Applying the Video Summarization Algorithm to Surveillance Systems

Yi-Nung Chung, Tun-Chang Lu, Ming-Tsung Yeh, Yu-Xian Huang, and Chun-Yi Wu
Department of Electrical Engineering, National Changhua University of Education, Changhua City, Taiwan
Email: ynchung@cc.ncue.edu.tw

***Abstract*—Surveillance systems are very popular today. Efficiently access the record image of surveillance system is an important research goal. This paper proposes a simple method to enhance the video image summarization. The video summarization production process in this paper is divided into intrusion detection of moving objects and time axis with animation output. The purpose is to improve the surveillance access efficiency of documentary video. In moving target detection section, a series of image processing method is applied for processing images. It is recognized as an invasion of moving object image storage. It will save the image production time code and mark the top left corner in the image. Finally, the summarization of both WMV and AVI video files are conducted in this paper.**

***Index Terms*—digital image processing, surveillance system, video summarization, image processing, time code**

I. Introduction

The development of surveillance system and digital image processing techniques are more popular day by day. They are almost in everywhere, such as public institutions, private shops, companies, and general houses. The evolution of surveillance system is very fast. It includes the digitization, network, wireless and the cloud computing. The home surveillance system even can be accessed by the mobile phone apps [1]. If it is necessary, such as the accident is happened, someone needs to check the video. It is usually need a lot of time to watch the video. So, the efficiency of examining the surveillance records is an important part of the surveillance system. The desired period of video maybe just a short time, but the total surveillance contents are several tens of hours in the video. A long period of video is useless but user has to spend a lot of time to search the desired period. The total recording video takes a lot of capacity of hard disk and the transfer time will be longer also. This paper proposes a new method for recording video by using image processing to summarize the video contents [2], [3]. It can save a lot of time and capacity of hard disk by the proposed video summarization algorithm. In order to verify the effectiveness of the proposed method, the image recognition rate and accuracy of the time axis are simulated by using the Matlab software to accomplish the above methods.

Manuscript received February 1, 2015; revised June 10, 2015.

II. Research Methods

A. Moving Objects Detection

Fig. 1 represents the flow diagram of the proposed method. First, based on total time of all frames to calculate the frames per second (FPS). Second, the system uses temporal differencing to produce the difference between two frames and convert image to binary. Third, it refines the binary image by morphology processing [4], [5]. Final, the status of image can be determined in the system.

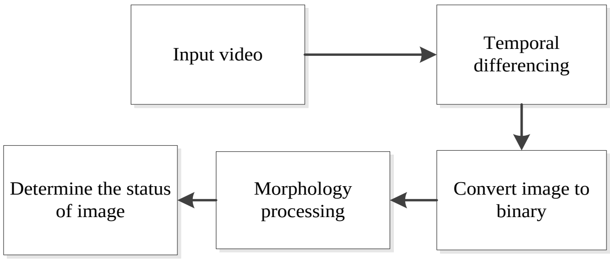

Figure 1. Flow diagram of invasion objects detection.

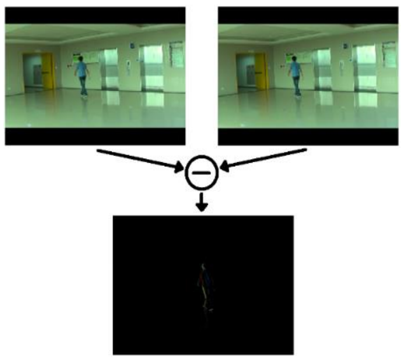

Figure 2. Temporal differencing result.

Temporal differencing [6], [7] is the fastest way to find the difference between two images. If the difference of pixel is under the criterion, which means this pixel is not a moving object. The difference of pixel is above the criterion, which means the pixel is a moving object. This method is mainly used to detect image changes in the video. It has good performance for adaptability of environment and does not interfere with brightness change. Its disadvantage is that the images need to further processing for more complete foreground image. The Fig. 2 represents the temporal differencing result and it can

doi: 10.18178/joig.3.1.20-24

get the uncompleted contour. Compared to the background subtraction which needs to establish a set of background model for comparison with the current image, temporal differencing is not limited to the default background and too many complex operations. Therefore, this paper uses a temporal differencing to detect a moving object. The algorithm will improve the environment and reduce the fitness of processing speed to increase its usability.

The frequency of temporal differencing will affect the recognition accuracy and efficiency of the overall system. Higher processing frequency can improve accuracy, but it will increase the computational load of the system and reduce overall system speed. Otherwise, the lower frequency of the image with no excessive detection and reduce recognition accuracy rate. But it will significantly reduce the amount of computation system, and increase overall system speed. As shown in Fig. 3, the FPS is 30 but the processing frequency of the system is 3Hz. It is 90% reduction in the amount of calculation. As shown in Fig. 4, the FPS is 30 but the processing frequency of the system is 30Hz. It is very accurate but the computation complexity is too high and overall efficiency is poor.

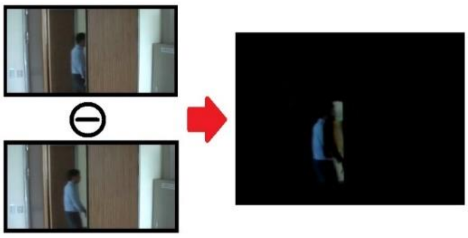

Figure 3. 3Hz Temporal differencing image.

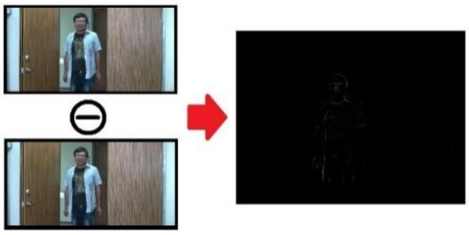

Figure 4. 30Hz Temporal differencing image.

Due to the temporal differencing has better adaptation of the brightness than background subtraction, therefore it does not need other color space to remove the shadow to improve the accuracy of the image [4], [8]. This paper uses the RGB color space [9]-[10], because the RGB image has faster processing speed. An RGB image is converted according to Equation 1 which points per pixel gray scale image into between 0-255. Then turn to binary image based on the threshold. In this paper the threshold is chosen 0.375, which represents when the pixel gray scale less than 96 is set to 0 that is black. Otherwise, the pixel higher than 96 is set to 1 that is white. As shown in Fig. 5, this is a RGB image which after temporal differencing. Then the system converts the RGB image to the binary image and it can get the black-white image.

$$\text{Gray} = (R * 0.299) + (G * 0.587) + B * 0.114) \quad (1)$$

Figure 5. RGB image transfer to binary image after temporal differencing.

The threshold must be chosen moderately. The lower threshold value is too sensitive to environmental that causes excessive noise, otherwise, the higher threshold value causes the object is unobvious, which is shown in Fig. 6.

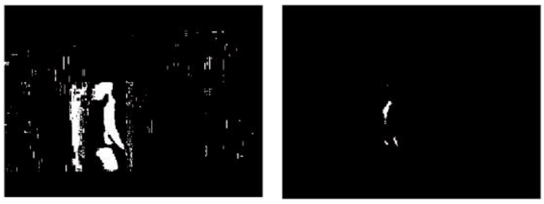

Figure 6. The image results for too low or too high threshold.

Morphology processing is an important part of image processing. The basic idea is using a set of structural elements to get the massage in the image. It can apply the geometry to change the shape, size and finally retain the basic shape of the characteristics of the image. The basic operation is dilation, erosion, closing and opening. As shown in Fig. 7, the small pieces will be connected to a solid block and it can get the foreground contour.

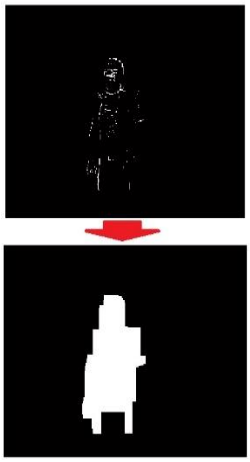

Figure 7. Morphology processing.

B. Determine the Status of Objects

The flow diagram of determining the status is shown in Fig. 8, the total number of the white pixels is X. X divided by the resolution Y to get Z. If Z is greater than the threshold that means moving objects in current frame, otherwise, if it is lower than threshold that means no moving objects in current frame. After determination, it can save the frame which has moving objects, and remove the other frame. It stores the image and marks the value of N which is the number of frame in the video.

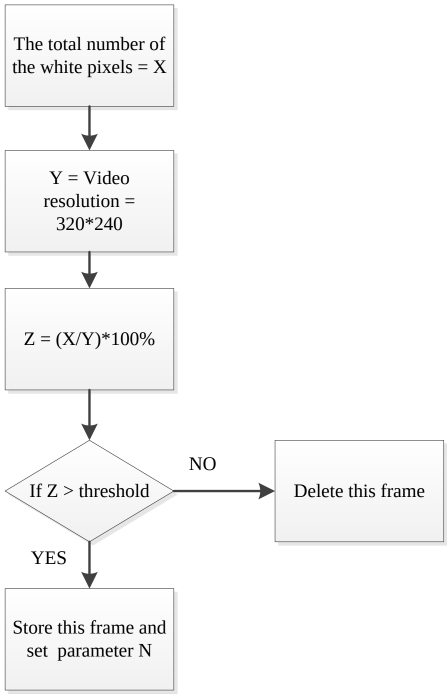

Figure 8. Flow diagram of determining the status.

III. The Time Axis and Output of Animation

The main idea of time axis and the output of animation are to save fragments of determined moving objects, which are called abstract image and giving time codes in the original video in order to access it. As shown in Fig. 9, first, it reads the stored frame of determined invasion of the moving object in the previous paragraph, calculates the time axis based on each stored frame tag N, and displays in the video. Then it products a video by using the principle of making animation and perform the same set according to the original video's FPS. Finally, it gets an AVI video file which contains the time axis and video summary.

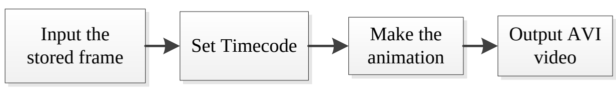

Figure 9. The flow chart of time axis and the output of animation.

A. Time Code

The determination of the time axis is depended on the value of N in the original frame, which is shown in (2). X for Xth seconds and N for Nth frames in the original video. For example, a video is up to 1000 seconds and it is 30 FPS, which represents the video contains 30000 frames. The frame appears at the 500th in the original video if N is 15000.

$$X(\text{sec}) = \frac{N}{Number\ of\ frames} \times Number\ of\ seconds \tag{2}$$

Time axis aims to access the original video as well as to confirm the time point conveniently. It can clearly indicate the time point of the abstract image. As shown in Fig. 10, the time axis is displayed in upper left corner in the video. H means hours, M means minutes, and S means seconds.

Figure 10. The schematic diagram of time axis.

B. The Process of Making Animation

The advantages of Audio Video Interleave (AVI) is that the process of such format only describe the data structure of frames in a video, no specific side compression code to convert any types of format, that is why we use such format of a video. As shown in Fig. 11, there are 20 continuous frames and the black screens means no moving objects in those screens. If the original video is 20 FPS, we will also set the output video to 20 FPS, so we can know that the process starts from the frame tagged m3321 at 0h 2m 46.05s and ends at 0h 2m 46.8s.

Figure 11. The animation pane.

IV. Experimental Results

Because the final results of this experiment are AVI video files and each image file final summary ranging from 10 minutes to 30 minutes. It will be split up to over ten thousand pictures, so only take out a part of the summary images to verify the proposed method. The experimental results include two parts which are image recognition rate and accuracy of the time axis, respectively.

A. Image Recognition Rate

In the experiment the system uses both Windows Media Video (WMV) files and AVI files. When a moving object enters the picture, it starts abstract images and records. The video 1 is a WMV file shown as Table I in which there are eight screens were lost. The video 2 is an AVI file shown as Table II and the correct rate of abstract images recorded quite well.

TABLE I. VIDEO 1

Original film format	Original film length	
320*240 WMV	02 hr. 48min. 23sec.	
Starting time code	Ending time code	FPS film
02:27:29:05	02:27:31:08	22.
Correct frames	40	
Incorrect frames	8	

TABLE II. VIDEO 2

Original film format	Original film length	
176*144 AVI	01 hr. 07min. 53sec.	
Starting time code	Ending time code	FPS film
00:24:17:11	00:24:18:28	30
Correct frame	48	
Incorrect frame	0	

B. Time Axis Correct Rate

In this experiment the system still uses both WMV files and AVI files. When a moving object enters the picture, it starts abstract images and records. The video 3 is a WMV file shown as Table III in which there are some errors for time axis. The video 4 is an AVI file shown as Table IV and the time axis is exactly correct.

TABLE III. VIDEO 3

Original film format	Original film length	
320*240 WMV	02 hr. 48min. 23sec.	
Original film time code	Summarization Time code	FPS video
00:47:08:0	00:53:28:21	22.07
Time error	+ 00:06:20:00	

TABLE IV. VIDEO 4

Original Film format	Original film length	
176*144 AVI	01 hr. 07min. 53sec	
Original film time code	Summarization time code	FPS video
00:26:43	00:26:43:15	22.07
Time error	+ 00:00:00:00	

C. Experiment Analysis

The statistical results of moving objects detection on both WMV and AVI files are shown in Table V. The WMV video files have total 222,910 scenes and it has 10212 false pictures. The image recognition rate of WMV files is 95.41%. AVI video files have total 122,190 scenes and it has 6996 false pictures. The image recognition rate is 94.30%. The time axis errors for both WMV and AVI files are shown in Table VI. The time axis error in the WMV file is from 0 second to 6 minutes and 42 seconds. The time axis error of AVI file is always zero.

TABLE V. VIDEO 5

	WMV Format	AVI Format
Entire Frame	222910	122190
Correct Frame	212698	115224
Incorrect Frame	10212	6966
Recognition Rate	95.41%	94.30%

TABLE VI. VIDEO 6

	WMV Format	AVI Format
Maximum Time Error (Sec)	402	0
Minimum Time Error (Sec)	0	0

V. CONCLUSIONS

In order to enhance the efficiency of the video image surveillance, this paper proposed a summarization method which can detect the moving object and summarize such images. This algorithm can help users to check whether it has moving objects during a long time period. It can save a lot of time and computer capacity. Based on the experimental results, the detection rates for both WMV and AVI video image are over 94.3%. The time axis is 100% correct for AVI files and the accuracy of time axis of WMV is 93.8%.

ACKNOWLEDGMENT

This work was supported by the National Science Council under Grant MOST 103-2221-E-018-017-

REFERENCES

[1] H. Yang, A. C. Kot, and X. Jiang, "Binarization of low-quality barcode images captured by mobile phones using local window of adaptive location and size," *IEEE Trans. on Image processing*, vol. 21, no. 1, pp. 418-425, Jan. 2012.

[2] Z. Zhou, X. Chen, *et al.*, "Activity analysis, summarization, and visualization for indoor human activity monitoring," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1489-1498, Nov. 2008.

[3] A. H. Meghdadi and P. Irani, "Interactive exploration of surveillance video through action shot summarization and trajectory visualization," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 12, pp. 2119-2128, Dec. 2013.

[4] Y. X. Huang, G. W. Jhao, Y. N. Chung, and C. H. Hsu, "Application of image processing technology in detection and tracking of vehicles on the road," *ICIC Express Letters*, vol. 8, no. 5, pp. 1301-1305, May 2014.

[5] J. S. Kim, D. H. Yeom, and Y. H. Joo, "Fast and robust algorithm of tracking multiple moving objects for intelligent video surveillance systems," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 5, Aug. 2011.

[6] H. Yoo, U. Yang, and K. Sohn, "Gradient-Enhancing conversion for illumination-robust lane detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 3, Sep. 2013.

[7] C. Y. Chang and C. H. Lin, "An efficient method for lane-mark extraction in complex conditions," in *Proc. IEEE International Conference on Ubiquitous Intelligence and Computing*, Sep. 2012, pp. 330-336.

[8] G. Wang, D. Xiao, and J. Gu, "Review on vehicle detection based on video for traffic surveillance," in *Proc. IEEE International Conference on Automation and Logistics*, Sep. 2008, pp. 2961-2966.

[9] G. H. Na, K. S. Shim, K. W. Moon, S. G. Kong, E. S. Kim, and J. Lee, "Frame-Based recovery of corrupted video files using video codec specifications," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 517-526, Feb. 2014.

[10] V. D. Silva, H. K. Arachchi, E. Ekmekcioglu, and A. Kondoz, "Toward an impairment metric for stereoscopic video: A full-reference video quality metric to assess compressed stereoscopic video," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3392-3404, Sep. 2013.

Yi-Nung Chung received the B.S. degree in electrical engineering from National Cheng Kung University, Tainan, Taiwan, in 1981, the M.S. degree in electrical engineering from Lamar University, Beaumont, TX, in 1986, and the Ph.D. degree from the Department of Electrical Engineering, Texas Tech University, Lubbock, TX, USA, in 1990. He is currently a Professor in the Department of Electrical Engineering, National Changhua University of Education, Changhua, Taiwan. His research interests include signal processing, image processing, renewable energy, and solar energy system.

Tun-Chang Lu received the B.S. degree in electrical engineering from Tamkang University, Taipei, Taiwan, in 2009, the M.S. degree in electrical engineering from the Tamkang University, Taipei, Taiwan, in 2011. He is currently working toward the Ph.D. degree with the Graduate Institute of Electrical Engineering, NCUE. His research interests include image processing, signal processing, and target tracking.

Ming-Tsung Yeh was born in Changhua, Taiwan, in 1968. He received the B.S. degree in information management from the Da Yeh University, Changhua, Taiwan, in 2000 and the M.S. degree in electrical engineering from the National Changhua University of Education (NCUE), Changhua, Taiwan, in 2012. He is currently working toward the Ph.D. degree with the Graduate Institute of Electrical Engineering, NCUE. His research interests include image processing, object tracking, and pattern recognition.

Yu-Xian Huang received the B.S. degree in industrial education and technology from National Changhua University of Education (NCUE), Changhua, Taiwan, in 2008, the M.S. degree in electrical engineering from the NCUE, Changhua, Taiwan, in 2011. He is currently working toward the Ph.D. degree with the Graduate Institute of Electrical Engineering, NCUE. His research interests include image processing, digital signal processing, and object tracking.

Chun-Yi Wu received the B.S. degree in electrical engineering from National Changhua University of Education (NCUE), Changhua, Taiwan, in 2011, the M.S. degree in electrical engineering from the NCUE, Changhua, Taiwan, in 2014. His research interests include digital image processing and video summarization.