A Statistical Approach to Background Subtraction for Surveillance Systems

Naoya Ohta
Computer Science Department, Gunma University
1-5-1 Tenjin-cho, Kiryu, Gunma 376-8515 Japan
ohta@cs.gunma-u.ac.jp

Abstract

Background subtraction is a commonly used process in surveillance systems. One difficult problem when using the process is maintaining a correct background image against changing illumination conditions. Most methods for maintaining the background image are based on intuitive definitions about the illumination change and are implemented as somewhat ad hoc algorithms. In contrast, we first define mathematical models representing the relation between the illumination intensity, a reflection index of objects and a pixel value. We also mathematically define an assumption about illumination, which requires that the distribution of the illumination intensity in a small region does not change. Then we formalize the background subtraction problem as a statistical test (χ^2 test) based on the models and assumption. The experiments show that our models appropriately express the imaging process of a camera and our method[1] *provides stable detection performance for foreground objects.*

1. Introduction

A system that monitors an area by camera and detects persons or objects entering the area is called a surveillance system. One image processing technique often utilized in such methods is background substitution. The system keeps a background image and subtracts it from input images. If persons or objects appear, a large difference in the images is observed and the system can detect them. This technique is simple and exhibits reasonably good performance if the background image is correct. However, the major difficulty of this method lies in maintaining a correct background image. The background of the scene changes with illumination conditions. It is very difficult to distinguish intensity changes caused by illumination change from those caused by entering objects.

One approach to overcome this problem is to use information other than intensity of a gray image, such as color or range [1, 2]. If stereo cameras or a color camera is available, this approach is promising. However, when the situation allows us to use only a monochrome camera, we cannot adopt this strategy. In that case, we have to tackle the original problem, of how to keep the correct background image.

Many algorithms maintaining the background image have been proposed in the past [8]. Those algorithms adapt the illumination change and update the background image. Some of them perform well, but the problem many of them have is that the algorithm is designed in a somewhat ad hoc manner and assumptions they rely on are not clear. This becomes a problem when we want to enhance performance by combining those methods with others. An important aspect of an algorithm which enables us to use it as a part of a system is that models or assumptions that the algorithm is based on are defined in an explicit and clear way. In this paper, we explicitly define the models and assumptions the proposed method adopts and deductively construct the algorithm from them.

We define models of the illumination change, spatial distribution of illumination, and noise in pixel values of an image in the next section. Then, we construct a statistical test for the image difference detection referring to [4]. Our method is similar to [6, 7] in adopting a statistical formalization, but our models of illumination change and spatial distribution of illuimination are much simpler. The appropriateness of our models are verified by experiment.

2. Mathematical models for background subtraction

First, we define an explicit model of illumination change. Let the reflection index of a point on an object be r, intensity of the illumination at that object point be α and the bias of

[1] Computer programs written in C for the method are available from http://www.ail.cs.gunma-u.ac.jp/Labo/programs-e.html.

the imaging device be β. We assume that the pixel value q of an image, corresponding to the object point, is given by

$$q = \alpha r + \beta. \tag{1}$$

We also assume the bias β is known through calibration. This model captures the notion that change in the illumination intensity linearly affects the pixel value. We will discuss later the appropriateness of this model, together with a method to determine the bias β.

We convert the pixel value q to p via the equation

$$p = q - \beta = \alpha r. \tag{2}$$

The value of p is zero when the environment is completely dark, which makes the problem simpler.

So far, we have considered one pixel in one image. From here, we suppose that we have two images, these being a reference image and an input image. The reference image is an image which was taken when only background objects were present. The input image is one where persons or objects entering the monitoring area may have been imaged. We want to detect the persons or objects by computing the difference between the reference image and the input image. Let us indicate reference image quantities by using the superscript (r) and input image quantities by using the superscript (i). In addition, we consider a small region S in the images and assume that the objects are evenly illuminated in region S. Similar assumptions and their verification are found in [3, 5]. We use the subscript i in order to index each pixel in the region. For instance, $p_i^{(r)}$ expresses the converted ith pixel value of the reference image.

We define a noise model of pixel values. Let the noise of the reference image and that of the input image at position i be $\Delta p_i^{(r)}$ and $\Delta p_i^{(i)}$, respectively. We assume $\Delta p_i^{(r)}$ and $\Delta p_i^{(i)}$ are independent zero mean Gaussian noise with variances $\sigma^{(r)2}$ and $\sigma^{(i)2}$, respectively. If we denote true pixel values, which would be measured if no noise were present, by $\bar{p}_i^{(r)}$ and $\bar{p}_i^{(i)}$, then the measured pixel values $p_i^{(r)}$ and $p_i^{(i)}$ are given by

$$\begin{aligned} p_i^{(r)} &= \bar{p}_i^{(r)} + \Delta p_i^{(r)} = \alpha^{(r)} r_i + \Delta p_i^{(r)}, & (3)\\ p_i^{(i)} &= \bar{p}_i^{(i)} + \Delta p_i^{(i)} = \alpha^{(i)} r_i + \Delta p_i^{(i)}. & (4) \end{aligned}$$

Let us define the ratio $\bar{t}$ of the illumination intensities $\alpha^{(i)}$ by $\alpha^{(r)}$ as follows

$$\bar{t} = \frac{\alpha^{(i)}}{\alpha^{(r)}}. \tag{5}$$

If the reflection index r_i is the same between the reference image and the input image (in other words, if the same objects are imaged on the two images), then $\bar{t}$ must be constant within the area S because $\alpha^{(i)}$ and $\alpha^{(r)}$ are constant over that area. In that case, we have, for all subscripts i, that

$$\bar{p}_i^{(i)} = \bar{t}\,\bar{p}_i^{(r)}. \tag{6}$$

However, if the reflection index r_i is different because of the entrance of persons or objects, then eq. (6) does not hold. From this observation, we can formulate the detection problem of an entering person or object as a statistical decision problem as follows. First, we statistically estimate the illumination ratio $\bar{t}$ from measured pixel values $p_i^{(r)}$ and $p_i^{(i)}$ under the assumption that the reflection index r_i is the same. Second, we investigate whether this assumption is reasonable in a statistical sense. We explain the details in the following sections.

3. Problem formalization as statistical estimation

We first discuss the statistical estimation of the illumination ratio $\bar{t}$, which we regard as the parameter to be estimated. We define two dimensional vectors $\boldsymbol{m}_i$, $\bar{\boldsymbol{m}}_i$ and $\Delta\boldsymbol{m}_i$ as follows:

$$\boldsymbol{m}_i = \begin{pmatrix} p_i^{(r)} \\ p_i^{(i)} \end{pmatrix}, \quad \bar{\boldsymbol{m}}_i = \begin{pmatrix} \bar{p}_i^{(r)} \\ \bar{p}_i^{(i)} \end{pmatrix}, \quad \Delta\boldsymbol{m}_i = \begin{pmatrix} \Delta p_i^{(r)} \\ \Delta p_i^{(i)} \end{pmatrix}. \tag{7}$$

Using these vectors, eqs. (3) and (4) are expressed by

$$\boldsymbol{m}_i = \bar{\boldsymbol{m}}_i + \Delta\boldsymbol{m}_i, \tag{8}$$

and the covariance matrix $\boldsymbol{V}_m$ of the noise vector $\Delta\boldsymbol{m}_i$ is given by

$$\boldsymbol{V}_m = \begin{pmatrix} \sigma^{(r)2} & 0 \\ 0 & \sigma^{(i)2} \end{pmatrix}. \tag{9}$$

If we define the parameter vector $\bar{\boldsymbol{u}}$ by

$$\bar{\boldsymbol{u}} = \begin{pmatrix} \bar{t} \\ -1 \end{pmatrix}, \tag{10}$$

then eq. (6) can be expressed as

$$(\bar{\boldsymbol{u}}, \bar{\boldsymbol{m}}_i) = 0, \tag{11}$$

where $(\cdot\,,\cdot)$ denotes the inner product of vectors. Using these expressions, the probability $\Pr(\{\boldsymbol{m}_i\})$ that a set of pixel values $\{\boldsymbol{m}_i\}$ $(i = 1 \ldots n)$ is observed in the region S may be expressed as

$$\Pr(\{\boldsymbol{m}_i\}) = \prod_{i=1}^{n} \frac{1}{2\pi\sqrt{|\boldsymbol{V}_m|}} \exp\{-\frac{1}{2}(\boldsymbol{m}_i - \bar{\boldsymbol{m}}_i, \boldsymbol{V}_m^{-1}(\boldsymbol{m}_i - \bar{\boldsymbol{m}}_i))\}. \tag{12}$$

Therefore, the maximum likelihood estimators of the parameter vector $\bar{\boldsymbol{u}}$ and the pure pixel values $\{\bar{\boldsymbol{m}}_i\}$ are defined as the values which minimize J_1 while maintaining the truth of eq. (11), where J_1 is given by

$$J_1 = \sum_i^n (\boldsymbol{m}_i - \bar{\boldsymbol{m}}_i, \boldsymbol{V}_m^{-1}(\boldsymbol{m}_i - \bar{\boldsymbol{m}}_i)). \qquad (13)$$

Next, we discuss how to minimize J_1. If we temporarily fix $\bar{\boldsymbol{u}}$, the problem becomes one of minimizing J_1 with regard to $\bar{\boldsymbol{m}}_i$ under constraint (11). Introducing the Lagrange multipliers λ_i, we define the quantity K such that

$$K=\sum_i^n \{(\boldsymbol{m}_i - \bar{\boldsymbol{m}}_i, \boldsymbol{V}_m^{-1}(\boldsymbol{m}_i - \bar{\boldsymbol{m}}_i)) + \lambda_i(\bar{\boldsymbol{u}}, \bar{\boldsymbol{m}}_i)\}. \qquad (14)$$

Differentiating K with regard to $\bar{\boldsymbol{m}}_i$ and λ_i and setting them to be zero, we then obtain

$$\bar{\boldsymbol{m}}_i = \boldsymbol{m}_i - \frac{\lambda_i}{2}\boldsymbol{V}_m\bar{\boldsymbol{u}}, \quad \lambda_i = 2\frac{(\bar{\boldsymbol{u}}, \boldsymbol{m}_i)}{(\bar{\boldsymbol{u}}, \boldsymbol{V}_m\bar{\boldsymbol{u}})}. \qquad (15)$$

These equations imply that J_1 assumes the minimum. Substituting $\bar{\boldsymbol{m}}_i$ of eq. (15) for that in J_1 of eq. (13), we obtain the objective function J_2 that only contains $\bar{\boldsymbol{u}}$ as variables:

$$J_2 = \sum_{i=1}^n \frac{(\bar{\boldsymbol{u}}, \boldsymbol{m}_i)^2}{(\bar{\boldsymbol{u}}, \boldsymbol{V}_m\bar{\boldsymbol{u}})} = \frac{(\bar{\boldsymbol{u}}, \boldsymbol{M}\bar{\boldsymbol{u}})}{(\bar{\boldsymbol{u}}, \boldsymbol{V}_m\bar{\boldsymbol{u}})}. \qquad (16)$$

Here, $\boldsymbol{M}$ is a moment matrix expressed by

$$\boldsymbol{M} = \sum_{i=1}^n \boldsymbol{m}_i\boldsymbol{m}_i^\top, \qquad (17)$$

where the superscript $^\top$ denotes the transpose of a vector. The minimizer $\bar{\boldsymbol{u}}$ of J_2 is given as the generalized eigenvector $\boldsymbol{v}_0$ corresponding to the smallest generalized eigenvalue μ_0 of the generalized eigenvalue problem

$$\boldsymbol{M}\boldsymbol{v} = \mu\boldsymbol{V}_m\boldsymbol{v}. \qquad (18)$$

Therefore, the maximum likelihood estimate $\hat{\boldsymbol{u}}$ of the parameter vector $\bar{\boldsymbol{u}}$ is given as $\boldsymbol{v}_0$ whose second element is -1. The first element of $\boldsymbol{v}_0$ is the maximum likelihood estimate $\hat{t}$ of the illumination ratio $\bar{t}$ and the smallest generalized eigenvalue μ_0 is the residual J_2. A numerical solution of the generalized eigenvalue problem (18) can be computed by solving two successive (regular) eigenvalue problems [4].

4. Statistical test

In the previous section, we described how to estimate the refraction index ratio $\bar{t}$ from the noisy pixel values $\{\boldsymbol{m}_i\}$. In this section, we formalize the problem of detecting entering persons or objects as a statistical test using the residual J_2. If the value of J_2 is consistent with the assumption represented by eq. (6) or (11), we can conclude that there are no entering objects. Otherwise, it is highly likely that they exist. In order to carry out the statistical test, we must know the probability distribution of J_2. We discuss this problem next.

The first step to computing the probability distribution function of J_2 is to approximate J_2 by a quadratic function around the true parameters $\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{m}}_i$. Let Δt be a perturbation of $\bar{t}$. We write the parameter vector $\boldsymbol{u}$ in the form

$$\boldsymbol{u} = \bar{\boldsymbol{u}} + \Delta t\boldsymbol{e}_x, \quad \boldsymbol{e}_x = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad (19)$$

and substitute $\boldsymbol{u}$ of eq. (19) and $\boldsymbol{m}_i$ of eq. (8) for $\bar{\boldsymbol{u}}$ and $\boldsymbol{m}_i$ in eq. (16), respectively. If we take into account that

$$\frac{1}{(\bar{\boldsymbol{u}} + \Delta t\boldsymbol{e}_x, \boldsymbol{V}_m\bar{\boldsymbol{u}} + \Delta t\boldsymbol{e}_x)} = \frac{1}{(\bar{\boldsymbol{u}}, \boldsymbol{V}_m\bar{\boldsymbol{u}})} + O^1 \qquad (20)$$

and eq. (11), then the residual J_2 can be expressed by

$$J_2 = \tilde{J}_2 + O^3, \qquad (21)$$

where

$$\tilde{J}_2 = \frac{1}{(\bar{\boldsymbol{u}}, \boldsymbol{V}_m\bar{\boldsymbol{u}})}\sum_{i=1}^n \{(\bar{\boldsymbol{u}}, \Delta t\boldsymbol{e}_x)^2 + 2\Delta t(\bar{\boldsymbol{u}}, \Delta\boldsymbol{m}_i)(\boldsymbol{e}_x, \bar{\boldsymbol{m}}_i) + \Delta t^2(\boldsymbol{e}_x, \bar{\boldsymbol{m}}_i)\}. \qquad (22)$$

In the above equations, O^k denotes terms of Δt and $\Delta\boldsymbol{m}_i$ whose order is equal to or higher than k. Therefore, the quadratic approximation of J_2 is $\tilde{J}_2$ in eq. (22). Given data $\{\boldsymbol{m}_i\}$, we compute Δt that gives the minimum of $\tilde{J}_2$. Differentiating $\tilde{J}_2$ with respect to Δt and setting the resultant equation to be zero, we have

$$\Delta t = -\frac{\sum_{i=1}^n \{(\bar{\boldsymbol{u}}, \Delta\boldsymbol{m}_i)(\boldsymbol{e}_x, \bar{\boldsymbol{m}}_i)\}}{\sum_{i=1}^n \{(\boldsymbol{e}_x, \bar{\boldsymbol{m}}_i)^2\}}. \qquad (23)$$

Substituting this for Δt in eq. (22), we obtain

$$\tilde{J}_2 = \frac{1}{(\bar{\boldsymbol{u}}, \boldsymbol{V}_m\bar{\boldsymbol{u}})}\left(\sum_{i=1}^n \{(\bar{\boldsymbol{u}}, \Delta\boldsymbol{m}_i)^2\} - \frac{(\sum_{i=1}^n \{(\bar{\boldsymbol{u}}, \Delta\boldsymbol{m}_i)(\boldsymbol{e}_x, \bar{\boldsymbol{m}}_i)\})^2}{\sum_{i=1}^n \{(\boldsymbol{e}_x, \bar{\boldsymbol{m}}_i)^2\}}\right). \qquad (24)$$

This is the approximated residual of J_2, which is a function only of $\Delta\boldsymbol{m}_i$.

The second step is to compute the probability distribution function of the approximated residual $\tilde{J}_2$ under the condition that the noise vector $\Delta\boldsymbol{m}_i$ has Gauss distribution whose covariance matrix is $\boldsymbol{V}_m$. We define a_i and b as follows:

$$a_i = \frac{(\bar{\boldsymbol{u}}, \Delta \boldsymbol{m}_i)}{\sqrt{(\bar{\boldsymbol{u}}, \boldsymbol{V}_m \bar{\boldsymbol{u}})}} \tag{25}$$

$$b = \frac{\sum_{i=1}^{n} \{(\boldsymbol{e}_x, \bar{\boldsymbol{m}}_i) a_i\}}{\sqrt{\sum_{i=1}^{n} \{(\boldsymbol{e}_x, \bar{\boldsymbol{m}}_i)^2\}}}. \tag{26}$$

Using a_i and b, $\tilde{J}_2$ can be expressed as

$$\tilde{J}_2 = \sum_{i=1}^{n} \{a_i^2\} - b^2. \tag{27}$$

Since the numerator of eq. (25) is a linear sum of the elements of $\Delta \boldsymbol{m}_i$, it has a normal distribution with variance $(\bar{\boldsymbol{u}}, \boldsymbol{V}_m \bar{\boldsymbol{u}})$. Thus, a_i has the standard normal distribution. The numerator of eq. (26) is a linear sum of a_i and has a normal distribution whose variance is the squared sum of the a_i's coefficients given by $\sum_{i=1}^{n} (\boldsymbol{e}_x, \bar{\boldsymbol{m}}_i)^2$. This means that b also has the standard normal distribution. Therefore, $\sum_{i=1}^{n} \{a_i^2\}$ and b^2 have χ^2 distributions whose degrees of freedom are n and 1, respectively. Since b is a linear sum of a_i, using Cochran's theorem, it can be proved that $\tilde{J}_2$ has a χ^2 distribution with $n-1$ degrees of freedom. From the above discussions, we can conclude that the residual J_2 has a χ^2 distribution with $n-1$ degrees of freedom to a first order approximation.

We formalize the entering object detection problem as a χ^2 test with the null hypothesis of eq. (6) or (11). Since the residual J_2 is given by the smallest generalized eigenvalue μ_0 of eq. (18), the condition rejecting the hypothesis is

$$\mu_0 > \chi^2_{1-\nu}(n-1). \tag{28}$$

In χ^2 tests, a value of 5% or 10% is typically adopted as the significance level ν. In this test, however, those values are too large because ν expresses the probability that the system falsely detects entering objects caused by image noise. In the experiments described later, we adopt 0.01% as the value of ν. Since we use a 5×5 pixel image patch as the area S, the degrees of freedom of the χ^2 distribution is 24 $(= 5{\times}5{-}1)$ and we detect entering objects by the condition

$$\mu_0 > 58.61 = \chi^2_{1-0.0001}(24). \tag{29}$$

5. Verification of model of illumination change

Previously, we assumed eq. (1) expresses the relation between pixel values of images and the reference index of an object. In this section, we verify this relation by experiment. The bias β in eq. (1) is also determined here. We first explain the processes by which the experimental images were taken. These images will also be used in the entering objects detection experiments described later.

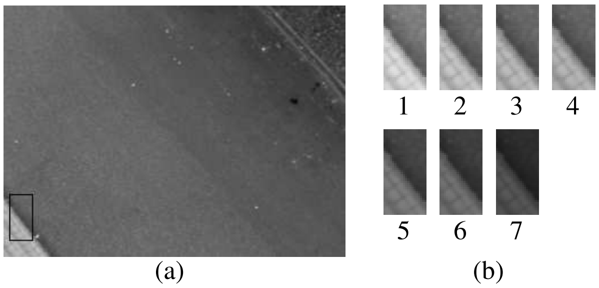


Figure 1. (a) Reference (background) image. (b) Images used to verify the model of the illumination change.

Figure 1 (a) shows a road scene from our campus, which we used for the experiments. The images were taken by a camera (SONY XC-75) mounted on the fourth floor of an adjacent building to the road, recorded on VHS videotape, and then digitized into a computer. The weather on that day was mainly cloudy but with occasional sun. The time of day was between 3 p.m. and 5 p.m. The illumination conditions changed as the clouds and the sun moved. Example images are shown in the first rows of fig. 3.

We verified the correctness of eq. (1) as follows. We selected a scene where only the background objects had been imaged and the illumination intensity had remained constant. We captured 10 adjacent frames from such a scene and averaged them in order to reduce noise. We repeated this procedure for 7 different scenes under different illumination intensities and obtained 7 averaged images. We set a rectangular area indicated by the black frame in Fig 1 (a) on each averaged image and used pixels in that area for the data. The 7 images of the area are shown in fig. 1 (b): image 1 is brightest and 7 is darkest. Next, we drew a two dimensional graph using those images as follows. We chose image 3 as a reference and regarded its pixel values as the horizontal (x) coordinate of dots. The pixel values of the other images at the corresponding image positions were used for the vertical (y) coordinate of the dots. If eq. (1) is a correct model of the illumination change, all dots form 6 lines and they intersect at one point. The resultant graph is shown in fig. 2 (a) and we can observe that this occurs.

Next, we determine the bias β from the data. Equation (1) suggests that dots in each line are expressed by

$$(x, y) = (\alpha_x r + \beta, \alpha_y r + \beta), \tag{30}$$

where α_x is the illumination intensity of image 3 and α_y is that of the other one image. Therefore, the lines in fig. 2 (a) are supposed to intersect each other at $(x, y) = (\beta, \beta)$ and this point gives the value of β. First, we fit a line to each group of dots as follows. Let (x_{ki}, y_{ki}) be the coordinate of the ith dot of the kth group. Define a 3-dimensional vector $\boldsymbol{x}_{ki} = (x_{ki}\ y_{ki}\ 1)^\top$ and compute the moment matrix

$$\boldsymbol{S}_k = \sum_i \boldsymbol{x}_{ki}\boldsymbol{x}_{ki}^\top. \tag{31}$$

If the unit eigenvector corresponding to the smallest eigenvalue of matrix $\boldsymbol{S}_k$ is $\boldsymbol{s}_k = (s_{xk}\ s_{yk}\ s_{zk})^\top$ then the fitted line to the kth group is expressed by

$$xs_{xk} + ys_{yk} + s_{zk} = 0. \tag{32}$$

We next compute the intersection point. Define another moment matrix

$$\boldsymbol{L} = \sum_k \boldsymbol{s}_k\boldsymbol{s}_k^\top \tag{33}$$

and compute the unit eigenvector $\boldsymbol{l} = (l_x\ l_y\ l_z)^\top$ corresponding to the smallest eigenvalue of $\boldsymbol{L}$. Using this eigenvector, the intersection point is given as $(x, y) = (l_x/l_z, l_y/l_z)$. We compute the point using the data of fig. 2 (a) and obtained $(-5.747, -4.334)$. Since its x and y coordinates must be the same in the ideal case, we adopted the averaged value as follows.

$$\beta = (-5.75 - 4.33)/2 = -5.04 \tag{34}$$

In order to execute the proposed method, we need the value of the covariance matrix $\boldsymbol{V}_m$ of eq. (9). To do this, we measured the variance of the noise Δp in the pixel value p expressed by eq. (2). In the above experiment, we used the averaged images of 10 frames under 7 different illumination conditions. Using them, we extracted the noise as follows. We subtract each averaged image from the 10 original images from which the averaged image was computed and determine the variance of the extracted noise. Figure 2 (b) shows the computed values of the variance for each illumination intensity. Although we can observe a slight tendency that the variance increases as the illumination becomes dark, the values do not exhibit a big fluctuation. So, we adopt the averaged value as the variance of the noise, specifically:

$$\sigma^2 = 2.75. \tag{35}$$

6. Experiments for entering object detection

We selected several scenes from the videotape where pedestrians or bicycles were present and captured their images into a computer. Those images are shown in the first rows of fig. 3. The figures under the images are the time of day of the scenes, computed based on the VCR's time counter. We also chose a scene where only the background objects appeared. We digitized 10 successive frames of that scene and averaged them into one image, which was used as the background image. This image is shown in fig. 1 (a). The time of that scene was 2:58:20 and just before the first scene in fig. 3 was observed.

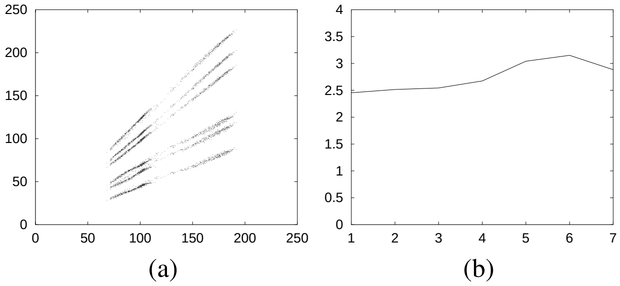
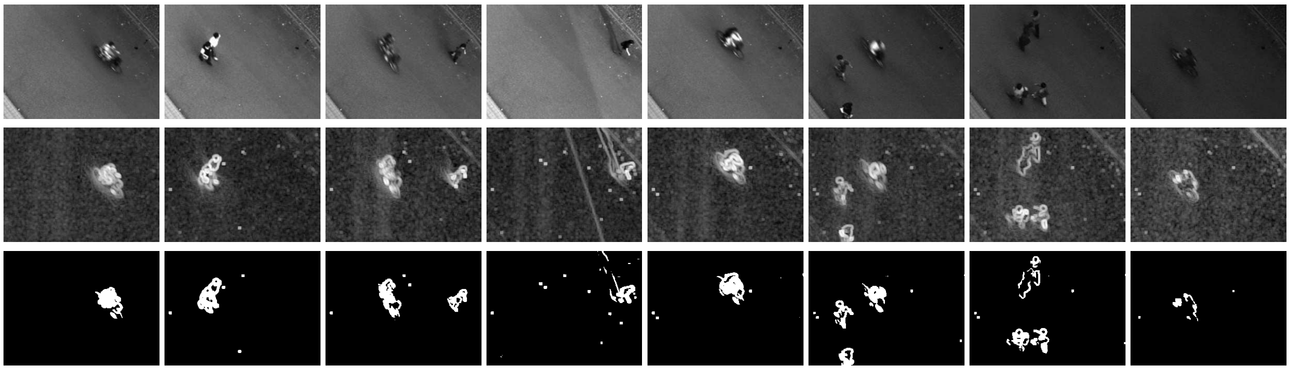


Figure 2. (a) Linearity of pixel value in different illumination intensities. The lines from the top to the bottom correspond image 1, 2, 4, 5, 6 and 7. (b) Variance of noise in pixel value vs. illumination intensities.

We set a 5×5 pixel area as the area S, where the illumination intensity is assumed to be constant, and converted the pixel values q in area S to p by eqs. (2) and (35). Then, we computed the moment matrix $\boldsymbol{M}$ of eq. (17), solved the generalized eigenvalue problem of eq. (18) and obtained its smallest eigenvalue μ_0. The image difference was detected by the condition of eq. (29) using μ_0. This procedure was applied at all pixel positions in the images. In the computation of eq. (18), we used eq. (35) for the value of $\sigma^{(i)2}$ in eq. (9). For that of $\sigma^{(r)2}$, we used the value of eq. (35) divided by 9 ($= 10-1$) because the reference image is an average of 10 images.

The results are shown in the middle and the bottom rows of fig. 3. The pixel values of the middle images are $\log \mu_0$. We used the logarithm because the values of μ_0 are spread over a wide range. The bottom images are the detection results for entering persons or objects. The white pixels indicate the regions where the objects were detected.

Given the results, we observe the following: (1) As a general observation, the proposed method stably detects the entering persons or objects for 2 hours using only one reference image. (2) Except for the first scene, white dots on the road appear in the bottom images. They are mainly caused by moving leaves on the road. (3) In the fourth scene, the building cast a shadow and the illumination condition was different within both sides of the border. Therefore, large values of μ_0 are observed on the border. (4) The seventh and eighth scenes were at dusk and the images are in low contrast. Therefore, the values of μ_0 are small, but the persons and bicycles are detected nevertheless. (5) In some regions within the boundaries of the persons or bicycles, the image difference was not detected. This is because those regions



(1) 3:00:10 (2) 3:11:50 (3) 3:36:30 (4) 3:40:03 (5) 4:05:40 (6) 4:18:10 (7) 4:51:05 (8) 4:59:24

Figure 3. Results of image difference detection. From the top to the bottom, original images, images with pixel value of $\log \mu_0$ and detected image difference.

had no texture both in the reference and the input images.

As a conclusion, our method provides a stable detection ability for a sustained period using one background image. This is a significant advantage of our method. However, if the background itself changes (remark 2) or some objects cast shadows (remark 3), our algorithm does not work as desired. But the important point here is that those issues are clearly predicted from the models or assumptions our method adopts. We defined the background as one whose reflection index was not changed and assumed that distribution of illumination in region S was unchanged. Any undesirable results occurred where these models were not valid. Of course, this renders the method imperfect, but a perfect one is obviously beyond reach. The characteristic that the behavior of the algorithm can be predicted is an important one when we combine it with other algorithms for making a practical system.

7. Conclusion

In this paper, we applied a statistical test to the background subtraction method to detect the difference between two images taken under different illumination conditions. We first established an explicit model of illumination change and of the noise contaminating the pixel values of images. Based on these models, we formulated a χ^2 test for image difference detection. Experiments were conducted first for verifying the appropriateness of the models that our method adopted, and second for evaluating the performance of our method. The second experiment showed that our method provides stable results for a long period of time. Like all techniques, the proposed method has defects. However, we can predict them from the models and assumptions the method adopted. This is a significant advantage of our method and is critical if our method is to be embedded into a larger, practical system.

Acknowledgments

The author thanks Professor Kenichi Kanatani of Okayama University and Michael J. Brooks of University of Adelaide, who discussed technical issues with him and helped to correct the English in this paper.

References

[1] C. Eveland, K. Konolige, and R. C. Bolles. Background modeling for segmentation of video-rate stereo sequence. In *Proc. CVPR'98*, pages 266–271, 1998.

[2] G. Gordon et al. Background estimation and removal based on range and color. In *Proc. CVPR'99*, pages 459–464, 1999.

[3] D. W. Jacobs et al. Comparing images under variable illumination. In *Proc. CVPR'98*, pages 610–617, 1998.

[4] K. Kanatani. *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier Science, Amsterdam, 1996.

[5] S. K. Nayar and R. M. Bolle. Reflectance ratio: A photometric invariant for object recognition. In *Proc. ICCV'93*, pages 280–285, 1993.

[6] P. L. Rosin. Thresholding for change detection. In *Proc. ICCV'98*, pages 274–279, 1998.

[7] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. CVPR'99*, pages 246–252, 1999.

[8] K. Toyama et al. Wallflower: Principles and practice of background maintenance. In *Proc. ICCV'99*, pages 255–261, 1999.