Proceedings of the 3rd IASTED International Conference
VISUALIZATION, IMAGING, AND IMAGE PROCESSING
September 8-10, 2003, Benalmadena, Spain

SEQUENTIAL MONTE-CARLO ESTIMATION OF BACKGROUND IMAGE FOR BACKGROUND SUBTRACTION UNDER CHANGING ILLUMINATION

Masaru Tsuchida, Takahito Kawanisi, Hiroshi Murase and Shigeru Takagi

NTT Communication Science Laboratories, NTT Corporation
3-1 Morinosato-Wakamiya, Atsugi, 243-0198, JAPAN

{tsuchida, kawanisi, murase.takagi}@eye.brl.ntt.co.jp

Abstract

This paper proposes a background image estimation method for background subtraction under changing illumination. In our method, a background image is generated as a weighted linear combination of the prepared background images. The weights change slowly as time advances. We therefore adopt the sequential Monte-Carlo method to track the weights, and background images are generated. A generated image in which area having pixel value equal to that of the input image is the largest is selected as the background image under the illumination. In the experiments, background estimation and subtraction were carried out under changing illumination conditions; i.e., geometrical setup, brightness, and color. The results show that background images are correctly estimated even if 80% of the area of the input image is occupied by a foreground object.

Key Words
tracking, background subtraction, background estimation, illumination change, sequential Monte-Carlo estimation.

1. Introduction

Background subtraction is a simple and effective technique for real time detection of a foreground object and is used for many applications, such as video surveillance systems. This technique adapts well to systems whose camera position and viewing angle are fixed. A background image is prepared in advance and subtracted from an input image. By evaluating the differences between the two images, objects in the foreground can be detected.

Here, let us consider the situation where the illumination condition (e.g. geometrical setup, brightness, and color) is changing. The appearance of the background is influenced by the illumination condition, and the background image changes with illumination variation. Using the background subtraction technique in that situation would require a background image for each illumination condition. However, it is infeasible to collect all background images under all possible illumination conditions.

One approach to overcoming this problem is background modeling. In this method, values invariable to illumination change are used. Recently, this method is becoming available even for non-stationary background [1-5]. However, most of the algorithms are not simple, and do not deal with illumination color change.

A second approach is to update the background image. Using this approach to track foreground objects requires accurate real-time estimation of background images. Updating methods are based on temporal blending. A new background image is generated as a weighted linear combination of the last background image and a new input image [6,7]. This technique can adapt to illumination change quickly. However, foreground objects on an input image become part of the updated background image, and it causes some errors. Moreover, the continuous illumination change also causes continuous errors.

A third approach is to estimate background image by projecting the input image onto an eigen-space [8]. The eigen-space is calculated from a set of background images taken under various illumination conditions. The background image for each input image is generated by projecting the input image onto the eigen-space [9]. However, when the area of foreground objects on an input image is large (e.g., when some people are around the camera), this method does not work well.

This paper proposes a novel approach in which we generate background images as weighted linear combinations of prepared background images and select one in which area having pixel value equal to that of the input image is the largest. The number of possible combinations of weights is huge and the weights change slowly as time advances. We therefore adopt the sequential Monte-Carlo method [10] to track the weights. This method is based on a probabilistic framework and enables us to reduce the computation. It has often been used for robust tracking of an agile object. Experiments show that our method works well even if foreground objects occupy a large area of the input image.

2. Background Estimation under Changing Illumination

In the proposed method, a background image for each illumination condition is generated as a weighted linear combination of prepared background images captured under sparsely sampled illumination conditions. Let us agree to the following three assumptions.

- Camera position and viewing angle are fixed.
- Exposure parameters are adjusted to avoid camera saturation.
- Auto correction functions (e.g. gain control, and white balance control) are turned off.

Determining the weights of the prepared images means tracking illumination variation. To track the weights, we adopt the sequential Monte-Carlo method [10], which is reviewed in Sec. 2.1. The method for determining the weights is presented in Sec. 2.2. In developing the method and in the experiments, omni-directional images [11] were used, but the proposed method is available for other types.

2.1 Sequential Monte-Carlo Method

The problem of tracking states of an object can be formulated in a probabilistic framework by representing tracking as a process of conditional probability density propagation. Let the state of the object at time t be $\mathbf{X}_t$, the observation result from a sensor at time t be $\mathbf{S}_t$, and the observation sequence from time 1 to t be $S_t = \{\mathbf{S}_1, \cdots, \mathbf{S}_t\}$. Let the prior probability density of states at time t after observation sequence S_{t-1} be $p(\mathbf{X}_t \mid S_{t-1})$, and the likelihood π_t of getting observation $\mathbf{S}_t$ if object is in state $\mathbf{X}_t$ be $p(\mathbf{S}_t \mid \mathbf{X}_t)$. Then the posterior probability density of states at time t after observation sequence S_t is calculated by the Baysian rule as

$$p(\mathbf{X}_t \mid S_t) = k_t p(\mathbf{S}_t \mid \mathbf{X}_t) p(\mathbf{X}_t \mid S_{t-1}), \tag{1}$$

where k_t is a normalization constant. If the state transition probability $p(\mathbf{X}_t \mid \mathbf{X}_{t-1})$ is known and the Markov property is satisfied, then the prior probability density can be calculated as

$$p(\mathbf{X}_t \mid S_{t-1}) = \int p(\mathbf{X}_t \mid \mathbf{X}_{t-1}) p(\mathbf{X}_{t-1} \mid S_{t-1}) d\mathbf{X}_{t-1}. \tag{2}$$

The sequential Monte-Carlo method samples the state space with N points at time t; $\mathbf{S}_t = \{\mathbf{s}_{t,1}, \cdots, \mathbf{s}_{t,N}\}$, calculates the probability density $p(\mathbf{X}_t \mid S_t)$ of the sample point by Eqs. (1) and (2), and decides new N sample points using the probability density and state transition probability as follows (Fig. 1):

Step 1. Weight each sampled state with a value corresponding to their relative likelihood, $\pi_t = \{\pi_{t,1}, \cdots, \pi_{t,N}\}$. The n th likelihood $\pi_{t,n}$ is calculated by comparing the observation result $\mathbf{s}_{t,n}$ with a known observation model of the object. This weighted set of samples represents the approximation of the posterior at time t.

Step 2. Select each sample of $\mathbf{S}_t$, $\mathbf{s}_{t,n}$ $N\pi_{t,n}$ times, and then obtain a new set of samples $\tilde{\mathbf{S}}_{t+1} = \{\tilde{\mathbf{s}}_{t+1,1}, \cdots, \tilde{\mathbf{s}}_{t+1,N}\}$. Some elements with relatively low weight may not be chosen at all.

Step 3. Shift sample $\tilde{\mathbf{s}}_{t+1,n}$ according to the state transition probability $p(\mathbf{s}_{t+1,n} \mid \mathbf{s}_{t,n}) = p(\mathbf{x}_t \mid \mathbf{x}_{t-1})$, and then obtain a set of samples at time $t+1$, $\mathbf{S}_{t+1} = \{\mathbf{s}_{t+1,1}, \cdots, \mathbf{s}_{t+1,N}\}$.

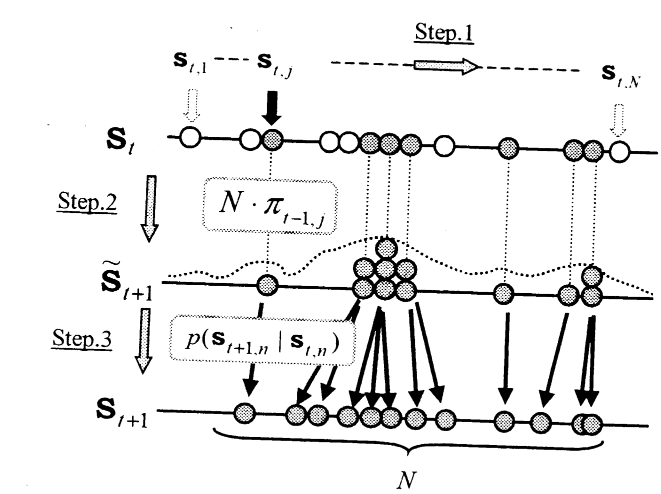


Fig. 1. Diagram of the sequential Monte-Carlo method.

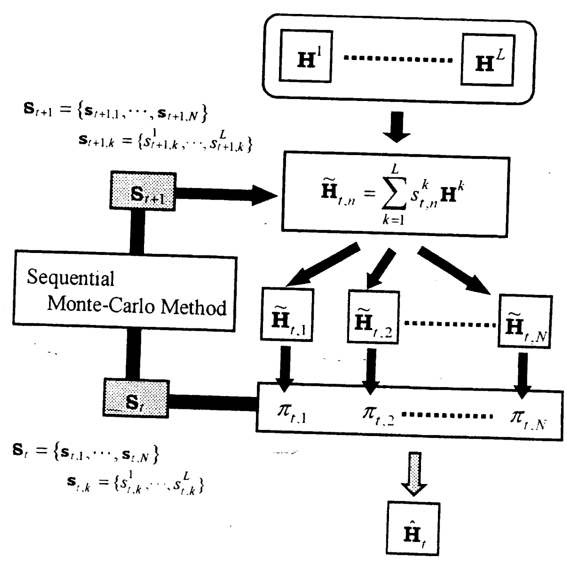


Fig. 2. Diagram of background image estimation.

2.2 Background Estimation

Figure 2 is a diagram of the proposed method. First, background images under sparsely sampled illumination conditions are .captured and stored. Let the stored background images be $\mathbf{H}^k$ ($k = 1,\cdots,L$), the set of weight vectors at time t be $\mathbf{S}_t = \{\mathbf{s}_{t,1},\cdots,\mathbf{s}_{t,N}\}$, and the weight vector be $\mathbf{s}_{t,n} = (s_{t,n}^1,\cdots,s_{t,n}^L)$. Then N background images are generated using $\mathbf{H}^k$ and $\mathbf{s}_{t,n}$. The n th generated background image at time t is defined as

$$\widetilde{\mathbf{H}}_{t,n} = \sum_{k=1}^{L} s_{t,n}^k \mathbf{H}^k . \quad (3)$$

Concerning the background area, a correct background image is the same as the input image. Then, we select an image from the background images based on the idea that the area on an estimated image, whose pixel value is equal to input image, becomes largest when the estimated image is correct. Likelihood $\pi_{t,n}$ of a generated background image $\widetilde{\mathbf{H}}_{t,n}$ is defined as

$$\pi_{t,n} = k_n \sum_y \sum_x f(i_{x,y} - \widetilde{h}_{x,y}) , \quad (4)$$

where k_n is a normalization constant, $i_{x,y}$ and $\widetilde{h}_{x,y}$ are the pixel value of an input image $\mathbf{I}_t$ and an estimated background image $\widetilde{\mathbf{H}}_{t,n}$ at (x, y). Let the threshold be ρ, and the function $f(i_{x,y} - \widetilde{h}_{x,y})$ in Eq. (4) is defined as

$$f(i_{x,y} - \widetilde{h}_{x,y}) = (\rho - | i_{x,y} - \widetilde{h}_{x,y} |)^2 \quad \text{when } | i_{x,y} - \widetilde{h}_{x,y} | \leq \rho , \quad (5a)$$

$$f(i_{x,y} - \widetilde{h}_{x,y}) = 0 \quad \text{when } | i_{x,y} - \widetilde{h}_{x,y} | > \rho . \quad (5b)$$

The generated image whose likelihood $\pi_{t,n}$ is maximum is selected as the background image at time t .

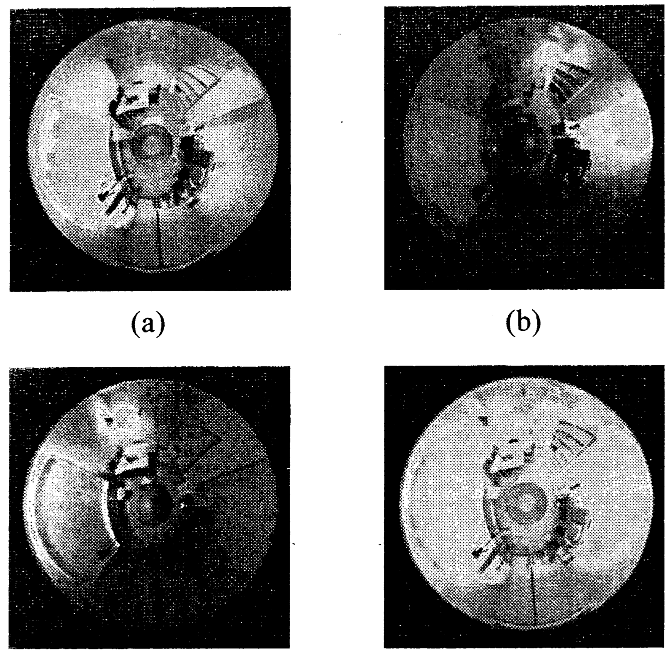
The set of weight vectors at time $t+1$ is generated from the set of weight vectors at time t as follows

1) Select a weight vector $\mathbf{s}_{t,n}$ at the probability proportional to its likelihood $\pi_{t,n}$.
2) Shift the $\mathbf{s}_{t,n}$ by the dynamic model, and let it be a weight vector at time $t+1$.
3) Repeat 1) and 2) N times.

The set of weight vectors at time 0 is generated randomly.

3. Experiments

We conducted experiments to show that our method works well even if a large area of the input image is occupied by foreground object. Images were captured in a room environment with a color omni-directional camera consisting of a hyperbolic mirror mounted above the camera lens [11]. This camera provides a 360-degree field of view. To change the illumination condition, we used three lights, a, b, c (a: fluorescent lamp; b, c: incandescent lamps). The color and brightness of the background was changed by changing the combinations of lamp brightness.



(a) (b) (c) (d)

Fig. 3. Prepared background images.
Four images captured under different illumination conditions are used for image synthesizing. (a) Fluorescent lamp, (b) incandescent lamp on the right wall, (c) incandescent lamp on the left wall, and (d) all lamps.

Fig. 4. Examples of input images
The percentage of foreground area is 20% (left), 50% (center), and 80% (right)

As background images for generating new background images, four images were collected when one of the lamps was on or all lamps were on (Fig. 3). Next, 150 images were collected under changing combinations of lamp brightness. To increase the area of a foreground object, part of the image was artificially modified to white. And these images were used as input images. Figure 4 shows three examples from 150 images. The percentages of foreground area are around 20, 50, and 80%. The proposed method is implemented on Pentium IV 1.8 GHz PC, and it runs at almost 3 frames/sec when 100 samples per time step ($N = 100$) are used on 320x240 images.

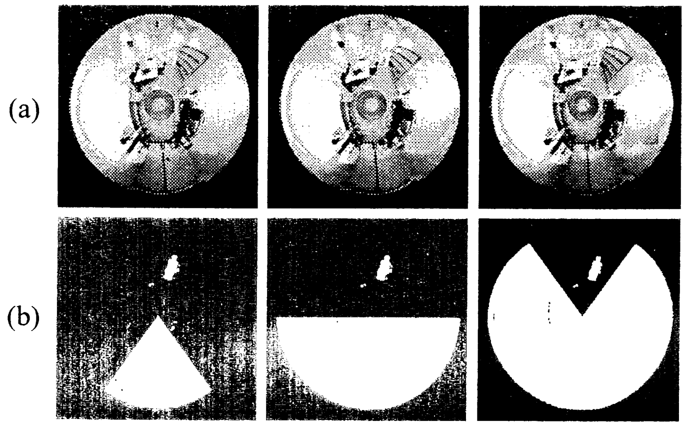


Fig. 5. Results for the proposed method.
(a) Estimated background image. (b) Detected image difference.

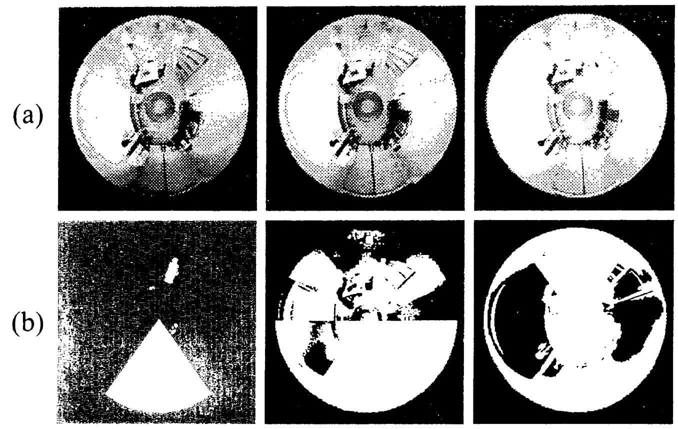


Fig. 6. Results for the conventional method.
(a) Estimated background image. (b) Detected image difference.

By our method, the background images in Fig. 4 were estimated as shown in Fig. 5(a) and subtraction images became as shown in Fig. 5(b). In Fig. 5(b), the fan-shaped area is the artificially modified area and the small area above it is the foreground object, in this case, a person. From the subtraction images, it may be said that background images are correctly estimated for all cases.

By the conventional method based on linear projection onto eigen-space [1], the background images in Fig. 4 were estimated as shown in Fig. 6(a) and subtraction images became as shown in Fig. 6(b). The conventional method fails to estimate background when foreground size is large (50%, 80%).

Figure 7 shows the root mean square error (RMSE) between the input image and the estimated background image at the background area. The RMSE of the proposed method is small and almost constant. On the other hand, the RMSE of the conventional method becomes large when the foreground area is over 40%.

Next, the person's area on each image was extracted by hand (Fig. 8), and true-positive and false-positive rate of person detection were calculated. As shown in Fig. 9, the false-positive rate for our method does not vary so much and seems to be little influenced by foreground area. On the other hand, the values for the conventional method become large when foreground ratio is over 40%.

These results show that the proposed method works well even if the foreground area is over 50% of the input image.

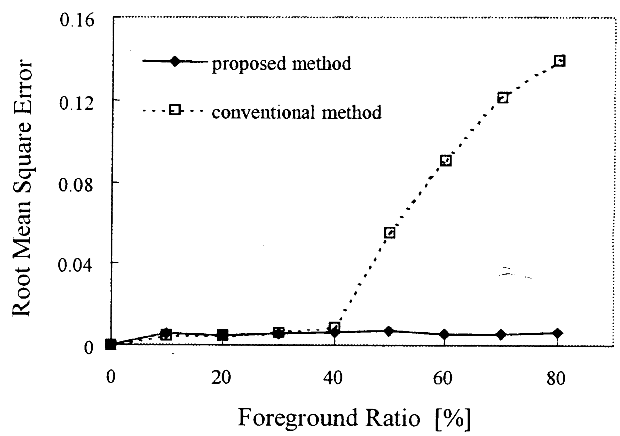
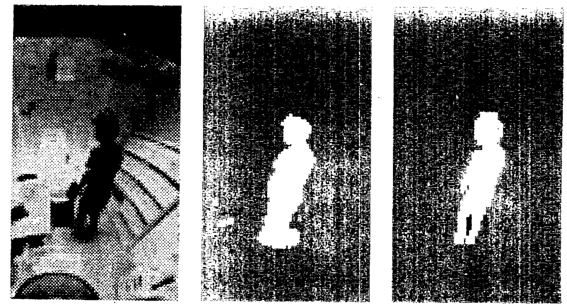


Fig. 7. Evaluation of estimated images.
Straight line: the proposed method, dash line: the conventional method.



(a) (b) (c)

Fig. 8. Extraction of foreground object.
(a) A part of input image, (b) experimental result, (c) extracted by hand.

4. Conclusion

This paper proposed a novel method to estimate a background image under changing illumination. A background image is generated as a weighted linear combination of background images collected at dispersed points in the illumination space. The weight vector used for estimation is determined by the sequential Monte-

Carlo method. Experiments were conducted in a room environment and under changing brightness of three lamps. The experiment shows that background images are correctly estimated even if 80% of the area of the input image is occupied by foreground object.

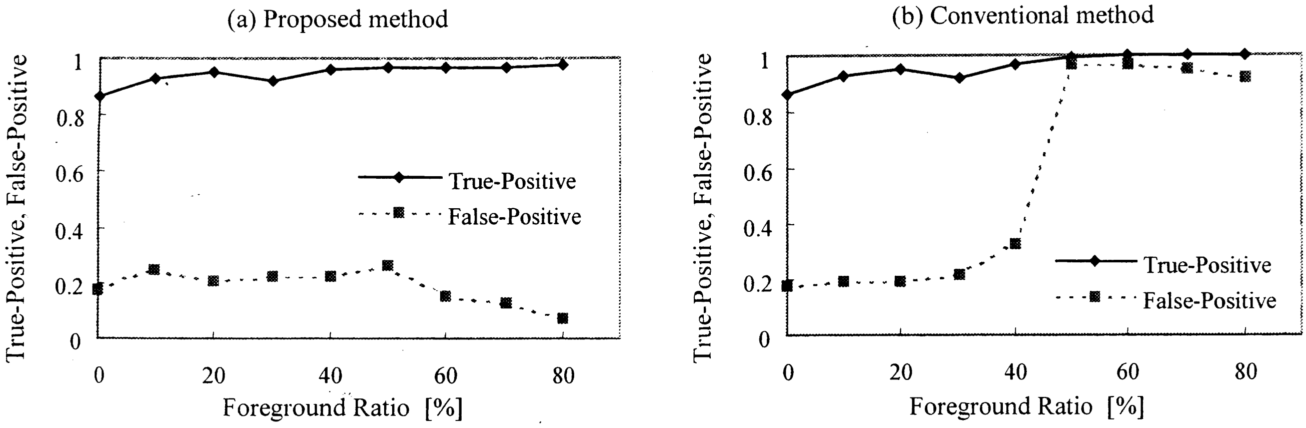


Fig. 9. Evaluation of person detection

References

[1] T. Mastuyama, T. Ohya and H. Habe, Background Subtraction for Non-Stationary Scenes, *Proc. IEEE 4th Asian Conference on Computer Vision*, 2000, 662-667.

[2] M. Seki, H. Fujiwara and K. Sumi, A Robust Background Subtraction Method for Changing Background, *Proc. WACV 2000*, 2000, 207-213.

[3] Y. Ren, C. Chua and Y.Ho, Motion Detection with Non-stationary Background, *Proc. ICIAP 2001*, 2001, 78-83.

[4] N. Ohta, A Statistical Approach to Background Subtraction for Surveillance Systems. *Proc. ICCV 2001*, 2001, 481-486.

[5] M. Lamarre and J. J. Clark, Background subtraction using competing models in the block-DCT domain, *Proc. ICPR 2002*, 2002, 299-302.

[6] L. Li, W. Huang, I. Y. H. Gu and Q. Tian, Foreground Object Detection in Changing Background Based on Color Co-Occurrence Statistics, *Proc. WACV 2002*, 2002, 269-274.

[7] C. Stauffer and W.E.L. Grimson, Adaptive Background Mixture Model for Real-time Tracking, *Proc. CVPR 1999*, 1999, 246-252.

[8] T. Mastuyama, T. Wada, H. Habe and K. Tanahashi, Background Subtraction under Varying Illumination, *IEICE Trans. D-II, Vol. J84-D-II, No.10*, 2001, 2201-2211 (in Japanese).

[9] P.N. Belhumeur and D.J. Kriegman, What is the set of images of an object under all possible lighting conditions? , *Proc. CVPR*, 1996, 270-277

[10] M. Isard and A.Blake, Condensation - conditional density propagation for visual tracking, *Int. J. of Computer Vision, 29*(1), 1998, 5-28.

[11] H. Ishiguro, Development of low-cost compact omnidirectional vision sensor and their applications, *Proc. Int. Conf. Information Systems, Analysis and Synthesis*, 1998, 433-439.