The Thirteenth International Symposium on Artificial Life and Robotics 2008(AROB 13th '08),
B-Con Plaza, Beppu, Oita, Japan, January 31-February 2, 2008

Extracting Moving Objects from a Video by Sequential Background Detection

Y. Sakai, M. Miyoshi, J. K. Tan, S. Ishikawa

Kyushu Institute of Technology, Sensuicho 1-1, Tobata, Kitakyushu 804-8550, Japan
({miyoshi, etheltan, ishikawa}@ss10.cntl.kyutech.ac.jp)

Abstract: This paper describes a technique for extracting moving objects from a video image sequence taken by a slowly moving camera. Background subtraction method is effective for extracting moving objects from a video provided by a fixed camera. But the latest background image should be employed for the subtraction in order not to be influenced by the light intensity change. A temporal median technique is proposed in this paper which detects the background sequentially. For a video image stream provided by a slowly moving camera, the camera motion is estimated using a registration algorithm and the temporal median filter is applied to the common image area among a set of successive image frames to extract the background. The technique was applied to the video images obtained from a hand-held camera and those taken from a camera set at the front seat of a car, and satisfactory results were obtained.

Keywords: Object detection, motion, background, temporal median filter.

I. INTRODUCTION

Automobile technology has come to the stage of realizing safe driving to prevent car accidents. Video cameras are set to observe outside of a car as well as inside of it to detect potential danger around the driving car. In particular, automatic detection of a human who is standing, walking or running close to the driving car is of great importance. In order to detect a human from a video image stream, various techniques have been developed in the computer vision field. Many of them assume a fixed camera and the fixed background [1], however, which cannot be employed in the video images taken from a running car. There are some techniques applicable to a mobile camera case, where they use optical flow to extract FOE [2]. But they are not very accurate when a car runs slowly at a junction, for example.

In this paper, we propose a temporal median technique for reconstructing the background from a video image sequence containing moving objects. Once the background is obtained, the moving objects including a human can be detected by the subtraction between the present image and the background. The proposed technique assumes that the observing camera is fixed or moving slowly such as a hand-held camera. But, at the same time, it simulates the observation from a car near a crossing or a junction where a car drives slowly, stops and starts again.

II. TEMPORAL MEDIAN FILTER

A median filter in image processing is a smoothing filter which makes the gray value of a chosen pixel change to a median value among the gray values of its 8-neighbor pixels and itself. This is a well known noise reduction filter. It often provides us with high performance in reducing random noise on an image in spite of its simple procedure. This median filter is employed in the proposed technique not in the spatial way but in a temporal way. It is therefore called a temporal median filter or a temporal median technique in the present paper.

Let us assume a fixed camera and denote a video image sequence taken by the camera by $f(m,n,t)$ ($m=1,2,\ldots,M$; $n=1,2,\ldots,N$; $t=1,2,\ldots,T$). Then a temporal median image $g(m,n,t)$ obtained from the original image $f(m,n,t)$ is defined by the following;

$$g(m,n,t)=\mathrm{med}\{f(m,n,t') \mid t'=t-F+1, t-F+2,\ldots,t-1,t\}. \quad (1)$$

Here, function med{} returns the median value of a set of numbers in {}. The idea of Eq.(1) is illustrated in **Fig. 1**.

Parameter F is a positive odd number not less than 3. It signifies that F sequential image frames are taken into account for the temporal median operation. Obviously, parameter F need be defined depending on how fast, or how slow, an object passes in front of a camera. If an object of M pixels wide move V pixels horizontally between successive image frames, α frames defined by

$$\alpha=\left[\frac{M}{V}+1\left(\frac{M}{V}-\left[\frac{M}{V}\right]\right)\right] \quad (2)$$

is necessary so that the object does not overlap anymore. Here $[x]$ is the largest integer not more than x (Gaussian bracket) and $1(x)$ is a step function defined by

$$1(x)=\begin{cases}1\ldots x>0\\0\ldots x\le 0\end{cases}. \quad (3)$$

The number α is thus a positive integer.

Parameter F should then satisfy the following inequality;

$$F \ge 2\alpha + 1 \quad (4)$$

This inequality claims that even a pixel on which an object stays the longest recovers the background gray value.

Since, from Eq. (1), the temporal median is a pixel-wise operation independent of adjacent pixels, it has nothing to do with the content of video images concerned. Hence no specific image processing is applied to the video before the application of Eq. (1). This makes the temporal median filtering fast and inexpensive.

III. VIDEO CAMERA CONDITION

If an observing camera is fixed at a certain location, Eq.(1) is solely applied to a video image stream and satisfactory experimental results are obtained. On the contrary, an image registration technique needs to be employed for a mobile camera case. We also need to assume that the movement of the camera is reasonably slow. Otherwise the temporal median operation does not make sense. If we consider a driving car, the above case video images are taken when a car slows down and stops at a crossing, for example.

We define some predefined small rectangle regions on an image to detect camera motion. Each region is found itself on the successive image by the employment of SSDA [3] and the displacement of the region between the two images gives a motion vector. Since it is not known if a region contains a moving object, we must find the most certain regions that may give camera motion. This is realized by employing RANSAC [4]. Once the camera motion is detected on the successive frames as a motion vector, common part on an image is identified among the successive F image frames and Eq.(1) is applied to the common part to yield the background.

In **Fig. 2**, (a) shows image frame sequence taken by a camera which is moving slowly to the right. A person is passing in front of the camera view from the right to the left. Then correlation of the background is computed between every successive image frames, and the matched regions are found as shown in (b), on which the temporal median is operated. The local regions for performing SSDA are placed on an image as illustrated in (c). After the calculation of motion vectors with respective regions, the most frequent motion vectors are chosen which is recognized to provide the camera motion.

IV. EXPERIMENTAL RESULTS

Some experiments have been done on recovering the background sequence from a video image under the existence of some persons passing in front of a video taking camera. The resultant background images are then used to extract the persons in the video.

Figure 3 is the result of person detection along with a bycicle employing hand-held camera images. Fig. 3(a) is the original images, (b) is the recovered background images, and (c) is the extracted moving objects. This results show expecting performance of the proposed technique.

The second experiment was performed on car video images. A video camera was fixed at a front seat of a car and took images of frontal road environments while the car was driving in the town. **Figure 4(a)** is the image including pedestrians crossing the road while the car is stopping. The proposed technique computed the background from the successive video frames containing the crossing people as shown in (b), and extracted them as given in (c).

V. DISCUSSION AND CONCLUSION

This paper proposed a temporal median technique for recovering the background under many passers in front of a camera. The obtained background was employed for extracting moving objects, in particular, persons from a video provided from a hand-held camera and also from the camera fixed at a front seat in a car. As is understood from the experiments, performance of the presented technique is reasonable with respect to the video images provided from a fixed or a slowly moving camera, whereas the results contain more noise when the car runs faster. Thus the proposed technique can be employed in particular in moving objects detection near crossings where a car slows down, stops and starts again. This is of great importance in a practical sense, since many human vs. car accidents happen near or in the crossings or junctions.

The technique has another advantage over existent moving object detection techniques employing background subtraction in that it is insensitive to light intensity change as it recovers the background from the latest video image frames. It is therefore insensitive to sun light change or weather change in a day.

If a person sits or stands still, he/she is part of the background and cannot be extracted. For a walking/ running person who stops in the video image, a former background can be employed for extracting him/her. Both a fast walker and a slow walker could be included in a video. Then the number of frames F to which frames the temporal median filter is applied needs to be defined considering the slow walker, resulting in a larger number of F.

The proposed technique can well be applied to those

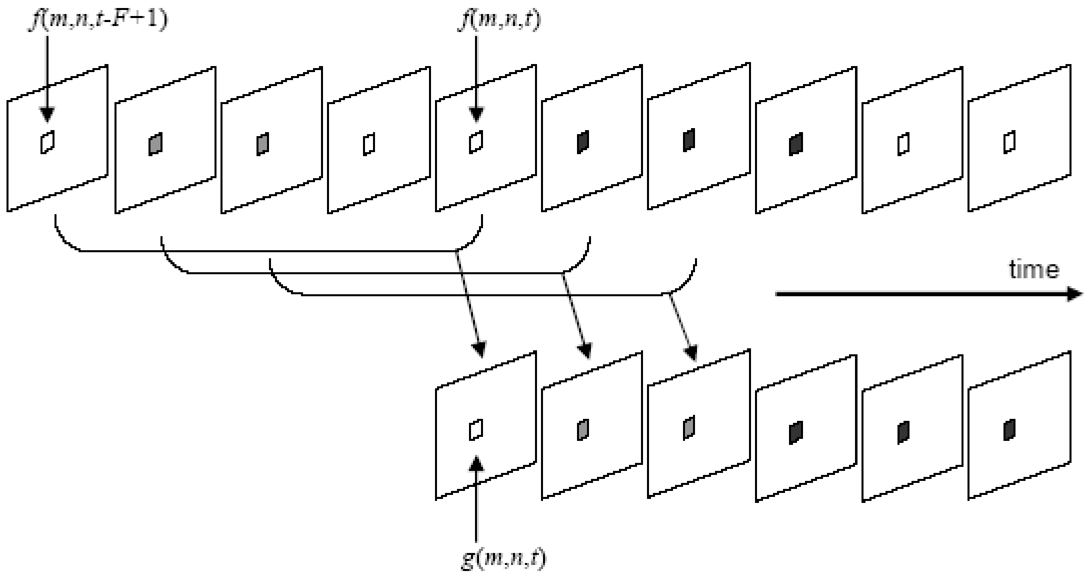

Fig. 1. Operation of the temporal median. In this example, pixel (m,n) is focused on throughout the time lapse.

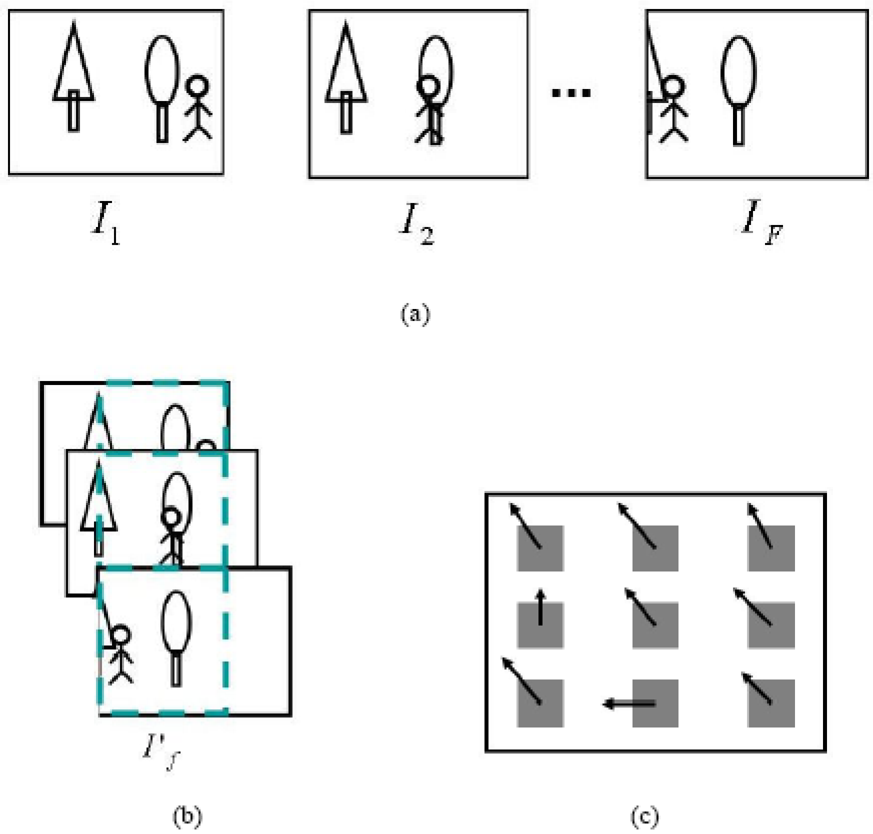

Fig. 2. Procedure for a mobile camera: (a) Image sequence; (b) common part on respective images; and (c) local regions to find motion vectors.

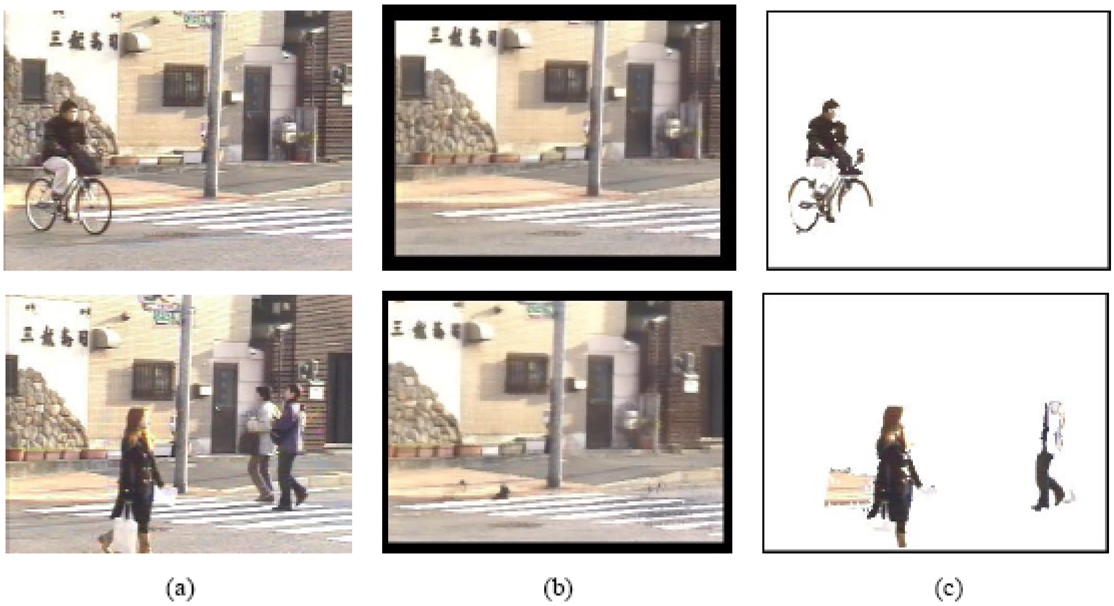

Fig. 3. Experimental result _1: Images from a hand-held video camera. (a) Original video images; (b) the obtained backgrounds; and (c) extracted moving objects.

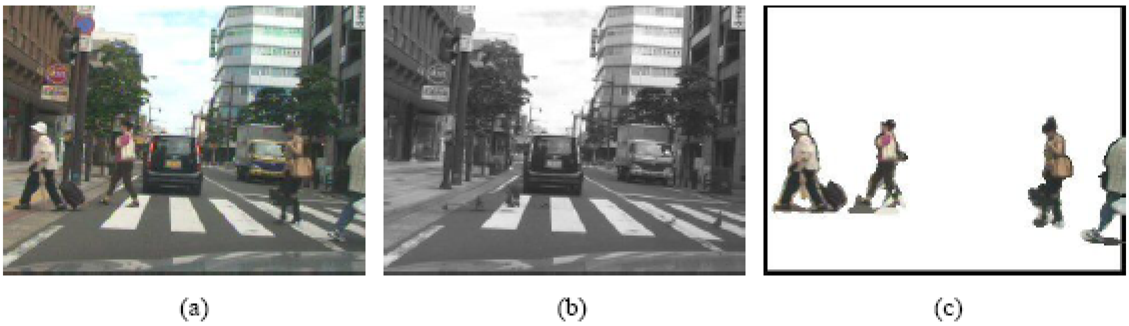

Fig. 4. Experimental result _2: Images from a video camera mounted on a car. (a) Original video images; (b) the obtained backgrounds; and (c) extracted moving objects.

objects moving perpendicular to the light axis of an observing camera, i.e., from one side to the other in an image. The technique is weak when objects move along the light axis. In a practical case, appropriate camera setting may solve this difficulty.

REFERENCES

[1] Satoh Y, Kaneko S, Niwa Y, Yamamoto K (2003), Robust object detection using a radial reach filter. IEICE Trans. Inform. Syst., J86-D-II(5):616-624

[2] Zhang Y, et al. (2006), Robust moving object detection at distance in the visible spectrum and beyond using a moving camera. Proc. of CVPR Workshop

[3] Barnea EI, Silverman HF (1972), A class of algorithms for fast digital image registration. IEEE Trans. on Comput., C-21:179-186

[4] Fischler MA, Bolles RC (1981), Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communication of ACM, 6(24):381-395

