effect when tracking high-speed moving objects. Because this algorithm combines the particle filter with mean shift, which enables the algorithm has the advantage of particle filter tracking. Particle filter tracking can achieve great tracking effect when objects are partially occluded, interfered and in complex scenarios, etc. To verify the tracking effect of the algorithm in the scenarios with interference, human body in the diving video is chose to be the tracking object.

(a) Camshift tracking results (b) the tracking results of our method

Figure 2. Comparison of video tracking effect in diving competition

Figure2 is the result of tracking the players in the video diving competition. We chose the athlete's torso as the tracking target. In Figure 2(a), the continuous image of third frame is followed by the Camshift algorithm. In Figure 2(b), the continuous image of third frame is the tracking result of this algorithm. Figure 2(a) the target of the third frame has completely deviated from the actual target. In addition to the reasons for the high speed of the target movement, it was also disturbed by the audience as a background. The above conclusion is easy to get from third frame to the target center and target size. In Figure 2(b) three consecutive frames, the target center and size of the algorithm are quite consistent with the actual target center and size. In the case of background interference, our tracking algorithm can also track the target of high-speed movement well.

5. Conclusion

After analyzing the principle of mean shift algorithm, we pointed out the mean shift algorithm is unable to realize the cause of the high speed moving target tracking, and for this reason, this paper proposes a new tracking algorithm, this algorithm first application of particle filter algorithm to estimate the location of the target in the current frame, and then through the mean shift algorithm to update the location of the object. Because of the mean shift algorithm can quickly converge, the mean shift optimization method can be used to extract the typical "particles" which reflect the probability characteristics of the system more quickly. With a small amount of particles, the tracking precision
can be satisfied, and the particle filter performance can be improved to meet the need of real-time tracking.

References

[1] G. H. Chen, L. Yang, “Moving Object Tracking Algorithm based on Camshift Kalman”. *Journal of Changchun University of Technology*, vol.38, no.3, pp.256-259, 2017

[2] J. J. Liu, Y. Yang, “Video based Moving Object Tracking Algorithm”, *Journal of Jilin University (Information Science)*, vol.35, no.3, pp. 347-353, 2017

[3] F. F. Wu, “Moving Target Detection and Tracking Algorithm based on Rhododendron and Gauss Mixture model” *Journal of Nanjing University of Science and Technology*, vol.41, no.1, pp. 1-46, 2017

[4] X. C. Gu, J. J. Liu, “Moving Target Tracking Algorithm based on Kalman Filter”, *Journal of Changchun University of Science and Technology (NATURAL SCIENCE EDITION)*, vol.38, no.5, pp.136-139, 2016

[5] X. Y. Liu, Y. Meng, “Research on Algorithm of Moving Object Detection and Tracking”, *Journal of Hebei United University (NATURAL SCIENCE EDITION)*, vol.37, no.1, pp. 65-70, 2015

[6] H. X. Xu, “Moving Target Tracking Algorithm based on Improved Mean Shift”, *Journal of ordnance and equipment engineering*, vol.37, no.2, pp. 127-130, 2016

[7] C. Liu, J. Hui, “Moving Object Tracking Algorithm based on Improved CAMShift”, *Computer Engineering and Application*, vol.50, no.11, pp.149-153, 2014

[8] Z. H. Ma, S. H. Gu, “CamShift Moving Target Tracking Algorithm based on SIFT Feature Matching”, *Computer Science*, vol.41, no.6, pp. 291-294, 2015

[9] G. David, “Object Recognition from Local Scale-invariant Features”, *Proceedings of the International Conference on Computer Vision*. IEEE Press, pp.1150-1157, 1999

[10] J. Zhang, X. B. Mao, “Summary of Research on Moving Object Tracking Algorithm”, *Computer Application Research*, vol.26, no.12, pp. 4407-4410, 2009