J Forensic Sci, September 2006, Vol. 51, No. 5
doi:10.1111/j.1556-4029.2006.00205.x
Available online at: www.blackwell-synergy.com

TECHNICAL NOTE

Shih-Hsuan Chiu,[1] Ph.D.; Chuan-Pin Lu,[1] M.S.; and Che-Yen Wen,[2] Ph.D.

A Motion Detection-Based Framework for Improving Image Quality of CCTV Security Systems

ABSTRACT: Closed-circuit television (CCTV) security systems have been widely used in banks, convenience stores, and other facilities. They are useful to deter crime and depict criminal activity. However, CCTV cameras that provide an overview of a monitored region can be useful for criminal investigation but sometimes can also be used for object identification (e.g., vehicle numbers, persons, etc.). In this paper, we propose a framework for improving the image quality of CCTV security systems. This framework is based upon motion detection technology. There are two cameras in the framework: one camera (camera A) is fixed focus with a zoom lens for moving-object detection, and the other one (camera B) is variable focus with an auto-zoom lens to capture higher resolution images of the objects of interest. When camera A detects a moving object in the monitored area, camera B, driven by an auto-zoom focus control algorithm, will take a higher resolution image of the object of interest. Experimental results show that the proposed framework can improve the likelihood that images obtained from stationary unattended CCTV cameras are sufficient to enable law enforcement officials to identify suspects and other objects of interest.

KEYWORDS: forensic science, closed-circuit television, motion detection, surveillance system

Closed-circuit television (CCTV) security systems have been widely used in banks, convenience stores, parking areas, automatic teller machines, and other facilities. If they are set up properly and follow generally accepted recommendations and guidelines (1), they are useful to deter crime and depict criminal activity. However, most of these systems are fixed focus and therefore are not very useful for object identification (e.g., vehicle plate numbers, persons, etc.), as they cannot be automatically refocused on objects of interest in motion within the field of view of the camera.

Real-time moving-object detection/tracking is an important technique for automatic surveillance systems. In Taiwan, some commercial CCTV systems already include this function. They use simple "intensity change" and statistical calculation to determine whether there are any object intrusions. However, this type of technique is easily influenced by noise and not suitable for object tracking. Recently, some object detection/tracking algorithms were proposed in research papers. Geiger et al. (2) proposed a dynamic programming method to detect deformable contours; Agbinya and Rees (3) proposed a multi-object detecting and tracking method; and Chen et al. (4) combine mosaic-based temporal segmentation and color-based spatial segmentation to extract moving objects. However, these algorithms are not suitable for real-time CCTV systems with a complex image background. In our framework, we propose a novel algorithm to detect moving objects for real-time CCTV systems.

In this paper, we propose a framework for improving the image quality of CCTV security systems. In this framework, we use two CCD cameras: one camera (camera A) is fixed focus with a zoom lens for moving-object detection, and the other one (camera B) is variable focus with an auto-zoom lens to capture higher resolution object images. Camera A is set up with the motion detection technology. We use it to find the image region of objects of interest. Camera B is equipped with our auto-zoom control algorithm for adjusting its focus. When camera A detects a moving object in the monitored area, camera B will take a higher resolution image of the object.

Methods

The framework includes three main procedures: (1) camera position translation analysis: we estimate the position relationship between two cameras, and determine a moving-object region. (2) Moving-object detection: we use camera A to detect and determine an image region for zoom controlling the focus of camera B. (3) Zoom controlling: we adjust the focus of camera B based upon the image region obtained from the procedure (2). Camera A is a fixed focus camera with a zoom lens (most cameras in current CCTV systems are fixed focus); the other one (camera B) is variable focus with a zoom lens (see Fig. 1). In this framework, all captured images are exported to nonproprietary AVI (audio/video interlaced) files.

[1]Department of Polymer Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan.
[2]Department of Forensic Science, Central Police University, Taoyuan, Taiwan.

Received 23 Oct. 2005; and in revised form 4 Mar. 2006; accepted 26 Mar. 2006; published 7 Aug. 2006.

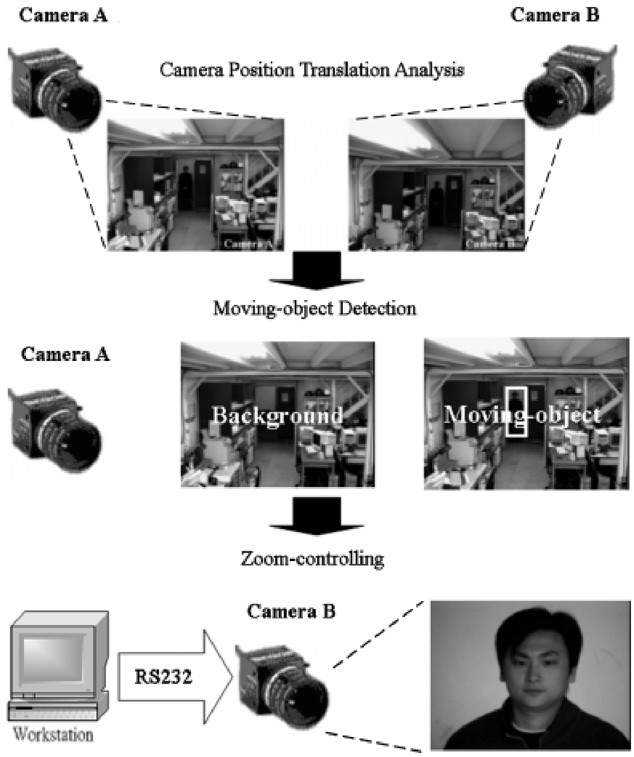

FIG. 1—*Diagram of the proposed framework.*

Camera Position Translation Analysis

In this procedure, we need to estimate a position translation value between two cameras (A and B), and find a superimposed region of images obtained from both cameras. We use a correlation method (3) to analyze the position translation between two images. Figure 2 illustrates our correlation method for measuring the similarity between two images. To obtain the optimal result, we use histogram equalization (3) to enhance images before the correlation matching process. This processing must follow Scientific Working Group on Imaging Technology (SWGIT) guidelines (5). In Fig. 2*a*, *g*, and *f* show the images from two cameras after histogram equalization; *h* is a sub-image of *g* (covers *c*. 1/4–1/8 of *g*). We move *h* around *f*; P_1 is the beginning point. We can obtain the cross-correlation coefficients $\gamma(x,y)$ by the following equation:

$$\gamma(x,y)=\frac{\sum_i\sum_j[f(i,j)-\bar{f}(i,j)][h(x+i,y+j)-\bar{h}]}{\left\{\sum_i\sum_j\left[f(i,j)-\bar{f}(i,j)\right]^2\sum_i\sum_j\left[h(x+i,y+j)-\bar{h}\right]^2\right\}^{1/2}}$$

$$x=0\sim(l_h-1),y=0\sim(l_w-1) \quad (1)$$

where $\bar{h}$ is the mean gray value of *h*, and $\bar{f}$ is the mean gray value of *f* within the *h* region. l_h and l_w are the image height and width, respectively. P_2 is the position that has the highest $\gamma(x,y)$ value, and the corresponding region is that *h* matches *f* in the optimal sense. We can obtain the image position translation by P_1 and P_2, as shown in Fig. 2*b*. Figure 3 shows a matching example by the correlation method. Figures 3*a* and *b* are two images from cameras A and B. Figures 3*c* and *d* show the enhanced images (i.e. *g* and *f* in Fig. 2) by histogram equalization. Figure 3*e* shows a sub-

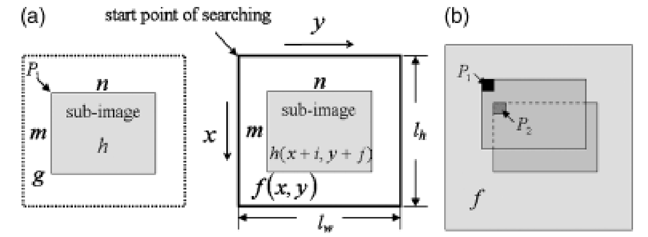

FIG. 2—*Correlation matching method: (a) g and f are two observed mixed images. h is a sub-image of g; n and m are sub-image's width and height, respectively; (b) we move h around f; P_1 is the beginning point. P_2 is the position that has the highest $\gamma(x, y)$ value.*

image *h* from Fig. 3*c*. Figure 3*f* shows correlation coefficients $\gamma(x,y)$ obtained by moving *h* around Fig. 3*d*. P_2 is the brightest point that has the highest $\gamma(x,y)$ value. Figure 3*g* shows the matched result.

Moving-Object Detection

In our framework, we proposed a novel object detection/tracking algorithm for real-time CCTV systems. In order to analyze variance of image sequences, the proposed algorithm defines three motion models: (1) a current motion model $\tilde{g}_C(t)$ (current image state), (2) a background motion model $\tilde{g}_B(t)$ (to analyze image background variance within a long time period), and (3) a main

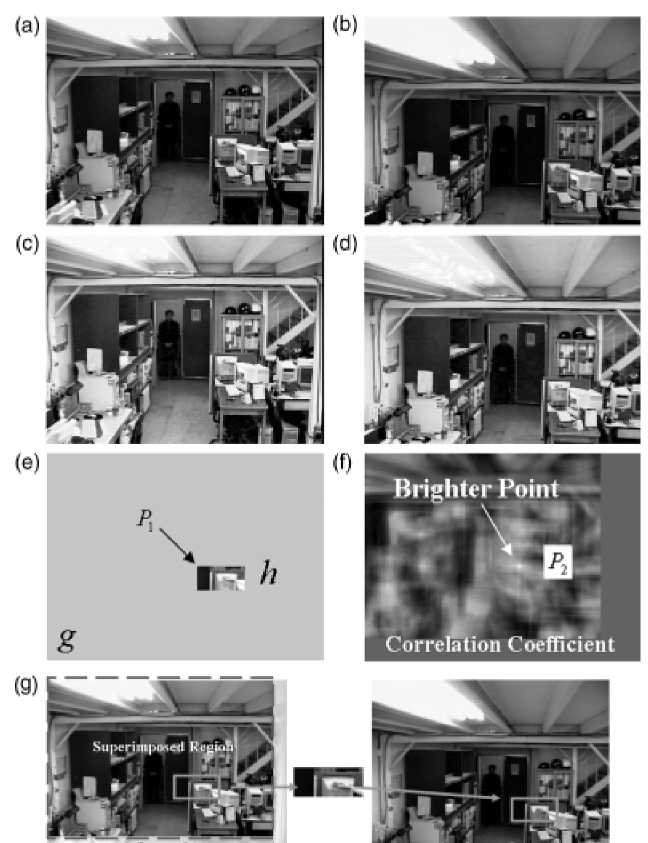

FIG. 3—*Example of the matching by the correlation method: (a, b) two images from cameras A and B; (c, d) the enhanced images (i.e., g and f in Fig. 2) after histogram equalization; (e) a sub-image h from (c); (f) correlation coefficients $\gamma(x, y)$ by moving h around (d); P_2 is the brightest point that has the highest $\gamma(x, y)$ value; (g) the matched result.*

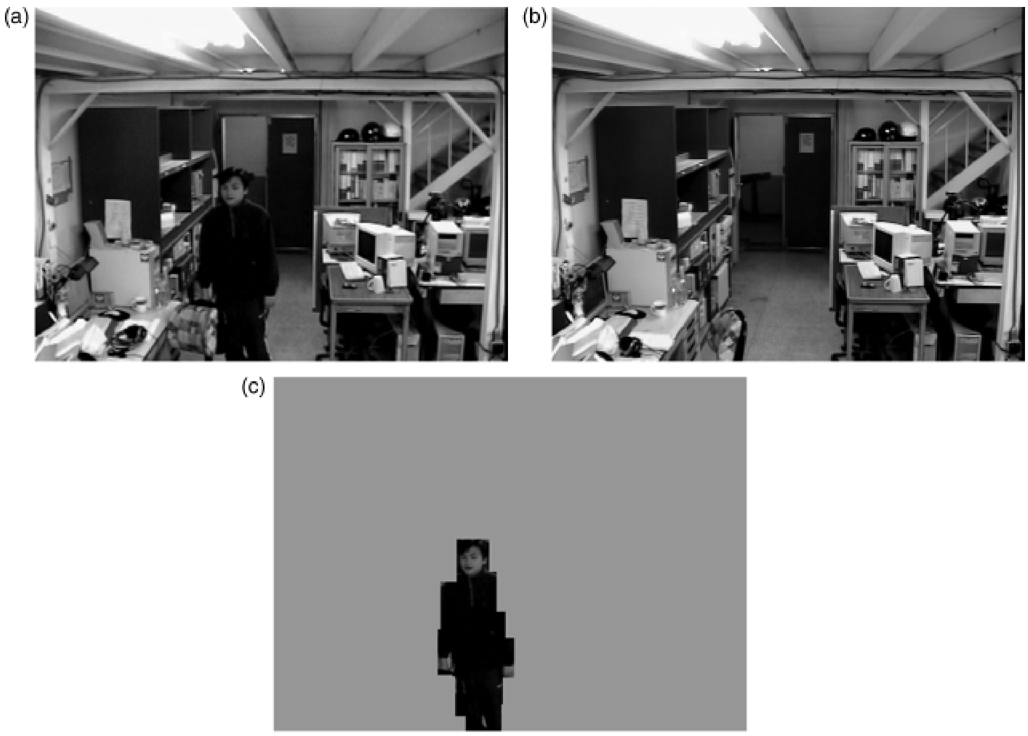

FIG. 4—*Example of moving-object detection: (a) an image with a moving object; (b) the background image; (c) the detection result.*

motion model $\tilde{g}_{\mathrm{M}}(t)$ (to analyze dominant variance within a short time period). Let $g(t)$ be a fusion image by "image edge" and "gray scale" as

$$g(t) = w_1 \text{ edge}(f_t) + w_2 \text{ gray}(f_t) \tag{2}$$

where w_1 and w_2 are the weight coefficients ($w_1+w_2 = 1$); f_t is the tth image frame from camera A. *edge* $(\cdot)$ is the edge information of f_t by the Sobel gradient operator (6), and *gray* $(\cdot)$ is the gray scale of f_t. Three motion models are defined as

$$\text{The current motion model: } \tilde{g}_{\mathrm{C}}(t) = g(t) \tag{3}$$

$$\text{The background motion model: } \tilde{g}_{\mathrm{B}}(t) = \frac{1}{A}\sum_{m=t-A+1}^{t} g(m) \tag{4}$$

$$\text{The main motion model: } \tilde{g}_{\mathrm{M}}(t) = \frac{1}{B}\sum_{m=t-B+1}^{t} g(m) \tag{5}$$

where A and B are time sequence numbers ($A > B$). We obtain $\tilde{g}_B(t)$ and $\tilde{g}_{\mathrm{M}}(t)$ by first averaging A and B frames, respectively. This process of averaging sequential video frames is used to find the moving objects from stationary scenes while $\tilde{g}_{\mathrm{C}}(t)$ is different from $\tilde{g}_{\mathrm{B}}(t)$ or $\tilde{g}_{\mathrm{M}}(t)$. We can also use this averaging process to reduce noise in stationary scenes (5). Our detection is based upon the following two rules (where g_T is a sensitivity parameter):

$$\text{Rule 1: } \tilde{g}_{\mathrm{C}}(t) > (\tilde{g}_{\mathrm{M}}(t) + g_T) \text{or } \tilde{g}_{\mathrm{C}}(t) < (\tilde{g}_{\mathrm{M}}(t) - g_T)$$

$$\text{Rule 2: } \tilde{g}_{\mathrm{C}}(t) > (\tilde{g}_{\mathrm{B}}(t) + g_T) \text{or } \tilde{g}_{\mathrm{C}}(t) < (\tilde{g}_B(t) - g_T)$$

If any pixel (x, y) in f_t satisfies either rule, it will be recorded. Moving objects should appear in the recorded pixel regions (such as Fig. 4*c*). Figure 4 demonstrates an example for this moving-object detection process. Figure 4*a* shows an image with a moving object, and Fig. 4*b* shows the background image of Fig. 4*a*. The moving object is successfully detected as Fig. 4*c*.

Zoom Controlling

After finding a moving object, camera B is refocused to obtain images with a better resolution. In order to execute this zoom-controlling process, we need to find a relationship between "object length" and "zooming time." We use Fig. 5 to explain the process. L_i is the length of a moving object. T is the time step. After several time steps, we can obtain a relationship between "object length" and "zooming time." With this relationship, camera B is refocused according to the moving object's length in the real-time tracking process.

Experiments

We use a personal computer with an Intel Pentium IV 3 GHz CPU and 1 Gbyte RAM to implement our framework. The

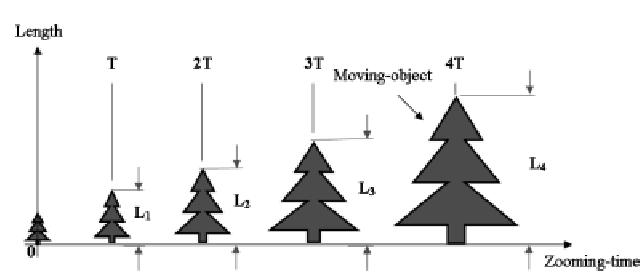

FIG. 5—*Process to find a relationship between "object length" and "zooming-time." L_i is the length of a moving object, which is measured from the image center to the object top. T is the time step.*

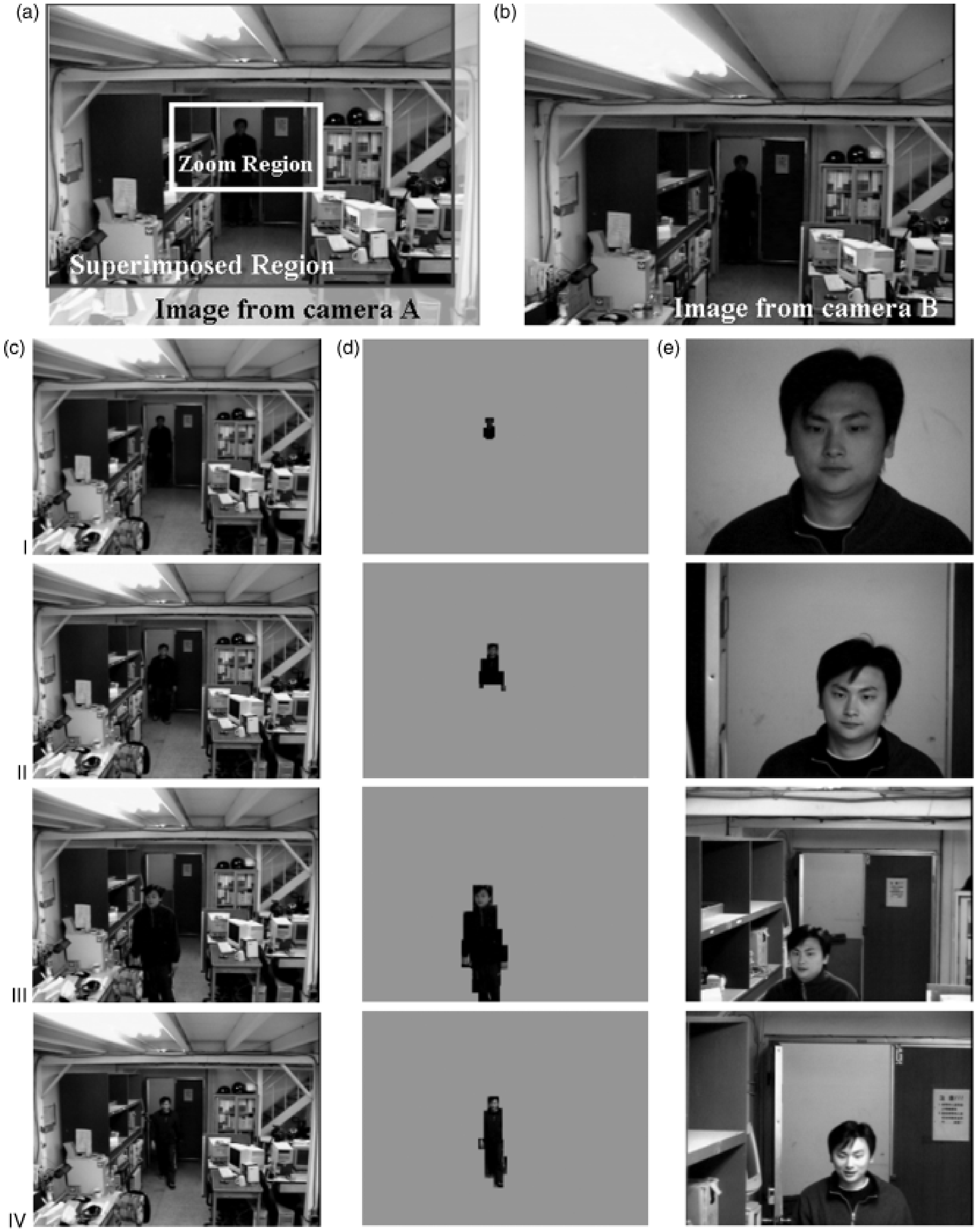

FIG. 6—*Example: (a) an image from camera A and the superimposed region obtained by the camera position translation analysis process; (b) an image from camera B; (c) some image sequences from camera A; (d) the moving-object detection results; (e) some image sequences from camera B.*

system contains two YOKO 210Z (NTSC) CCD cameras (Jhonghe City, Taiwan) and one Euresys Picolo image grab card (the image analytic rate is 15 frames/sec, image recording rate is 30 frames/sec, and the image size is 640 × 480 pixels). We use a Gaussian blur to reduce noise (7). All output video files are exported to nonproprietary AVI files and stored in a 200 Gbyte hard disk (nonremovable medium). We use four experiments to demonstrate the performance of our framework. These experiments include moving objects (human and vehicle) in indoor and outdoor environments. We set $w_1 = 0.9$, $w_2 = 0.1$, $g_T = 10$, $A = 200$, and $B = 50$ in all experiments.

Figure 6*a* shows an image from camera A and the superimposed region obtained by the camera position translation analysis process. Figure 6*b* shows an image from camera B. Figures 6*c* and *e* show some image sequences from cameras A and B, respectively. Figure 6*d* shows the results of moving-object detection. Figures 7 and 8 show experiments in indoor and outdoor environments, respectively. Figure 9 shows an experiment for vehicle monitoring. We set the vehicle appearing on the top left as our object of interest. We can recognize the vehicle plate number. From the experimental results, the object of interest (e.g., face, car) covers more than 15% of camera B's field of view (1).

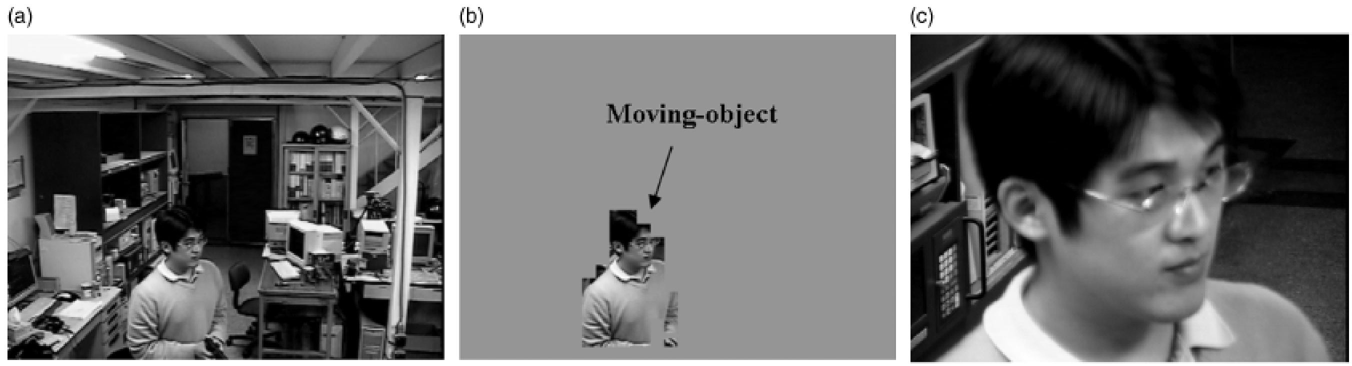

FIG. 7—*Experiment: (a) an image from camera A; (b) the moving-object detection result; (c) an image from camera B.*

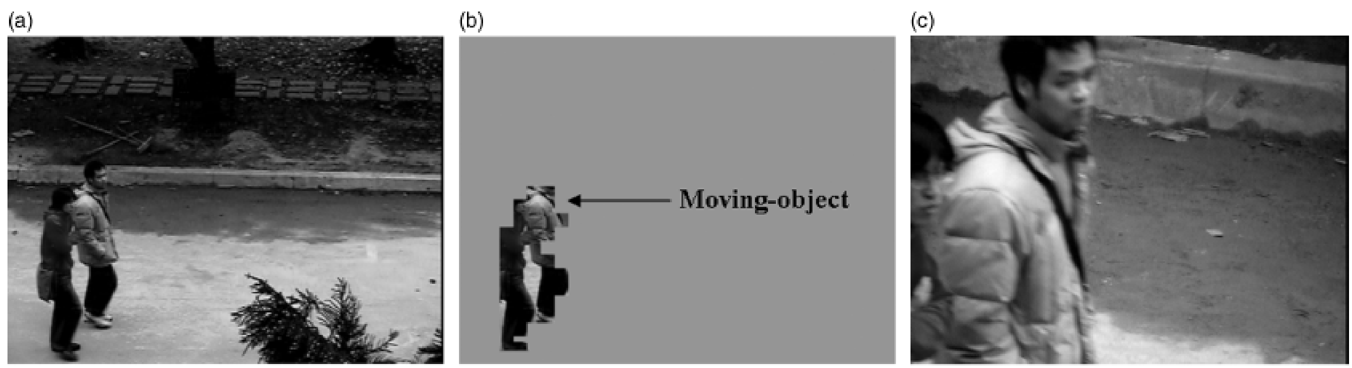

FIG. 8—*Experiment: (a) an image from camera A; (b) the moving-object detection result; (c) an image from camera B.*

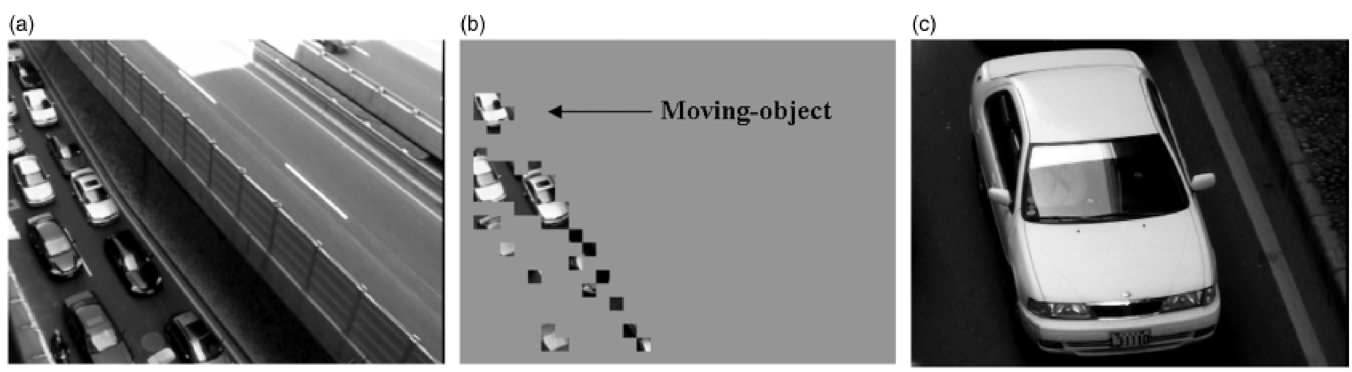

FIG. 9—*Experiment for vehicle monitoring: (a) an image from camera A; (b) the moving-object detection result (we set the vehicle appearing on the top left as the object of our interest); (c) an image from camera B (we can recognize the vehicle plate number).*

Conclusions

In this paper, we propose a framework for improving the image quality of CCTV security systems. It is based upon motion detection technology and follows SWGIT guidelines. Four experiments demonstrate encouraging results to show the performance and reliability of this framework. This framework is proposed for stationary unattended CCTV cameras. In the future, the framework will be extended to a speed dome camera (two-axis rotation) for dynamic tracking and capturing the moving objects.

References

1. Scientific Working Group on Imaging Technology (SWGIT). Recommendations and guidelines for using closed-circuit television security systems in commercial Institutions. Forensic Sci Commun 2005;7(1) (accessible from http://www2.fbi.gov/hq/lab/fsc/backissu/jan2005/standards/2005standards1.htm).
2. Geiger D, Gupta A, Costa LA, Vlontzos J. Dynamic programming for detecting, tracking, and matching deformable contours. IEEE Trans Pattern Anal Mach Intell 1995;17:294–302.
3. Agbinya JI, Rees D. Multi-object tracking in video. Real-Time Imag 1999;5(5):295–304.
4. Chen LH, Lai YC, Su CW, Mark Liao HY. Extraction of video object with complex motion. Pattern Recog Lett 2004;25(11):1285–91.
5. Scientific Working Group on Imaging Technology (SWGIT). Recommendations and guidelines for the use of digital image processing in the criminal justice system. Forensic Sci Commun 2003;5(1) (accessible from http://www.fbi.gov/hq/lab/fsc/backissu/jan2003/swgitdigital.htm).
6. Gonzalez RC, Woods RE. Digital image processing. 2nd revised ed. New Jersey: Prentice-Hall, 2002.
7. Scientific Working Group on Imaging Technology (SWGIT). Best practices for document image enhancement. Forensic Sci Commun 2005;7(3) (accessible from http://www.fbi.gov/hq/lab/fsc/backissu/july2005/standards/2005_07_standards01.htm).

Additional information and reprint requests:
Che-Yen Wen, Ph.D.
Department of Forensic Science
Central Police University
56 Shu-Ren Road
Kuei-Shan, Taoyuan 33334 Taiwan
E-mail: cwen@mail.cpu.edu.tw