

Pixel Histogram based Background Modeling for Moving Target Detection

Jiang Hua

School of Machinery and Automation
Wuhan University of Science and
Technology
Wuhan, China
huajiang@wust.edu.cn

Liangcai Zeng

School of Machinery and Automation
Wuhan University of Science and
Technology
Wuhan, China
zengliangcai@wust.edu.cn

Gongfa Li

School of Machinery and Automation
Wuhan University of Science and
Technology
Wuhan, China
ligongfa@wust.edu.cn

Hongwei Wang

Zhejiang University
Zhejiang, China
hongweiwang@intl.zju.edu.cn

Zhaojie Ju

Intelligent Systems & Biomedical
Robotics Group
University of Portsmouth
Portsmouth, United Kingdom
zhaojie.ju@port.ac.uk

Abstract— Existing moving target detection methods mainly include inter-frame differences, background differences, optical flow and so on. For the recognition of human motions in the process of human-computer collaboration, existing algorithms are usually difficult to meet the requirements of real-time processing and easily interfered by lighting or image noises. In this paper, a method for establishing a static background model based on pixel histogram is proposed. The effect of moving targets and noises on the background model is excluded due to the selectivity of the new algorithm to the gray values, so it can detect the real background more reliably. Compared with other moving target detection methods, this method has the characteristics of fast speed, strong anti-interference ability, and the ability to identify human body movement quickly and accurately.

Keywords—moving target detection; human motions; static background model; pixel histogram

I. INTRODUCTION

Moving target detection is the process of subtracting the redundant information in time and space of the image through the method of computer vision, and effectively identifying moving targets. It is the basis for applications such as target classification, motion recognition, and behavior understanding, and a key skill in video sequence image processing. At present, the methods of moving target detection mainly include inter-frame difference, background difference, optical flow and so on [1]. Due to the complexity of the optical flow method, it is generally difficult to meet the requirements of real-time processing and susceptible to interference from lighting and image noise. The inter-frame difference method directly compares the difference in gray values of corresponding pixels in two or three consecutive frames of the video sequence, and then sets the threshold to extract the motion area of the image [2]. However, the time interval between adjacent frames is so short that the overlapping part of the target is not easy to detect. Therefore it is generally used for simple real-time motion detection to avoid creating a large amount of holes.

The key of the background difference method is the establishment of the background model. Many studies have been done on this. Gloyer et al. adopts the median value of the pixels of consecutive frames as the background gray

value to establish a background model, but this method requires a large amount of calculation and deviations often occur with changing lighting [3]. In response to this limitation, Wren et al. assume that pixel values meet a certain distribution model, such as Gaussian distribution. Then you can get a better background update method by judging whether the pixel belongs to the foreground based on the threshold, and constantly update its distribution parameters through training [4]. However, the single Gaussian model cannot accurately describe the complex background model, so Stauffer et al. regard each pixel gray value as a weighted mixture of multiple Gaussian models to propose a hybrid Gaussian model background modeling method. This method cannot only adapt well to complex scenes, but also automatically calculate model parameters through adjusting the background model. But the actual complex backgrounds do not always conform to the Gaussian distribution, so the method based on Gaussian model will cause many problems [5]. Elgammal et al. proposed a parameter-free detection method based on kernel density estimation, which does not need to make assumptions about the density distribution of the background, but uses several image sample information to estimate pixels based on a standard kernel function [6]. However, there is no common known parameter model in different environments and the kernel density function is estimated from the data, so the consumption is too large and the real-time performance is not good [7].

At the same time as the development of moving target detection methods based on statistical theory, scholars have also proposed a variety of moving target detection methods with different theoretical foundations, such as cluster theory-based methods, fuzzy theory-based methods, background prediction methods and neural network methods [8-10]. These studies extend the application fields of the skill.

Generally, complex algorithms will naturally have better results, but the calculation time is too long to meet the real-time requirement [11]. Therefore, the commonly used target detection method in human-computer interaction is the mean method [12]. Using the average value of the current image sequence as a reference model, a background image similar to the current static scene is generated. However, if

the moving target differs greatly from the background, more frames of images are required to establish a reliable background model, and this method is more sensitive to noise [13]. In this paper, a method for establishing a static background model based on pixel histogram is proposed, to solve the problem of rapid recognition in the process of human-computer collaboration [14]. The basic idea is to count the gray distribution of each color component of the pixel and establish the corresponding histogram through normalization [15]. Based on the experimental results, the proposed method has a fast speed and strong anti-interference ability, and can identify human movements quickly and accurately.

The remainder of this paper is organized as follows. Section 2 describes the method for establishing background model based on pixel histogram and related technical points. Section 3 indicates the experimental results about the advantages of the proposed algorithm compared to the current mainstream method. Section 4 draws the conclusion and gives the discussion of future work.

II. METHODS

The algorithm proposed in this paper converts the gray values of the three color components of the pixels in the image to the level of $0 \sim (M-1)$ by (1), then the selected maximum gray value converted back to the level of $0 \sim 255$ by (2) is the corresponding background model at the pixel position. Although the perception-based color model (CIE, HVC, HSV) is more suitable for image expression and analysis, this paper chooses the RGB color model because the video used in this article is in AVI format that its internal color data is based on RGB space.

$$g_c(i, j) = \left\lfloor \frac{M-1}{255} f_c(i, j) \right\rfloor, c = r, g, b \quad (1)$$

$$f'_c(i, j) = \left\lfloor \frac{255}{M-1} g_c(i, j) \right\rfloor, c = r, g, b \quad (2)$$

In this paper, the selection of M is directly related to the efficiency of the algorithm and the quality of the background model. Therefore, how to choose the appropriate M is very important. In an image, the color of the background and the target is usually different, so the gray scale is different in the three color components. The gray distribution of an image on a pixel is usually composed of background gray, target gray, and random noise gray. The distribution of blue, green, and red color components of a pixel on an image is shown in Fig. 1.

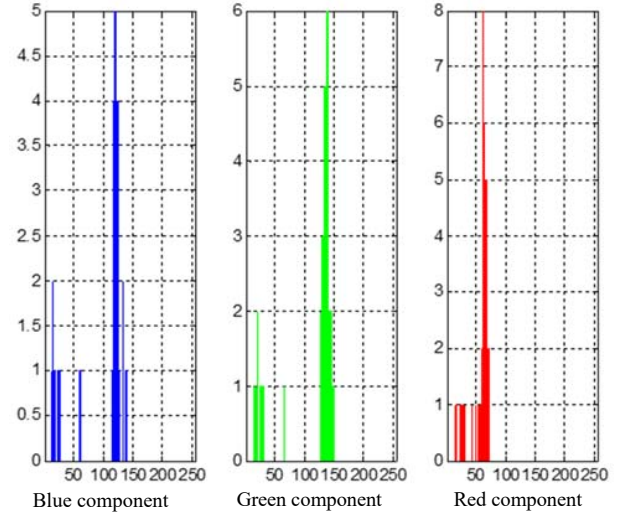


Fig. 1. Gray scale histogram of each color component of a pixel on an image

It can be seen from the figure that the histogram of all three color components are obviously divided into two clusters of concentrated histograms and a small number of isolated short histograms, which correspond to the background color gray, target color gray and noise color gray respectively. So the corresponding background gray scale of the color can be selected as the gray scale value of the color component on the pixel.

However, the calculation consumption is very large if each pixel on the image stores three 255-level gray scale histograms. As can be seen from Fig. 1, the gray distribution of the target and background are not concentrated on a certain gray value, but distributed within a certain gray range, due to the influence of environmental factors such as lighting and the camera's own factors. The peak in the histogram represents that the related gray value has the highest probability compared with other gray values. Those histograms around the peak indicate that the probability of related gray value is low. If the gray value corresponding to the peak is selected as the value of the corresponding component of the pixel in the background model at this time, the contribution of the other slightly lower gray value corresponding to the background model will be ignored.

To solve this problem, it can be solved by appropriately reducing the number M of gray levels. Therefore, the gray level from 0 to 255 is converted to $0 \sim M-1$ ($M = 26$), and the gray level histogram is established again, as shown in Fig. 2.

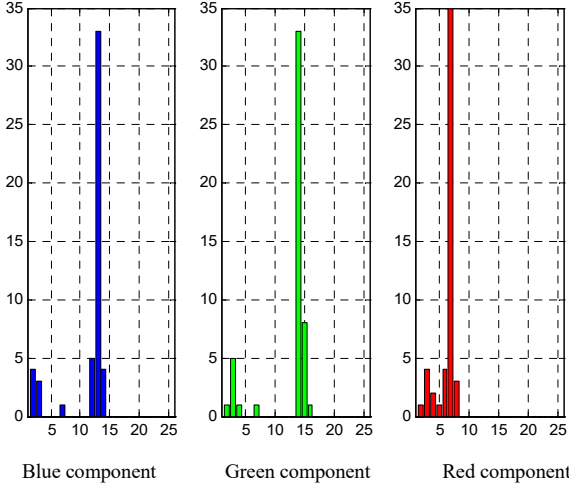


Fig. 2. Gray scale histogram of each color component of a pixel on an image ($M=26$)

Observed from Fig. 2, the background and target parts are still distinct in the histogram. Compared with the original histogram, the histograms in a certain range are merged and the peak value is greatly improved. Compared with the surrounding histograms, the gap is more obvious. The peak value is retransformed into the 0-255 area, and can be used as the gray value of the corresponding component of the pixel in the background model. After such processing, the amount of data calculation is the previous $M / 255$. At the same time, due to the histograms within a certain range near the peak are merged, more data around the peak can also contribute to the background model.

It can be seen from the above analysis that the smaller the value of M , the less the amount of calculation, and it will inevitably reduce the color resolution of the background model. However, in practical applications, color images are often converted into grayscale images, so excessive color resolution is unnecessary. In order to make the background model capable of representing a gray scale background of 0 to 255 levels, the minimum gray scale that can be expressed on the color component should be 3. According to (3), the value of M can be inferred:

$$\begin{aligned} \left\lfloor \frac{255}{M-1} g_c(i, j) \right\rfloor &= 3 \\ \frac{255}{M-1} * 1 &< 4 \\ M &> 64.75 \end{aligned} \quad (3)$$

That is, when M is greater than 64, the generated background model has the ability to express colors of 256 gray levels. Because in practical applications, the purpose of creating a background model is often to separate the background from the target, but not care much about how many colors it can show. Therefore, the selection of the M value can continue to decrease while ensuring a certain background performance, and the specific value should be sufficient to separate the background and the target in the histogram.

III. EXPERIMENTAL RESULTS

The experiment uses a video image with a length of 801 frames as the material to establish a background model based on the method proposed in this paper and divides the

target and the background, compared with the current mainstream method of background modeling.



Fig. 3. Original video image

In the experiment, a target person dressed in dark black clothes walked back and forth in a static scene. We choose the first 50 frames as the training images of the background model, and take $M = 26$. The background model obtained by the algorithm in this paper and the mean value method is shown in Fig. 4.

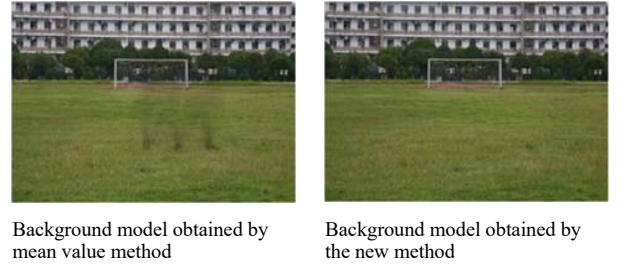


Fig. 4. Comparison of the background models obtained by the two algorithms

As can be seen from Figure 4, in the model obtained by the mean value method, the walking area of the person is slightly dark, and even traces of the person's movement are clearly visible. This is because the region is darker in the video, so the gray scale value of the RGB color component processed by the mean value method will appear smaller. The background model established by the algorithm in this paper will exclude the influence of the moving target and noise on the background model. It can reflect the real background better.

The following uses the difference image method to segment the target and the background based on the established background model. First the background image and the k th frame image are converted into gray scale images, the conversion formula is as (4):

$$I = \frac{1}{3}(R + G + B) \quad (4)$$

The converted gray image of the k th frame and the background gray image are obtained. According to (5), the difference image can be inferred:

$$D_k(i, j) = \begin{cases} 0, & |f_k(i, j) - B(i, j)| < T \\ 255, & |f_k(i, j) - B(i, j)| \geq T \end{cases} \quad (5)$$

T is the detection threshold, which can be set according to the specific situation to try to segment out the complete moving target. In this experiment, $T = 15$, and the obtained differential image is shown in Fig. 5 (a) and (b). However, these differential images often contain a lot of noises. To remove these noises, you can use the open operation to process the differential images to reduce these interferences. The processing results are shown in Fig. 5 (c) and (d).

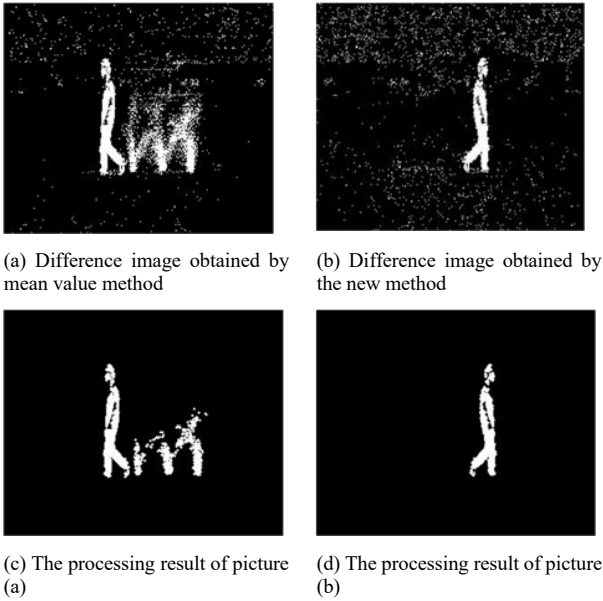


Fig. 5. Comparison of the detection results of moving targets by two algorithms (frame 60)

It can be seen from Fig. 5 that when $T = 15$, both algorithms can completely segment the moving target. The difference image obtained by the new algorithm in this paper is more ideal. Although there is also a lot of noise, the noise is almost completely eliminated after being processed to separate the target and background, as shown in Fig. 5 (d).

The background image obtained by the mean value method contains obvious moving traces, which are mistakenly identified as moving targets. And it is difficult to eliminate these false targets even if it is morphologically processed, as shown in Fig. 5 (c). To solve this problem, the number of images trained in the background model can increase, but this requires more training time and memory space. Another method is to increase the threshold T , but it may identify a part of the moving target as background image, causing the target to appear "holes" that will destroy the integrity of the target.

From the experimental results, we can see that the algorithm proposed in this paper has obvious advantages over the traditional mean value method in terms of memory consumption, processing time, anti-interference ability, etc. And it also lays a good foundation for the subsequent segmentation of the background and target, so as to define more accurate positioning of moving targets.

IV. CONCLUSION

Moving target detection has always been a hotspot in image processing. The establishment of background models is the basis and key to moving target detection. The traditional mean background method has a long training time and requires many training images. This article focuses on the average background method. Insufficient terms, a method for building a background model based on pixel histogram is proposed. Experiments show that compared with the average background method, this method requires fewer training image frames, faster model building, strong anti-interference ability, and has certain practical application value. In the future, other more complex algorithms will be combined to help reduce the amount of computation more effectively and improve the robustness of

the system. The proposed method will be further explored and applied into the applications, such human motion analysis, human robot collaboration, etc [16].

ACKNOWLEDGMENT

This work was supported by grants of National Natural Science Foundation of China (Grant Nos.51575407, 51575338, 51575412, 61733011).

REFERENCES

- [1] P. Chen, L. Zheng, X. Wang, H. Li, and L. Wu, "Moving Target Detection Using Colocated MIMO Radar on Multiple Distributed Moving Platforms," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4670–4683, Sep. 2017.
- [2] R. Jain and H.-H. Nagel, "On the Analysis of Accumulative Difference Pictures from Image Sequences of Real World Scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-1, no. 2, pp. 206–214, Apr. 1979.
- [3] B. Gloyer, H. K. Aghajan, K.-Y. Siu, and T. Kailath, "Video-based Freeway Monitoring System using Recursive Vehicle Tracking," p. 8.
- [4] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 19, no. 7, p. 6, 1997.
- [5] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, Fort Collins, CO, USA, 1999, pp. 246–252.
- [6] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric Model for Background Subtraction," in *Computer Vision — ECCV 2000*, vol. 1843, D. Vernon, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 751–767.
- [7] Kyungnam Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Background modeling and subtraction by codebook construction," in *2004 International Conference on Image Processing, 2004. ICIP '04.*, Singapore, 2004, vol. 5, pp. 3061–3064.
- [8] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-Time Imaging*, vol. 11, no. 3, pp. 172–185, Jun. 2005.
- [9] L. Jiao, Ed., *Advances in natural computation: second international conference, ICNC 2006, Xi'an, China, September 24-28, 2006: proceedings*. Berlin ; New York: Springer, 2006.
- [10] F. H. C. Tivive, A. Bouzerdoum, and C. Abeynayake, "GPR Target Detection by Joint Sparse and Low-Rank Matrix Decomposition," *IEEE Trans. Geosci. Remote Sensing*, vol. 57, no. 5, pp. 2583–2595, May 2019.
- [11] Jiang, D. , Zheng, Z. , Li, G. , Sun, Y. , Kong, J. , Jiang, G. , Xiong, H. , Tao, B. , Xu, S., Yu, H. Liu, H., and Ju, Z., *Gesture recognition based on binocular vision*, *Cluster Computing*, 22(6): 13261–13271, 2019.
- [12] Liu, B., Cai, H., Ju, Z., Liu, H., *RGB-D Sensing based Human Action and Interaction Analysis: A Survey*, *Pattern Recognition*, 94:1-12, 2019.
- [13] Gao, Q., Liu, J., Ju, Z., Zhang, X., *Dual-hand detection for human-robot interaction by a parallel network based on hand detection and body pose estimation*, *IEEE Transactions on Industrial Electronics*, 66(12): 9663-9672, 2019.
- [14] Yinfeng. Fang, Nalinda. Hettiarachchi, Dalin. Zhou and Honghai. Liu, "Multi-modal Sensing Techniques for Interfacing Hand Prostheses: a Review," *IEEE Sensors Journal*, vol. 15, pp. 6065–6076, November, 2015.
- [15] C. Yang, C. Zeng, P. Liang, Z. Li, R. Li, C-Y. Su, *Interface design of a physical human-robot interaction system for human impedance adaptive skill transfer*. *IEEE Transactions on Automation Science and Engineering*, 2018, 15(1): 329-340.
- [16] Ogenyi, U., Liu, J., Yang, C., Ju, Z., Liu, H., *Physical human-robot collaboration: robotic systems, learning methods, collaborative strategies, sensors and actuators*, *IEEE Transactions on Cybernetics*, 2019, in press.