

Surveillance Video Synopsis Based on Moving Object Matting Using Non-interactive GrabCut

Jianrong Cao*, Xuemei Sun and Lili Hao

Department of Information and Electrical Engineering, Shandong Jianzhu University, Jinan, Shandong 250101, China

Abstract: An algorithm of surveillance video synopsis is presented in this paper. At first, the detection method of moving object for surveillance video is proposed. Then the ROI (Region of Interest) of moving object is extracted and a noninteractive matting algorithm of moving object based on GrabCut is presented. These matting moving objects are placed without overlapping in a background image, so a frame that several moving objects have been placed in a background image can be obtained. Finally, a series of these frame images can be obtained in timeline and a single camera surveillance video synopsis can be formed. The experimental results show that our detection method of moving object is superior to those methods such as Gaussian mixture model, background subtraction, frame difference method and Li's method. The synopsis obtained contains all objects appeared in surveillance video and is a concise expression for original video.

Keywords: Video monitoring, Background modeling, Object detection, Video synopsis.

1. INTRODUCTION

With the development of social economy and computer technology, the demand of social public security is also proposed. Many surveillance cameras are installed in many big public places and used to monitor and record the state of the places, so a massive video data is formed. If we need query and retrieve the recorded video data, the traditional forward and backward method can be used to browse the surveillance video, but it is time-consuming and may omit the important information. It has become a hotspot and difficult problem for the researchers to describe briefly, browse, query and analyze the video content in video monitoring field.

Video synopsis is a video summary technology that concentrates the original video information to the greatest extent by analyzing and processing a long video file and extracts the key information. The traditional video summaries only compress the video content in timeline. However, video synopsis compresses the video content not only in time but also in space. The length of video synopsis is shorter than that of the original video and the storage space of surveillance video can be greatly saved. Meanwhile, the basic content of the original video is retained and the user can quickly browse and search the video content. The video synopsis technology has the broad application prospects and the important research significance.

The surveillance video synopsis technology based on object can be divided into two categories. One is the conventional video summarization method that the surveillance video is divided into a so-called shots (the actual

surveillance video has not the shot), and then the key frames are extracted to form a video summary. Another is a new video analysis method proposed in recent years. The frame image can be changed by this new method while the frame image as the minimum basic unit of video can not be divided in the traditional video summarization technology, so the length of the video synopsis is further shortened and the storage space is significantly saved. The representative methods are the video montage technology and video synopsis technology. Y Pfitz, A Rav-Acha, S Peleg [1-3] proposed the overall framework of a video synopsis technology in their papers. It is a video to video conversion mode. In recent years, Teng Li *et al.* [4] have proposed a multiple video summary method in which many auxiliary videos are placed in a main video in timeline. Yael Pritch *et al.* [5] have proposed a video summary using the similar activity clustering method. The human eye focus position of surveillance video operator has been used to obtain a real-time surveillance video synopsis presented by U. Vural [6]. Uros Damjanovic [7] has adopted a method of detecting and clustering the important events of surveillance video to form the synopsis, and the so-called important event here is the moving object with more energy change in successive frames. Xu Min [8] has also presented a video synopsis technology based on target motion information.

In this paper, our algorithm of video synopsis is different from those methods mentioned above. After background modeling and detecting of moving object, the ROI of moving object is extracted and a noninteractive matting algorithm of moving object based on GrabCut is used to obtain the matting moving object. These matting objects are placed without overlapping in a background image and a series of these frame images can form a single camera surveillance video synopsis.

*Address correspondence to this author at the Department of Information and Electrical Engineering, Shandong Jianzhu University, Jinan, Shandong 250101, China; Tel: (086) 053186361988; E-mail: jrcao@126.com

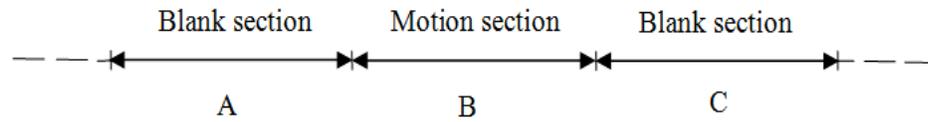


Fig. (1). Segmentation of surveillance video.

2. SELECTION AND UPDATE OF BACKGROUND IMAGE

The surveillance video synopsis is a procedure of processing the recorded video. In order to improve the efficiency of generating video synopsis, the surveillance video will be segmented into a blank section without motion information and motion section with motion information according to the features of monitoring video as shown in Fig. (1).

Video segmentation can be performed by the histogram difference of two adjacent frames. When the difference value of histogram is less than a predetermined threshold, the frame is in blank section and there is not the moving object between the two frames. If the difference value is greater than the threshold, the motion section begins. When the difference value is less than the threshold again, the motion section is ended and the blank section starts. In the blank section, the posterior frame can be used as the background image. The background image of motion section can use this background image as the initial background image, and then the Gaussian mixture model is utilized to model and update the background image.

Fig. (1) is an example. The A and C are the video section without containing the moving object and the section B contains the moving object. After background modeling for the A, B and C respectively, the background images of A and B video clip can be compared. If the difference is within a certain threshold range, the change of background image in two video clips is little. The background image of video clip A is selected as the background of A and B. If the difference exceeds the threshold value, the change of two video background images is large. Then the background image of B is compared with that of C. If the difference is within a certain threshold range, the background image of video clip C is selected as background. If the difference exceeds the threshold value, the background image of video clip B is different with A and C, then the background image of B is chosen as the background. The background image of each motion section is selected by this method for the whole surveillance video.

3. DETECTION OF MOVING OBJECT

The detection of moving object is a foundation of surveillance video analysis. Many algorithms have been proposed for object detection in video surveillance application such as background subtraction, optical flow, Gaussian mixture model (GMM), frame difference method, and so on [9].

Background subtraction is the most popular choice to detect the stationary foreground object because it works well when the camera is static and the illumination changes gradually. This method detects the foreground object by measuring the difference between the current image and the background image. Even though the background subtraction technique works well to extract the moving object, it may

suffer to dynamic background changes such as the entrance of a new background object or sudden illumination change.

Optical flow based motion segmentation uses characteristics of flow to detect independently the moving object even in the presence of camera motion. However most optical flow methods are computationally complex and very sensitive to noise.

Gaussian mixture model is widely used in moving object detection in image sequences. However, the existing moving object detection methods based on Gaussian mixture model are not so efficient especially when dealing with noise or illumination mutation.

Frame difference method is a temporal differencing technique. It uses the pixel-wise difference between two or three frames in frame sequences to extract the moving object. It is computationally simple, fast and adaptive to dynamic environment. But the frame difference method is most sensitive to the threshold value when determining pixel-wise difference between consecutive frames and it may produce the holes in the foreground object. When a foreground object moves slowly or stops moving, the temporal differencing technique fails in detecting a change between consecutive frames and loses the target object information. A survey of moving object detection method can refer to literature [10, 11].

In this paper, we propose a detection method of moving object that combining the background subtraction with the Gaussian mixture model method.

At first, the median smooth filter and image equalization are adopted for the video frame in order to reduce the influence of noise and luminance change. Then the region filling is used to fill the holes of object detected by the background subtraction. Considering the real-time requirements, we utilize the following method to fill the holes.

The background is firstly filled by the foreground white color (point (0, 0) as the seed point), so the black holes can be obtained. Then this binary image is reversed (0 is changed to 1 and 1 is changed to 0). The image with white color holes is achieved and it is added with the original binary image in order to fill the holes of object. Fig. (2) is an example at frame 364. Fig. (2a) is a video frame and Fig. (2b) is a foreground image without image filling. Fig. (2c) is a foreground image using our image filling method and the number of hole in this image is decreased appreciably.

Finally, the detection of moving object is achieved by the logical AND operation of the binary foreground image of GMM and the binary foreground image of background subtraction.

We compare our method with GMM, three frames difference method, background subtraction and Li's method [12]. The performance of our method is similar to that of Li's

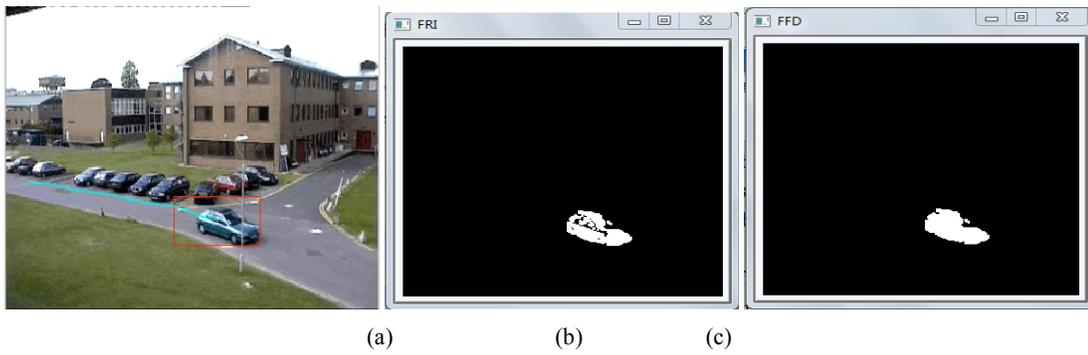


Fig. (2). An example of filling the holes.



Fig. (3). Segmentation results of Li's method at frame 15.

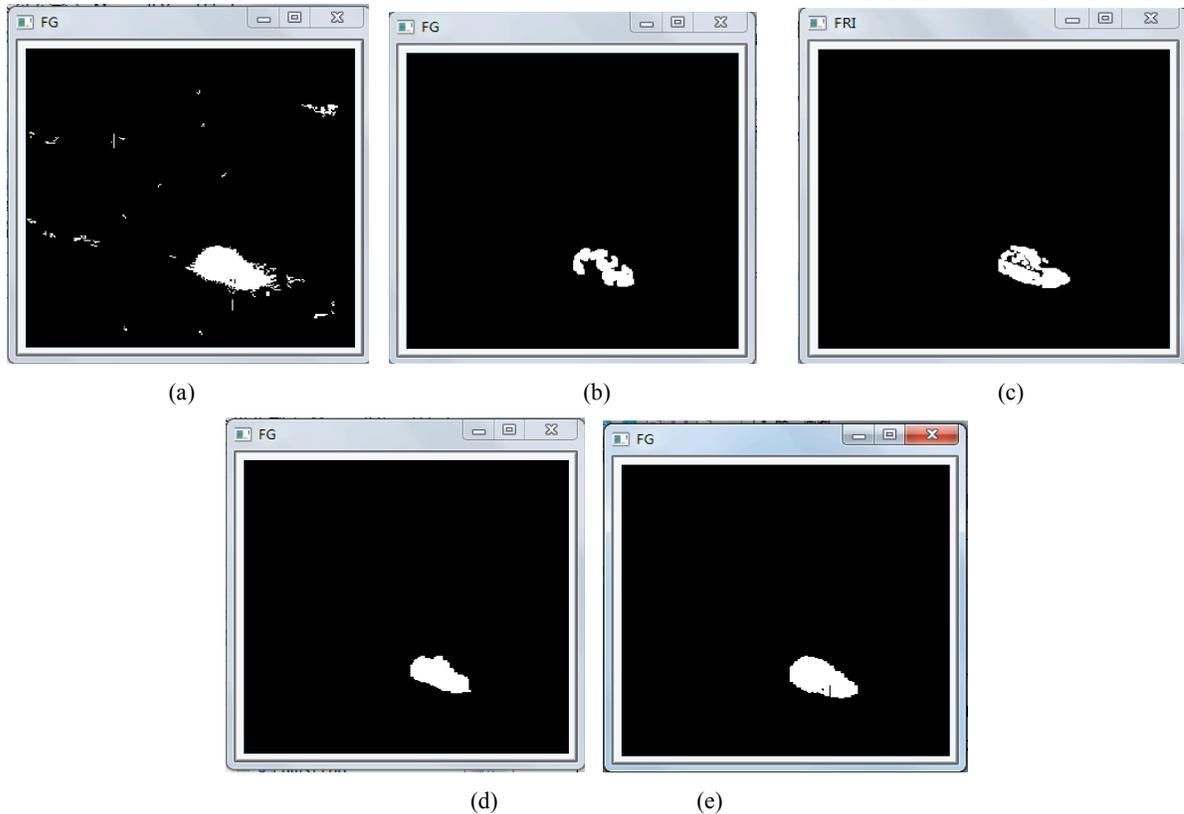


Fig. (4). Segmentation results of difference method.

method, but Li's method has many error detections in the beginning section of video because of the image noise as shown in Fig. (3). Fig. (4) shows the results of detection by GMM (a), three frames difference method (b), background subtraction (c), Li's method (d) and our method (e) at frame 364 respectively.

In Fig. (4), we can find that the foreground detected by GMM method has many disturbed areas that may produce many object detection errors. The moving object detected by three frames difference and background subtraction method may contain many holes. Li's method and our method have the better results.



Fig. (5). Frame image, ROI and matting moving object.

4. ALGORITHM OF MOVING OBJECT MATTING

Our moving object matting method uses the noninteractive GrabCut algorithm to extract the objects from the ROI image.

4.1. Extraction of the Roi Image

After the moving objects are detected by our method, we can obtain the binary foreground mask image. This image is processed in order to filter the smaller blobs and reserve the larger blobs. The position of centroid and the size of these blobs (moving objects) are computed and utilized to locate the moving object and determine the area of ROI. The center of rectangular ROI is the centroid of moving object and the width and height of rectangular ROI is 1.5 times width and height of the moving object respectively. The image area of rectangular ROI is extracted from the frame image.

4.2. Moving Object Matting

Moving object can be extracted by comparing the current frame with the background image, but the edge of the extracted moving object is not accurate and the holes in moving object are often produced. In order to extract more accurately the moving object, GrabCut algorithm [13] can be used. But this algorithm is an interactive image segmentation method and the user must indicate some pixels of segmentation boundary between the foreground and background. In order to avoid the necessary manual interaction, this method is improved and a noninteractive matting algorithm of moving object is proposed based on GrabCut.

After the moving object has been successfully detected and tracked, a rectangular region is drawn up around the moving object tracked. This rectangular region is called the ROI of moving object. The area outside ROI is directly considered as the background. The center part pixels of ROI are the foreground pixels and the edge position pixels of ROI are considered as the background. So some background and foreground pixels will be determined and GrabCut algorithm is used to segment the moving object in the continuous image of video. The edge of the matting image has a lot of burr, so edge smooth is used and the moving object boundary can be obtained accurately. This method avoids the common holes of moving object in the traditional object detection method. The procedure of our algorithm is as following.

Initialization steps:

① After detecting and tracking the moving objects, the rectangular region around moving objects is determined.

② The area outside rectangular region and the edge position pixels of rectangular ROI are considered as the back-

ground T_B to initialize the trimap T , the foreground

$$T_F = \Phi, T_U = \bar{T}_B,$$

③ for $n \in T_B, \alpha_n = 0$; for $n \in T_U, \alpha_n = 1$,

④ Set $\alpha_n = 0$ and $\alpha_n = 1$ respectively to initialize the foreground and background GMMs.

Iterative minimization steps:

① Compute GMM parameter k_n for each pixel n in T_U ,

$$k_n = \arg \min D_n(\alpha_n, k_n, \theta, z_n) \quad (1)$$

② Learn GMM parameter from data Z ,

$$\theta = \arg \min_{\theta} U(\alpha, k, \theta, z) \quad (2)$$

③ Estimate segmentation: use min cut to solve,

$$\min_{\alpha_n, n \in T_U} \min_K E(\alpha, k, \theta, z) \quad (3)$$

④ Repeat from step ①, until convergence,

⑤ Obtain result of segmentation.

⑥ Postprocessing steps.

Although our algorithm of moving object matting based on GrabCut can extract more accurately the moving object, the region of image with lower contrast at the transition from foreground to background may be lost. In Postprocessing steps we change the matting moving object image into a binary mask image and add it with the binary mask image obtained by our detection method in the same ROI position. The new binary mask image is used to extract the moving object from the original frame image.

Fig. (5) is an matting example of moving object using our method in surveillance video. The moving object detected is wrapped in a rectangular frame and the center pixels in the rectangular region are considered as the foreground, while the edge pixels of the rectangular region are considered as the background, and then the Grabcut algorithm is used to further segment. The accurate segmentation result can be obtained.

Fig. (5) shows that our algorithm can accurately achieve the matting of moving object when the probable location of moving object has been obtained. Because the moving object tracking has been made and the moving object has been enclosed in a rectangular frame in the process of tracking, the manual reference point label of foreground and background needed in the Grabcut algorithm is avoided. So it is more convenient for moving object matting.



Fig. (6). Surveillance video images in different time.



Fig. (7). Matting moving objects from surveillance video.



Fig. (8). Background image selected and the video synopsis formed.

5. ALGORITHM OF VIDEO SYNOPSIS BASED ON OBJECT

The matting moving objects are placed in a selected background image by our algorithm to form the final surveillance video synopsis. The specific procedure is as follows:

① The background image is selected according to the method described in this paper.

② The selected background image is reproduced.

③ Place the moving objects into the background image according to their order of appearance and their movement path.

④ When a new object is arranged, the rectangular area of object is removed from the background image and the next object will be placed into a residual background image in order to avoid the overlap of moving object.

⑤ Comparing the rectangular area of all moving objects placed with the area of the background image, if the difference value is less than a preset threshold, it is considered to have been filled with moving objects; Or if the path of a new moving object has been filled with other moving objects, the moving object can not be arranged in its moving path and it means that the background image has been filled with other moving objects.

⑥ If the background image has been filled, return step ①; If the video is at the end, go to step ⑦.

⑦ The image frames formed above are combined together in timeline and the video synopsis based on object can be obtained.

6. EXPERIMENTAL RESULTS AND DISCUSSIONS

There is an example of surveillance video which has 448 frames and is 17 seconds duration. According to our algorithm, the formation process of surveillance video synopsis is shown as Fig. (6-8).

Fig. (6) is a set of images of the original monitoring video in different time and shows the different moving object tracked. The matting results of each moving object are shown in Fig. (7). Fig. (8) (a) is a background image selected using the method described in this paper. The video synopsis formed finally is presented in Fig. (8b).

Fig. (8b) shows that the moving objects in the original video appear in different time and the matting moving objects are placed in a background image along the moving path. So the moving objects appeared in the monitoring video at different time can be seen in a frame image and such image is one of video synopsis images. Obviously, the video synopsis is a concentration of surveillance video in time and space and a concise expression of the monitoring video con-



Fig. (9). Another example of synopsis.

tent. It can greatly improve the efficiency of video browse, query and analysis.

Fig. (9) shows another example of synopsis. The video has 2822 frames and is 94 second duration. In this video a car stops and becomes the part of the background but the background image difference between the motion section and the blank section is less than the setting threshold, so the background image is not changed.

7. CONCLUSIONS

This paper presents a video synopsis algorithm for surveillance video. After background modeling and detecting of moving object for surveillance video, the proposed noninteractive moving object matting method based on GrabCut is used to obtain accurately the moving object from the video frame, and then these objects will be placed without overlapping in the same background image along their motion path. So many moving objects are concentrated into a background image frame. Finally a series of such images can be obtained along the time axis and a single camera surveillance video synopsis is formed by these images. The experimental results show that our detection method of moving object is superior to those methods such as Gaussian mixture model, background subtraction, frame difference method and Li's method. The video synopsis is concise and it can greatly improve the efficiency of browsing, querying or previewing.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

Support for this study is provided by the university and college independent innovation project of Jinan Science and Technology Bureau (201202002), Shandong province development project of science and technology (2013GGX10131)

and Shandong Provincial Key Laboratory of Intelligent Building Technology.

REFERENCES

- [1] Y. Pfitch, A. Rav-Acha, S. Peleg, "Nonchronological video synopsis and indexing," *IEEE Trans. PAMI*, vol. 30, no.11, pp.1971-1984, 2008.
- [2] Y. Pritch, A. Rav-Acha, A. Gutman, and S. Peleg, "Webcam Synopsis: Peeking Around the World," in 11th International Conference on Computer Vision, pp. 1-8, 2007.
- [3] A. Rav-Acha, Y. Pritch, and S. Peleg, "Making a Long Video Short: Dynamic Video Synopsis," in IEEE Conf. Computer Vision and Pattern Recognition, pp. 435-441, 2006.
- [4] Teng Li, "VideoM: Multi-Video Synopsis," in IEEE International Conference on Data Mining Workshops, pp. 854-861, 2008.
- [5] Y. Pritch, "Clustered Synopsis of Surveillance Video," in 6th IEEE International Conference on Advanced Video and Signal Based Surveillance, pp. 195-200, 2009.
- [6] U. Vural, Y.S. Akgul, "Eye-gaze based real-time surveillance video synopsis," *Pattern Recognit Lett*, vol. 30, no. 12, pp. 1151-1159, 2009.
- [7] U. Damjanovic, "Event Detection and Clustering for Surveillance Video Summarization," in IEEE Ninth International Workshop on Image Analysis for Multimedia Interactive Services, pp. 63-66, 2008.
- [8] Xu min, "Research of video synopsis technology based on object motion information," M. S. thesis, University of Science and Technology of China, Hefei, China, 2009.
- [9] J. Nascimento and J. Marques, "Performance evaluation of object detection algorithms for video surveillance," *Multimedia, IEEE Trans*, vol. 8, no. 4, pp. 761-774, 2006.
- [10] W. Hu, T. Tan, L. Wang, and Steve Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. Syst, Man, Cybernet – Part C: Appl. Rev.*, vol. 34, no. 3, pp. 334-352, 2004.
- [11] M. Antony, J. Anitha, "A survey of moving object segmentation methods," *Int. J. Adv. Res. Electron. Commun. Eng. (IJARECE)*, vol. 1, no. 4, pp. 73-80, 2012.
- [12] L. Li, W. Huang, I.Y.H. Gu, and Q. Tian, "Foreground object detection from videos containing complex background," in Proc. ACM Int. Conf. Multimedia, Berkeley, CA, pp. 2-10, 2003.
- [13] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut-interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph*, vol. 23, no. 3, pp. 309-314, 2004.