Hybrid Moving Object Detection System Based on Key Frame Extraction

Yu Du[1*], Xueyao Wang[1] , Shi-Ting Wang[2], Shuangshuang Xue[2], and Zengchang Qin[1,3 #]
[1] Intelligent Computing and Machine Learning Lab, School of ASEE
[2] School of Automation Science and Electronic Engineering (ASEE)
Beihang University (BUAA), Beijing, 100191, China
[3] Robotics Institute, Carnegie Mellon University, USA
Corresponding Emails: *ydu1989@gmail.com, #zcqin@buaa.edu.cn

Abstract: **In this paper we construct a hybrid moving object detection system. In this system, we first use the frame difference method to extract key frames in a given video sequence, then use the optical flow method and the HSV background subtraction method to extract the moving objects, respectively. We propose two hybrid methods: Improved Optical Flow method and Improved HSV Background Subtraction method. Experimental results have shown that our proposed system has strong robustness and be effective to detect moving objects in various conditions. We also introduce a human-computer interaction tool for selecting the focus area for users. This system will largely benefit for real-world video surveillance applications.**

Keywords: moving object detection, key frame, frame difference method, improved optical flow method, improved HSV background subtraction method.

I. INTRODUCTION

Moving object detection is currently one of the most active research topics in the domain of computer vision [1]. Among traditional motion detection methods, frame difference method [2, 3], optical flow method [4] and background subtraction method [5, 6] were widely used to detect moving objects. Frame difference method is computationally inexpensive due to its low programming complexity, and it has insensitivity to illumination variations and can adapt to various conditions; however it can only extract the boundary, but not the full area of the moving object. Optical flow method does not require any prior knowledge of scene information and can handle dynamic tracking, while it is time-consuming and it can be greatly influenced by even small illumination variations. HSV background subtraction method is closer to human eye perception compared to other methods, and it can reduce the impact of dynamic scene chances using background updating. However, it is difficult to adjust subtraction threshold adaptively and excessive operations of converting pixels from RGB (Red, Green and Blue) color space to HSV (Hue, Saturation, and Value) color space [9] can slow down its computation speed. Based on their superiorities and shortcomings, we design a system to combine the advantages of these three methods to improve the detection accuracy. The main contribution of this paper is designing a new real-world application by combining existing methods rather than a theoretical treatment of moving object detection.

In this paper we present frame difference method to solve the first task - the key frame extraction. During this process, we propose novel judgment rules to search for key frames corresponding to fast motions in the given video. Then we use optical flow method and HSV background subtraction method, respectively, to extract the moving object based on key frames. A brief overview of the two combined methods is given in Section 2. Section 3 introduces our hybrid moving object detection system with a novel interactive user interface. Section 4 presents our experimental validations.

II. COMBINED MOVING OBJECT DETECTION METHODS

A. Frame difference method

Frame difference method carries out frame difference operation of two adjacent frames in a collection of image sequences (usually decomposed from a video). To effectively avoid the high time cost, we use inferior moving object detection of traditional frame difference method for motion detection to extract the key frame, that is, the frame corresponding to a fast moving object.

Traditionally, the Euclid distance and variance are used to quantify the frame difference. Frame difference-Euclid distance is defined by:

$$\Delta I_n = \sqrt{\sum_{i,j} (\Delta f_{n+1}(i,j) - \Delta f_n(i,j))^2} \qquad (1)$$

Frame difference-variance is defined as follows:

$$Var = \sum_{i,j} (\Delta f_n(i,j) - \overline{\Delta f_n})^2 \qquad (2)$$

and

$$\Delta f_n(i,j) = f_{n+1}(i,j) - f_n(i,j)$$

where $f_n(i,j)$ is the pixel value of point (i,j) in the nth frame. $\overline{\Delta f_n}$ denotes the average value of all frame difference values which are formally defined by:

$$\overline{\Delta f_n} = \frac{1}{ij} \sum_{i,j} \Delta f_n(i,j) \qquad (3)$$

Through a series of experiments we found that compared to frame difference-Euclid method, frame difference-variance method is less sensitive to noise and can adapt to various dynamic environments with good stability (see Section IV).

Thus we take frame difference-variance as evaluation criteria to quantify the frame difference. To accelerate the computation efficiency, we set sampling extraction interval to extract one frame from each several frames, in which way we compress the video. Also we set the number of key frames to be extracted according to users' demands. Then we subtract adjacent frames in image sequences and calculate frame difference variance in high-dimensional space. Afterwards, we draw the frame difference-variance curve and find the local maximum point in each certain length (e.g. from point *A* to *J* in Fig.1). We set up the following judgment rules to search for the key frames.

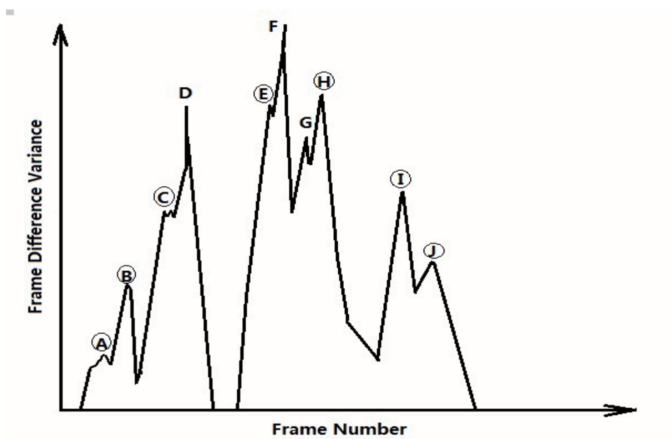

Figure 1. The frame difference-variance curve.

Rule 1. *For the points that represent sudden changes in frame difference variance, we consider them caused by illumination variation or other external factors and not consider them as key frames (e.g., point D, F).*

Rule 2. *For two adjacent local maximum points within 20 frames in the video sequence, we set the maximal point as the key frame. (E.g., point H is the key frame while point G is not.)*

Rule 3. *The remaining points were set to key frames in accordance with descending frame difference variance value.*

For instance, given the frame difference-variance cure in Fig. 1, we set point *H, E, I, C, J* as key frames if we need to extract five key frames from a given video sequence, while point *H* is the *1st* key frame and point *J* is the *5th* key frame based on the rules given above.

We consider all frames in a particular region when extracting the key frames, otherwise, setting rules of using certain but not all frames would increase computation complexity. As we will see below, the proposed system can select a number of regions of interest. Therefore we are able to extract key frames with several moving objects corresponding to different regions of interest.

B. Improved optical flow method

Optical flow method is one of the key methods to determine the speed and direction of moving objects. Numerous previous studies have been focused on using time-domain data and relevance of image sequence pixel intensity change to determine the movement of each pixel location [7].

The optical flow method is time-consuming and even small illumination variations can greatly reduce the detection accuracy. On the other hand, for the frame difference method, small illumination variations can hardly influence detection accuracy, while large illumination variations cause an abrupt maximum variance value that can be usually ignored.

Due to frame difference-variance method's insensitivity to illumination variations, we first use frame difference-variance method to extract key frames in the given video sequence, and then use the Horn-Schunck algorithm [7] under optical flow method to extract moving objects based on the key frames, in which way the detection operation speed would be greatly improved. We also take morphological operations [8] guided by the edge gradient to produce close object contours.

C. Improved HSV background subtraction method

HSV color space [9] is formulated and used based on the human visual system, using Hue, Saturation and Value to describe the color. Compared to the RGB color space, the HSV color space is closer to human eye perception and therefore well-used in image processing and pattern recognition.

A common problem for moving object detection is that illumination variations can reduce detection accuracy. To solve this problem, Stauffer and Grimson [10] proposed a background update process by setting up a subtraction threshold to improve its insensitivity to illumination variation. In the background update process, we first set up a subtraction threshold and then compute the difference value between key frame and background frame of each pixel. For the pixels with larger difference value than the threshold, we regard them as moving object pixels and keep the original background frame pixels unchanged. For those pixels with smaller difference values than the threshold, we consider them caused by illumination change, and then replace the background frame pixels with corresponding key frame pixels. As we can see, the subtraction threshold is crucial to the detection accuracy. However, by far there is no general method to determine the threshold in this problem setting, so that we set it empirically through trial-by-trial experiments. Then we subtract key frame and background frame in three-channel (H, S, V) and set three-channel weights to integrate the final composite image. [11]

Considering the accelerating effect of frame difference-variance method to improve the detection operation speed, we also introduce key frame extraction process mentioned above into HSV background subtraction method to enhance its computation efficiency.

III. HYBRID MOVING OBJECT DETECTION SYSTEM

In our proposed system, we firstly use frame difference method to extract key frames in the given video sequence, then use the optical flow method and the HSV background subtraction method, respectively, to detect moving object for a sample sequence intercepted from the video. Given the detection outcomes by two algorithms, we retain the method that has higher detection accuracy and use it to detect moving objects for the whole video sequence. The schematic flowchart of the hybrid moving object detection system is shown in Fig.2.

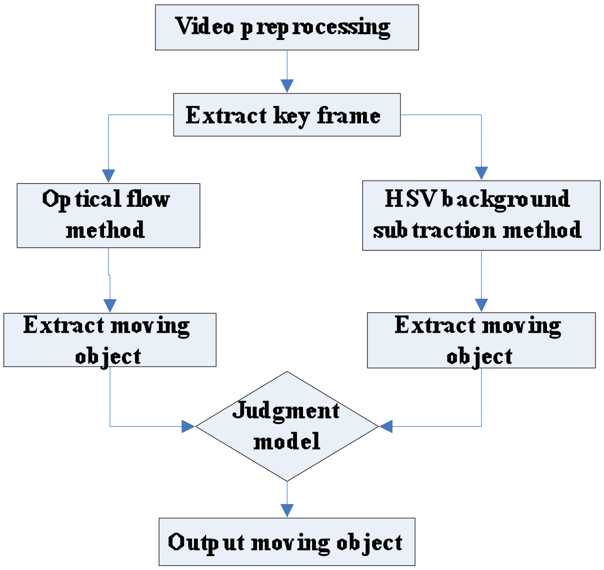

Figure 2. Hybrid moving object detection system flowchart.

The detailed process of the hybrid moving object detection system can be described by the following steps:

Step 1. Video preprocessing: video format conversion, image sharp processing, with the purpose of reducing computational complexity.

Step 2. Set essential parameters: e.g. the number of key frames to be extracted, sampling extraction interval, and subtraction threshold in HSV background update process. Set region of interest according to specific situations and user's demands. For example, the user can input the coordinates of the interest region or using mouse to draw out a square of interests region from the scene. [1]

Step 3. Key frame extraction: Use frame difference method to subtract adjacent frames and calculate frame difference variance in high-dimensional space, then get key frames based on our proposed judgment rules mentioned above.

Step 4. Use optical flow method and HSV background subtraction method respectively to detect moving object based on key frames for a sample sequence. Adequate morphological processing is necessary to remove noise and improve the detection accuracy.

Step 5. Compare two detection outcomes by two combined methods, we retain the method that has the higher detection accuracy and use it to detect moving object for the whole video sequence. Then output detection outcomes.

In the design of the hybrid moving object detection system, we develop a friendly and convenient man-machine interaction software interface compiled and displayed by the Matlab. The Matlab code and the GUI files are downloadable from the project webpage [12]. The system can be roughly divided into three parts: the main display area, the control area and the region of interest. The interface layout with a sample scene of BUAA Main Library security video is shown in Fig.3. To use this interface, users can read and play the video in the main display area and select region of interest. Click 'Run' button to start unseen running process to extract key frames by the order of importance. We then can obtain the starting time point of each abnormal segment. Afterwards we use the optical flow method and the HSV background subtraction method, respectively, to extract moving objects and choose preferable outcomes of these two methods. We can jump directly to the abnormal video segment corresponding to key frames and replay the video segment, as well as store abnormal pictures for the purpose of comprehensive monitoring for later use.

Our system has a good human-computer interaction tool, which satisfies the users in the following aspects: Users can select region of interest by adjusting the horizontal and vertical ratio of the selected area. Besides, users can set essential parameters to improve detection accuracy in various conditions: e.g. the number of key frames to be extracted, sampling extraction interval, and subtraction threshold in HSV background update process and etc. As there is still no general method to perform well in all various conditions, using various detection methods can enhance overall detection and robustness of the system. Due to its strong scalability, we can also embed other key frame extraction and moving object detection methods to further improve the system.

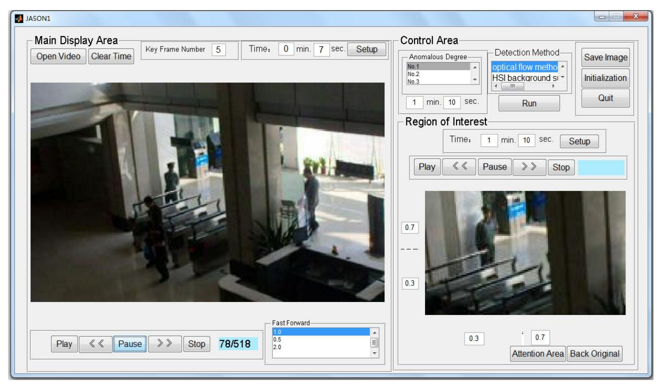

Figure 3. User interface of moving object detection system. The left-hand window is the Main Display Area, the upper-right window with parameter settings is the Control Area and the bottom-right window is the Region of Interest selected from the main scene by users through interactive tools.

IV. EXPERIMENTAL VALIDATIONS

In this section, we verify the key frame extraction process and moving object detection process to demonstrate the effectiveness of the proposed hybrid system.

A. Key frame extraction

In order to investigate feasibility and effectiveness of the frame difference method, we use it to extract key frames in two video surveillance applications for detecting the number of people

[1] We assume that the camera is fixed so that the interest region is unchangeable. This assumption is true for most of the real-world video security surveillance applications.

going across the door. We set five key frames for each video sequence. The first video has a background without illumination variations while in the second video with the intentionally changed illumination intensity.

In the first experiment, we keep the illumination intensity unchanged. Its frame difference variance curve and corresponding key frames were shown in Fig. 4. Key frames were selected in the middle of each section containing obvious motions, usually corresponding to local maximum frame difference variance values. As the moving object shifts a large distance in fixed camera lens in each unit time, the overlap part of two adjacent frames reduces and frame difference variance increases suddenly.

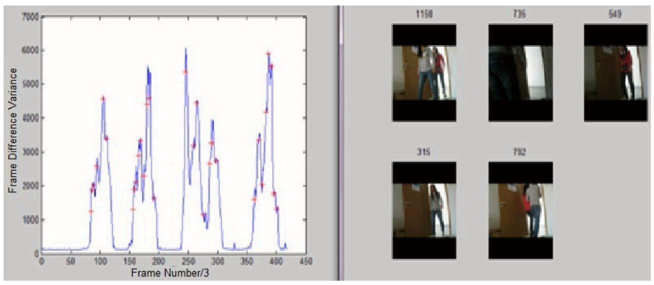

Figure 4. Frame difference variance curve and corresponding key frames without illumination intensity change.

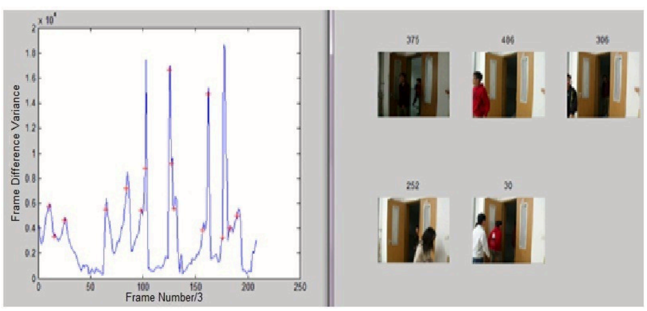

Figure 5. Frame difference variance curve and corresponding key frames with intentional illumination intensity change.

In the second experiment, we intentionally change the illumination intensity. Its frame difference variance curve and corresponding key frames were shown in Fig. 5. Due to illumination intensity changes, we find some frames turning bright or dark suddenly, corresponding to certain abrupt maximum values in the frame difference variance curve. According to our key frame judgment rules, we consider those points as anomalous points and ignore them. In this way, key frames correspond to time points when the moving objects shift a large distance in fixed camera lens. For both of the above experiments, various sampling extraction interval settings have little influence on key frame extraction process and its accuracy, and video compression accelerate the computation efficiency to a large extent.

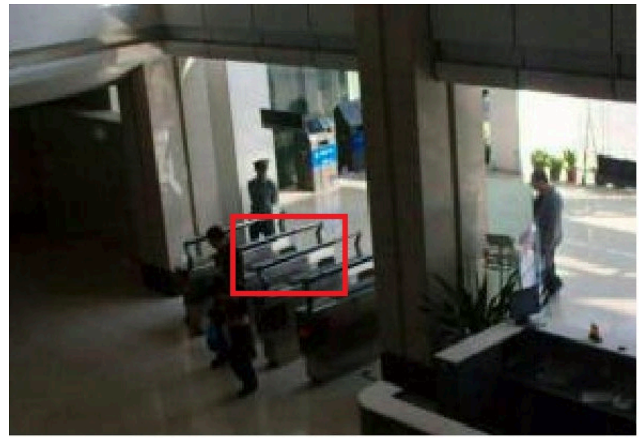

Figure 6. Focus on the region of interest. In many practical applications, we may only be interested in a small region of the video scene. For example, given the above library security camera scene, the focus is the security gate.

B. Moving object detection

Our proposed hybrid moving object detection system can be widely used in various surveillance applications. In the following experiment, we use the system to analyze a video sequence that people go through the controlled entry in a library. We analyze the video sequence obtained by a surveillance camera hanging on the floor ceiling of the BUAA Main Library. Because the camera is with a wide-angle, it captures a large part of the library entrance. We are only interested in the region where the security gates located. To systematically verify our two detection methods, we intercept a video sequence which contains large flow density. The focused region of interest was shown in Fig. 6.

We first extract key frames in the video sequence and then use the optical flow method and the HSV background subtraction method, respectively, to extract the moving object, i.e. people going across the security gates. Experimental results were presented in figures from Fig. 7 to 10. Fig. 7 shows the generated optical flow image of the first key frame, and connected domains by image binarization with their centers marked blue stars. Fig. 9 shows three-channel composite image by the HSV background subtraction method and connected domains of the first key frame by image binarization. In Fig. 8 and Fig.10, we present six extracted objects by the two detection methods.

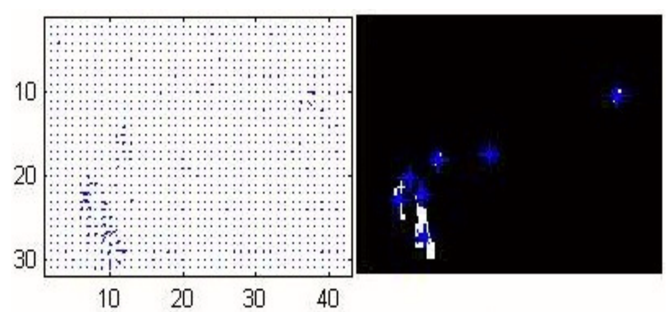

Figure 7. Generated optical flow figure and connected domains by the optical flow method.

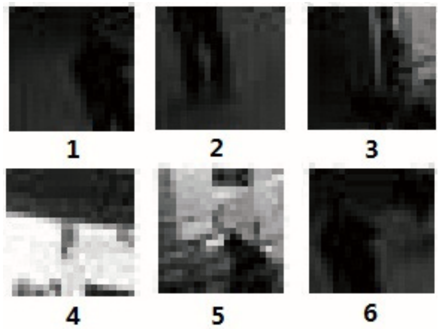

Figure 8. Six extracted objects by the improved optical flow method.

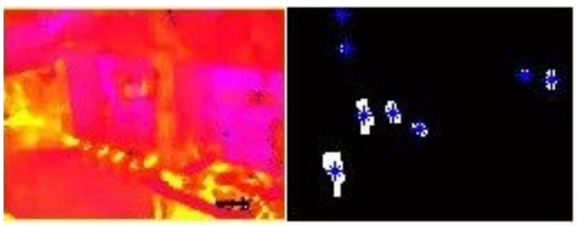

Figure 9. Generated three-channel composite image and connected domains by HSV background subtraction method.

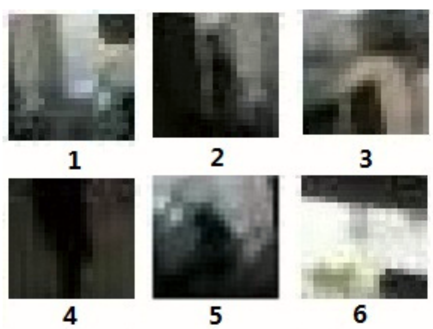

Figure 10. Six extracted objects by the improved HSV background subtraction method.

Based on a series of experiments containing illumination variation control, we found that each combined method has its own superiority compared with the other method in certain specific conditions. In the video sequence without illumination variation, the optical flow method has high detection accuracy, no missing objects and repeated detection situations. While the HSV background subtraction method often detect repeatedly, due to overlapped connected domains in the image binarization.

In the video sequence with frequent illumination variations, the optical flow method's performance drops down suddenly for its intense sensitivity to illumination variations, thus causes frequent missing and repeated detections. In contrast, with background update process, the HSV background subtraction method performs superiorly due to its insensitivity to illumination variations and can achieve high detection accuracy.

V. Conclusions

In this paper, we propose a hybrid moving object detection system with stability, feasibility and validity. We conducted a serial of experiments and the results showed that our proposed system has strong robustness and effectiveness to detect moving object in various illumination conditions and being computationally inexpensive. We also design a human-computer interactive interface for selecting the regions of interests for users. This tool will be practically useful for any security surveillance applications.

Our future work focuses on detecting multiple arbitrary-shaped regions in real-time complex scenes. The system is extendable to embed other motion detection methods. We also hope to adaptively adjust the subtraction threshold in the HSV background update process to further improve the detection accuracy in the future work.

Acknowledgment

This work is partially funded by the Aerospace One Hundred Project from Beihang University, the NCET of MOE and the SRF for ROCS, SEM.

References

[1] R. Collins, A. Lipton, T. Kanade, "Introduction to the special Section on video Surveillance," Proceeding of IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, pp.745-746.

[2] C. Zhan, X. Duan, S. Xu, Z. Song, and M. Luo, "An Improved Moving Object Detection Algorithm Based on Frame Difference and Edge Detection". Fourth International Conference on Image and Graphics, 2007, pp.519-523.

[3] D. Migliore, M. Matteucci and M. Naccari, "A revaluation of frame difference in fast and robust motion detection". Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks, 2006, pp.215-218.

[4] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High Accuracy Optical Flow Estimation Based on a Theory for Warping". Proceedings of 8th European Conference on Computer Vision, Springer LNCS 3024, 2004, vol.4, pp.25-36.

[5] C. Wren, A. Azarbayejani, T. Darrel, and A. Pentland, "Pfinder: real-time tracking of the human body". Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, pp.780-785.

[6] M.Piccardi, "Background subtraction techniques: a review", IEEE International Conference on Systems, Man and Cybernetics, 2004, vol.4 pp.3099-3104.

[7] B. Horn, and B. Schunck, "Determining optical flow". Artificial Intelligence, 1981, vol. 17, pp.185-203.

[8] H. Elbehiery, A. Hefnawy, and M. Elewa, "Surface Defects Detection for Ceramic Tiles Using Image Processing and Morphological Techniques". Proceeding of IEEE Transactions on Egnieering, Computing and Technology, 2005, vol.5, pp.158-162.

[9] http://en.wikipedia.org/wiki/HSL_and_HSV

[10] C. Stauffer and W. Grimson, "Adaptive background mixture models for real-time tracking". Proceeding of IEEE Computer Vision and Pattern Recognition, 1999, pp.246-252.

[11] K. Riza, K. Gwangwon, K. Youngbae, P. Sungbum, and P. Jongan, "Efficient Image Retrieval Using Adaptive Segmentation of HSV Color Space", International Conference on Computational Sciences and Its Applications", 2008, pp.491-496.

[12] http://icmll.buaa.edu.cn/projects/object_detection