Hierarchical Image Gathering Technique for Browsing Surveillance Camera Images

Wataru Akutsu, Tadasuke Furuya, Hiroko Nakamura Miyamura, and Takafumi Saito

Graduate School of Bio-Applications and Systems Engineering,
Tokyo University of Agriculture and Technology
wataru@vc.base.tuat.ac.jp

Abstract. We propose an image gathering and display method for efficient browsing of surveillance camera images. The proposed method requires large cost to inspect lengthy image sequences taken by a surveillance camera. The proposed method involves generating a still image by gathering the moving parts from image sequences captured by a fixed camera. The gathered images are generated for several intervals of time and are displayed hierarchically. A user can easily browse the scene by observing the images with moving parts. Since detection and recognition of the target objects are performed by a human operator, efficient and reliable browsing can be established.

Keywords: gathered images, temp oral resolutions, hierarchical display.

1 Introduction

In recent years, the surveillance camera has been used extensively in crime prevention, and surveillance cameras have become ubiquitous. As such, the costs associated with monitoring and/or browsing captured images has increased greatly. For efficient use of surveillance cameras, it is necessary to reduce these costs. Surveillance cameras are used in various situations. For example, the view angle may be fixed or variable (i.e., panned or zoomed). The amount of traffic at the capture location may be significant or slight. The image may be monitored in real-time or recorded for browsing afterward. In the present study, we focus on the situation with fixed view angle cameras set up in areas having a low volume of traffic, for browsing afterward. Such a situation is not special, and several low-cost surveillance camera systems may be used in this situation, especially at night.

In such a situation, a large number of images are captured and stored, but most of these images are of the background only. Usually, users only want to view scenes containing moving objects, especially humans. Therefore, it is necessary to browse these images efficiently. As naive browsing methods, a large number of images can be checked simultaneously by displaying several thumbnail-size images, or viewed in a shorter time by fast-forwarding through video sequences. However, both methods still have enormous costs and are not practical for lengthy image sequences. In order to search, extract, and track humans automatically in surveillance camera images, various techniques have been proposed and developed. However, it is difficult to

M.J. Smith, G. Salvendy (Eds.): Human Interface, Part I, HCII 2007, LNCS 4557, pp. 383–390, 2007.

recognize humans without error in various situations. Since many objects such as small animals, shaking trees, and lighting changes, could be detected as moving objects, and it is sometimes difficult to recognize whether a detected object is human. A number of methods recognize humans well, but require precise parameter settings for each situation.

For efficient image browsing, we propose a new display method with image gathering. A gathered image is a still image that contains all moving objects in all of the frames during a given period of time. The gathered image is a kind of photographic playback, and the user can easily determine whether human(s) exist in the image.

In the present paper, we also propose a hierarchical image gathering and display method. The gathered images are generated for several intervals of time length, and are displayed hierarchically. A user can easily browse the scene by looking at the moving parts of the images. Since detection and recognition of the target objects are performed by a human operator, efficient and reliable browsing can be established.

2 Related Research

2.1 Automatic Recognition Methods

Several techniques have been proposed for efficient browsing of surveillance camera images. One such method is automatic human detection and recognition by image processing. Background subtraction is a simple method by which to detect moving objects[1,2]. Techniques for distinguishing human have also been studied extensively. These techniques are based on object size[3] or pattern training and neural-network learning of human behavior[4,5]. In addition, techniques for detecting the position of humans[6] and contour definition of moving objects[7,8,9] have been proposed. The major problem in automatic detection and recognition is detection leakage and misrecognition. In addition, a number of methods require precise parameter setting for each situation. However, perfect detection and recognition of humans in various situations is difficult.

2.2 Hierarchical Display

Multi-resolution video[10], proposed by Finkelstein,*et al.*, efficiently displays a video sequence. This technique uses the hierarchical change of spatial and temporal resolutions. We applied the hierarchical change of temporal resolutions to a surveillance camera image. However, since the multi-resolution video uses an image that smoothes each pixel in the time direction of the axis, moving objects that appear only for a short time become thin and disappear in low time resolutions. Our method , on the other hand, extracts the moving parts and displayed them without smoothing.

3 Generation of the Gathered Image

A gathered image is a still image that contains all moving objects in all of the frames during a given period of time. The generation strategy of the gathered image is as follows. The intensity of an image rapidly changes when a moving object comes into the view. Therefore, if the intensity change is examined during the time period at a fixed pixel position, a moving object at a pixel can be detected by examining the frame with the maximum intensity change. By applying this process to all pixels, the gathered image of moving objects can be obtained (Fig. 1). For better results, approximately one frame per second is used for gathering. To reduce noise, the average intensity of a small rectangular block is used instead of the single pixel intensity. The average intensity of the small rectangular block is as follows:

$$L_n(Q_i) = \frac{\sum_{(x,y)\in Q_i} |B_{n+1}(x,y) - B_n(x,y)| + |B_n(x,y) - B_{n-1}(x,y)|}{2S}$$

Q_i : ith block

$B_n(x,y)$: luminance of picture element (x,y) in the nth frame

S : number of pixels in each Q_i

The resulting image includes almost all moving objects, unless multiple objects are located in the same position. The user can easily determine whether human(s) are present during the time period. If only background is present in a gathered image, it is no longer necessary to check the corresponding time period. The image might be complicated if too many moving objects exist (Fig. 2). In such cases, we need the gathered images for shorter time periods. In addition, we can easily confirm the presence of human(s) in such images.

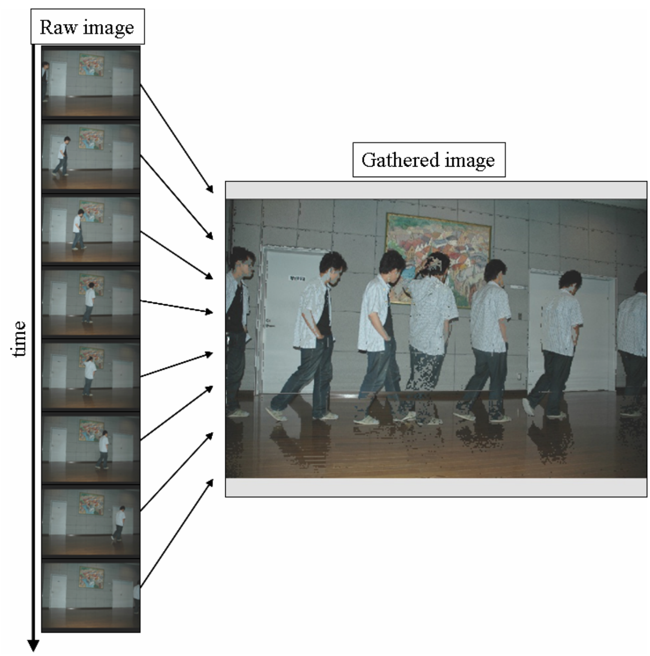

Fig. 1. This figure shows a time sequence image in which a human walks across the captured area. Information from several row images is gathered into one still image. As a result, the user can easily confirm the behavior of the subject in the image.

(a) Sample1 (b) Sample2

Fig. 2. The gathered image is obfuscated when too many moving objects exist

4 Hierarchical Image Gathering and Display Method

For easier management of the gathered images for various time periods, we propose a hierarchical image gathering and display method. The gathered images are generated for several intervals of time length and are displayed hierarchically. When browsing the images captured for a twelve-hour period, for example, the period is divided into twelve periods, and the gathered images for each one-hour period are displayed. Some of the one-hour gathered images might include humans, while others may contain the background only. When the user clicks an image containing humans, the one-hour period is divided into six images, and the gathered image for each ten-minute period is displayed. If the user clicks one of the gathered images with humans, then, ten one-minute gathered images are displayed. In this way, a user can efficiently browse the scenes containing humans.

5 Experiment

We applied the proposed method to a one-hour image taken by a surveillance camera (Fig. 3). The experimental environment is shown in Table 1. The gathered image was hierarchically managed in three stages (one hour, ten minutes, and one minute.)

First, the user can confirm the presence of human(s) in a gathered image of one hour. This shows that human(s) are present between 16:00 and 17:00. Next, six gathered images of ten minutes in the next hierarchy are confirmed. Here, a human

Table 1. Experimental environment

Camera type	Fixed view angle cameras
Image resolution	704×480 pixels
Serial shoot duration	1/second
Block size	2×2 pixels

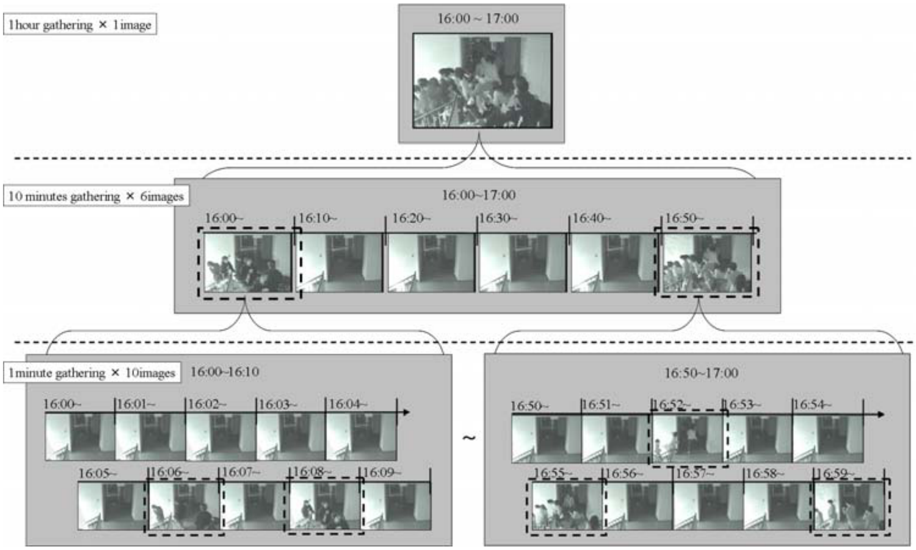

Fig. 3. Hierarchical image gathering and display. This figure shows that the user can easily select only images in which human(s) appear. As a result, effective browsing is established.

can be confirmed to be present from 16:00 to 16:10 and from 16:50 to 16:60. Similarly, only gathered images of human(s) are confirmed by the image of the subordinate position hierarchy. As a result, the presence of human(s) between 16:00 and 17:00 can be confirmed at 16:05, 16:06, 16:08, 16:52, 16:55, and 16:59.

6 Discussion

6.1 Browsing Efficiency of Hierarchical Display

A user can easily browse only necessary parts, i.e., times at which human(s) appear, by hierarchically managing the gathered image. In the experiment, 3,600 images can be confirmed through a number of gathered images. Efficient browsing of a large number of images is established. When you manage the image far into the future, only the chosen part need be manage, and an improvement in the efficiency of image management is achieved. Moreover, equivalent results are obtained in other environments.

6.2 Appropriate Block Size for Consolidating Image

Gathered images of each block size (2×2, 4×4, 8×8, 16×16 pixels) are generated. It is predicted that the outline of the object will be preserved when the block size is small. However, the influence of the block noise is considered, and the processing time of the gathered image increased. The image quality and processing time have a trade-off

Table 2. Processing time for each block size

Block size	Processing time(s)
2×2 pixels	96.44
4×4 pixels	93.77
8×8 pixels	92.42
16×16 pixels	89.66

Fig. 4. This figure shows gathered images in which the block size changes. The top images are magnified areas of the lower images.

relationship. Then, the size of the block is changed to generate the gathered images, and the influence of the change is investigated. The processing time for each block up to five minutes (300 pieces) is shown in Table 2.

The influence of the noise is negligible in the minimum block size of 2×2 pixels. The processing time of the minimum block size is 12.1% larger than that of the maximum block size. However, the processing time for the minimum block size is only approximately 32% of the captured time interval, and thus, real-time processing is also possible. Considering the clarity (Fig. 4) of the gathered image and the processing time, 2×2 pixels is the best block size in the proposed method.

6.3 Situation with Rapid Illumination Change

There is an environment in which a sensor light operates when a human approaches (Fig. 5). In such an environment, the luminance value of the entire image changes rapidly. Therefore, there is a possibility that human(s) information may be overwritten by information on lighting. We therefore conduct an experiment in such an environment. The gathered image of a time period that contains lighting is generated (Fig. 6). Most of the time, humans appear on the gathered image, although human(s) information might not appear, depending on the human(s) color. However, information of lighting is plainly shown. As a result, the human(s) appearance is not overlooked by browsing the hierarchy below.

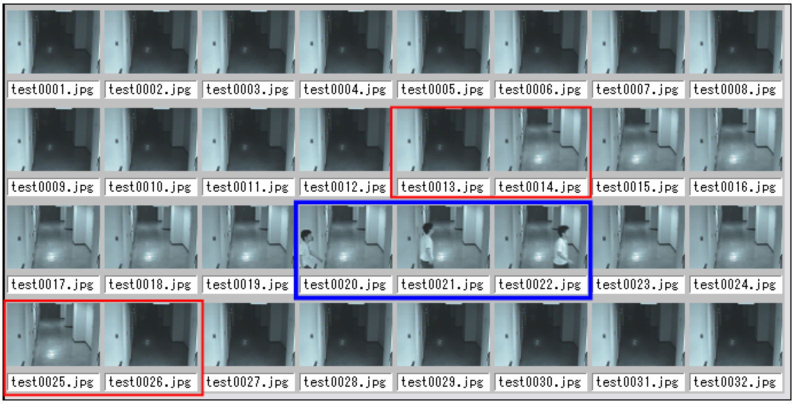

Fig. 5. This figure shows the environment with lighting. The part in which a rapid luminance changes is enclosed by a red frame. The part in which the human appeared is enclosed by a blue frame.

Fig. 6. Gathered image of time period that contains lighting. Information of lighting can be obtained, although there is a part in which the person disappears, depending on their color.

7 Conclusion

In the present study, we proposed an image gathering and display method for efficient browsing of surveillance camera images. The greatest advantages of the proposed method, compared to automatic recognition methods, are its robustness and reliability. There is no need for threshold setting in the proposed method. If small animals or other meaningless moving objects appear during a time period, the gathered image does include these objects. However, the user can easily recognize that the moving objects are meaningless and can discard them. The obtained experimental results show that the proposed method is also robust for changing lighting conditions. Therefore, the proposed method is effective for use in various situations.

References

1. Cucchiara, R., Grana, C., Piccardi, M., Prati, A., Sirotti, S.: Improving Shadow Suppression in Moving Object Detection with HSV Color Information. IEEE Intelligent Transportation Systems Conference Proceedings, pp. 334–339 (2001)
2. Haritaoglu, I., Harwood, D., Davis, L.: W4: Real-Time Surveillance of People and TheirActivities. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 809–830 (2000)
3. Jacinto, C., Jorge, M.: Performance Evaluation of Object Detection Algorithms for Video Survillance. IEEE Transaction on Multimedia 8(4), 761–774 (2006)
4. Gian, F., Lucio, M., Carlo, R.: Automatic Detection and Indexing of Video-Event Shots for Surveillance Application. IEEE Transction on Multimedia 4(4), 459–471 (2002)
5. Hu, W., Tan, T., Wang, L., Maybankl, S.: A Survey on Visual Surveillance of Object Motion and Behaviors. IEEE Transaction on CyberneticsPart C: Application and Reviews, vol. 34(3) (2004)
6. Hayashi, K., Haga, T., Seki, M., Sasakawa, K.: Detecting Human Position and Action in Surveillance Scenes. IPSJ CVIM14, 47(9), 12–20 (2006)
7. Kass, M., Witkim, A., Terzopoulos, D.: Snakes, Active Contour Models. Int. J. Computer Vision 1.1(4), 321–331 (1988)
8. Sethian, J.: Level set Methods, 1st edn. Cambridge University Press, Level set Methods (1996)
9. Osher, S., Sethian, J.A.: Fronts Propagating with Curvature Dependent Speed, Algorithm based on Hamilton-Javobi formation. J. Computational Physics 79, 12–49 (1998)
10. Finkelstein, A., Jacobs, C.E., Salesin, D.H.: Multiresolution Video. In: Proceedings of SIGGRAPH 96, pp. 281–290 (1996)