Background Subtraction for Non-Stationary Scenes

Takashi Matsuyama† Takashi Ohya‡ Hitoshi Habe†

† Department of Intelligence Science and Technology
Graduate School of Informatics, Kyoto University
Sakyo, Kyoto, JAPAN 606-8501
e-mail: tm@i.kyoto-u.ac.jp
‡ Canon Inc., Kawasaki, Japan

Abstract

In this paper, we propose a robust background subtraction method for non-stationary scenes. The non-stationarities modeled by the method are (1) variations of overall lighting conditions and (2) local image pattern fluctuations caused by waving leaves, fluttering flags, flickering CRTs and so on. First we propose a correlation measure between two blocks in images (SMNVD) to realize the robust background subtraction against varying illuminations. To characterize local image pattern fluctuations, we propose a two dimensional histogram (TNVDCM), where the distribution representing the temporal fluctuation pattern in a block is recorded. Experimental results of the background subtraction using SMNVD and TNVDCM demonstrate their robustness and effectiveness for real world scenes.

1 Introduction

Although the background subtraction is a useful method to detect and track moving objects in video images, its effectiveness is limited; the stationary background scene assumption does not hold always in the real world. To augment the background subtraction for non-stationary scenes, [1], [2] and [3] employed probability distributions to model intensity variations at each pixel and defined subtraction operators to detect anomalous pixel values corresponding to non-background moving objects.

In this paper, we propose a novel robust background subtraction method for non-stationary scenes. The non-stationarities modeled by the method are (1) variations of overall lighting conditions and (2) local image pattern fluctuations caused by waving leaves, fluttering flags, flickering CRTs and so on.

First we propose the *spatially modulated normalized vector distance* (SMNVD, in short) to realize robust background subtraction against varying illumination. Intuitively speaking, the normalized vector distance (NVD, in short) proposed in [4] measures the relative angle between a pair of N^2 dimensional vectors formed by scanning corresponding $N \times N$ blocks in the background and observed images respectively. While such angular difference is insensitive to changes of absolute values of the vectors (i.e. overall intensity changes in the blocks), it becomes unstable when the absolute values decrease. To cope with this problem, we propose SMNVD, which is computed by biasing NVD according to spatial characteristics in the blocks. Experimental results demonstrated its superiority to NVD when the level of the illumination is low.

To characterize non-stationary background objects such as waving leaves, fluttering flags, and flickering CRTs, we propose the *temporal NVD co-occurrence matrix* (TNVDCM, in short). It represents the temporal fluctuation pattern in a block and is defined by a two dimensional histogram where (i, j) element records the co-occurrence frequency of a pair of NVD values i and j observed in the block at the time interval of Δt. By analyzing the population distribution in TNVDCM, we categorize non-stationary background objects into five classes and define a specific subtraction operator for each class. Experimental results showed that while the class categorization sometimes becomes incorrect, the results of the background subtraction are very robust against fluctuations caused by non-stationary background objects.

While SMNVD and TNVDCM are separately implemented, their integration will realize the robust background subtraction method for real world scenes.

2 Spatially Modulated Normalized Vector Distance

In this section, we propose SMNVD to perform the robust background subtraction against varying illumination. Following the definition of NVD, the concept of SMNVD is introduced. Then a practical implementation of SMNVD is described.

2.1 Normalized vector distance

Let $f(x, y)$ be the reference image that was computed from the *a priori* taken background image sequence[1], and $g(x, y, t)$ be the observed image at time t. We divide these images into a set of blocks with $N \times N$ pixels respectively. Let $f_i(x, y)$ and $g_i(x, y, t)$ denote corresponding blocks in $f(x, y)$ and $g(x, y, t)$

[1]The reference image records the median value at each pixel position in the background images. We call this image *median image*

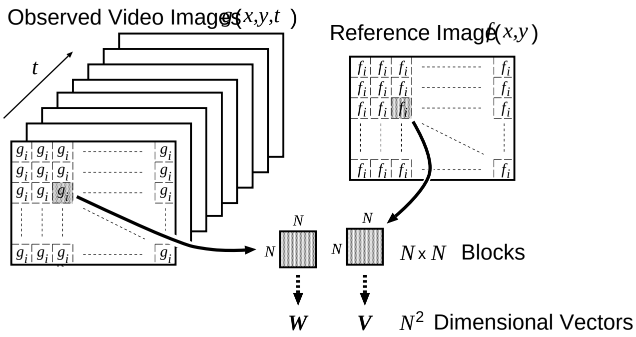


Figure 1: Blocks in the reference and observed images

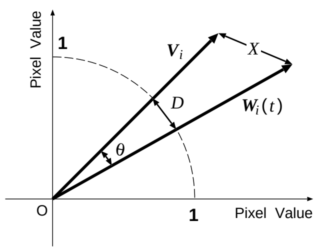


Figure 2: Distance between a pair of vectors

respectively. Then, we generate N^2 dimensional vectors $\boldsymbol{V}_i$ and $\boldsymbol{W}_i(t)$ by scanning $f_i(x,y)$ and $g_i(x,y,t)$ respectively. The components of these vectors represent pixel values in the blocks. These vectors indicate local image pattern in the reference and observed images respectively.

To detect anomalous regions corresponding to moving objects, we need to compute correlation between the reference and observed images. Since the correlation between images can be computed based on the distance between vectors $\boldsymbol{V}_i$ and $\boldsymbol{W}_i(t)$, we need to define the vector distance which is independent of absolute values of the vectors (i.e. illumination conditions). Possible definitions of the vector distance are X, θ and D in Figure 2. Obviously X cannot be used because it depends on absolute values of the vectors. Although the relative angle between the vectors can be used, the sensitivity of $\cos\theta$, which is computable, degrades when θ comes close to zero.

In [4], a new vector distance (NVD) insensitive to varying illuminations was proposed. NVD is defined as follows.

$$D_i(t) = \left| \frac{\boldsymbol{V}_i}{|\boldsymbol{V}_i|} - \frac{\boldsymbol{W}_i(t)}{|\boldsymbol{W}_i(t)|} \right|, \tag{1}$$

$D_i(t)$ and θ satisfy

$$D_i(t) = \sqrt{2(1-\cos\theta)}. \tag{2}$$

$D_i(t)$ is insensitive to changes of absolute values of the vectors. In general, however, since $\boldsymbol{W}_i(t)$ is corrupted with random noise, the mean of NVD does not become zero and varies approximately in inverse proportion to $|\boldsymbol{W}_i(t)|$ and the variance of NVD varies approximately in inverse proportion to $|\boldsymbol{W}_i(t)|^2$. This means that the NVD value becomes unreliable as the absolute values of the vectors decrease.

As is well known, illumination invariant measures based on the color information (e.g. [5]) also become unreliable when the level of the illumination decreases.

2.2 Incorporation of Spatial Characteristics

Our idea for alleviating this problem is to *incorporate spatial characteristics in the blocks into the definition of the vector distance.* While NVD measures the relative angle between vectors, it does not reflect any spatial structures in the blocks. Hence, augmenting NVD by taking spatial characteristics into account, the performance of the background subtraction will be improved; overall noise and moving objects can be clearly distinguished from the background.

Generally speaking, let $SP_i(t)$ denote spatial characteristics in block i. Then, *spatially modulated normalized vector distance* (SMNVD, in short) is defined as:

$$S_i(t) = F(D_i(t),\ SP_i(t)), \tag{3}$$

where F is the function biasing $D_i(t)$ according to the spatial characteristics.

2.3 An Implementation of SMNVD

Here we describe an implementation of SMNVD for the robust background subtraction against varying illuminations. Although it is a very simple implementation, experimental results in Section 4.1 show its effectiveness.

The algorithm of the background subtraction using the implemented SMNVD is as follows:

1. Create the reference image from an *a priori* taken background image sequence.

2. Calculate a set of NVD values for each block i using the reference image and the background image sequence, and record the mean $\overline{D}_i$ and variance σ_{D_i} of the NVD value set as the background model. Assuming that the distribution of NVD obeys the Gaussian distribution, the probability density function $P_{\mathrm{NVD}}(D_i(t))$ of NVD is defined by

$$P_{\mathrm{NVD}}(D_i(t)) = \frac{1}{\sqrt{2\pi}\sigma_{D_i}} \exp\left\{ -\frac{(D_i(t)-\overline{D}_i)^2}{2\sigma_{D_i}^2} \right\}. \tag{4}$$

3. For each block i, define a group of small windows w_i^k $(k = 1, \ldots, n)$ in the internal area of the block, where n denotes the number of the windows (In Figure 3, five windows are defined.) Then calculate a set of NVD values C_i^ks for each window w_i^k using the reference image and the background image sequence, and record the mean $\overline{C_i^k}$ of C_i^ks as the background model.

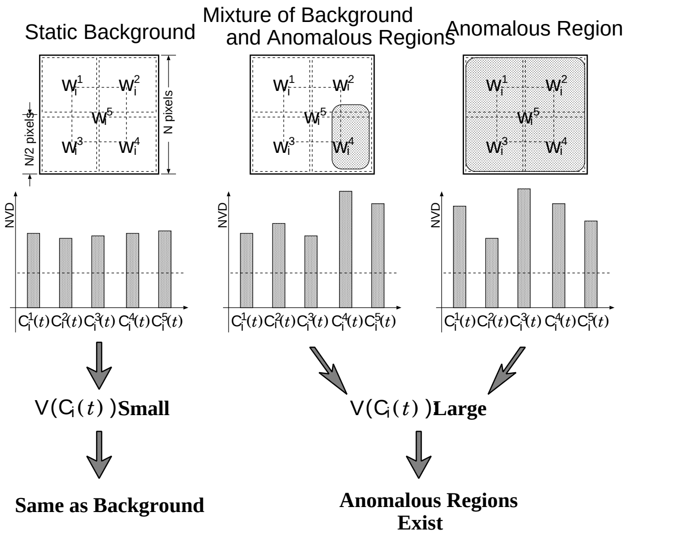


Figure 3: Spatial characteristics in block i

4. Given an observed image $g(x, y, t)$, compute NVD values for block i and its internal windows w_i^k, i.e. $D_i(t)$ and $C_i^k(t)$ $(k = 1, \ldots, n)$.

5. Figure 3 illustrates three different types of situations observed in block i: the block is covered by the background region, mixture of background and anomalous (i.e. moving object) regions , or the anomalous region. To discriminate these situations, we employ the *spatial* variance of NVD values in the windows, $V(C_i(t))$, which is defined by

$$V(C_i(t)) = \frac{1}{n} \sum_{k=1}^{n} (C_i^k(t) - \overline{C_i^k})^2. \qquad (5)$$

Static background Under normal illumination conditions, the value of each $C_i^k(t)$ $(k = 1, \ldots, n)$ is almost equal to zero and hence their variance $V(C_i(t))$ is almost equal to zero as well. Although each $C_i^k(t)$ becomes unstable when the level of illumination is low, $C_i^k(t)$ is almost equal to each other since the spatial distribution of noise in a block is uniform. Consequently, their variance $V(C_i(t))$ stays small irrespectively of illuminating conditions.

Mixture of background and anomalous regions In those windows covered by an anomalous region, their NVD values $C_i^k(t)$ becomes large. So, $V(C_i(t))$ gets larger.

Anomalous region Unless moving objects have the same texture as the background scene, each $C_i^k(t)$ takes a large random value. So, the $V(C_i(t))$ gets larger.

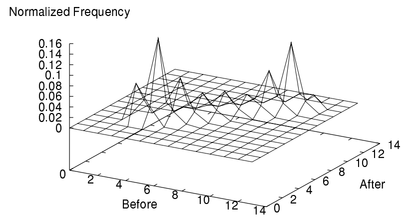


Figure 4: An example of TNVDCM

In summary, we can consider that $V(C_i(t))$ represents the spatial characteristics in block i, so that we define $SP_i(t)$ in equation 3 by $V(C_i(t))$.

6. Integrating $SP_i(t)$ into $D_i(t)$, we will get $S_i(t)$. Here F in Equation 3 is defined as follows:

$$F(D_i(t), SP_i(t)) = \begin{cases} 1 & if \quad P_{\mathrm{NVD}}(D_i(t)) < Th1 \\ & \quad and \quad SP_i(t) > Th2, \\ 0 & otherwise. \end{cases} \qquad (6)$$

where $Th1$ and $Th2$ are threshold values.

7. If $F(D_i(t), SP_i(t))$ is 1, the block i is regarded as containing anomalous regions.

3 Temporal NVD Co-occurrence Matrix

Real world scenes usually include various types of locally fluctuating objects such as waving leaves, fluttering flags, flickering CRTs and so on. Whereas they cause local image pattern fluctuations, they should be regarded as the background. Thus the model and operator of the background subtraction should be augmented so that it can cope with such non-stationary background scenes. Our idea for characterizing temporal image fluctuation patterns is to *record co-occurrence frequency of a pair of NVD values in a block separated by time interval* Δt.

In this section, we propose a robust background subtraction method for background scenes containing locally fluctuating objects. The method uses the Temporal NVD Co-occurrence Matrix (TNVDCM), whose population distribution represents the temporal fluctuation pattern in a block. By analyzing the distribution pattern, we can categorize local image pattern fluctuations into five classes and apply an optimized background subtraction operator for each class.

3.1 Definition

Let $D_i(t)$ denote NVD computed for block i using the reference image and *a priori* taken background image sequence, where t denotes the time covering the period of the image sequence and takes discrete values from 1 to T. Following the quantization of $D_i(t)$ into

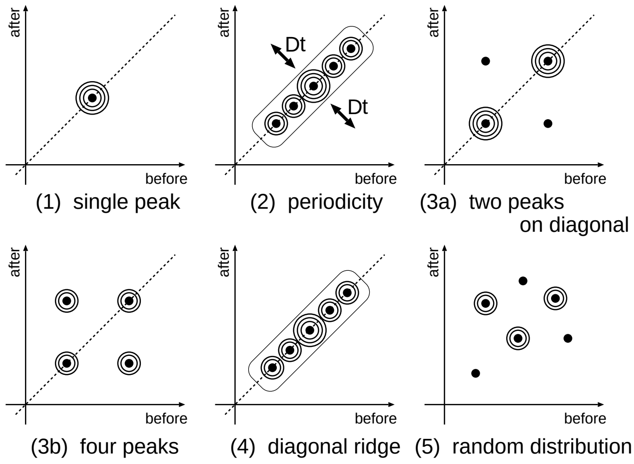


Figure 5: Distribution patterns in TNVDCM

$D_i'(t)$, which ranges from 1 to M, we compute $M \times M$ matrix $A_i = \{a_{jk}^i\}(j,k = 1, \cdots, M)$, which is defined as follows:

$$a_{jk}^i = \frac{1}{T - \Delta t} \sum_{t=1}^{T-\Delta t} P(i,j,k,t), \tag{7}$$

where

$$P(i,j,k,t) = \begin{cases} 1 & \mathit{if} \quad D_i'(t) = j \ \mathit{and} \ D_i'(t + \Delta t) = k, \\ 0 & \mathit{otherwise}. \end{cases} \tag{8}$$

and Δt represents the time interval. We call this matrix the *temporal NVD co-occurrence matrix* (TNVDCM, in short). An example of TNVDCM is shown in Figure 4.

3.2 Distribution Patterns in TNVDCM

The population distribution in TNVDCM reflects the fluctuation pattern of non-stationary background objects. We identify the following five different types of distribution patterns (Figure 5):

(1) Single peak TNVDCM includes a single peak. This pattern corresponds to a static background region, where NVD values do not vary dynamically.

(2) Periodicity If the image fluctuation pattern in a block has periodicity, the diagonal moment $M_{\rm d}$ of the distribution in TNVDCM[2] varies as Δt changes. When Δt is equal to the period, the moment becomes minimal. Using this property, we can examine if the fluctuation pattern is periodic and if so, compute the period of the fluctuation ω_i.

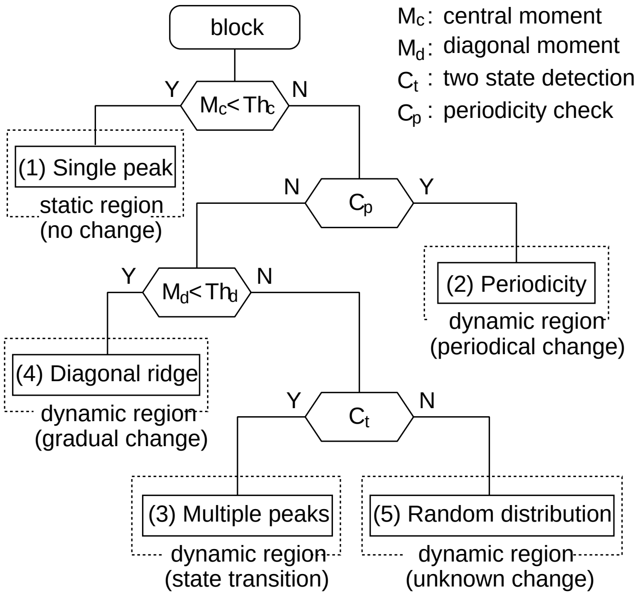


Figure 6: Decision tree for the categorization

(3) Multiple peaks When the image pattern fluctuation can be modeled by the temporal transition between discrete states (e.g. vibrating high contrast edges), multiple peaks appear in TNVDCM. Figure 5(3a)(3b) illustrate possible distribution patterns for two state transition models.

(4) Diagonal ridge When a block represents waving leaves or swinging tree branches, NVD varies gradually and widely. Then a diagonal smooth ridge appears in TNVDCM.

(5) Random distribution Randomly moving objects such as fluttering flags causes a random distribution in TNVDCM.

3.3 Optimized Background Subtraction Operator

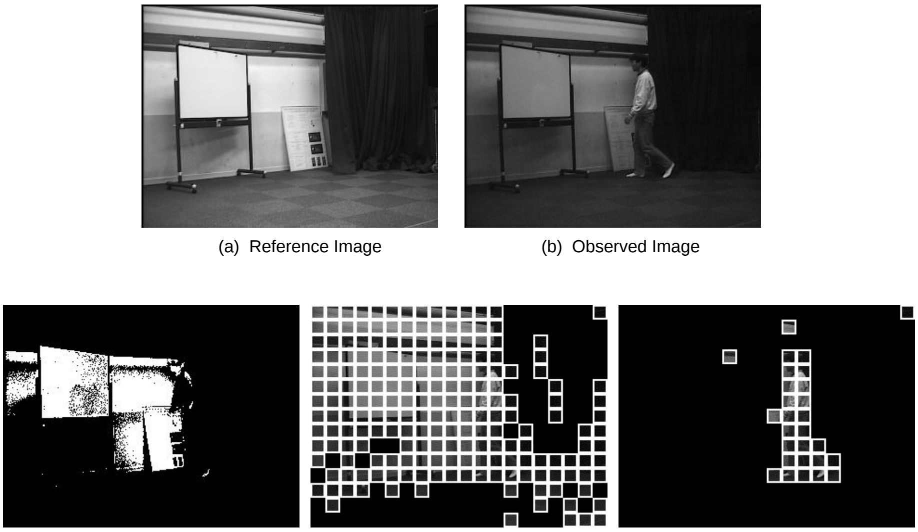
To design a background subtraction method for non-stationary scenes, we first classify each block i into the above mentioned five categories based on its TNVDCM. Given an observed image $g(x,y,t)$, we select and apply the optimized subtraction operator according to the category of block i.

3.3.1 Categorization of non-stationary background objects

The procedure of categorizing non-stationary background objects is described by the decision tree in Figure 6, where $M_{\rm c}$, $M_{\rm d}$, $C_{\rm t}$, and $C_{\rm p}$ represent the central moment[3], the diagonal moment, the result of

[2] The diagonal moment is defined as $M_{\rm d} = \sum_{j,k} a_{jk}|j-k|$.

[3] The central moment is defined as $M_c = \sum_{j,k} a_{jk}(|j - c_x| +$



(a) Reference Image (b) Observed Image

(c) Detection Result by Pixelwise Subtraction (d) Detection Result by NVD (e) Detection Result by SMNVD

Figure 7: Experimental Result 1. In (c), white pixels are detected regions. In (d) and (e), square regions are the detected blocks

examining two state transition[4], and the result of the periodicity check respectively. The examination of two state transition is implemented by applying clustering to TNVDCM. The periodicity check is performed by examining variations of M_{d} with different time interval Δts.

3.3.2 Optimized Subtraction Operators

Given an observed image $g(x, y, t)$, the optimized subtraction operator is selected and applied to block i according to its classified category.

(1) Single peak The distribution of NVD values in block i computed from the reference and the background image sequence is modeled by the Gaussian distribution and the threshold operation for $D_i(t)$ (i.e. NVD value computed for block i in $g(x, y, t)$) using Equation 4 is applied.

(2) Periodicity In the categorization process, the period of the fluctuation ω_i is estimated. Assuming that the difference between a pair of NVD values separated by time interval ω_i obeys the Gaussian, the threshold operation for $D_i(t-\omega_i)-D_i(t)$ is applied.

(3) Multiple peaks The state transition is modeled by a hidden Markov model and the probability of the series of $D_i(t-l)(l = L, \cdots, 0)$ calculated from the observed image sequence is computed based on the model and is thresholded (not implemented yet).

(4) Diagonal ridge The distribution in TNVDCM is normalized to represent the probability density function and the likelihood of a pair of NVD values, $D_i(t-\Delta t)$ and $D_i(t)$, is computed and thresholded.

(5) Random distribution Same as (4).

4 Experiments

In this section, we show two experimental results to examine the effectiveness of the implementations of SMNVD and TNVDCM. Although these implementations are very simple, the experimental results demonstrate their effectiveness.

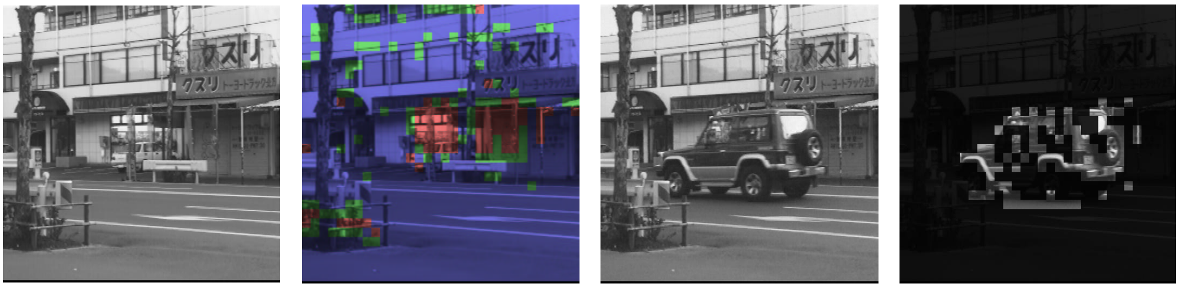
4.1 Experiment 1

This experiment was conducted to show that SMNVD is superior to NVD under varying illumination conditions. Images are captured from a fixed black and white camera. First the median image (Figure 7(a)) is computed from the background image sequence and is used as the reference image for the background subtraction. Video images which contain a moving person are captured under the relatively dark illumination (Figure 7(b)).

Figures 7(c), (d) and (e) shows the anomalous regions detected by the ordinary pixelwise intensity subtraction, the threshold operation for the probability density function of NVD, and the method using SMNVD described in Section 2.3 respectively. As is obvious from these results, the proposed background subtraction method using SMNVD is much robust against varying illuminations.

$|k - c_y|$), where (c_x, c_y) denotes the centroid of the population distribution in TNVDCM.

[4] In the current implementation, we assume that the periodic fluctuation pattern in a block has two states.



(a) Reference Image (b) Categorization Result (c) Observed Image (d) Detection Result

Figure 8: Experimental Result 2

4.2 Experiment 2

This experiment was conducted to demonstrate the effectiveness of the method described in Section 3.3. Its aim is to detect moving objects from indoor and outdoor scenes including flickering CRTs, waving leaves, fluttering flags, and swinging wires. Several different sequences of video images were taken and analyzed. Figure 8 shows one of the experimental results, where swinging tree branches and wires, and fluttering flags are included and a car moves from right to left.

Figure 8(b) illustrates the result of the block categorization, where three colors, blue, green, and red, represent “single peak”, “diagonal ridge”, and “random distribution” respectively. While no block is classified into “periodicity” and “multiple peaks” in this scene, these two classes appeared in other scenes used in the experiment.

Figure 8(d) shows the result of the background subtraction using the optimized subtraction operators. The bright blocks denote detected anomalous regions corresponding to moving objects. Although some moving regions are missing, no erroneous block is detected. This demonstrates that the proposed optimized background subtraction is very robust against fluctuations caused by non-stationary background objects.

5 Conclusions

In this paper, we proposed a robust background subtraction method for non-stationary scenes. The non-stationarities modeled by the method are variation of overall lighting conditions and local image pattern fluctuations: as for the former, we proposed *spatially modulated normalized vector distance* (SMNVD) and as for the latter, *temporal NVD co-occurrence matrix* (TNVDCM). Although the implementations described in this paper are very simple, the experimental results demonstrated their effectiveness.

After writing this paper, Toyama et al [6] conducted extensive experiments to compare the performance of various background subtraction methods. According to their result, SMNVD attained rather high performance, while TNVDCM was not evaluated. Moreover, we have made substantial improvement of SMNVD, where illumination conditions of an observed image are estimated by the principal component analysis and the threshold values for the background subtraction (i.e. $Th1$ and $Th2$ in equation (6)) are adjusted based on the estimated illumination conditions. We verified that the performance is improved significantly[7].

As for TNVDCM, we should develop a more concrete algorithm by studying its theoretical characteristics. In addition, we should study the integration of SMNVD and TNVDCM, which will realize the robust background subtraction for real world scenes.

This work was supported by the Research for the Future Program of the Japan Society for the Promotion of Science (JSPS-RFTF96P00501).

References

[1] H.Nakai: Robust Object Detection Using A-Posteriori Probability, *Tech. Rep. of SIG-CV90-1*, Information Processing Society Japan, 1994 (in Japanese)

[2] E.Grimson: A Forest of Sensors, *Proc. of VSAM Workshop*, 1997.11

[3] L.Davis: Visual Surveillance and Monitoring, *Proc. of VSAM Workshop*, 1997.11

[4] S.Nagaya, T.Miyatake, T.Fujita, W.Ito, and H.Ueda: Moving Object Detection by Time-Correlation-Based Background Judgement Method, *IEICE Trans. D-II*, Vol. J79-D-II, No. 4, pp. 568-576, 1996 (in Japanese)

[5] K.Ohba, Y.Sato, and K.Ikeuchi: Appearance Based Object Recognition with Illumination Invariance, *In Proc. DARPA Image Understanding Workshop*, pp. 1229-1236, 1997.

[6] Toyama, K. et al: Wallflower: Principles and Practice of Background Maintenance, Proc. of ICCV, pp. 255-261, 1999

[7] H. Habe, T. Wada, and T.Matsuyama: A Robust Background Subtraction Method under Varying Illumination, Tech. Rep. of IPSJ, SIG-CVIM115-3, 1999 (in Japanese)