Moving Target Tracking based on Adaptive Background Subtraction and Improved Camshift Algorithm

Gang XU, Dong ZHAO, Qi ZHOU, Ding HUANG
North China Electric Power University
zhaodongff@163.com

Abstract

Object tracking is one of the most important techniques in digital image processing. Considering the weakness and shortage of traditional moving target tracking method, this paper designed and realized a simple and effective object tracking method. At first, proposed a pixel intensity classification and the single Gaussian model based background reconstruction algorithm. Then use the background subtraction method to extract the foreground target as the initialization of the tracking algorithm. Finally, apply the Kalman filter based camshift algorithm to track the moving target, which combined with a simple algorithm to solve the shadow problem, narrowing the search range of the target to achieve fast and real-time tracking, and making the tracking result more accurate.

1. Introduction

At present, moving target detection and tracking in video sequences does not only have a theoretical value, but also contains a huge commercial value in the military security, intelligent video surveillance, intelligent human-computer interaction, and in image storage and retrieval. It has inspired the world of scientific research workers a strong interest, especially in the United States, Britain and other countries large-scale related projects have already been conducted [1]. The commonly used methods of moving target detection are background subtraction, frame difference method and optical flow method. The main research methods of tracking moving targets are wave gate tracking, optical flow, Kalman filter tracking, active contour tracking, template matching tracking and multi-mode tracking method [2]. This paper proposes a background reconstruction algorithm based on pixel gray-scale classification and single Gaussian model to detect the foreground target, and on this basis, using the background subtraction algorithm to conduct the detection of moving target. Finally, use the target detected ahead as the initialization of the tracking method, and apply the Kalman filter based camshift algorithm to track the moving target, which combined with a simple algorithm to solve the shadow problem, narrowing the search range of the target to achieve fast and real-time tracking, and making the tracking result more accurate.

2. Adaptive background updates for detection of moving objects

Camshift (Continuously Adaptive Mean Shift) is a kind of tracking algorithm based on color histogram, which can track the target effectively. But one of the main shortages is that manually initialization is needed [3]. To solve this problem this paper proposed an adaptive background subtraction algorithm to detected and extract the target automatically.

Background model can be divided into two categories, namely model-based methods and methods based on the reconstruction. On the basis of analysis of these two methods, this paper proposed a background reconstruction algorithm based on pixel gray-scale classification and single Gaussian model. For a group of image sequence that contains foreground target, firstly each pixel value in the image sequence is classified by the method of pixel gray-scale classification, then select background related pixel grayscale value to start the model according to the single Gaussian model. This method takes the characteristics of the two types of background modeling method into account, and is able to take advantage of the multi-frame image that contains foreground target to establish an accurate background model.

2.1. Background model initialization

The background reconstruction algorithm in this paper is a kind of background reconstruction algorithm based on pixel gray-scale classification. The basic basis of pixel gray-scale classification method is that in

978-1-4673-0174-9/12/$31.00 ©2012 IEEE

the image sequences the background is always the most often observed. The specific implementation procedure is progressed as follows.

1) Choose K frames freely from the image sequences at first for background reconstruction.

2) Divided each pixel grayscale value range (0~255) of the image into N same intervals.

3) For the selected K frames, calculate per-pixel the gray-scale interval each point belongs to via the formula.

$$q_k(x,y) = \left[f_k(x,y) / (256/N) \right] \tag{1}$$

Where $f_k(x,y)$ is the grayscale value of pixel $m(x,y)$ in the kth frame, $q_k(x,y)$ is the interval the pixel $m(x,y)$ belongs to, symbol [] means round the number.

4) Calculate the appearance frequency of pixel in each gray-scale interval, confirm the highest frequency gray-scale interval $q_m(x,y)$ of every pixel point.

5) Choose the pixel point whose grayscale value belongs to the related highest frequency gray-scale interval for background reconstruction.

$$B_g(x,y) = \begin{cases} 1, & if\left(q_k(x,y) = q_m(x,y)\right) \\ 0, & other \end{cases} \tag{2}$$

$B_g(x,y) = 1$ means that this pixel can be used for background reconstruction.

The algorithm continuously calculates the corresponding gray-scale difference in the adjacent two frames, classifying the pixel grayscale value according to the gray-scale difference, and finally selects the pixel grayscale value with highest frequency as the background pixel grayscale value. This algorithm retains the advantages of such methods, does not need to build models for the background and target in the scene, directly reconstructs the background from images containing foreground moving target, effectively avoids aliasing phenomenon. At the same time, the algorithm can simply adjust one parameter with clear physical meaning to get the satisfactory results in a large range of parameter changes.

When adopting the pixel gray-scale classification method, after the selection process of pixel that background reconstruction used, the grayscale value range of the selected pixel is more concentrated, so modeling can be directly progressed using the single Gaussian model [4]. For any pixel $m(x,y)$ in the background image, the Gaussian distribution function of this pixel can be expressed as

$$P\left(m_{x,y}\right) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{\left[m(x,y)-\mu(x,y)\right]^2}{2\sigma^2(x,y)}} \tag{3}$$

The estimated value of the mean $\mu(x,y)$ and variance $\sigma(x,y)$ can be calculated by

$$\hat{\mu}(x,y) = \sum_{k=0}^{k-1} f_k(x,y) B_k(x,y) \Big/ \sum_{k=0}^{k-1} B_k(x,y) \tag{4}$$

$$\hat{\sigma}(x,y) = \sqrt{\sum_{k=0}^{k-1} [f_k(x,y) B_k(x,y)]^2 \Big/ \sum_{k=0}^{k-1} B_k(x,y) - \mu^2(x,y)} \tag{5}$$

Where $\mu(x,y)$ and $\sigma(x,y)$ were used to represent the estimated value of the parameter, and $f_k(x,y)$ was used to represent the pixel grayscale value of pixel $m(x,y)$ in the kth frame. The background modeling process is to estimate $\mu(x,y)$ and $\sigma(x,y)$ pixel by pixel. It can be seen that the mean image of the model is the background image generated via the pixel gray-scale classification based background reconstruction method, the variance gives the dispersion of the background pixel gray-scale, provides a basis for the set of detection parameters.

2.2. Adaptive background updates and target detection

Background update is the key part of the moving target detection algorithm. This paper adopts an adaptive background update method based on pixel-level detection and frame-level detection. Pixel-level detection can solve the issue that the target moves at slow and fast speed, and the problem of target moving in and out, on the other hand frame-level detection can solve the impact of background change in a large area. The algorithm solved most of the movement ways of target through the two level updates algorithm, and it can better meet the actual demand. The idea of the algorithm is to make use of the fact that foreground pixels change faster than the background pixels. That is to consider the pixel not changing in a long time as the background pixel, and update it into the background model. The algorithm flowchart is shown in Figure 1.

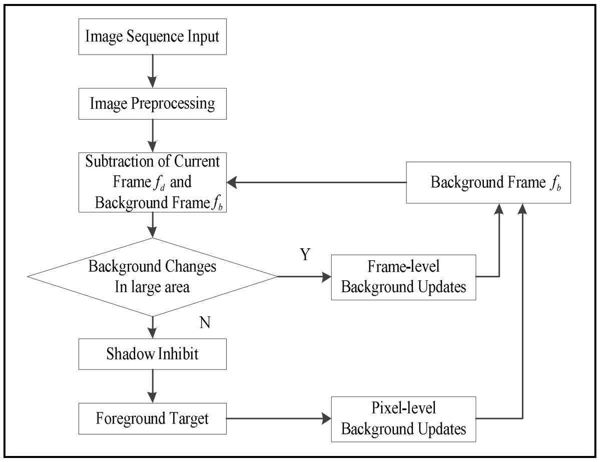

Figure 1. Moving target detection algorithm flowchart

The detailed process of the algorithm is as follows.

1) Firstly, preprocess the input image sequence.

2) Use background subtraction algorithm to detect the moving target and get the binary image. The background subtraction algorithm formula can be expressed as

$$D_k(x,y) = \left| f_k(x,y) - f_b(x,y) \right| \tag{6}$$

$$R_k(x,y) = \begin{cases} 1, & D_k(x,y) > T \\ 0, & D_k(x,y) \le T \end{cases} \tag{7}$$

Where $D_k(x,y)$ is the differential image, $R_k(x,y)$ is the binary image, $f_k(x,y)$ is the current frame image, $f_b(x,y)$ is background image, T is the threshold. When doing background subtraction and segmentation, it's difficult to find a bimodal distribution of D_k, especially under the circumstances that the number of moving target is large. So this paper select a fixed threshold, rather than a dynamic adaptive threshold segmentation, the threshold is mainly obtained through the empirical method, this paper uses T=15 experiments.

3) The frame level background update determines whether the background changes in a large scale, such as clouds cover, lights off or camera jitter and so on. Specific judgments are as follows, adopt the percentage target area accounted of the overall image area to determine whether illumination changes. If the percentage is higher than the predefined threshold M, then the illumination changed, and updates the background model with this frame. And if not, process the next step.

4) Binarize the detected target, then remove the existed isolated noise in the image using the median filter, fill the hole in the target area by morphological dilation operation, Finally, using the processed template image to outline the new target area in the images.

5) The pixel-level background update. To update the background image in real-time after getting the foreground target by background subtraction, Specific method is that, Set a dynamic matrix $D_{i,j}(k)$, which records the changing state of the pixel.

$$D_{i,j}(k) = \begin{cases} D_{i,j}(k-1) - 1 & if \quad R_k(x,y) = 0 and D_{i,j}(k-1) \ne 0 \\ 0 & if \quad R_k(x,y) = 0 and D_{i,j}(k-1) = 0 \\ \lambda & if \quad R_k(x,y) \ne 0 \end{cases} \tag{8}$$

The background update can be used the following formula.

$$B_{i,j}(k) = \begin{cases} \alpha I_{i,j} + (1-\alpha) B_{i,j}(k-1) & if \quad D_{i,j}(k) = 0 \\ B_{i,j}(k-1) & if \quad D_{i,j}(k) \ne 0 \end{cases} \tag{9}$$

where λ records the time length of pixel in moving state, specified according to the movement speed of foreground target, $B_{i,j}(k)$ is the background at time k, Matrix $D_{i,j}(k)$ is the probability that the pixel belongs to background, the smaller the value of D the higher the probability of that it is the background, on the contrary, the lower. When $D_{i,j}(k) = 0$, the background will update this pixels by the above equation. α is the update rate, Generally, to foreground target, its background updates rate can be set to zero. when the value is too small, it will not keep up with the actual background update rate, the value is too large it may update some undetected pixel into the background, appearing empty and delay phenomenon In the background, and possibly lose the foreground target. In this paper α is set to 0.15.The value of α is usually between 0 and 1,if it was set to 0, the algorithm degenerates to be the basic background subtraction method. If it was set to 1, then became the frame difference method. So the value of α in this algorithm can be set to different values according to the different actual situation to meet demand.

2.3. Shadow inhibit process

The defect of the background subtraction method is that the shadow is often estimated as the target in false, this seriously interferes the segmentation and extraction. As shadow has the same motion characteristics with the moving target, the shadow elimination becomes difficult.

In this paper, a simple method (gradient filter) is adopted to reduce the shadow impact to the object detection. After noise filtering the input image is handled with the morphological gradient method to remove the shadow, the specific algorithm is

$$G = (I \oplus B) - (I \Theta B) \quad (10)$$

Where I is the original input image, B is the 3×3 structure operator of morphological operation, G is the gradient image, $\oplus$ is the morphological inflation operator, and Θ is the morphological corrosion operator. After gradient the effect of shadow will be reduced [5].

3. Kalman filter based camshift algorithm

Another shortage of the Camshift algorithm is that if there meets a large area which has the similar color with the tracking target, or the target is obstructed, the track will fail. Considering this problem, we combined Kalman filtering to improve the Camshift algorithm.

3.1. Camshift tracking algorithm

Camshift target tracking algorithm is based on the color characteristics. As RGB color space is more sensitive to changes in illumination brightness, so camshift algorithm transformed the images from the RGB color space into the HSV color space to reduce the impact of changes in light intensity on the tracking performance, and adopted the H component to create the target color histogram.

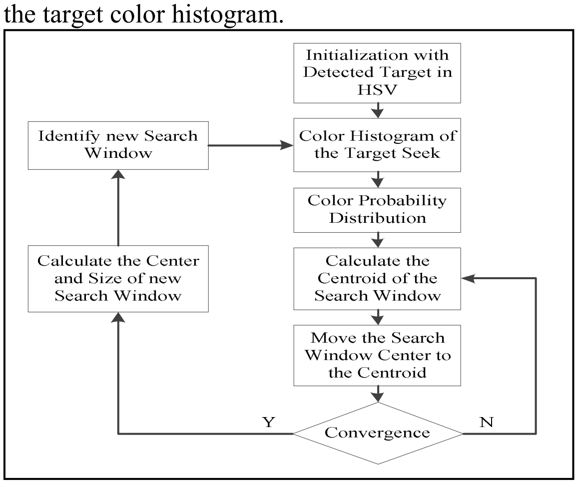

Figure 2. Camshift algorithm

As shown in Figure 2, the mainly procedures of the camshift algorithm are as follows.

1) According to the extracted target from the background subtraction above, set the search window with size of s in the color probability distribution.

2) Calculate the zero-order matrix M_{00} of the search window

$$M_{00} = \sum_x \sum_y I(x, y) \quad (11)$$

The first-order matrix of x and y is

$$M_{10} = \sum_x \sum_y xI(x, y) \quad (12)$$

$$M_{01} = \sum_x \sum_y yI(x, y) \quad (13)$$

Where $I(x, y)$ is the pixel value at the point (x, y), and the change range of (x, y) is inside the search window s.

3) Calculate the centroid of the search window as the centroid of the target,

$$x_c = \frac{M_{10}}{M_{00}}; y_c = \frac{M_{01}}{M_{00}} \quad (14)$$

Reset the search window size s as the color probability distribution function of the above search window.

4) Repeat the 1), 2), 3) steps until convergence (centroid changes less than a given threshold).

5) The major axis l, minor axis w and direction angle θ of the target can be obtained by calculating the second-order matrix.

$$\begin{cases} M_{11} = \sum_x \sum_y xyI(x, y) \\ M_{20} = \sum_x \sum_y x^2 I(x, y) \\ M_{02} = \sum_x \sum_y y^2 I(x, y) \end{cases} \quad (15)$$

$$l = \sqrt{\frac{(a+c) + \sqrt{b^2 + (a-c)^2}}{2}} \quad (16)$$

$$w = \sqrt{\frac{(a+c) - \sqrt{b^2 + (a-c)^2}}{2}} \quad (17)$$

$$\theta = \frac{1}{2} \arctan\left(\frac{2b}{a-c}\right) \quad (18)$$

Where $a = \frac{M_{20}}{M_{00}} - x_c^2, b = 2(\frac{M_{11}}{M_{00}} - x_c y_c), c = \frac{M_{02}}{M_{00}} - y_c^2$.

In the image sequence, the next two frames do not change much, so as to reduce calculated amount, it is not need to calculate the color probability distribution of every pixel in each frame. In a simple environment Camshift algorithm can achieve better tracking results, but as it does not make any prediction of moving objects, the algorithm can't solve problems in a complex conditions, such as target color is similar to background interference in a large area, the target is obscured, the target moves too fast to lead to tracking failure and so on [6]. Therefore, this paper adopts the Kalman filter to predict target position in the next frame.

3.2. Improvement based on kalman filter

To solve the problem that camshift tracking can't work well when the target is similar to the background or is obstructed partly, this paper combined the Kalman filter with the camshift tracking algorithm as a prediction system of the target position.

Kalman filter is a linear recursive filter, it makes the optimal estimation of the next state based on the previous states sequence of the system and the prediction has characteristics as unbiased, stability and optimal [7]. Kalman filter algorithm consists mainly of state and observation equations

$$X_k = A_k X_{k-1} + W_k \tag{19}$$

$$Z_k = H_k X_k + V_k \tag{20}$$

where X_k is the state vector, Z_k is the observed state vector of the system, A_k is the state transition matrix from time t_{k-1} to t_k , H_k is the observation matrix at time t_k , Dynamic noise W_k and the measurement noise V_k is uncorrelated white noise sequences. Assuming that Q_k and R_k is the covariance matrix of dynamic noise W_k and measurement noise V_k . A_k is calculated by

$$A_k = E\left\{W_k W_k^{-1}\right\} \tag{21}$$

To estimate X_k according to Z_k is called Kalman filtering, and estimate X_{k-1} according to Z_k is known as the Kalman prediction or extrapolation. To update the current system state X_k using Kalman filter, to estimate the future state X_{k-1} of the system using Kalman prediction.

As the system have been identified A_k and H_k are known, W_{k-1} and V_k are satisfied with certain assumptions, are also known. Let P_k be the covariance matrix of $\hat{X}_k$, P_k' be the error covariance matrix of X_k and $\hat{X}_k$ [8].The Kalman filter algorithm is calculated as follows, initialize X_0 at time t_0 with the mean value of $\hat{X}$,then solve P_0 . At time t_k , the system state prediction equation is

$$\hat{X}_k = A_k \hat{X}_{k-1} + K_k (Z_k - H_k A_k \hat{X}_{k-1}) \tag{22}$$

Where K_k is the gain coefficient matrix and

$$P_k' = \Phi_{k\cdot k-1} P_{k-1} \Phi^T_{k\cdot k-1} + Q_{k-1} \tag{23}$$

$$K_k = P_k' H_k^T (H_k P' H_k^T + R_k)^{-1} \tag{24}$$

$$P_k = (1 - K_k H_k) P_k' \tag{25}$$

The mainly steps of the improved camshift algorithm are as follows

1) Target location prediction in current frame. Mainly to use Kalman filter to predict the location of the target of the current frame, the Kalman filter prediction is based on reliable predictions based on objective historical movement information.

2) Target matching. According to the principle of Kalman filtering, if the position of the moving target in the next image can be predicted, then the history tracking positions and the positions unlikely to appear in can be filtered out in the next image. Depending on the color probability distribution of the target, use Camshift algorithm search the most similar goals with the target template in the neighborhood of the Kalman filter prediction, the scope of the search can be reduced.

3) Improvement. When meeting a large area of background interference which has the similar color with the target, the target calculated by camshift algorithm expands rapidly, however the normal deformation between the adjacent two frames varies slowly. If the difference between the adjacent two calculated target becomes larger than the threshold, at this time, we drop the calculated target, instead of, we use the area predicted by Kalman filter to reinitialize the camshift algorithm. If the current target suddenly becomes much smaller than the last target out of the threshold, then we can consider that the target is obstructed. At this time we take the prediction value calculated by Kalman filter instead of the camshift algorithm and reinitialize it with the original detected target.

4) Kalman filtering state Update. To update the state of the Kalman filter using the matched target location as the observed values of the Kalman filter, and thus estimate a more accurate predictive value of the next frame target.

Figure 3 shows the ideal above.

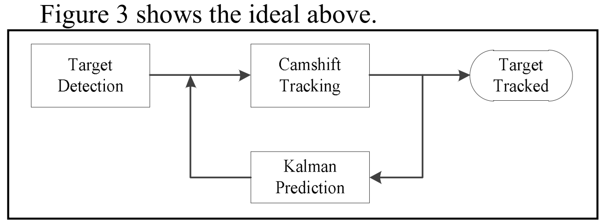

Figure 3. Kalman filter based camshift algorithm

4. Experimental results and analysis

We adopt real video sequence to validate proposed algorithm in camera fixed condition. The size of video image is 640×480. The experiments achieved the tracking of moving target. Using the adaptive background subtraction and Kalman filtering based Camshift algorithm in this paper, the moving target can be detected and tracked effectively. Figure 4 shows target tracking under color complex background. Figure 5 is tracking results when target is obstructed. Row 1 is the original image sequence, row 2 is the tracking results.

Figure 4. Tracking with color interference

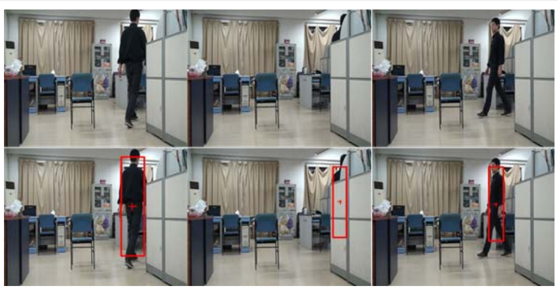

Figure 5. Tracking when obstructed

Seen from the above operation results, when there is large area has the similar color with the moving target, it can still be detected accurately by the algorithm in this paper. And when obstructed the tracking target will not be lost.

5. Conclusions

This paper utilized a kind of adaptive background subtraction and improved camshift algorithm to realize the moving target detection and tracking. Through the experimental results, we know that the algorithm can track the target efficiently and with high robustness. But if the conditions become worse, further improvement is needed.

References

[1] Yinghong LI, Yigui PANG, Zhengxi LI, and Yali LIU, "An Intelligent Tracking Technology Based on Kalman and Mean Shift Algorithm" *International Conference on Computer Modeling and Simulation(ICCMS)*, 2010, Vol.1, 107-109.

[2] D. Comaniciu, V. Ramesh, P. Meer, "Kernel-Based Object Tracking", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2003, Vol.25, No.5.

[3] Olson T, Brill F. "Moving Object Detection and Event Recognition Algorithms for Smart Cameras"[J].*Proc. DARPA Image Understanding Workshop*,1997, Vol.20(5):205-208.

[4] Hou ZQ, Han CZ. "A background reconstruction algorithm based on pixel intensity classification". *Journal of Software*, 2005, Vol.16(9):1568−1576

[5] Eric A. Wan and Rudolph van der Menve. "The Unscented Kalman Filter for Nonlinear Estimation" ,*IEEE*, 2000.

[6] Ridder C, Munkelt O, Kirchner H. "Adaptive background estimation and foreground detection using Kalman-filter". In: *Proc. of the Int'l Conf. on Recent Advances in Mechatronics*, ICRAM'95. UNESCO Chair on Mechatronics, 1995. 193−199.

[7] Magee D. "Tracking multiple vehicle using foreground, background and motion models". *Image and Vision Computing*, 2004, Vol.22(2):143-155.

[8] Elgammal A, Duraiswami R, Harwood D, Davis L. "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance". *Proc. of the IEEE*, 2002, Vol.90(7):1151−1163.