13-29

MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

Robust Small-Object Detection for Outdoor Wide-Area Surveillance

Daisuke ABE †, Eigo SEGAWA ‡, Osafumi NAKAYAMA §, Morito SHIOHARA ‡
Shigeru SASAKI ‡, Nobuyuki SUGANO §, Hajime KANNO §

† : FUJITSU KYUSHU NETWORK TECHNOLOGIES LIMITED
‡ : FUJITSU LABORATORIES LTD, § : FUJITSU LIMITED

E-mail: abed@jp.fujitsu.com

Abstract

In this paper, we present a robust small-object detection method, which we call "Frequency-Pattern-Enhanced Subtraction", for wide-area surveillance such as that of harbors, rivers, and plant premises. For achieving robust detection under changes in environmental conditions, such as illuminance level, weather, and camera vibration, our method distinguishes target objects from background and noise based on the differences in frequency components between them. The evaluation results demonstrate that our method detected more than 95% of target objects in the images of large surveillance areas ranging from 30 – 75 meters at their center.

1. Introduction

The recent international situation causes the need for increased attention to security measures. In particular, automatic surveillance systems for large-scale public facilities such as harbors, airports, rivers, and roads have become increasingly important. This has resulted in efforts to develop object detection methods required to implement such surveillance systems. However, conventional methods used for wide, open areas are not sufficiently robust for practical uses because of a number of difficulties: objects appearing small and blurred in input image, environmental variability such as weather and sunlight conditions, and camera vibration caused by the wind. This paper describes a new method of detecting objects, such as people and vehicles, in wide, open areas.

For wide-area surveillance, a multiple camera approach could be taken, which divides a whole surveillance area into several subareas and watches each area using different cameras. However, such an approach is costly with respect to introduction, running, and maintenance. For commercial uses, watching a whole area with a single camera is desired. Kawanishi et al. [1] developed a system that detected objects and tracked them using a Pan-Tilt-Zoom (PTZ) camera. Their system focuses on an object during the tracking, so it might overlook the other objects. Kobayashi et al. [2] and Nagai et al. [3] proposed a detection method based on optical flow using a single, fixed camera. However, it is not suitable for wide-area surveillance because little movements of objects in images are observed between frames. Fukui et al. [4] presented a detection method based on background subtraction, which also used a single, fixed camera. Their method, however, has difficulty detecting objects in images shooting wide-area because the objects are small and blurred, and there are little differences in brightness between the object and the background in the image. Detecting objects in open areas is especially difficult because the difference between the object and the background does not remain steady. This is caused by environmental variables or changes in imaging conditions including camera vibration caused by the wind.

2. Object Detection in Wide, Open Areas

2.1 Problems

In wide-area surveillance by a single normal camera, the objects in an image are small and blurred. For example, in an image with a range of 50 meters (width at center of image) obtained by using a VGA-size camera, the size of a person in the image is very small, that is, approximately 5 by 10 pixels, as shown in Figure 1. As a result, distinguishing the objects from the background is difficult. In addition, in outdoor surveillance, it is influenced by envi-

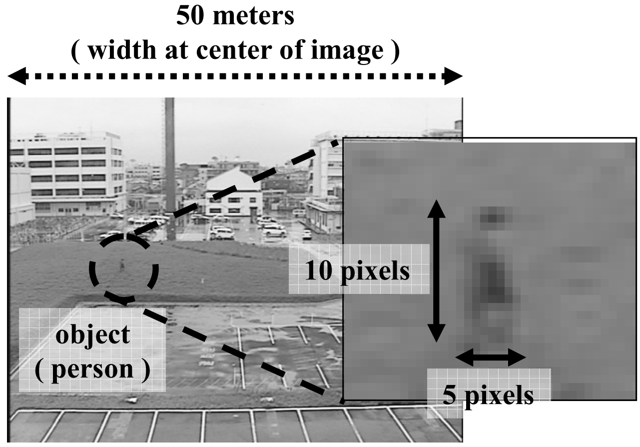


Figure 1. Image used for wide-area surveillance
(weather conditions: rain, time: daytime)

ronmental conditions such as weather, sunlight, and camera vibration caused by the wind. For example, if a wind of 60 m/s blows toward a camera installed at a height of 10 meters in a typical surveillance system, camera vibration causes the image to shift about 20 pixels compared with calm, windless conditions. Therefore, there are two challenges for object detection in wide, open areas as follows.

1) Distinguish small blurred objects from background and noise under changes in the environment.
2) Deal with camera vibration caused by the wind.

2.2 Approach to Challenges

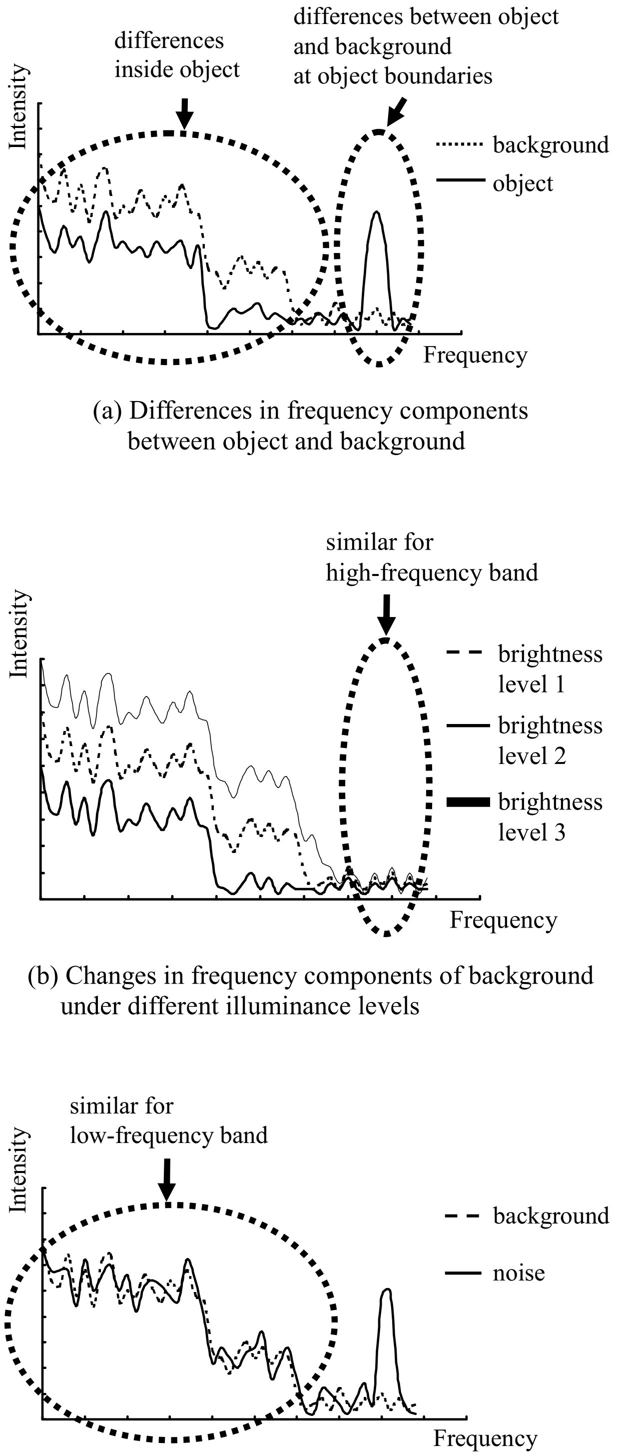
First, to solve challenge 1, we examined the frequency characteristics of background and foreground objects under several conditions as follows: (a) when a target object is present; (b) when environmental changes take place including changes in sunshine; and (c) when noise, which is caused by physical description of image sensor such as white, random noise etc., is present in the image. The results are shown in Figure 2. If an object is present, changes can be seen in the frequency components over a wide range from low to high frequency, as shown in Figure 2 (a). That is because the differences between object and background at object boundaries cause changes in the high-frequency band, and the differences between pattern inside the object and background cause changes in the low-frequency band. On the other hand, if an object is not present, either low-frequency components or high-frequency components change. As shown in Figure 2 (b), the changes in illuminance levels affect only the low-frequency band, and noise causes changes in the high-frequency band, as shown in Figure 2 (c). These characteristics are summarized in Table 1.

Table 1. Changes in frequency characteristics between background and foreground (○ = difference)

	changes in high frequency	Changes in low frequency
object	○	○
environmental changes	—	○
noise	○	—

According to Table 1, we detect target objects without detecting noise and environmental changes by extracting regions that have changes in both high and low frequencies. We named this method “Frequency-Pattern-Enhanced Subtraction”.

As a solution to challenge 2, an image stabilizer is widely used to correct the image misregistration caused by camera vibration. However, a camera field of view is more than 40 degrees in wide-area surveillance, so the image shot by such a camera suffers strong lens distortion, and the stabilized image still has about ±2 pixels of misregistration (called "residual error"). The residual error is very small, but detection of small objects, as required in wide-area surveillance, is disturbed. Therefore, we developed a method that reduces the residual error to 1 pixel or less. Each developed method is discussed in detail below.



(a) Differences in frequency components between object and background

(b) Changes in frequency components of background under different illuminance levels

(c) Differences in frequency components between noise and background

Figure 2. Frequency distribution in wide, open area (schematic patterns)

2.3 Frequency-Pattern-Enhanced Subtraction

As shown in Figure 3, the Frequency-Pattern-Enhanced Subtraction consists of two parts: the region-dividing process and high- and low-frequency-enhancing processes. First, the two images, the input and background image, are divided into small regions, which are 5 × 5 pixels. Next, high and low frequencies are enhanced in each small region. Each region is enhanced by a 3 × 3 Sobel operator for enhancing high frequencies and by using a 3 × 3 smoothing filter for enhancing low frequencies. Next, the enhanced frequency patterns in each small region are compared between the two images. The comparison of the frequency patterns is performed by computing the normalized correlation. By enhancing high frequencies, object shown in Figure 3 is extracted, and in some cases, noise may also be extracted, but environmental changes are not extracted. On the other hand, by enhancing low frequencies, the object is extracted, and in some cases, the region of the environmental changes may also be extracted, but noise is not extracted. Finally, the two results are integrated, and the regions that are extracted in both processes are detected as objects. As mentioned above, objects in a wide, open area can be detected by this method, which consists of enhancing multifrequency bands and comparing the small regions of the enhanced images, without being influenced by noise and environmental changes.

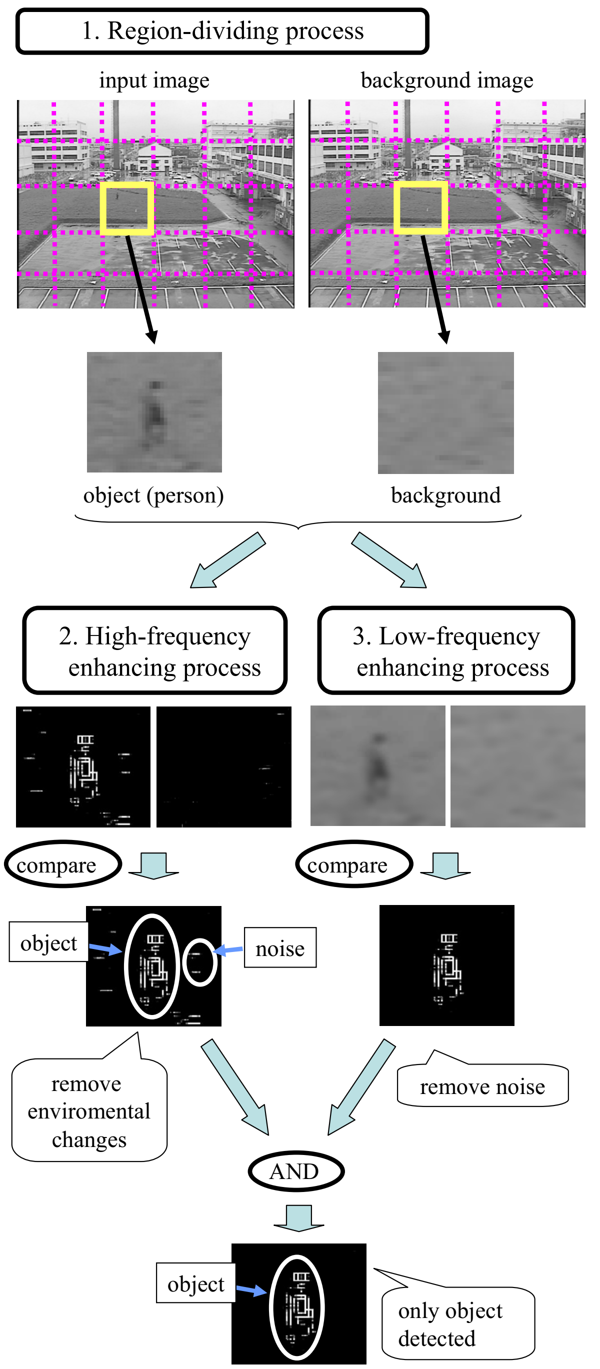


Figure 3. Frequency-Pattern-Enhanced Subtraction

2.4 Measures against Camera Vibration

The residual error (about ±2 pixels) of the corrected image by an image stabilizer is corrected in both high- and low-frequency enhancing processes in the Frequency-Pattern-Enhanced Subtraction. The residual error results in displacement of corresponding positions of the background and input images, as shown in Figure 4 (upper pictures), because the differential process produces a small amount of difference even for the pattern that actually belongs to the background. This small amount of difference causes both overlooking a small object about 5 pixels square in the image and false detection. Therefore, to solve this, the displacement of each small region is corrected so that the background and input image are best aligned in each frequency-enhancing process, as shown in Figure 4 (lower pictures). As mentioned above, objects in a wide, open area are detected robustly without being influenced by the camera vibration.

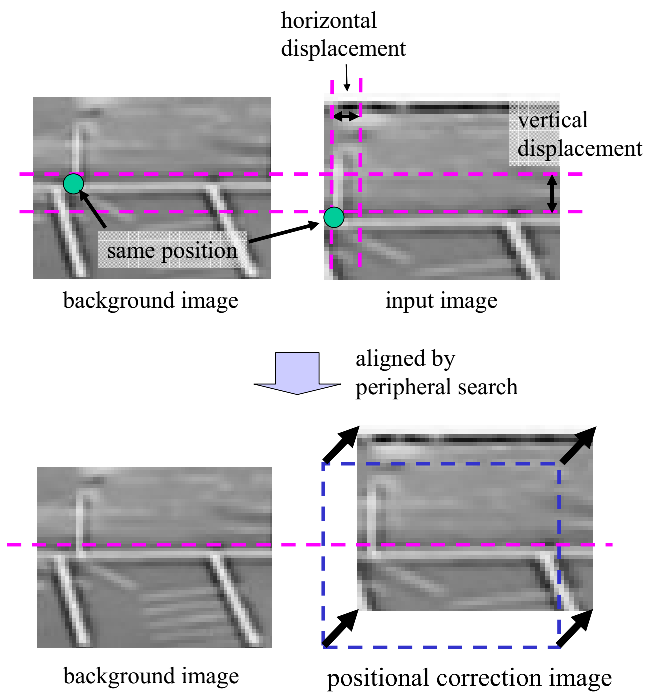


Figure 4. Positional correction of each small region

3. Evaluation

The presented method is evaluated on the typical real images. The images used for evaluation are collected under various environmental conditions, as listed below, and they last about 18 hours in total.

- surveillance range (ten types): 5 - 110 meters (width at center of image)
- weather: clear, cloudy, rainy
- time: day, night
- camera vibration (combination of horizontal, vertical, and rotational): maximum of ±20 pixels

The relationships between the surveillance range and detection rate under different background-to-object brightnesses are shown in Figure 5. In this graph, the X-axis represents the surveillance range and the Y-axis represents the detection rate; the line markers represent different background-to-object brightnesses. The processing time and specifications of the evaluation machine are shown in Table 2. The processing time was 300 ms/frame on a PC (with a Pentium IV CPU @ 2.2 GHz). That is sufficiently fast for wide-area surveillance because the objects move little in 300 ms, but it would need to be made faster if there is a consideration about extending into narrow-area surveillance. The presented method exhibited a detection rate of more than 95% in the 30 – 70 meter surveillance range when the object has a brightness difference compared to the background greater than or equal to 20 steps, where people can perceive generally, and a false detection rate of 5% or less at the time of surveillance. However, as shown in Figure 6, the detection rate decreases greatly for objects that have a background-to-object brightness difference of less than 10 steps. This problem is our future work.

Table 2. Processing time and specifications of evaluation machine

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

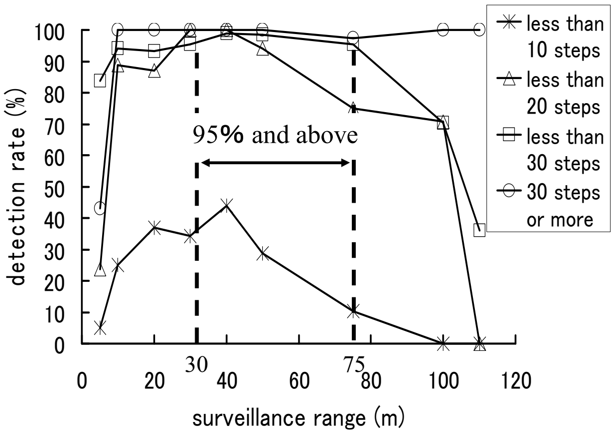


Figure 5. Detection rate as function of surveillance range at different brightnesses

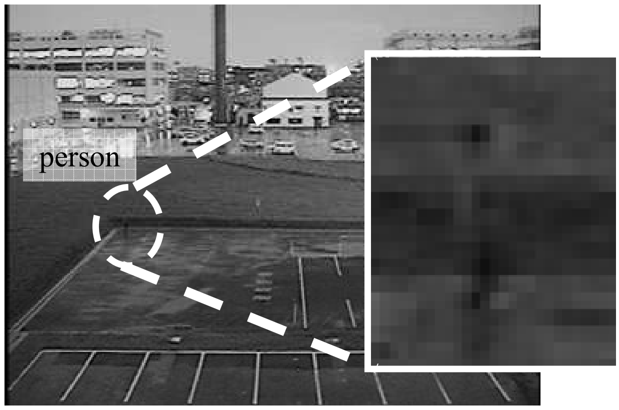


Figure 6. Primary factor in performance degradation (low-contrast object)

4. Conclusion

We presented the "Frequency-Pattern-Enhanced Subtraction", object-detection method, for surveillance of wide, open areas, such as harbors, rivers, and plant premises. Our method is a new background subtraction technique that enhances multifrequency bands and comparison between small regions in the enhanced images, and enables robust detection of small blurred objects without being influenced by noise and environmental changes. The evaluation results demonstrated that our method detected more than 95% of target objects in images of a large surveillance area ranging from 30 to 75 meters at their center if the background-to-object brightness difference is more than 20 steps. These results suggest that our method can be applied to wide-area, outdoor surveillance for such places as rivers and harbors and to emerging object-detection systems for use 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," ICIP, vol.3, pp. 716-719, 2001.

[2] N. Kobayashi, et al., "Human Tracking in Outdoors Using Spatio-Temporal Information," Proc. Second Symposium on Sensing via Image Information, pp. 173-176, 1996

[3] Y. Mae, et al., "Object tracking in cluttered background based on optical flow and edges," ICPR, vol.1, pp. 196-200, 1996.

[4] S. Fukui, et al., "Extraction of moving objects by estimating background brightness," the journal of the IIEEJ, vol.33, no.3, pp. 350-357, 2004