Image Recognition Wide-Area Surveillance Technology for Safety and Security in Society

● Eigo Segawa ● Masaki Miura ● Daisuke Abe

(Manuscript received November 9, 2006)

Surveillance is needed to create a safe and secure society. To meet this need, we developed a small-object detection technology for 24/7 wide-area outdoor surveillance based on an image recognition technology that is functionally similar to human vision. It can detect people and other small objects that are represented by as little as five pixels in an area of up to 75 m square and report their locations. This technology, called frequency pattern emphasis subtraction (FPES), detects a small object by analyzing the differences between the spatial frequencies of the background and the object. Because FPES emphasizes the spatial frequency component of objects, the detection precision is unaffected by changes in the weather. In experiments, this technology detected 95% or more of the objects in various areas ranging from 30 to 75 m square. This paper describes FPES, certain problems posed by using automatic surveillance, and possible solutions.

1. Introduction

Since the coordinated terrorist attacks that occurred in the US on September 11, 2001, the world has been enveloped in a fear of terrorist attacks that could occur at any time and in any place. Even today, people are still concerned about this fear. However, rather than a fear of terrorism, which lacks a sense of reality, the Japanese are more concerned with problems closer to home, such as the large number of cases involving indiscriminate homicide and arson. Accordingly, a growing increase in crisis-management consciousness can be seen in Japan.

Against this background, interest in security has increased, and the probability of crime occurring at large-scale public facilities such as rivers, harbors, airports, and roads used by large numbers of people is particularly high. Consequently, expectations regarding surveillance systems that can monitor the comings and goings of people and motor vehicles in and around public facilities 24/7 are rising. One of the elemental technologies that support surveillance systems is image recognition technology.

This paper describes certain problems posed by using automatic surveillance in place of the human eye and the technology developed to solve these problems.

2. Expectations toward automatic surveillance and related issues

In conventional surveillance systems, heavy reliance is placed on the human eye. However, particularly in large-scale facilities such as amusement centers, roving patrols made by security guards are inadequate. Surveillance is conducted by several specialist surveillance personnel who watch images from cameras installed in the facility on monitors in a surveillance room. However, it is difficult for surveillance personnel to watch the monitors 24/7 without missing anything, and there is a limit to what people can see with the naked eye. Moreover, as the surveillance area widens, not only are there more locations to watch, but also an increase in the number of camera blind spots. Even if the surveillance personnel concentrate hard, continued surveillance of a wide area is difficult at best. It is therefore no surprise that technology that replaces the human eye, for example, a surveillance infrastructure designed to support a safe and secure society and which adopts image recognition technology to automatically recognize details in an image, is gaining attention.

To realize this type of automatic surveillance, technology that can detect an intruding person or object and automatic detection technology must be researched and developed. However, there are considerable technical issues regarding automatic detection in wide, outdoor areas such as rivers, harbors, and parks that are more than 50 m square. The first issue is that the objects in the area are small and difficult to distinguish from the background. For example, in an image of an area 50 m square , a person is represented by only about five pixels. With conventional technology, it is impossible to distinguish between a person and the noise included in the image. As a result, detection using conventional technology is only possible in areas up to 10 m square. The second issue is that variations in environmental factors such as the weather, sunlight and camera shake due to the wind and other factors cause variations in the recording conditions, making it difficult to achieve stable detection of people and objects. In particular, variations in the recording conditions are important issues at the practical-use stage, and even for surveillance of areas 10 m square, there are practically no cases where these issues have been resolved.

The conventional technologies employed for detecting objects in a wide area can be generally classified into three types. The first type uses multiple fixed cameras to detect an object, the second tracks an object using a pan/tilt/zoom (PTZ) camera,[1] and the third detects an object from images recorded by a single fixed camera. The first two types require a high expenditure for cameras and cannot track multiple objects. On the other hand, the single fixed camera type uses optical flow detection to denote a dynamic area as an object.[2] However, the movement of an object in a wide-area image is very small, unless it is traveling at high speed; therefore, the single fixed camera type is not suited for wide-area surveillance. When background subtraction[3] is used to detect an object from its difference with the background, because an object against a wide background is relatively small, the difference in luminance between the background and object is also small, making detection difficult. When only one camera is used, it cannot keep up with changes in the recording conditions caused by environmental changes and camera shake, thereby making it difficult to realize stable object detection with these technologies.

3. Wide-area surveillance issues and our approach to development of new technology

With the aim of providing a powerful surveillance infrastructure for a safe and secure society, we developed a new technology that enables stable detection of objects outdoors 24/7 in an area of 75 m square using a single camera. This section describes the issues that arise during wide-area surveillance and our approach toward developing the new technology to solve these issues.

3.1 Wide-area surveillance issues

Before we could develop this technology, we had to resolve the following two issues that were mentioned in the previous section :

1) How to detect and differentiate from noise objects that are represented as small blurs, even when there are environmental changes such as changes in the weather or sunlight.
2) How to detect objects that are represented as small blurs, even when there is camera shake caused by the wind or other factors.

Figure 1 shows an example of the difficulties regarding the first issue. The image is from an NTSC camera and has a central width of 50 m. The person in the center of the image is represented as a small, blurred object of about 5 × 10 pixels, making it difficult to differentiate the object from noise. Moreover, the recording conditions were unstable due to variations in the weather and sunlight.

Regarding the second issue, images are blurred due to camera shake caused by the wind and other factors. For a general surveillance system with a camera at a height of 10 m, a typhoon-force wind with a velocity of 60 m/s can cause a shift of up to 20 pixels in a VGA-size (640 × 480 pixels) image, making the stable detection of objects difficult.

3.2 Approach to the development of new technology

To solve issue 1), we defined three types of outdoor conditions that are likely to occur: a target object is present; changes in the weather or other factors cause a change in the environment; and image noise is present. We then researched the frequency characteristics of the foreground and background images for each of these conditions. **Figure 2** shows the results. When a target object is present, the difference caused by the presence or absence of object boundaries and the difference in luminance of the object's interior cause variations in the frequency components over a wide range [Figure 2 (a)]. Conversely, when there is an environmental change such as a change in the weather, there is a change in overall luminance, which means that changes only occur in the low-frequency range [Figure 2 (b)]. Moreover, because noise causes changes to appear as spikes in the image, changes only occur in the high-frequency range [Figure 2 (c)]. **Table 1** shows the results of our research.

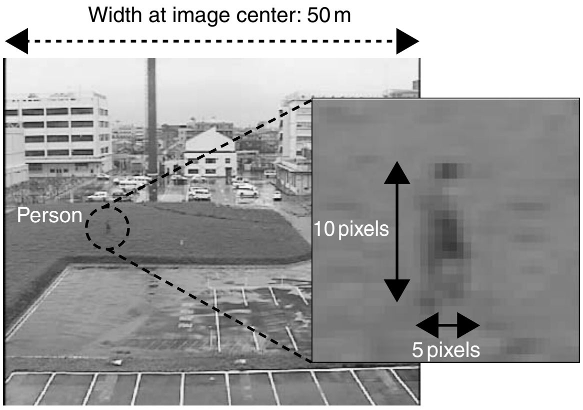

Figure 1
Outdoor wide-area surveillance image (daylight with rain).

The results indicate the following. If we study the differences in the changes in these frequency components, by detecting the regions that have changes in both the high-frequency and low-frequency components, we can discriminate an object regardless of changes in the weather and noise. We call this method frequency pattern emphasis subtraction. By using a day and night camera, this method enables discrimination even at night between an object and the luminance noise generated by the night camera. Objects can therefore be detected even if there is no large-scale lighting system such as that used in baseball stadiums.

Regarding issue 2), there are several methods available for correcting camera shake. The methods often used in general surveillance systems are mechanical correction performed by devices such as gyroscopes and image stabilization (vibration reduction) through optical correction. However, a large lens distortion occurs with wide-angle cameras that have an angle of view of 40° or more. This distortion causes a problem where vibration of around two pixels, called residual error, remains

Table 1
Changes in background and foreground frequency characteristics.

	High-frequency changes	Low-frequency changes
Object	Yes	Yes
Changes in sunlight	No	Yes
Noise	Yes	No

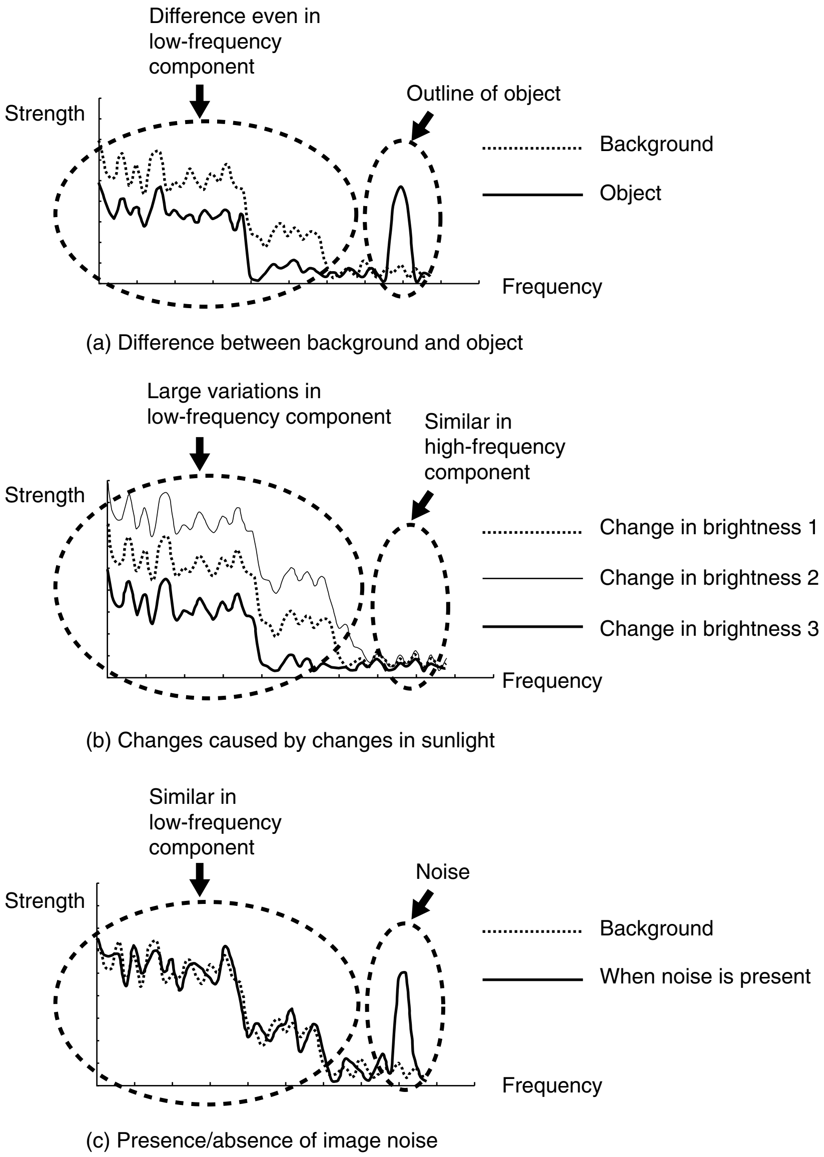

Figure 2
Frequency distribution in wide-area outdoor images.

in images for which vibration reduction has been performed. In wide-area surveillance, because a target object is represented by only about 5 × 5 pixels, the effect of this residual error is considerable. We therefore developed a multi-frequency based stabilization method to control, to within one pixel, the shift in position that occurs whenever positional correction is performed for each frequency using the above-mentioned frequency pattern emphasis subtraction method. The next section describes these methods in detail.

4. Frequency pattern emphasis subtraction

In frequency pattern emphasis subtraction, high-frequency pattern emphasis is performed first, followed by low-frequency pattern emphasis (**Figure 3**). After each of the frequenncy emphasis operations, the background and input images are divided into small areas, which are then used to make a comparison between the emphasized background and input image patterns.

By dividing the background and input

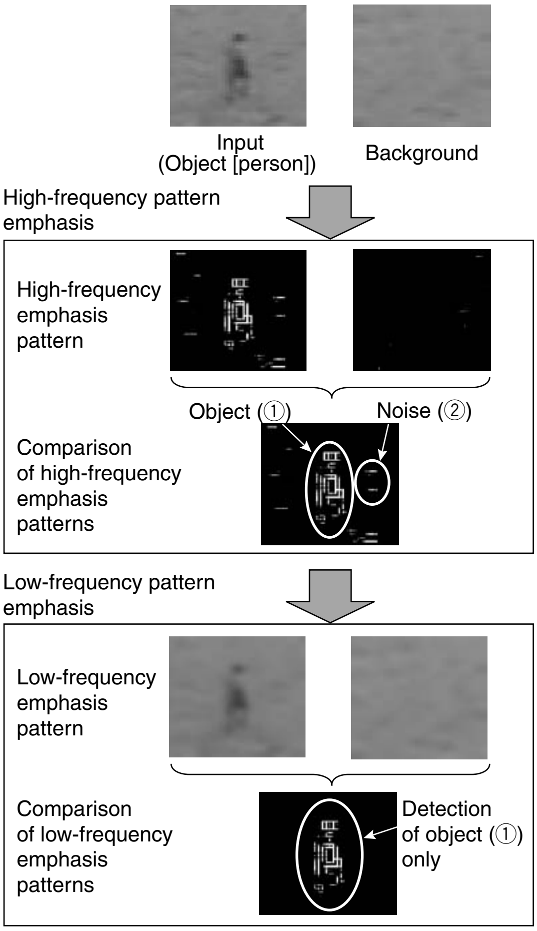
Input
(Object [person])
Background
High-frequency pattern emphasis
High-frequency emphasis pattern
Object (①)
Noise (②)
Comparison of high-frequency emphasis patterns
Low-frequency pattern emphasis
Low-frequency emphasis pattern
Comparison of low-frequency emphasis patterns
Detection of object (①) only

Figure 3
Frequency pattern emphasis subtraction.

images into small areas, even when the brightness in parts of the images has changed due to an environmental change, it is a uniform change when viewed in each of the small areas. Moreover, because the increase or decrease in the low-frequency component mentioned earlier can be used to create models, there is no adverse effect.

With high-frequency pattern emphasis, we first perform frequency emphasis using a 3 × 3 Sobel operator. Normalization correlation is used to compare the emphasized patterns using the small areas corresponding to the background and input images, and the small areas that are dissimilar to the background are detected. This emphasis process enables detection of object candidates (①) and noise (②) in the images. By using the high-frequency component, which is not easily affected by changes in brightness to narrow down the object candidates for local comparison in small-area units, the effect of environmental changes can be eliminated at this stage.

Next, we use a 3 × 3 averaging filter to perform low-frequency pattern emphasis of the detected object candidates. In the same way as for high-frequency pattern emphasis, the emphasized patterns are used to make a comparison between the background and input images. In this way, because the noise is similar to the background in the low-frequency components, it is eliminated, thereby enabling us to detect only objects.

5. Multi-frequency based stabilization

To further reduce the residual error of ±2 pixels that cannot be resolved with conventional image stabilization (vibration reduction), positional adjustments are made for each of the frequency pattern emphasis operations in the frequency pattern emphasis subtraction process.

If there is residual error due to a shift in position with respect to the background and input images [**Figure 4 (a)**], subtraction processing generates a small difference, even though it is essentially the background pattern. A distinction cannot be made between a small detected object of about 5 × 5 pixels in the images and this small difference, so the difference is detected as an object. With subtraction processing for each small area, however, the difference is taken after positional adjustment has been made to ensure that the background and input image are as well-matched as possible [**Figure 4 (b)**]. Because individual adjustments are made for each small area, stable detection of an object is possible even when there is camera shake.

6. Performance evaluation

We evaluated the detection performance in images having a central width of 50 m. Moreover,

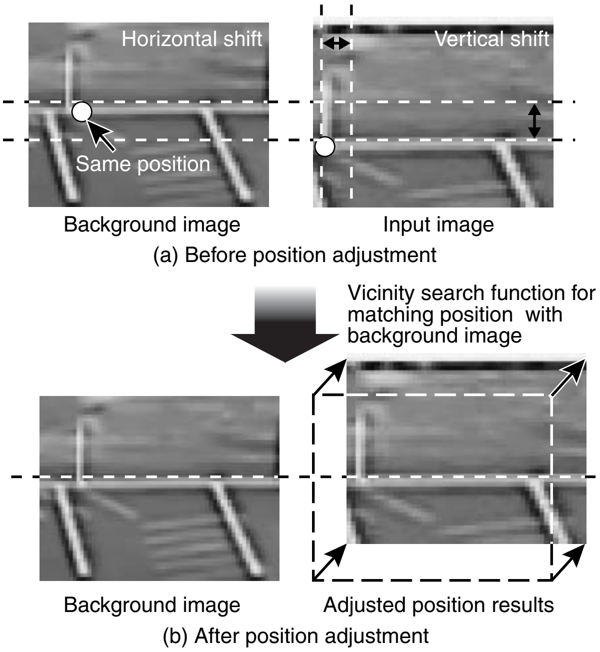
Horizontal shift
Vertical shift
Same position
Background image
Input image
(a) Before position adjustment
Vicinity search function for matching position with background image
Background image
Adjusted position results
(b) After position adjustment

Figure 4
Adjustment of each area's position.

Table 2
Detection performance for various interval lengths.

Interval length (Number of frames)	Detection rate (%)	Misdetection rate (%)
10	93.4	60.9
20	95.9	30.7
30	96.9	7.1
40	97.3	3.7
50	97.5	0.6

Shaded areas indicate detection rate of 95% or more and misdetection rate of 5% or less

Table 3
Frame processing time and specifications of PC used for measurement.

Processing time (average)	303 ms/frame
Processor	Pentium 4, 2.2 GHz
Memory	512 MB
OS	Linux kernel 2.4.9

to determine the objects that can be detected with this technology and the practical surveillance area, we changed the difference in luminance between the background and object, as well as the size of the surveillance area, and also evaluated the performance limit of the recorded images.

6.1 Evaluation of surveillance area with central width of 50 m

The evaluation data comprises eight hours of images that were recorded in a changing environment. The images have a central width of 50 m, similar to the image shown in Figure 1.

The main environmental changes in these images are described below.

1) Weather changes: fine weather and rain
2) Time of day changes: day (30 000 lx) and night (0.3 lx)
3) Camera shake (horizontal, vertical, and various other directions): up to ±20 pixels

The following two parameters were used for performance evaluation.

1) Detection performance

Detection rate: percentage of objects detected

Misdetection rate: percentage of non-objects detected

2) Processing time: This is the time-stability parameter for the processing time per frame.

Table 2 shows the relationship between the detection rate and misdetection rate when the interval length was changed. **Table 3** lists the frame processing time and specifications of the PC that was used. As shown in Table 2, setting the interval length to 40 frames or higher yields a detection rate of 95% or more and a misdetection rate of 5% or less.

As shown in **Figure 5**, when the causes of degraded performance were analyzed, we found it is not possible to perform stable detection of objects that are difficult to discern with the naked eye when the difference in luminance with the background is 10 gradations or less.

6.2 Evaluation of limit performance

We used 10 hours of image data obtained when both the size of the surveillance area and the difference in luminance between the background and object were changed to investigate the

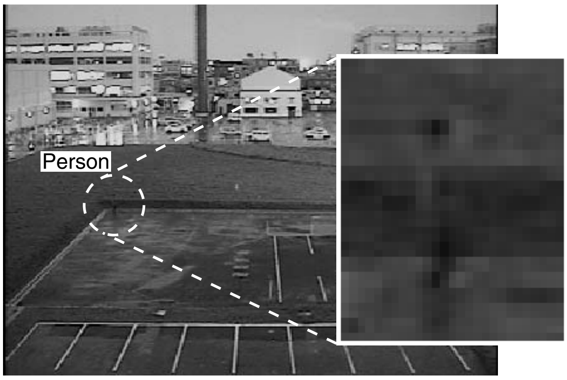

Figure 5
Main cause of misdetection is low image contrast.

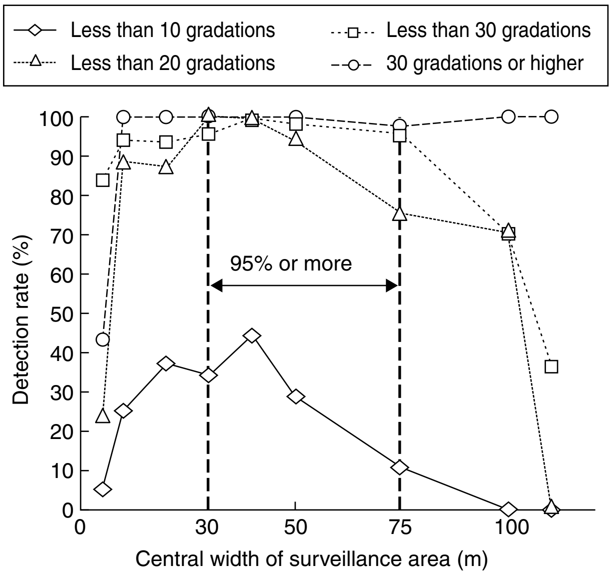

Figure 6
Relationship between central width of surveillance area and detection rate.

relationship between the detection performance (detection rate only) and changes made to both parameters. The data used for this evaluation was obtained from a surveillance area with a central width of 5 to 110 m, with the difference in luminance between the background and object changing from 5 to about 100 gradations. **Figure 6** shows the relationship between the central width of the surveillance area and the detection rate with regard to the difference in luminance between the background and object. The horizontal axis in the diagram represents the central width of the surveillance area, and the vertical axis represents the detection rate. The sequences of observation points represent the differences in luminance between the background and object.

From the results shown in Figure 6, in the same way as shown in the results of our evaluation in the preceding section, we can see that performance drops considerably when the difference in luminance between the background and object is less than 10 gradations. However, if the object can be discerned by the naked eye when the difference in luminance is 20 gradations or higher, we found that a detection rate of 95% or more can be achieved in a surveillance area with a central width of 30 to 75 m.

7. Conclusion

This paper proposed a new background subtraction method called frequency pattern emphasis subtraction that uses multiple-frequency emphasis to emphasize and integrate differences in background and object patterns for the purpose of detecting objects over wide outdoor areas. This is difficult to accomplish with conventional methods, which represent objects on a small scale and cannot provide stable images. In eight hours of images recorded in a surveillance area with a central width of 50 m, our system detected at least 95% of the objects that were moving in the area and had a misdetection rate of 5% or less. Furthermore, if an object could be discerned by the naked eye when the luminance difference between the background and object was 20 gradations or higher, we achieved a detection rate of at least 95% in the 30 to 75 m range.

We have initiated business expansion activities using this technology as the core technology for image surveillance systems and developed software for centralized management not only of surveillance images from multiple locations, but also of intruder information. Our new technology is expected to contribute to the creation of a safe and secure society by not only enabling wide-area surveillance of outdoor areas such as rivers and harbors, but also surveillance systems for detecting intruders in large-scale indoor facilities such as airports.

References

1) T. Kawanishi et al.: Dynamic Active Search for Quick Object Detection with Pan-Tilt-Zoom Camera. Proc. of the IEEE 2001 International Conference on Image Processing (ICIP2001), **3**, 2001, p.716-719.

2) A. Nagai et al.: Detection of Moving Object in Changing Background. (in Japanese), *Trans. IEICE*, **J80-D-**II, 5, p.1086-1095 (1997).

3) S. Fukui et al.: Extraction of Moving Objects by Estimating Background Brightness. *The Journal of the IIEEJ*, **33**, 3, p.350-357 (2004).

Eigo Segawa, *Fujitsu Laboratories Ltd.*
Mr. Segawa received the B.S. and M.S. degrees in Computer System Engineering from Osaka University in 1991 and 1993, respectively. In 1993, he joined Fujitsu Laboratories Ltd, where he has been engaged in research on image processing algorithms and systems. He is a member of the IEEE and IEICE.

Masaki Miura, *Fujitsu Ltd.*
Mr. Miura received the B.S. and M.S. degrees in Electronic Engineering from Tohoku University, Sendai, Japan in 1988 and 1990, respectively. He joined Fujitsu Limited, Kawasaki, Japan in 1990, where he has been engaged in the development of video recognition and video transmission systems.

Daisuke Abe, *Fujitsu Kyushu Network Technologies Ltd.*
Mr. Abe received the B.S. and M.S. degrees in Electrical and Computer Engineering from Nagasaki University, Nagasaki, Japan in 1998 and 2000, respectively. He joined Fujitsu Kyushu Network Technologies Ltd. (formerly Fujitsu Kyushu Digital Technologies Ltd.), Fukuoka, Japan in 2000, where he has been engaged in research and development of image recognition systems. He is a member of the Information Processing Society of Japan (IPSJ).