

Adaptive Background Estimation and Shadow Removal in Indoor Scenes

Junya Morita*, Yoshio Iwai*, Masahiko Yachida*
Graduate School of Engineering Science
Osaka University

Abstract

Background subtraction algorithm is susceptible to both global and local illumination changes such as shadows, sunlight and reflection. These changes sometimes cause failures of object tracking, gesture recognition and posture estimation. In this paper, we propose a method for detecting objects casting shadows in an indoor scene by modeling a pixel value as the total energy received from light sources. Our proposed method can also estimate a rate of illumination change in order to improve background subtraction. The usefulness of the method is demonstrated by experimental results.

1 Introduction

Background subtraction method is one of the effective methods for detection of moving objects in static scenes because of fast processing. The method is, however, susceptible to both global and local illumination changes such as shadows, sunlight and reflection. These changes often cause failures of object tracking. In order to track moving objects stably, the object region should be correctly extracted against the effects of such illumination changes.

Many methods for removing illumination effects have been proposed. Gordon et al. use stereo images and color information for object detection. This method integrates depth estimation with background subtraction[1]. There is a method using the difference between two frames, however the method is based on heuristic assumptions[2, 3].

On the other hand, many statistical approaches are also proposed[4, 5, 6, 7, 8]. Stauffer et al. use a mixture Gaussian model for changes of background pixel values[7]. The relation between reflection on an object surface and a value of pixel is often statistically modeled by using Gaussian. Ohta uses chi-square test for detecting outliers[8]. These methods implicitly assume that there is one light source in a scene, or that there are one colored lights in a scene. However, there are many colored lights in real indoor scenes: tungsten-light (3500K), natural daylight color (5000K). Therefore, we need a method which can deal with multiple colored light conditions in order to apply background subtraction in indoor scenes.

In this paper, we propose a method for detecting objects casting shadows in an indoor scene by modeling a pixel value as the total energy received from lights. Our proposed method can deal with multiple lights and multiple colored lights as well. In order to deal with illumination changes, our method initially estimates the ratio between radiance of input image and that of background image estimated in advance.

*Address: 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan.
E-mail: {morita, iwai, yachida}@yachi-lab.sys.es.osaka-u.ac.jp

2 Background Model

Let the reflection coefficient matrix of an object surface be \mathbf{r}_j , and intensity of the illumination at that point j from light source i be \mathbf{L}_{ij} . We assume that a pixel value E_j is given by the following equation:

$$\mathbf{E}_j = \mathbf{r}_j \mathbf{L}_{ij}, \tag{1}$$

where \mathbf{r}_j includes albedo and specular reflection of an object surface, and \mathbf{L}_{ij} also includes rays reflected by walls and a ceiling. We introduce a parameter S_{ij} in order to deal with shadows. Equation 1 is rewritten as follows:

$$\mathbf{E}_j = \mathbf{r}_j S_{ij} \mathbf{L}_{ij}, \tag{2}$$

where S_{ij} expresses the degree of shut of light source i at point j and the range of S_{ij} is from 0 to 1. The light source i is shut-out at point j if S_{ij} is 0. In the case of multiple light sources, the relation between lights and a pixel value E_j is defined as follows:

$$\mathbf{E}_j = \mathbf{r}_j \sum_i S_{ij} \mathbf{L}_{ij}. \tag{3}$$

The above equation can be rewritten by substituting \mathbf{L}_{ij} for $\mathbf{r}_j \mathbf{L}_{ij}$ as follows:

$$\mathbf{E}_j = \sum_i S_{ij} \mathbf{L}_{ij}. \tag{4}$$

By using the above equation, we can use a background subtraction method in indoor scenes illuminated by multiple lights. \mathbf{L}_{ij} can preliminary be estimated in indoor scenes because lights in a room can often be controllable by using dimmer controller or switches. Once \mathbf{L}_{ij} is known, we can estimate S_{ij} from equation 4.

3 Background Subtraction in Consideration of Illumination Changes

For simplicity, we only describe the proposed algorithm in the case of two light sources. Our algorithm is also used in the case of three or more light sources. The flow of the estimation processes is shown in figure 1.

3.1 Estimation of Background

In the case of two light sources, equation 4 can be written as follows:

$$\mathbf{E}_j = S_{1j} \mathbf{L}_{1j} + S_{2j} \mathbf{L}_{2j}. \tag{5}$$

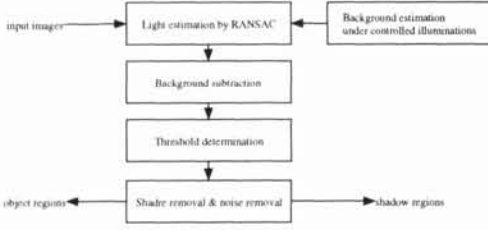


Figure 1: System flow of estimation process

As described in section 2, \mathbf{L}_{1j} and \mathbf{L}_{2j} are required in order to synthesis background images. From equation 5, if we get the two or more different pairs of (S_{1j}, S_{2j}) , we can estimate the background components: $\mathbf{L}_{1j}, \mathbf{L}_{2j}$. The shutting parameter of S is also considered as the parameter of brightness of a light source in a static scene. Therefore, we can estimate \mathbf{L}_{1j} and \mathbf{L}_{2j} from the two or more background images where the brightness of light sources is controlled by using dimmer controller and switches. Generally, in the case of N light sources, we can estimate N background components when we get N background images illuminated by various known brightness of light sources.

3.2 Estimation of Light Condition

In real indoor scenes, light condition is easily changed by turning on/off lights. The algorithm should be able to deal with changes of light condition. There are two changes of light condition: color and brightness. We only deal with the change of brightness.

S_{1j} and S_{2j} in equation 5 uniformly change at each pixel when the brightness of lights is changed. Therefore, if two or more pixels at arbitrary points are selected from an input image, we can estimate \tilde{S}_1, \tilde{S}_2 from the estimated background components, $\tilde{\mathbf{L}}_{1j}, \tilde{\mathbf{L}}_{2j}$. However, there is a problem of pixel selection from an input image for estimation. In order to reduce effects of incorrect selection, we use RANSAC algorithm for selection of pixel pairs. The algorithm is below:

1. Select two pixels ($j = 1, 2$) randomly.
2. Calculate S_1 and S_2 by using the following equation:

$$\begin{aligned} T_1 &= S_1 \tilde{T}_{11} + S_2 \tilde{T}_{21}, \\ T_2 &= S_1 \tilde{T}_{12} + S_2 \tilde{T}_{22}, \end{aligned} \quad (6)$$

where T is the sum of RGB components.

3. Accept (S_1, S_2) if the following condition is satisfied, and otherwise drop them:

$$0 \leq S_1, S_2 \leq 1. \quad (7)$$

4. Iterate the above process until the number of accepted points is greater than a threshold.
5. Make histograms of S_1, S_2 and select the maximum peak point for the estimates of S_1, S_2 .

The example is shown in figure 2. In this case, the step of histogram is 0.01 and $(\tilde{S}_1, \tilde{S}_2)$ are estimated to be (0.87, 0.74).

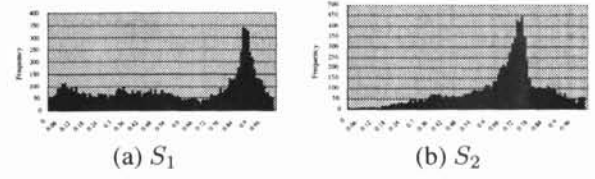


Figure 2: Histogram of S_1 and S_2

3.3 Background Subtraction

3.3.1 Shadow Removal

In the case of two light sources, an input image is classified into 4 regions:

1. background illuminated by two lights: $\mathbf{L}_1, \mathbf{L}_2$
2. shadow illuminated by \mathbf{L}_1 (\mathbf{L}_2 is blocked)
3. shadow illuminated by \mathbf{L}_2 (\mathbf{L}_1 is blocked)
4. object

The background and shadow vectors can be expressed by using $\tilde{\mathbf{L}}_{1j}$ and $\tilde{\mathbf{L}}_{2j}$. The object vector is different from these vectors. Soft shadows are expressed by the linear combination of $\tilde{\mathbf{L}}_{1j}$ and $\tilde{\mathbf{L}}_{2j}$. In real scenes, many soft shadows are distributed around the edge of parallelogram of $\tilde{\mathbf{L}}_{1j}$ and $\tilde{\mathbf{L}}_{2j}$. Therefore, we assume that soft shadows exist around the four segments of the parallelogram. From this assumption, We classify pixels into object and background by the following rules:

1. background if $\|\mathbf{I}_j - (\tilde{\mathbf{L}}'_{1j} + \tilde{\mathbf{L}}'_{2j})\| \leq T_a$ is satisfied.
2. shadow if $\|\mathbf{I}_j - (\tilde{\mathbf{L}}'_{1j} + \tilde{\mathbf{L}}'_{2j})\| > T_a$, $f \leq T_b$ and $\|\mathbf{I}_j\| > T_c$ are simultaneously satisfied.
3. otherwise pixels are classified into object.

We calculate feature value f as follows:

$$\begin{aligned} d_{min} &= \min_{0 \leq \alpha \leq 1, 0 \leq \beta \leq 1} \left[\|\mathbf{I}_j - \alpha \tilde{\mathbf{L}}'_{1j}\|, \right. \\ &\quad \|\mathbf{I}_j - \beta \tilde{\mathbf{L}}'_{2j}\|, \\ &\quad \|\mathbf{I}_j - (\alpha \tilde{\mathbf{L}}'_{1j} + \tilde{\mathbf{L}}'_{2j})\|, \\ &\quad \left. \|\mathbf{I}_j - (\tilde{\mathbf{L}}'_{1j} + \beta \tilde{\mathbf{L}}'_{2j})\| \right], \end{aligned} \quad (8)$$

$$f = \frac{d_{min}}{\|\mathbf{I}_j\|}, \quad (9)$$

where \mathbf{I}_j is a pixel value of an input image and $\|\mathbf{I}_j\|$ is a distance from the origin.

3.3.2 Threshold Determination

T_a, T_b, T_c should be assigned to valid values in order to distinguish objects from the background correctly. We describe a method for detection of T_b . The distribution of the histogram of feature f is double-peaked of object and background (shadow). We use Otsu's method[9] for automatic threshold determination. This method is used for binarization of gray level images, but is also used for threshold determination of feature used for our method by regarding feature values as gray level.

Valid values of T_a, T_c are slightly changed under various indoor scene and we use the values determined empirically.

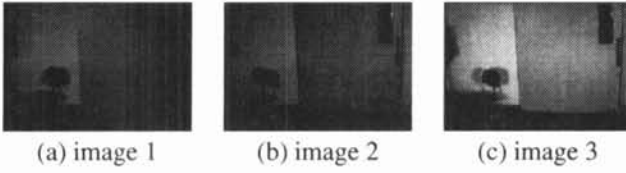


Figure 3: Backgrounds captured under various light conditions

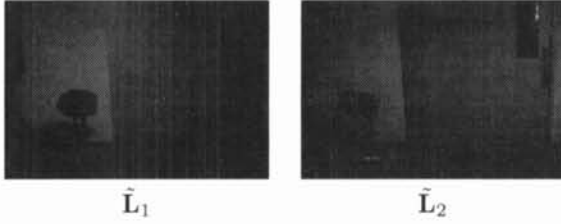


Figure 4: Estimated background components

4 Experiments

We conduct experiments for evaluation. The light source used in the experiments has three fluorescent lights and turn on/off each light independently. The change of brightness is made by switching these three fluorescent lights. The direction and position of lights and camera are fixed. In order to avoid the effect of other light sources, the experiments are conducted at night.

4.1 Experimental Results

Figure 3 shows background images illuminated by two different-colored lights. Figure 4 shows the background components estimated from the background images shown in figure 3. Figure 5 shows an experimental result. Figure 6 also shows an extraction result in the case of the same-colored two lights. Figure 7 shows an experimental result using three different-colored lights.

From the results, effects of shadows are largely removed and the object is correctly extracted. Moreover, even the colors of lights are different from each other, the object regions are accurately detected.

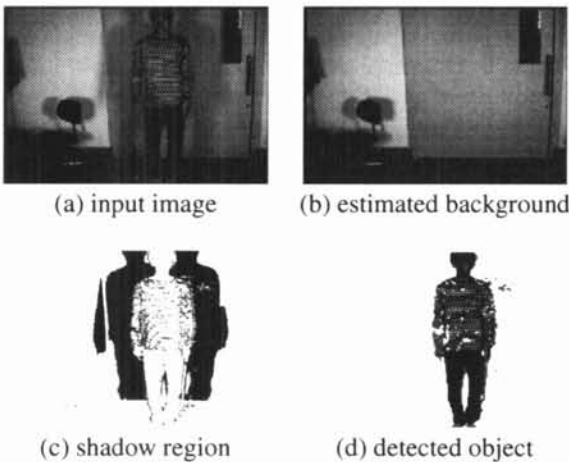
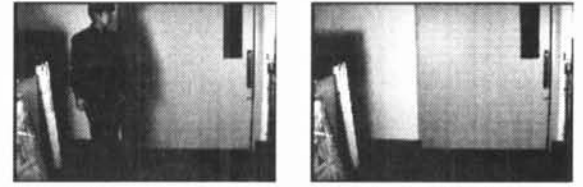
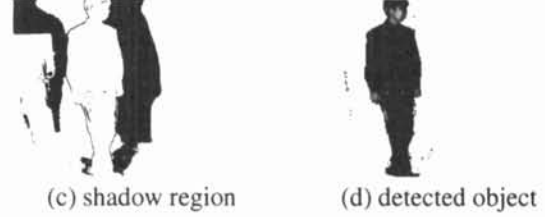


Figure 5: Result images (2 lights, different color)



(a) input image (b) estimated background

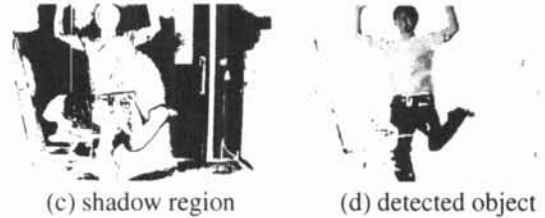


(c) shadow region (d) detected object

Figure 6: Result images (2 lights, same color)



(a) input image (b) estimated background



(c) shadow region (d) detected object

Figure 7: Result images (3 lights, different color)

4.2 Stability of Light Estimation

We also conduct an experiment for evaluation under various light conditions. We calculate the difference between the input image shown in the left column of figure 9 and the synthesized image shown in the right column of the figure. The background components used for this experiment are shown in figure 8. The values of $(\tilde{S}_1, \tilde{S}_2)$ and D_r are estimated by the proposed method described in section 3.2. D_r is the average of the difference between a synthesized background and a background in an input image (excluding objects).

From these results, even object and shadow regions occupy large areas in an input image, our proposed method can estimate background components accurately. The accuracy of estimation is worse when an input image is darker. The reason is that SNR of image is worse and quantization error is relatively larger when an input image is darker.

4.3 Accuracy of Object Extraction

We compare the results of automatic extraction by our method with that of manual extraction. Let A, B be object regions and background regions determined by our method, respectively. Let C, D be object regions and background regions determined by hand, respectively. Ob-

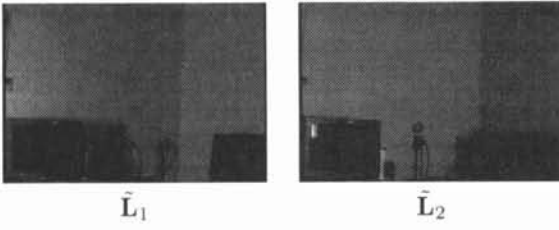
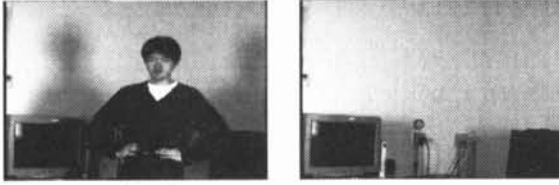
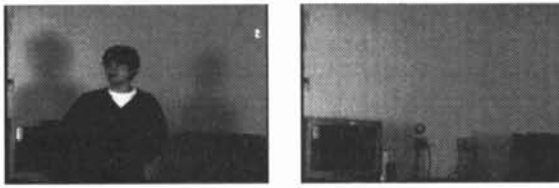


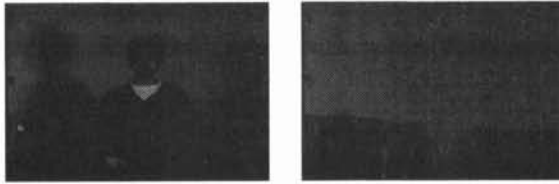
Figure 8: Estimated background components



(a) $(\tilde{S}_1, \tilde{S}_2) = (0.99, 0.98)$, $D_r = 1.83$



(b) $(\tilde{S}_1, \tilde{S}_2) = (0.65, 0.69)$, $D_r = 3.27$



(c) $(\tilde{S}_1, \tilde{S}_2) = (0.33, 0.36)$, $D_r = 3.71$

Figure 9: Estimated background images under various light conditions (left:input image, right:estimated background)

ject detection rate R_D and incorrect detection rate R_I are calculated from the following equation:

$$R_D = \frac{\#\{A \cup C\}}{\#C} \times 100, \quad (10)$$

$$R_I = 100 - \frac{\#\{B \cup D\}}{\#D} \times 100, \quad (11)$$

where $\#$ is the number of pixels in a region. The results using 10 images illuminated two lights are shown in table 1. The averages of object detection rate and incorrect detection rate are 90.9% and 0.55%, respectively. Shadow removal is correctly performed by the proposed method.

5 Conclusion

We proposed a background model in consideration of illumination changes and shadows. We also proposed a method for automatic background estimation. We conducted experiments with real images and confirmed the effectiveness of our proposed method. As preprocessing, our proposed method can be applied to various situations such as person tracking, gesture recognition, and object detection.

Table 1: The accuracy of object extraction (2 lights)

image	$R_D(\%)$	$R_I(\%)$
(a)	97.7	0.25
(b)	71.3	0.05
(c)	98.5	0.57
(d)	91.0	1.58
(e)	94.2	0.08
(f)	91.1	0.02
(g)	92.9	0.00
(h)	93.9	0.33
(i)	86.5	0.95
(j)	91.8	1.70
ave.	90.9	0.55

Acknowledgment

This research was partially supported by the Japan Society for the Promotion of Science under Grant-in-Aid for Creative Scientific Research (Project No. 13GS0018).

References

- [1] G. Gordon, T. Darrell, M. Harville, and J. Woodfill. Background estimation and removal based on range and color. In *Proc. of CVPR*, pp. 459–464, June 1999.
- [2] J. Stauder. Segmentation of moving objects in presence of moving shadows. In *Proc. of International Workshop on VLBV'97*, pp. 41–44, Linköping, Sweden, July 1997.
- [3] J. Stauder and R. Mech. Detection and tracking of moving cast shadows. In *Proc. of Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'99)*, Berlin, Germany, May 1999.
- [4] A. Elgammal, D. Harwood, and L. S. Davis. Non-parametric model for background subtraction. In *Proc. of IEEE ICCV'99 Frame-rate Workshop*, Greece, September 1999. IEEE.
- [5] T. Horprasert, D. Harwood, and L. S. Davis. A statistical approach for real-time robust background subtraction and shadow detection. In *Proc. of IEEE ICCV'99 Frame-rate workshop*, Greece, September 1999. IEEE.
- [6] I. Mikic, P. Cosman, G. Kogut, and M. Trivedi. Moving shadow and object detection in traffic scenes. In *ICPR*, volume 1, pp. 321–324, 2000.
- [7] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. of CVPR*, pp. 246–252, 1999.
- [8] N. Ohta. A statistical approach to background subtraction for surveillance systems. In *Proc. of ICCV*, pp. 481–486, Vancouver, Canada, July 2001.
- [9] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. on System, Man, and Cybernetics*, 9(1):62–66, 1979.