

# **A Moving Object Extraction Algorithm Based on Hybrid Background Subtraction and Pixel Mean Technique**

Chen Jun-Xiang, Lu Xing-Hua\*, Ye Ming-Ming, Ming Zhong, Zhou Fang, Luo Yu-Lin

*Huali College Guangdong University of Technology, Guangzhou Zengcheng, 511325, P. R. China*

\*Corresponding author E-mail: xhlu@gdut.edu.cn

**Abstract.** In this paper, a moving object extraction algorithm based on hybrid background subtraction and pixel mean is proposed to solve the problem that the existing object extraction algorithm can not adapt to the moving environment. The algorithm operation process includes target location identification, domain pixel mean value processing, image difference operation, background noise processing. Firstly, the processing frame is intercepted from the motion video and the thermal technology is used to recognize position of the moving target and get the preprocessing region. Then morphological recognition technique is used to recognize the target and establish the orthogonal rectangle enclosing the moving target. By means of domain pixel averaging, the background of the area in the orthogonal rectangle is modeled to obtain the background frame. Finally, the differential operation of processing frame and background frame is carried out, and then the target extraction results are obtained after the gray processing and binary processing. The experimental results show that this algorithm can effectively extract moving targets when it works in the motion environment, successfully merge the background subtraction method and the pixel mean value technique, and update the performance of the background subtraction method in the motion environment.

**Key words.** Pixel mean technique; Morphological recognition technology; Background modeling; Background subtraction method

## **1 INTRODUCTION**

With the development of the times, various products in people's life tend to be intelligent. Video surveillance system has been widely used in the financial, transportation, military and other fields. The detection and tracking of moving objects in video sequence has also become a very important research topic in the computer field.

In video sequence detection, the detection algorithm for moving targets include: interframe difference method, optical flow method and background subtraction method. Among them, the optical flow method calculates the vector of the optical flow between the frames by calculating the frame vector in the image, and realizes the detection of the target area. Its computation is so complex that it can not identify the target in real time and is affected by changes in lighting. It is proposed in reference[3] that a multi-frame optical flow model and its linearization method, to solve the real-time optical flow problems by selecting the GPU platform to do algorithm migration. Inter-frame difference method is to find the pixel difference between them and realize the moving target extraction through the differential processing of the image. The algorithm is simple, but when the

relative speed between moving target and environment is low, the extraction result will introduce a lot of environmental noise. In reference [2], a target extraction algorithm based on inter-frame difference and optical flow method is proposed. The algorithm firstly performs discontinuous frame difference on the sequence and uses the non-zero pixel value in the differential chart to calculate the optical flow. The method is faster than the optical flow method and has the advantages of reliable extraction and high accuracy. Background subtraction method can get the moving target through the differential operation of current image and the background image. The algorithm is simple, fast detection, suitable for the applications which need high instantaneity. However, the test result depends on the background modeling of the system. When the system is in motion, its robustness will be greatly reduced. Literature [1] proposed a combination of background subtraction method and inter-frame moving target detection and updating method of adaptive background. This method can automatically detect the change of environment and divide the region of variation, but it still can not adapt to the motion environment.

In this paper, the detection algorithm proposed above is analyzed and demonstrated in detail. It is found that the above algorithm can not extract the real-time feature of the moving target in the motion environment. Based on this, this paper proposes a moving target extraction algorithm based on hybrid background subtraction and pixel mean. Based on the principle of background subtraction, this algorithm uses the "domain pixel averaging technique" to model the background and adapt it to the background motion scene. Finally, the morphological processing algorithm is used to improve the system's extraction effect of moving targets and enhance the recognition performance.

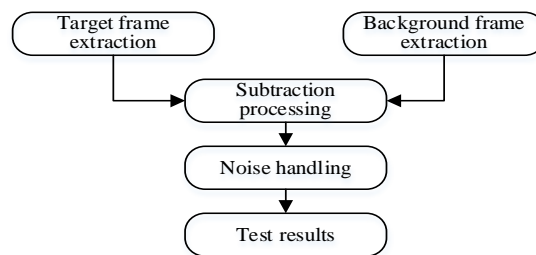
## 2 ALGORITHM ANALYSIS

### 2.1 Background subtraction method

Background subtraction method is to achieve the morphosis extracting of the moving target through the differential operation of current image and the background image, it can be expressed as:

$$D_i(x, y) = |f_i(x, y) - b_i(x, y)|; \quad (1.1)$$

Where  $f_{i(x, y)}$  is the current frame,  $b_{i(x, y)}$  is the background frame, and  $D_{i(x, y)}$  is the target extraction result. Algorithm flow is shown as below:



**FIGURE 1.1.**Flow chart of background subtraction

According to the working principle of background elimination method, the background subtraction method can not be applied to the sports field because the detection system can not accurately model the background for the extraction target in the sports field. Based on this, background modeling has been studied .

### 2.2 Adaptive analysis of background modeling algorithm

Common background modeling methods are the method of mean value, median method, Gaussian mixed model, etc. The method of mean value is that the system extracts a certain period of inter-frame image from the

video information extracted by the camera, and then achieves the purpose of background modeling by calculating the pixel average value of multi-frame images. The formula is as follows:

$$B(x_i, y_j) = \frac{\sum_{j=1,2,3,\dots,n} P_j(x_i, y_j)}{N}; \quad (1.2)$$

(Where N is the total number of the processed frames,  $B_{(xi, yj)}$  is the average pixel value of all the frames, and  $P_{j(xi, yj)}$  is the pixel value of each frame.)

The process of the median method is to arrange the corresponding pixels in N images and finally to take the median of the array as the modeling result. In literature [4], the video image is modeled by the median method. The formula is as follows:

$$B(x_i, y_j) = \text{Median}(P_j(x_i, y_j)); \quad (1.3)$$

Compared with the above algorithms, the Gaussian mixed method can adapt to the scene of small changes, but its adaptability is still poor in non-stationary scenes. The principle is as follows:

When the system is working, each pixel of the background image can be modeled by a Gaussian mixed model consisting of K Gaussian distributions. Which is:

$$P(x_j) = \sum_{i=1}^K \omega_{j,t}^i \cdot \eta(x_j; u_{j,t}^i, \sum_{j,t}^i); \quad (1.4)$$

$X_j$  denotes the value of pixel j at time t,  $X_j$  is the vector if the pixel is an RGB three-channel,  $X_j = [X_{jR} \ X_{jG} \ X_{jB}]$  denotes the estimated value of the weighting coefficient of the ith Gaussian distribution in the Gaussian mixed model at time t, and denotes the mean vector and the covariance matrix of the ith Gaussian distribution in the Gaussian mixed model respectively at time t (here, it is assumed that the red, green and blue components of the pixel are independent of each other);  $\eta$  denotes the Gaussian distribution probability density function.

$$\eta(x_j, u_{j,t}^i, \sum_{j,t}^i) = \frac{1}{(2\pi)^{\frac{d}{2}} |\sum_{j,t}^i|^{\frac{1}{2}}} \cdot \exp[-\frac{1}{2}(x_j - u_{j,t}^i)^T (\sum_{j,t}^i)^{-1} (x_j - u_{j,t}^i)]; \quad (1.5)$$

$$u_{i,t} = (u_{i,t}^r, u_{i,t}^g, u_{i,t}^b); \quad (1.6)$$

$$\sum_{i,t} = \begin{bmatrix} \sigma_r^2, 0, 0 \\ 0, \sigma_g^2, 0 \\ 0, 0, \sigma_b^2 \end{bmatrix}; \quad (1.7)$$

### 2.3 Analysis conclusion of the algorithm

It can be found from the algorithm principle introduced above, the mean value method, the method of mean value, Gaussian mixed model method in the case of camera motion can not be normal for a certain image background modeling is due to the modeling results from the multi-frame image. When there is a large spatial-domain pixel difference between the frames, the background noise and background distortion appear in the background modeling result. If the background modeling process can ignore the change of the inter-frame image, that is, the extraction result is determined by only one image, the modeling algorithm can adapt to the motion environment.

Therefore, based on the background subtraction algorithm, this paper uses the "domain pixel averaging technique" for background image modeling. The algorithm avoids the problem of poor modeling effect due to the difference of the background before and after the background modeling, and enhances the stability of the system for extracting moving targets.

### 3 MOVING TARGET DETECTION BASED ON PIXEL MEAN VALUE TECHNIQUE AND BACKGROUND SUBTRACTION METHOD

The principle of this algorithm is to differentiate the current picture from the background picture, namely:

$$D_{i(x,y)} = |f_{i(x,y)} - b_{i(x,y)}| \quad (2.1)$$

The detection process is to pre-process the extracted object through the average pixel modeling method, then the pixel processing result is differentiated from the original picture, and finally the background noise elimination processing is performed on the difference results. The block diagram is as follows:

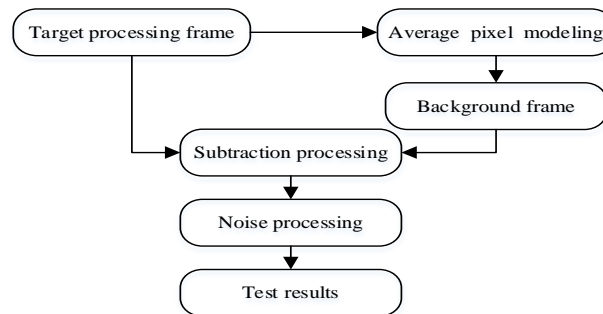


FIGURE 2.1. Flow chart algorithm

#### 3.1 Average pixel modeling method

The principle of the average pixel modeling method is to first extract the target pixel average processing, and then do the difference processing of the average result and the original image. Since the extraction target is always composed of a plurality of colors, when the average value is taken, the part with the higher pixel value can be retained after being differentiated and the pixel value can be reduced to zero. Therefore, the extraction result of the average pixel modeling method will save the pixel with a higher pixel value and feed back the result as a recognition result. The operation process is shown as below:

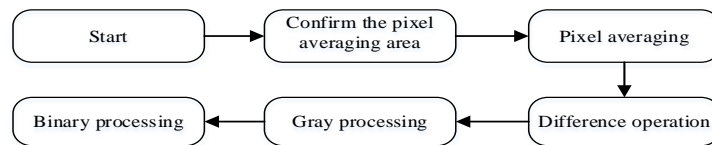


FIGURE 2.2. Average pixel modeling

##### 3.1.1 Determination of the pixel average processing area

The pixel-averaging area consists of an orthogonal rectangle created by the system for the extraction target. The process of establishing the rectangle includes: confirming the location of the extracted target and determining the coordinates of the orthogonal rectangle.

Since this system is targeted at cars and pedestrians, cars and pedestrians emit heat to the air when working. Therefore, the infrared detection technology can be used to confirm the position of the extracted object and

establish the target area. Then, a  $3 \times 3$  two-dimensional matrix structure is created to perform pixel contrast processing on the target area to confirm the coordinates of the orthogonal rectangle. The processing sequence is from left to right and from bottom to top. The processing mode is that the structure first performs pixel extraction on the first scanning target of each row, and then performs pixel comparison processing with the next scanning target. If the average value of the pixel contrast is less than  $X$ , the structure will extract the pixel value of the second scan target and proceed to the next comparison. If the average value of pixel contrast difference is greater than  $X$ , the difference between the pixel value of the scanned object and the average pixel value of the structure body is processed. Finally, the pixel coordinate value with the largest difference of the structure pixel is recorded as the coordinate of the orthogonal rectangle. Scanning is continued in this way until the four coordinate values of the orthogonal matrix are obtained, and the processing result is shown in Figure 2.3. The formula is as follows:

$$\begin{cases} \text{Continue scanning} & |B(x_i, y_j) - B(x_{i+1}, y_{j+1})| < X \\ \text{Record coordinates} & |B(x_i, y_j) - B(x_{i+1}, y_{j+1})| > X \end{cases} \quad (2.2)$$



FIGURE 2.3. Determination of processing area

### 3.1.2 Pixel averaging

The pixel averaging process is to fill a pixel average of the orthogonal rectangular areas. In the set of orthogonal rectangles, which contains the main extraction of the target pixel information. The color of the extracted target is called the key pixel. When performing pixel averaging, there are often a few white pixels and black pixels and other pixels in addition to the key pixels. So the pixel mean will be closer to the key pixel value. The processing results are shown in Figure 2.4, the formula is as follows:

$$E(x, y) = \left\{ \frac{\sum B(x_i, y_i)}{N} \mid i \in (x_n, x_m), j \in (y_n, y_m) \right\} \quad (2.3)$$

(Where  $E(x, y)$  is the pixel average,  $B(x_i, y_i)$  is the pixel value of each pixel in the region, and  $N$  is the total number of pixels)



FIGURE 2.4. Pixel averaging results

### 3.2 Difference processing of images

Through the mean value treatment of the pixel, perform the differential processing of the original picture and the average picture, the formula is as follows:

$$D(x, y) = B(x_{n1}, y_{n1}) - B(x_{n2}, y_{n2}); \quad (2.4)$$

In the operation result, since the image area other than the orthogonal rectangle is not subjected to the pixel processing, the pixel value after the difference operation between the partial image and the original image is zero. The last remaining key pixel is not completely deleted. After which it is processed in grayscale and binarized. Finally, after setting the appropriate threshold, the complete moving target contour can be obtained. The result is shown in Figure 2.5.



**FIGURE 2.5.**Differential Results

### 3.3 Background noise handling

It can be found from Figure 3.4 that the basic information of cars and pedestrians has been extracted, but there is a small amount of background noise in the extraction results. Therefore, we need to de-noising the extraction results.

In the literature [4], background de-noising was done using morphological swelling and erosion operations. Among them, the expansion operation is used to make up for the foreground holes. The computing process is as follows: First, a  $3 \times 3$  two-dimensional matrix structure is constructed, and the structure body performs pixel scanning on the target area and ANDs the two-bit matrix with the pixel points in the image. If at least one of the pixel values in the range operation result is 255, all the pixel values in the range are set to 255; if none of the pixel values in the range operation result is 255, all of the pixel values in the range Set to 0; continue scanning until the end, the formula is as follows:

$$E(x, y) = \begin{cases} 0 & | B(x_i, y_j) \oplus (alltheSare\ 255) \\ 255 & | B(x_i, y_j) \oplus (notalltheSare255) \end{cases}; \quad (2.5)$$

Erosion operation is used to eliminate the background noise. The principle of the operation is opposite to that of the expansion operation. The operation formula is as follows:

$$E(x, y) = \begin{cases} 0 & | B(x_i, y_j) \oplus (alltheSare\ 255) \\ 255 & | B(x_i, y_j) \oplus (notalltheSare255) \end{cases}; \quad (2.6)$$

The results after processing are as below:



**FIGURE 2.6.**Noise elimination results

#### 4 EXPERIMENTAL COMPARISON AND ANALYSIS

This experiment uses MATLAB to carry out the experiment simulation to the algorithm, simultaneously carries out the experiment contrast to the interframe difference method and the background subtraction method.

In this paper, the simulation experiments using Shenzhou K610D-i5 computer with memory 4.0G, frequency 2.6GHz. The experimental video image size of each frame is  $1280 \times 720$ ; sampling rate of 30fps; experimental video is obtained by the field.

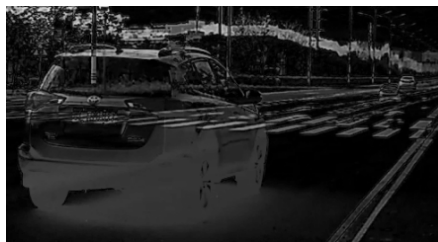
In the simulation experiment of inter-frame difference method, experiments are performed on the video respectively 0.01s, 0.02s inter-frame extraction and gray-scale processing, after the inter-frame difference operation. Finally, the binarization and noise elimination of the differential results. There are a lot of background noise in the extraction results, the distortion occurs when extracting the target, the results are shown below.



**FIGURE 3.1**Interframe difference processing results of 0.01s **FIGURE 3.2**Interframe difference processing results of 0.02s

In the background elimination method, the system needs the background modeling of the moving object first. Commonly used background modeling methods can not complete the background modeling. Literature [5] proposed a background modeling method based on long-time video sequence. Through three steps of background training, retrieval and updating, the algorithm achieves the integration of short-space spatial information background and completes the background estimation and moving target detection results for non-stationary change scenarios. The algorithm increases the stability of background modeling, but it needs long time video sequence training, which can not meet the experimental operation.

Based on this situation, two sports video sequences are re-shot in the same location. One of them is car-free and one has a car. After the car-free series of inter-frame average pixel processing, the differential background is taken as a result. Using difference background and target image to carry out difference operation, the final difference result is shown as follows.



**FIGURE 3.3.**Background subtraction processing results

Finally, under the same conditions, using the moving target extraction algorithm proposed in this paper to extract the target, the following figure shows the processing result:



**FIGURE 3.4.**Background difference processing results

According to the comparison of detection results, the proposed algorithm can restrain most of the background noise in the motion scene, and has good effect on moving objects in the motion scene. Compared with the background elimination method and the inter-frame difference method, the proposed algorithm has more superiority in motion scenarios stability. However, there is still a small amount of background noise in the extraction result, and the distortion of the target is extracted at the same time.

## 5 CONCLUSION

In this paper, by studying the principle and application of background subtraction method, interframe difference method and optical flow method, aiming at the defect that the above algorithm can not adapt to the motion environment, a moving target extraction algorithm based on hybrid background difference and pixel mean is proposed. The algorithm preprocesses the image by using the average pixel modeling method and establishes a new computing model by using the principle of background subtraction. The experimental results show that the proposed algorithm can extract the moving object smoothly in the motion environment and successfully merge the background elimination method with the pixel mean value method, and improve the working characteristics of the background elimination method.

In the experimental results, we can also find that there are still some background noise and target morphological distortion in the moving target extraction results, especially in the complex background and high-speed motion environment. At this point, it needs to consider the camera jitter, shooting delay and so on. It will improve the stability of the algorithm through the study of camera shake, visual delay and other cases.

## ACKNOWLEDGMENTS

This project is supported by and 2017 Undergraduate Scientific and Technological Innovation Project Fund of Guangdong Province (pdjh2017b0940).

## REFERENCES

1. Qu Jingjing, Xin Yunhong. Combined Continuous Frame Difference with background subtraction method for Moving Object Detection [J]. *Acta Photonica Sinica*, 2014, 43(07): 219-226
2. Journal of Computer Engineering and Applications, 2013, 49 (18): 200-203 Comment for this article: Feedback Author Email Title Code Content Copyright by Computer Engineering and Applications 2013,
3. WANG Xiao-wen. Vehicle Detection and Tracking Based on Optical Flow [D]. Jiangsu University of Science and Technology, 2015.
4. Lv Xian-Guo. Multi-perspective video sequences of human features extraction and rapid identification algorithm [D]. Shandong University, 2013.
5. Ding Jie, Xiao Jiangjian, Kuang Liqun, Song Kangkang, Peng Chengbin. Study on background modeling



- based on long time video sequence [J] .Acta Press, 2017-01-03.
6. Ma Guocui,Li Qingwu,Huo Guanying Wang Min,Sheng Huixing. NSCT Domain Modulus Maxima Vehicle Image Edge Detection Combined with Background Differencing[J]. International Journal of Advancements in Computing Technology,2013,5(7):.
  7. Ma Guocui,Li Qingwu,Huo Guanying Wang Min,Sheng Huixing. NSCT Domain Modulus Maxima Vehicle Image Edge Detection Combined with Background Differencing[J]. International Journal of Advancements in Computing Technology,2013,5(7):.
  8. WANG Si-si, REN Shi-qing.An Improved Moving Object Detection Algorithm Based on Hybrid Gaussian Model [J] .Computer Science, 2015,42 (S2): 173-174 + 178.
  9. YUAN Guo-wu, CHEN Zhi-qiang, GONG Jian, XU Dan, LIAO Ren-jian, HE Jun-yuan.A Moving Object Detection Algorithm Combining Optical Flow and Three-Difference Method [J]. Microcomputer Systems, 2013,34 (03): 668-671
  10. LIU Yan-jiang, ZHI Min.A double-layer background modeling method based on codebook and run-time mean [J] .Computer Engineering and Sciences, 2016,38 (06): 1220-1224.