



Information Mining on Indoor Surveillance Video

Yu Ming and Liu Wei

School of Computer Science and Engineering

Hebei University of Technology

Tianjin, China

E-mail: yueeming@126.com, lw306@163.com

Abstract: Intelligent video monitoring system has been used widely in daily life. In order to avoid the casualties, as well as to predict potentially dangerous situations, real-time monitoring of crowd activities indoor has become an urgent requirement. In recent years, the video-based passenger flow counting systems have been improved a lot with constant equipment update. However, these systems are mostly aimed at a particular scene. In crowded circumstances, the statistical precision is not very satisfactory. The analysis of abnormal situations is thus imperfect. This paper improves the main algorithms, such as: the extraction of human, the segmentation of crowd and the judgment of human moving direction. Finally, the improved algorithms integrate a system which achieves two-way passenger flow counting. Consequently it first makes the early warning of abnormal conditions possible. The system's statistical accuracy is remarkably improved, and testing result under different scenes is shown.

Keywords: passenger flow counting; people conglutination; movement history image; virtual line; information mining

1. Introduction

In the real world, there are many signals that we can perceive which can help us fully understand the surrounding environment. Visual signal plays an important role among these signals. Human and the other higher animals have well-developed visual organs. They can identify the target, and trigger further processing. For humans, this function is innate. However it is a very difficult task for the machine vision systems [1].

A complete intelligent video processing system consists of two interrelated and distinctive processes. The first is extraction about the prospect of moving object and the second is intelligence judgment of the conduct of detection target. For indoor surveillance video, the main target of intelligent video information mining system for analysis and processing is passenger. Therefore, the analysis and

statistics of the passenger are the key issue of such video information mining technology.

This paper researches video information mining technology through the analysis of indoor surveillance video. We target at passenger flow counting, crowd density estimates and abnormal behavior detection.

2. Detection of moving foreground in indoor surveillance video

2.1 De-noising of indoor surveillance video

In indoor surveillance video, noise is inevitable because of a variety of inherent reasons. Generally, noise can be divided into external noise and internal noise based in terms of causes. External noise refers to the noise generated by the external disturbances that enter the system, such as the noise caused by lightning. Internal noise is defined as that caused by

equipment or materials within the system, such as the noise [2] caused by interface loosening.

To date, there have been a lot of video sequence de-noising methods. In a variety of algorithms, according to approach and scope, de-noising methods can be divided into time domain de-noising algorithm, spatial domain de-noising algorithm and time-space domain de-noising algorithm [3].

Given the characteristics of indoor surveillance video and such factor as real-time, Gaussian filtering method is selected as the video de-noising processing method in this paper. Gaussian filter is a low-pass filter which is mainly used to filter out the noise and smooth the image, whose formula is as follows (2.1):

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2.1)$$

De-noising results shown in Figure.1



(a) Original video.



(b) De-noise result.

Figure.1 De-noise processing.

2.2 Shadow processing

2.2.1. Shadow generation

A shadow is often cast in indoor surveillance because of the block between the moving objectives and indoor uneven illumination or other reasons. Usually the shadow can be divided into two categories; one is called the umbra and the other is a moving shadow. Umbra is the moving target which has not been light shined on its areas. Moving shadow is the goal of movement of the projection in the background.

In the moving target detection, the umbra takes little effect on the extracting of the moving targets. However, the moving shadow can be easily mistaken for the former spot, which has a serious impact on the moving object segmentation and extraction. Therefore, in order to further enhance the detection accuracy of moving prospects target, we must first remove the shadow of the moving target to reduce the effects of the shadow on the target detection and segmentation.

2.2.2. A shadow detection method based on HSV space

Covered by the shadow, the brightness of the background will be reduced and become darker. The color of the background points covered by the shadow will become darker. Based on this feature, we can consider the image analysis by using color information change to analyze the moving objects and moving shadows. As the HSV color space will reflect the full color feature more clearly than RGB color space, this paper chooses to take into account the color features of shadow in HSV color space. In HSV color space, H means hue, S means saturation and V brightness. Based on a large number of experimental results, researchers found that the change of background pixels' hue and saturation covered by the shadow is not particularly large and the brightness will be slightly lower. Therefore, we can take into account the shadow of change with V component and the characteristics of H component to distinguish the

shadow with the prospect of moving targets and effectively improve the accuracy of target detection.

The formula of RGB space to HSV space is as follows (2.2):

$$V = \frac{1}{3}(R + G + B)$$

$$S = 1 - \frac{3}{R + G + B}[\min(R, G, B)] \quad (2.2)$$

$$H = \arccos\left\{\frac{[(R - G) + (R - B)]/2}{[(R - G)^2 + (R - B)(G - B)]^{1/2}}\right\}.$$

2.3 Detection of moving foreground in surveillance video

This paper discusses indoor surveillance video information processing system which is a fixed real-time monitoring of video capture device. The objective is to separate the background and moving target, and then analyze them. An improved subtraction background method brings better results.

Some extraction methods we usually take are as follows: optical flow, background subtraction, frame difference and wavelet method. A variety of detection methods have certain limitation in such applicable occasions as the detection of objects. In certain conditions, some algorithms can detect rigid body accurately, but their results are not ideal in other condition.

Optical flow method [4], for example, is an important method of motion image analysis. This method can be used in variational background. However, its arithmetic is complicated and the iterative process is so slow that it is inapplicable to real-time application.

Background subtraction [5] is a useful method for moving objects detection. The basic idea of this method is a subtraction of background model which is stored beforehand or observed in time from the current frame. And it gives movement area by threshold operation. In such circumstances arise such problems as the abstraction and update of background model, the effect of background disturbance, illumination changes and shadows. This method is

thus inapplicable to moving objects detection in motion background occasion.

Frame difference [6] makes use of the continual inter-frame difference to extract the moving object. Frame difference method has strong self-adaptability in dynamic environment but it could not extract all correlative characteristic pixels [7]. And this method also brings inanition phenomenon in moving entity's interior.

The wavelet method is mainly applied to detect the smaller moving object in the complex scenes. However, this detection algorithm is not only tedious but also for more application experience.

As can be seen from the above description, a variety of moving object detection algorithms have their own advantages and disadvantages. Therefore, in practical application, we should take different occasions into consideration, and avoid weaknesses to select the appropriate detection algorithm which will reduce the negative effects and improve detection accuracy.

In application, the scene of passenger flow detection is usually selected at the entrance or the main passenger flow channel in buildings. The scene mentioned above belongs to inner category whose illumination is stability. The camera of our system shots vertically, with the camera sitting above the entrance, which is easy to install and can avoid occlusion. From the above characteristics, we can see our system improves the background subtraction method. We use weighted update image to update the background. Accumulator is defined by (2.3).

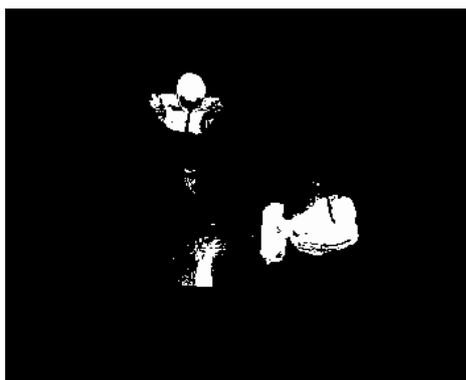
$$a(x, y) = (1 - u) \times a(x, y) + u \times Image(x, y) \quad (2.3)$$

Image(x, y) is the pixel value of input image. The letter of u means parameter of regulation update speed. It is defined as the rate of accumulator update pixel. If u chooses a smaller value, illumination and other illusion have a little effect on the result. Nevertheless the effect of such illusion will stay on the background for a long time. If u chooses a larger value, illusion takes a heavy effect for an instant. However, the effect disappears quickly when illusions disappear. The update rate is set to 0.003 according to the system which is used indoor. The

background of video changes little with the camera fixed and this algorithm is in smaller operational volume. The system could ensure the accuracy of the extraction of human activities by using this algorithm. Fig.2 (b) is the final extraction result of Fig.2 (a).



(a) Original image.



(b) Extraction result.

Figure.2 Human extraction.

3. Segmentation of moving foreground in indoor surveillance video

3.1 Segmentation algorithms of moving foreground

Image segmentation is one of the problems in computer vision research [8]. The purpose of image segmentation is to segment the image into meaningful regions. Image segmentation is the basis for further analysis and processing of the image [9].

In the second chapter, the algorithm has extracted the passengers in the image but they did not get the number of passengers. To get this

information is the goal to be reached.

At present, widely used segmentation methods are template matching method and artificial neural network method and so on.

Template matching method can get a correlation value by comparison operation through several standard templates image and matching image. The traditional template matching method involves collecting each pixel in the template. So this is a tedious method that can not segment and recognize the image quickly.

Artificial neural network is a way to simulate the human mind. Artificial neural network must learn a certain criterion before it works. However, human body information is more complex, so artificial neural networks should take a very long training time. Therefore, the accuracy of segmentation is not high by using this method and it is inapplicable to real-time application.

There is no universal theory for moving object segmentation, so we should adopt effective approach according to the specific situation. For the indoor surveillance video discussed in this paper, the images are collected by the camera from the vertical angle. The collection has almost no skin color characteristics. In recent years, it is proposed to use shape analysis method, but the test result is not ideal.

3.2 Distance transform algorithm

According to statistics, the crowd conglutination includes two patterns: (1)Following; system takes two passengers as a single person when they are very close to each other (2)Crisscross; the system usually drain logging easily when crowds pass the monitor area from different directions. Therefore, the accurate segmentation of the conjunction crowd into individuals is the key task. To better the situation, an improved horizontal distance transformation algorithm is used in the system.

Distance transformation algorithm [10] is a kind of operation for binary image. Its purpose is to translate a binary image into a gray one. In the

improved image, the gray level of every pixel represents the relative distance to the nearest background pixel. To calculate the nearest distance between foreground and background pixel we need to operate the whole image entirely. In other words, we must compute the distance between every foreground pixel and all background pixels to find the minimum value. The whole operations are so heavy that if every image size is very big there will be many foreground moving objects. However, computation about the minimum distance of all pixels will affect the efficiency of the real-time system. Our method calculates the distance between foreground pixel and nearest horizontal neighbor background pixel, with the main conglutination part being human limbs. The first scanning is from left to right and we compute the distance between foreground pixel and left margin. We compute the distance between foreground pixel and right margin when the second scanning is from right to left, then we compare it with the first results to get the minimum value. We can get hold of approximate distance image by scanning twice. This method reduces the operational volume of the system and ensures the real-time application. The scanning process is given in Fig 3.

0	0	1	1	1	1	1	1	1	0	0
---	---	---	---	---	---	---	---	---	---	---

(a) The original binary image

0	0	1	2	3	4	5	6	7	0	0
---	---	---	---	---	---	---	---	---	---	---

(b) The first scan from left to right

0	0	1	2	3	4	3	2	1	0	0
---	---	---	---	---	---	---	---	---	---	---

(c) The second scan from right to left

Figure.3 Scanning process.

According to the statistics, the gray peak values in crowd distance image concentrate in a certain range. We can process the image with threshold method. This method compares peak value with detection value to determine whether the target belongs to the person or not and avoid the disturbance from other targets. It can segment crowd image to individual image. Fig.4 (b) is the final segmentation result of Fig.4 (a).



(a) Original image.



(b) Segmentation result.

Figure.4 Processing result.

4. Passenger flow counting in indoor surveillance video and abnormal behavior detection

4.1 Passenger flow counting

Passenger flow counting (also called as client flow statistics or statistics on the passenger number) [11] is an important application in the indoor surveillance video field. China's rapid economic development booms all kinds of public buildings, large-scale public activities and hence grows crowd density all the time. How to get real-time monitoring in a large group activity and ensure the safety of human activities have become an important issue. In a spacious indoor place, accurate statistics on the number of passengers and accurate estimates of crowd density become the focus of attention.

Passenger flow counting system requires clear

classification of the moving directions of the people who pass the monitoring area to achieve the accuracy of passenger flow counting. Template matching method or detection line method is usually used in the system. However, the template matching is in large operational volume and inapplicable to real-time application, which achieves a good result to rigid-targets only. This method is not applicable for human target tracking, because human image changes a lot. The detection line method has a good result to one-way passenger flow counting but it is not applicable to two-way system.

Our system improved the movement history image method [12] by adding the virtual detection region and the determine direction algorithm. Because the interval of two adjacent frames is very short, the movement distance of people is very short and so the transferred image changes a bit. By using centroid, this method achieves a good result in this case. This algorithm is not in large operational volume, so the system can track the location of the target in real-time mode. The continuous multi-frames are analyzed by the use of video image resource in this method which is in high efficiency.

The movement history image method makes use of the relativity of target outline in continuous image and superposes continuous image into the final movement history image through every image in different time. We can receive partial derivatives toward x and y directions and compute grads direction $O(x, y)$ according to it when the movement history image has been obtained. The formula is as follows (4.1):

$$O(x, y) = \arctan\left(\frac{S_y(x, y)}{S_x(x, y)}\right) \quad (4.1)$$

We assume that the input image is as $f(x, y)$ (4.2):

$$f(x, y) = \begin{cases} 1 & \text{target} \\ 0 & \text{background} \end{cases} \quad (4.2)$$

We can calculate the target's centroid; and the formula is as follows (4.3), (4.4):

$$X_0 = \frac{\sum x \times f(x, y)}{\sum f(x, y)} \quad (4.3)$$

$$Y_0 = \frac{\sum y \times f(x, y)}{\sum f(x, y)} \quad (4.4)$$

To determine the direction of human motion, steps are as follows:

(1) To determine whether there are people or not in the current frame monitor region.

(2) Return the above step if there is nobody, or compute the target's centroid if there are some people.

(3) Compute the moving direction of people and make judgment about the condition of people.

(4) Cumulate the number of people with two directions and count the passenger flow in this scene.

4.2 Crowd density estimation

In order to take full advantage of the information contained in video and reduce the dangers with high dense of indoor passengers, the system measures the density of the crowd to do real-time monitoring, and analyzes in advance the insecurity that may arise to alarm early warning.

Generally speaking, indoor surveillance video collection cameras are supposed to be installed at the top of a relatively high position. First, it can guarantee the security of collection devices. Second, there are very little occlusion phenomenon of the human body in the video data collection. In its real-time operation, the system estimates the crowd density through comparison of the number of pixels, which is obtained by estimating the value of crowd density using the ratio of foreground pixel and background pixel. After the analysis of different experiments, these methods can achieve accurate results.

4.3 Analysis of abnormal behavior

The analysis of abnormal behavior using indoor surveillance video is a part of artificial intelligence

field, which is achieved through imitation the process of human analysis. If the camera and the coding transmission system can function as the human "eyes", the video analysis system then serve as "human brain".

Video analysis techniques can establish the mapping relationship between images and images description so that the computer system can carry out image processing and analysis to understand the implied video information. In this way, the computer system will be able to "understand" the video content and feedback to the monitoring officer timely. For example: the operator can set parameters and rules in dangerous condition before the computer analysis and processing of visual information. The computer matches the analysis result and rules. If the analysis result is fully consistent with the rules, the surveillance system can prompt alarm.

The system can analyze many dangerous conditions, such as illegal detention, reversing and so on.

5. Experiment result and analysis

The experimental platform of this system is Intel Pentium 2.4 GHz, 1G memory and Windows XP operation system.

In the train station, library, supermarket and many other occasions, there are commonly two passageways including entrance and exit in order to facilitate smooth passenger flow and guarantee passenger safety. However, for some reasons, there are some conditions where normal passenger flow is impeded. Therefore, we can install the video capture devices at one-way channel, using the video mining system for real-time analysis of passenger flow information and making a timely warning in the case of dangers. Through that monitoring, the working intensity of the staff can be reduced and the utilization of indoor surveillance video will be improved.

Processing steps that can help judge passengers' backward going behavior are as follows:

- (1)To set the direction of one-way channel.
- (2)To detect moving object and moving direction.

(3)To judge whether the passengers are moving backward or reversing and send warning messages.

Through analysis of indoor surveillance video in the libraries, video mining system can accurately judge if there are readers going backward and send warning message immediately.

The system in this paper uses surveillance video in the main passage of a building for experiment. The experimental results are shown in table 1. According to the analysis of the results, this system has stronger robustness and accuracy in different occasions and scenes.

Table 1. Experimental results

Video	ingoing person	outgoing person	detection value of ingoing person	detection value of outgoing person	average accuracy rate
scene1	57	45	51	48	91.4%
scene2	42	45	45	39	89.8%
scene3	36	42	33	39	92.2%
scene4	46	32	40	38	85.1%
scene5	63	67	58	60	90.5%
scene6	56	50	49	45	88.7%
scene7	39	32	34	28	87.3%

In some important areas, there are many surveillance cameras aiming at monitoring the passengers to prohibit their illegal residence. Video mining system analyses the real-time surveillance video information, and it can detect who act illegally in time.

Processing steps that judge passenger's illegal residence behavior are as follows:

- (1)To set up an area with virtual detection line that does not allow passenger access.
- (2)To detect the moving object by scanning this area.
- (3)To send warning message when mining system finds passenger access to this area.

Through analysis of many surveillance video on computers the experimental results show that the

operation of this treatment method is simple and the result is accurate, which effectively meet the strong real-time request of the system.

6. Conclusion

In this paper, there is a detailed analysis of the preprocess of indoor surveillance, including the noise processing, shadow elimination, moving foreground extraction, segmentation and other operations. At last, we designed for indoor surveillance video of the passenger flow statistics analysis system with the improved algorithms. We made a lot of tests in different scenes to reach the results, which in the end shows that this approach can segment multiple moving targets in video efficiently. The conglutination problem was solved to a certain extent. And this system has fast processing speed and high accuracy. Besides, the system has strong robustness in different scenes.

Acknowledgments

This work was supported by the Key Project Program (2009) from Hebei Provincial Education Ministry.

References

- [1] Xiong Zhihua, Study on Machine Vision System Based on DSP, Chongqing University, Master of Science thesis, 2005.
- [2] Liu Wei, Video Denoising Based on Non-local Means, Xidian University, Master of Science thesis, 2009.
- [3] Shen Weiming, Guo Xiaoyun, Yang Lidong, Real-time Video Denoising Arithmetic Based on Adaptive Multi-layers Background, *Computer Engineering and Applications*, Vol.44, pp.230-234, 2008
- [4] D. Meyer, J. Denzler, and H. Niemann, "Model Based Extraction of Articulated Objects in Image Sequences for Gait Analysis", In: *Proc. of IEEE International Conf. On Image Processing*, Santa Barbara, California, pp.78–81, 1997.
- [5] I. Haritaoglu, D. Harwood, and L. Davis, Real-time Surveillance of People and Their Activities, *IEEE Trans Pattern Analysis and Machine Intelligence*, Vol.22, pp.809–830, 2000.
- [6] A. Lipton, H. Fujiyoshi, and R. Patil, Moving Target Classification and Tracking from Real-time Video, *Proc IEEE Workshop on Applications of Computer Vision*, Princeton, NJ, pp.8–14, 1998.
- [7] Jia Yunde, Machine Vision, Science Press, Beijing, pp.235–244, 2000.
- [8] Bao Qingjie, Video Motion Targets Detection and It's Application in Video Compression, Zhejiang University, Master of Science thesis, 2008.
- [9] Yao Guozheng, Wang Yunjiu, D.Marr and Visual Computing Theory, *Overseas Automatic*, pp.55-57, 1984.
- [10] M. Lalonde, M. Beaulieu, and L. Gagnon, Fast and Robust Optic Disc Detection Using Pyramidal Decomposition and Hausdorff-based Template Matching, *IEEE Transactions on Medical Imaging*, Vol.20, pp.1193–1200, 2001.
- [11] Zheng Shibao, Video Surveillance Technology and Market Trends. *Video Engineering*, Vol.32, pp.37-41, 2008.
- [12] Wang Liang, Hu Weiming, and Tan Tieniu, A Survey of Visual Analysis of Human Motion, *Chinese Journal of Computers*, Vol.25, pp.225–235, 2002.