Scientific
Research

Algorithm Research on Moving Object Detection of Surveillance Video Sequence*

Kuihe Yang, Zhiming Cai, Lingling Zhao

Institute of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang050018, China
Email: yangkuihe@126.com, jczmdeveloper@gmail.com

ABSTRACT

In video surveillance, there are many interference factors such as target changes, complex scenes, and target deformation in the moving object tracking. In order to resolve this issue, based on the comparative analysis of several common moving object detection methods, a moving object detection and recognition algorithm combined frame difference with background subtraction is presented in this paper. In the algorithm, we first calculate the average of the values of the gray of the continuous multi-frame image in the dynamic image, and then get background image obtained by the statistical average of the continuous image sequence, that is, the continuous interception of the N-frame images are summed, and find the average. In this case, weight of object information has been increasing, and also restrains the static background. Eventually the motion detection image contains both the target contour and more target information of the target contour point from the background image, so as to achieve separating the moving target from the image. The simulation results show the effectiveness of the proposed algorithm.

Keywords: Video Surveillance; Moving Object; Detection; Frame Difference; Background Subtraction

1. Introduction

Compared with the relative static image, the movement image contains more information, and we can extract the contained information from the movement image by means of image processing. The purpose of moving object detection is to extract the required information from the image. Considering the complexity of the environment during the image acquisition process, the quality of moving object detection depends on the following characteristics: ability to adapt to changes in ambient light, ability to adapt to maintain good results in a variety of weather conditions, ability to avoid interference from detecting similar fluctuations and jitter exists in the scene, ability to accurately identify the erratic movement of large areas, and the change of the object's quantity in the movement scene.

The paper first introduces the principle of frame difference and background subtraction, and then researched the common two-frame difference and three-frame difference [1], lastly put forward a background subtraction based on codebook model and Bayesian classification.

2. Frame Difference

2.1. The Principle of Frame Difference

The frame difference is the most effective method for detecting change of two adjacent frames in the video image [2]. Suppose the video frame at time t is given by $f(x, y, t)$, then the next frame at $t + 1$ is $f(x, y, t+1)$. The binary image operation results of frame difference can be defined as:

$$D(x,y,t+1) = \begin{cases} 1 & \left| f(x,y,t) - f(x,y,t+1) > Th \right| \\ 0 & otherwise \end{cases} \quad (1)$$

where Th represents the threshold for decision. If the frame difference image value is greater than the threshold, then put the point as a foreground pixel. Similarly, when less than the threshold, regarding the point as a background pixel.

2.2. Two Neighbour Frames Difference

Firstly, calculate the difference of two adjacent frame images, then get image result $D_k(x, y)$:

$$D_k(x,y) = \left| f_k(x,y) - f_{k-i}(x,y) \right| \quad , i = 1,2,3\cdots, \quad (2)$$

where $f_k(x,y)$ and $f_{k-i}(x,y)$ respectively represent the two adjacent frame images, $D_k(x,y)$ represents difference im-

age. The so-called adjacent frame difference is the case when set $i = 1$, and can be defined as:

$$R_k(x,y) = \begin{cases} 0 & background & D_k(x,y) \leq T \\ 1 & foreground & D_k(x,y) > T \end{cases}, \quad (3)$$

where T represents the threshold for binarization processing. According to the formular (3), $D_k(x,y)$ is binarized and processed by mathematical morphology filtering. Then according to the connectivity of the calculated result $R_k(x,y)$, if the area of the connected region is greater than the threshold value, the area is judge as the object area to determine the minimum enclosing rectangle [3].

2.3. Three-Frame Difference

This method is an improved method based on two-frame difference. The method extract three consecutive frame images from the video stream, and the difference of three frame images is used to get the contour of moving object. Then this method can be applied to remove effects on background because of object movement, so as to obtain more precise object contour [4-6], and its operation process is expressed as follows:

$$b_{k-1,k} = \begin{cases} 0 & background & |f_k(x,y) - f_{k-1}(x,y)| > T_1 \\ 1 & foreground & |f_k(x,y) - f_{k-1}(x,y)| \leq T_1 \end{cases} (4)$$

$$b_{k,k+1} = \begin{cases} 0 & background & |f_{k+1}(x,y) - f_k(x,y)| > T_2 \\ 1 & foreground & |f_{k+1}(x,y) - f_k(x,y)| \leq T_2 \end{cases} (5)$$

where $f_k(x,y)$ is the k-th frame, T_1 and T_2 are both thresholds, $b_{k-1,k}$ and $b_{k,k+1}$ are the binarization of two adjacent frames difference. Then logic and operation is done to $b_{k-1,k}$ with $b_{k,k+1}$, and get the difference image. The operation process is given as follows:

$$d_k(x,y) = \begin{cases} 0 & background & b_{k-1,k}(x,y) \cap b_{k+1,k}(x,y) = 1 \\ 1 & foreground & b_{k-1,k}(x,y) \cap b_{k+1,k}(x,y) = 0 \end{cases}$$
$$(6)$$

3. Background Subtraction

3.1. The Principle of Background Subtraction

In moving object detection, background subtraction is a frequently-used detection method, which carries out difference calculation by the current image and background image to detect the area of the moving object [7]. The algorithm can be described as:

Step one: Firstly, we need to calculate the absolute gray image of the current frame image and the background image:

$$d_k(x,y) = |I_k(x,y) - B_k(x,y)|, \quad (7)$$

Step two: By setting the threshold Th, we can obtain the binarization image, thereby extracting the moving

object region from the image:

$$b_k(x,y) = \begin{cases} 0 & if \quad d_k(x,y) > Th \\ 1 & otherwise \end{cases}, \quad (8)$$

Step three: using mathematical morphology to filter the difference image, and then analysis communicating region. And if the area of connected region is larger than a given threshold value, that is a moving object.

3.2. Background Subtraction Based on Codebook Model

This object detection based on the codebook model is based on vector quantization and clustering [8-9]. By using the quantization and clustering idea, the changed pixel after image sequence analysis is classified and the code word set in a pixel called codebook.

Building a codebook for each pixel is the key of moving object detection based on the codebook mode. Based on Kim algorithm [10], the paper design the motion detection method based on the codebook model, and the basic steps include codebook creating, codebook filtering, codebook updating, foreground detection and background updating. The color model for evaluating color distortion and brightness distortion is shown as:

In this process, First we need to build a codebook for each pixel. Because the length of the codebook for each pixel is not the same as the others, so set $C = c_1, c_2, ..., c_L$, here c_L represents the codebook which length is L. Codebook consists of code c_i, $i = 1, 2, ... L$. The component of codebook include RGB vector $v = (\bar{R}, \bar{G}, \bar{B})$ and 6-tuple array $aux = \langle \bar{I}, \sigma^2, f, \lambda, p, q \rangle$, here $\bar{R}, \bar{G}, \bar{B}$ represents the average of code's pixel value, $\bar{I}$ and σ^2 respectively represent the average of brightness and variance of brightness for each code' pixel, and last we get the result $I = \sqrt{R^2 + G^2 + B^2}$, here f is frequency of code, λ is maximum time interval, p and q are respectively firt time and last time.

Figure 1 is the model for evaluating color distortion and brightness distortion used in the algorithm. Considering time t, input pixel $x_t = (R,G,B)$ and code c_i are given, then the RGB vector is $v_i = (\bar{R}_i, \bar{G}_i, \bar{B}_i)$. Set θ is the tangle of x between x_t and v_i, δ is the distance of x between x_t and v_i.

$$\cos^{-1}\theta = \frac{<x_t, v_i>}{\|x_t\| \cdot \|v_i\|}, \quad (9)$$

$$\|x_t\| = \sqrt{R^2 + G^2 + B^2}, \quad (10)$$

$$\|v_i\| = \sqrt{\bar{R}_i^2 + \bar{G}_i^2 + \bar{B}_i^2}, \quad (11)$$

$$<x_t, v_i> = \bar{R}_i R + \bar{G}_i G + \bar{B}_i B, \quad (12)$$
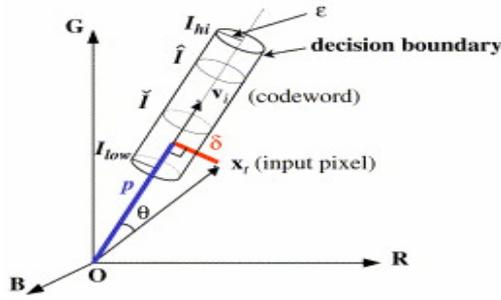
Then color distortion can be defined as:

Figure 1. The color model for evaluating color distortion and brightness distortion.

$$colordist(x_t, v_i) = \delta = \|x_t\| \sqrt{1 - \cos^2 \cdot \theta} \qquad (13)$$

The brightness distortion in the Kim algorithm is redefined as:

$$brightness(I, <\bar{I}, \sigma>) = \begin{cases} true & if\ |I - \bar{I}| \le 2\sigma \\ false & otherwise \end{cases} \quad (14)$$

If codebook contains L codes at time t, then RGB input vector expressed: $x_t = (R_t, G_t, B_t)$. Judge that whether x_t is matched with c_i in accordance with color distortion and brightness similarity.

$$colordist(x_t, c_i) \le \varepsilon_1, \qquad (15)$$

$$brightness(I, <\bar{I}_i, \sigma_i^2>) = true, \qquad (16)$$

where ε_1 means threshold of color distortion.

If the above two formulas hold, then x_t is matched with c_i, and code is updated as:

$$v_m = \left(\frac{f_i\bar{R}_i + R_t}{f_i + 1}, \frac{f_i\bar{G}_i + G_t}{f_i + 1}, \frac{f_i\bar{B}_i + B_t}{f_i + 1} \right) \qquad (17)$$

$$\begin{aligned} aux_i =& < (1-\alpha)\bar{I}_i + \alpha I_t, \sigma_i^2 \\ =& (1-\alpha)\sigma_i^2 + \alpha(I_t - \bar{I}_i)^2, \\ & f_i + 1, \max\{\lambda_i, t - q_i\}, p_i, t >, \end{aligned} \qquad (18)$$

And new code is created as:

$$V_{L+1} = (R_t, G_t, B_t), \qquad (19)$$

$$aux_{L+1} = <I_t, \Sigma^2, 1, t-1, t, t>, \qquad (20)$$

where Σ^2 is a custom larger variance. After creating codebook, let $\lambda = \max\{\lambda, (N-q+p-1)\}$, then to get the maximum time interval which any code is not exist during overall codebook build process.

By setting the threshold value λ, background model after filtering the codes out of background will be obtained as follows:

$$\mu = \{c_i \in C \cap \lambda_i \le T\}, \qquad (21)$$

where T is the threshold, which may be the half of the number of video frames.

After background model μ is gotten, it is detected by foreground to simply compare respective pixel code to the present pixel value and the current frame. If we find matching code word, then the current pixel should belong to the background; otherwise, it belongs to the foreground. Color distortion and luminance distortion as its matching criteria:

$$colordist(x, c_m) \le \varepsilon_2, \qquad (22)$$

$$brightness(I, <I_m, \sigma_m^2>), \qquad (23)$$

where ε_2 is allowed constant threshold value of the pre-set color distortion. If the above two formulas at the same time established, x matches the m-th code word, and code word is updated, and it is determined that the current pixel is a background pixel; otherwise, it is determined that the current pixel as a foreground pixel, at the same time to create a new code word to the codebook.

4. Moving Object Recognition Based on Improved Background Subtraction

How to get a relatively ideal background is one of the key issues of background subtraction points system, and it is need to consider how accurately to update background in response to a series of confounding factors such as the light changes. Under the case that the background image can be changed as the light changed accordingly, and achieve a certain accuracy of detection, under the premise of this topic in view of the strengths and weaknesses of background subtraction and frame difference method, here we combine two methods so that it can play each of the characteristics of the target recognition, and improve the effect of moving object detection.

Therefore, based on the background subtraction and frame difference method, this paper presents an improved background establishment method based on multi-frame images with background subtraction and then average of the results. In the method, we first calculate the average of the values of the gray of the continuous multi-frame image in the dynamic image, and then get background image obtained by the statistical average of the continuous image sequence, that is, the continuous interception of the N-frame images are summed, and find the average. In this case, weight of object information has been increasing, and also restrains the static background. Eventually, the motion detection image contains both the target contour and more target information of the target contour point from the background image, so as to achieve separating the moving target from the image. The specific steps are as follows:

Step one: in the streaming video acquisition, set each image five frames, filter sequence image media, get rid of the image random noise. Thus, we reduce the com-

plexity, overcome noise interference. And these images are, respectively, the marked as: f_0 - f_t.

Step two: to use frame difference to two-two difference with frames around current frame, then average the sum of difference image, set the result as background of current frame, denoted as $f_b(x,y)$.

Step three: if a point gray value in the frame difference image is larger than or equal to the threshold value set in advance, to use the average of image gray value instead of the gray value of the image here; otherwise, we can use the value of last frame instead of corresponding position. The equation is expressed as:

$$f_b(x,y) = \begin{cases} f_k(x,y) & |f_k(x,y) - f_{k-1}(x,y)| < Th \\ mean & |f_k(x,y) - f_{k-1}(x,y)| \geq Th \end{cases}, \quad (24)$$

where *mean* denotes the average of image gray value, *Th* denotes the threshold set in advance, thereby get background image $f_b(x,y)$.

Step four: to extract the background image $B_k(x,y)$ from the video image sequence, so that it only include the stationary background image. The current frames $f_k(x,y)$ and $f_b(x,y)$, are respectively difference with the background image $B_k(x,y)$, then get the results $FD(x,y)$ and $FG(x,y)$:

$$FD(x,y) = \begin{cases} 1 & |f_k(x,y) - B_k(x,y)| > T \\ 0 & |f_k(x,y) - B_k(x,y)| \leq T \end{cases}, \quad (25)$$

$$FG(x,y) = \begin{cases} 1 & |f_b(x,y) - B_k(x,y)| > T \\ 0 & |f_b(x,y) - B_k(x,y)| \leq T \end{cases}, \quad (26)$$

where T is the threshold set in advance.

Step five: to get the moving object area image by calculating intersection of the difference $FD(x,y)$ and $FG(x,y)$, then use Mathematical morphology to process motion region, and exclude the interference of background noise. The gradation of the background in the proposed method is closer to the real video of the background, thereby reducing the interfusion degree of moving object in the background.

In addition, because the video sequence should be changed image, we need IIR filter [11] to update background. The IIR filtering method is relatively common. In this method, the current frame image is represented by $F(x,y)$, and the background image is represented by $B(x,y)$, so the IIR filter updating method is expressed as follows:

$$B_i(x,y) = (1-\alpha)B_{i-1}(x,y) + \alpha F(x,y) \quad (27)$$

And in the preliminary stage of background estimation, we first select an initial background, and then it is updated by the following formula:

$$B_{k+1}(i) = B_k(i) + (\beta_1 \times M_k(i) + \beta_2 \times (1 - M_k(i))) \\ \times (C_k(i) - B_k(i)), \quad (28)$$

where $B_k(i)$ indicates the value of current background image, and is $B_{k+1}(i)$ the updated value. And $C_k(i)$ is the value of current frame, β_1 and β_2 are both coefficients of updating. $M_k(i)$ is expressed as:

$$M_k(i) = \begin{cases} 1 & if(|C_k(i) - B_k(i)| < T_b) \\ 0 & otherwise \end{cases}, \quad (29)$$

Because the presence of background noise, if $|C_k(i) - B_k(i)| < T_b$ holds, the current image is used as background; otherwise, it is used as foreground. Use different update coefficients β_1 and β_2 to update the current image background or foreground. If the current image belongs to background, β_1 is used to update; otherwise, β_2 is used to update. In addition, consider the change of β_1 and β_2 with time change, we get values $\beta_1(t)$ and $\beta_2(t)$; consider different image areas and parameters, $\beta_1(t,x,y)$ and $\beta_2(t,x,y)$ are obtained, where t is the time coordinate, x and y denote the coordinates of pixel in the image.

We combine background model updating with previous moving object detection results, and set larger update rate $\alpha_1 = 0.1$ for the relatively point of background. While the moving object information of foreground uses a smaller update rate $\alpha_2 = 0.01$. After improved, the influence of moving target to background model is obvious reduced, at the same time can be a timely and effective response to the background change.

5. Experimental Results and Analysis

In this paper, experiments were conducted to basic background subtraction in experimental environment of indoor laboratory. Using the fixed camera to real-time capture frames from video, thereby realizing moving object recognition based on background subtraction method. The experimental results are shown in **Figure 2**.

Figure 2 shows the progressive realization of algorithm, and ultimately achieves the effect of object detection based on background subtraction. Picture a) is the background image from video; picture b) is the difference of current frame and background image; picture c) is the effect of moving object detection. The experimental results show that the basic background subtraction algorithm can only obtain the contour of object but cannot effectively obtain the full area of the object. And Background model established in the static region by this method and realistic background have higher similarity. However the background model and the true background produce a deviation in the motion area. Because the mean results are affected by the gradation changes in the motion area, therefore, this method is only applicable to the case with fewer moving objects and longer background. Because long time video sequences is calculated, background updating should be slow down in order to ensure estimation bias of the minimum sequence mean.
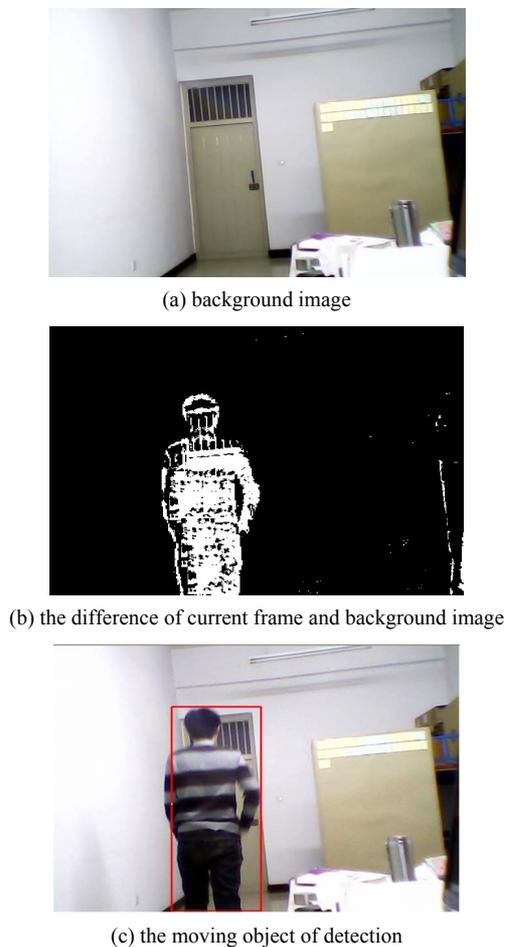
(a) background image

(b) the difference of current frame and background image

(c) the moving object of detection

Figure 2. The recognition process of background subtraction. (figure caption).

6. Conclusions

This paper described the moving object detection algorithms in video surveillance, provided an overview of the three methods for motion detection including frame difference and background subtraction method. Firstly frame difference such as the common two-frame difference and three-frame difference is introduced. Then, background subtraction based on codebook and bayesian classification is researched and analyzed. In this paper, based on background subtraction, we proposed an improved background subtraction method for detecting moving object, and described in detail the establishment

and update of the background model. Finally, experiments were conducted to the frame difference method, background subtraction method and improved method, and the results were analyzed and compared.

REFERENCES

[1] N. He, "Moving Object Detection and Shadow Removal Based on Video Analysis," master's degree thesis of Beijing Forestry University, 2012.

[2] J. J. Du, "Research on Detection and Tracking of Moving Object in Intelligent Video Surveillance System," master's degree thesis of Southwest Jiaotong University, 2009.

[3] Y. Chen, "Research on Detection and Tracking of Moving Object in Intelligent Video Surveillance System", master's degree thesis of Jiangsu University, 2010.

[4] F. Gao, G. J. Jiang, H. X. An and M. S. Qi, "A Fast Moving Object Detection Algorithm", *Journal of Hefei University of Technology(Natural Science)*, Vol. 2, 2012.

[5] X. Jin, "Moving Target Detection, Track and Application on Surveillance System," Master's Degree Thesis of Zhejiang University, 2010.

[6] Q. Wan, "Research on Methods of Multiple Moving Objects Detecting and Tracking in Intelligent Visual Surveillance," Master's Degree Thesis of Hunan University, 2009.

[7] Y. Y. Wu and X. K. Yue, "Image Segmentation for Space Target Based-on Watershed Algorithm," *Computer Simulations*, Vol. 2, 2011, pp. 300-303.

[8] X. J. Ren, S. J. Xiao and X. P. Peng, "Building Extraction from Remote Sensing Image using Improved Watershed Transform," *Computer Applications and Software*, Vol. 28, No. 12, 2011, pp. 250-252.

[9] J. M. Zhang, J. Zhang, and J. Wang, "Watershed Segmentation Algorithm Based on Gradient Modification and Region Merging", *Journal of Computer Applications*, Vol. 31, No. 2, 2011, pp. 369-371. doi:10.3724/SP.J.1087.2011.00369

[10] M. Y. Ren, "Study on Video Moving Object Segmentation Based on Spatio-Temporal Information," Master's Degree Thesis of University of Electronic Science and Technology of China, 2010.

[11] L. Lian, "Moving Target Detection and Tracking based on Frame Difference Method and Image Block Matching Method," *Guide of Sci-Tech Magazine*, 2011. Vol. 27, pp. 56-57.