Available online at www.sciencedirect.com
ScienceDirect
Pattern Recognition 41 (2008) 2253–2262
PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/pr

Recognition of camera-captured low-quality characters using motion blur information

Hiroyuki Ishida[a,b,*], Tomokazu Takahashi[a,b], Ichiro Ide[a], Yoshito Mekada[c], Hiroshi Murase[a]

[a] *Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan*
[b] *Japan Society for the Promotion of Science, Japan*
[c] *School of Life System Science & Technology, Chukyo University, 101, Tokodachi, Kaizu-cho, Toyota, Aichi 470-0393, Japan*

Received 11 August 2006; received in revised form 9 January 2008; accepted 10 January 2008

Abstract

Camera-based character recognition has gained attention with the growing use of camera-equipped portable devices. One of the most challenging problems in recognizing characters with hand-held cameras is that captured images undergo motion blur due to the vibration of the hand. Since it is difficult to remove the motion blur from small characters via image restoration, we propose a recognition method without de-blurring. The proposed method includes a generative learning method in the training step to simulate blurred images by controlling blur parameters. The method consists of two steps. The first step recognizes the blurred characters based on the subspace method, and the second one reclassifies structurally similar characters using blur parameters estimated from the camera motion. We have experimentally proved that the effective use of motion blur improves the recognition accuracy of camera-captured characters.

Keywords: Motion blur; Digital camera; Low-quality character; Character recognition; Generative learning method

1. Introduction

Character recognition technologies using portable digital cameras have gained attention in recent years in proportion to the diffusion of portable digital imaging devices [1]. However, even with the improvement of the devices, the quality of captured images is still not sufficient for recognizing characters in many practical cases. For example, characters in such images tend to be small and blurred even from a slight vibration of the hand holding the camera. This problem becomes more serious when the photographer backs the camera away from the target document, trying to capture a larger part of it. In this paper, an approach to recognize low-quality characters in such blurred image is presented.

As described above, image degradation is an unavoidable problem peculiar to camera-based character recognition. One of the main approaches to cope with such degradation is image restoration [2]. Various attempts have been made for image restoration. In Ref. [3], Hobby proposed a super-resolution method for small characters, and in Ref. [4], Li et al. used multiple images to create super-resolution image, while in Ref. [5], Mancas-Thillou et al. used a Teager filter to enhance low-resolution text. The PSF (point spread function) can also be applied to remove optical blur. The compound method proposed by Tsunashima et al. [6] is a simple but effective way to obtain the optical blur PSF because it allows us to estimate it by simply averaging multiple captured images. Another form of degradation to be removed is motion blur, which requires the identification of blur parameters [7]. Ben-Ezra et al. proposed a method for de-blurring motion blurred images using PSF [8].

In practical applications, however, restoring an image is not always effective for character recognition because small characters are difficult to de-blur. This paper proposes a recognition method that does not need any restoration. It instead, copes with the degradations by learning artificially degraded images and using estimated motion blur information.

* Corresponding author at: Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8601, Japan. Tel.: +81 52 789 3310; fax: +81 52 789 3807.
E-mail address: hishi@murase.m.is.nagoya-u.ac.jp (H. Ishida).

0031-3203/$30.00 © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2008.01.009

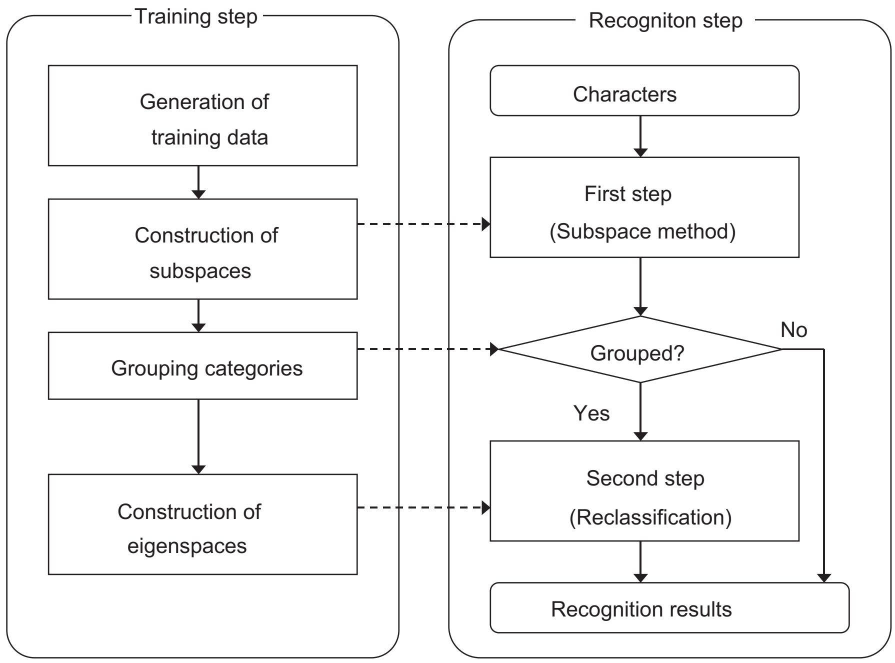

Fig. 1. Flow of the proposed camera-based character recognition.

In our previous papers [9,10], models for generating such images were presented and used for constructing subspaces. However, the use of generation models was limited to obtaining training sets. Once the subspaces were constructed in the training step, test images were compared only with them regardless to the individual training images with various degradation parameters. The proposed method also uses the subspace in the first step of recognition, but in the second step, the test images are directly compared with the generated images by means of eigenspace method [11] in order to correct the misclassifications in the earlier step.

Fig. 1 illustrates the flow of the proposed method. The training step is based on the generative learning method [9,10], where training images undergoing various speeds and orientations of motion blur are generated. The recognition method consists of two steps. The first employs the subspace method [12], whose effectiveness for low-resolution character recognition is demonstrated in Ref. [13]. However, the subspace method constructs a single subspace from the training images with various speeds and orientations of blur, which often yields the misclassification among structurally similar characters. The eigenspace method [11] is more effective for such characters, since the similarity to each training image is evaluated. A reclassification based on the eigenspace method is then introduced as the second step to improve the recognition accuracy of such characters. This second step reclassifies characters by effective use of the motion blur. For this purpose, motion blur parameters are estimated from camera motion; the similarity between the characters and training images simulated with the motion blur parameters is evaluated in the recognition step.

This paper is organized as follows: Section 2 explains the generation process of the training images which are used in all stages of the recognition. The parameters used both for the image generation and for the recognition are also introduced first in this section. Section 3 describes the first step of the recognition using the subspace method. Section 4 details the second step of the recognition using the eigenspace method and estimated blur parameters. Section 5 demonstrates the performance of the method through an experiment, and Section 6 concludes this paper.

2. Generation of training images

The generative learning method was developed to generate degraded patterns by simulating actual degradation. Traditionally, this synthesis-based approach has often been used for learning distorted characters in handwritten character recognition [14,15]. We have been applying a generative learning method to camera-based character recognition, and have so far investigated the effectiveness of an optical blur model and a motion blur model in Refs. [9,10], respectively. Also, the effectiveness of using a resolution transformation model to simulate low-resolution characters has been demonstrated by Sun in Ref. [16]. In contrast with the collection-based approach such as that introduced in Ref. [17], the generative learning method eliminates the exhaustive collection of training images. It enables us to acquire parametrically degraded character images in accordance with the actual degradations.

2.1. Generation models

To simulate various degradations, four generation models are defined along with parameters that control the degradation degree of images. The generation models used for this work are listed below.

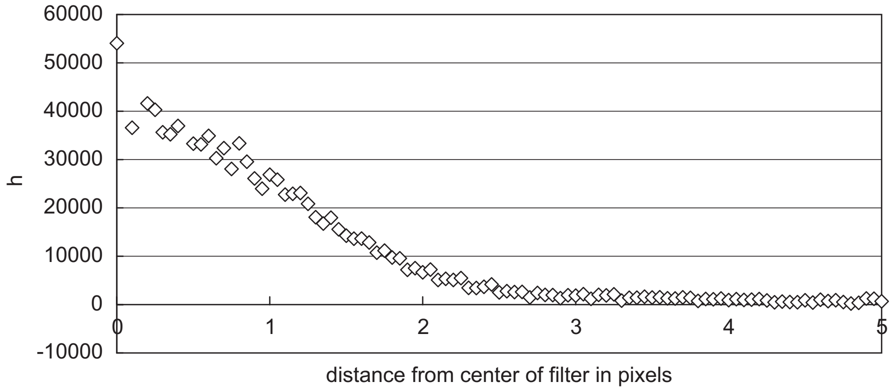

Fig. 2. Optical blur PSF. The horizontal axis indicates the distance from the center of a PSF filter.

- Optical blur model:
An optically blurred image z_1 is generated from the original image z_0 by convoluting it with an optical blur PSF h_{opt} as

$$z_1(x, y) = z_0(x, y) * h_{opt}(x, y), \tag{1}$$

h_{opt} needs to be estimated beforehand using the compound method [6]. This compound method enables us to estimate the optical blur PSF from an original image and its camera-captured images.
Let us briefly describe the estimation process. Assume that image degradation by the optical blur is represented as

$$g_i(x, y) = f(x, y) * h_{opt}(x, y) + n(x, y), \tag{2}$$

where f is the original image, g_i is the optically blurred image of f, h_{opt} is the optical blur PSF, and n is the noise function. Applying two-dimensional Fourier transformation to this equation, we obtain

$$H_{opt}(u, v) = \frac{G(u, v)}{F(u, v)} - \frac{N(u, v)}{F(u, v)}. \tag{3}$$

Since the noise component is unknown, h_{opt} cannot be estimated from a single g_i. The compound method averages multiple optically blurred images to restrain this noise. Assuming that we have I images g_i $(i = 1, 2, \ldots, I)$, the optical blur component in spatial frequency $\hat{H}_{opt}(u, v)$ is estimated by

$$\hat{H}_{opt}(u, v) = \frac{1}{I} \sum_{i=1}^{I} \frac{G_i(u, v)}{F(u, v)} - \frac{1}{I} \sum_{i=1}^{I} \frac{N_i(u, v)}{F(u, v)}. \tag{4}$$

Provided that I is large enough, the second term of this equation converges to 0 because no relation exists among the noise components $N_i(u, v)$ of each image, and then

$$H_{opt}(u, v) \approx \frac{1}{IF(u, v)} \sum_{i=1}^{I} G_i(u, v). \tag{5}$$

The optical blur PSF $h_{opt}(x, y)$ is obtained by inverting $H_{opt}(u, v)$. Fig. 2 shows an example of a PSF estimated from a digital camera.

- Motion blur model:
To deal with motion blur concisely, both speed and orientation of the motion blur appearing on one frame are assumed to be constant. This assumption allows us to employ the motion blur model proposed by Potmesil in Ref. [18] for computer-generated images. A motion blurred image z_2 is generated from the image z_1 from a blur extent parameter b and a blur angle parameter θ $(0 \leqslant \theta < \pi)$ as

$$z_2(x, y) = \int_{-1/2}^{1/2} z_1(x - bt\cos\theta, y - bt\sin\theta)\,\mathrm{d}t. \tag{6}$$

This operation can also be simplified in the form of a convolution with a motion blur PSF $h_{mot(b,\theta)}(x, y)$. The two-dimensional Fourier transformation is used to separate the blur component from the term z_1 as

$$H_{mot(b,\theta)}(u, v) = \frac{\sin[\pi b(u\cos\theta + v\sin\theta)]}{\pi b(u\cos\theta + v\sin\theta)}, \tag{7}$$

with $h_{mot(b,\theta)}(x, y)$ obtained by inverting $H_{mot(b,\theta)}(u, v)$. Consequently, Eq. (6) is represented as

$$z_2(x, y) = z_1(x, y) * h_{mot(b,\theta)}(x, y). \tag{8}$$

Unlike h_{opt} in the optical blur model, this $h_{mot(b,\theta)}$ is determined by the given two parameters.

- Segmentation model:
The segmentation model involves translation and expansion of the character in the image. A character area is defined as the minimum square region that contains the whole character. Let (x_0, y_0) be the center of the character area, and l be its side length. An image area to be segmented is defined by a horizontal gap parameter Δx, a vertical gap parameter Δy, and an expansion rate a as illustrated in Fig. 3.

- Resolution transformation model:
Here a synthetic degradation parameter d corresponding to the camera distance is introduced. The d is identical to the expansion rate of the PSF filter. If $d = 0$, the spatial resolution of the generated image is equivalent to that of the original image, while if $d = 1$, the generated image is identical to the convoluted image with the PSF.

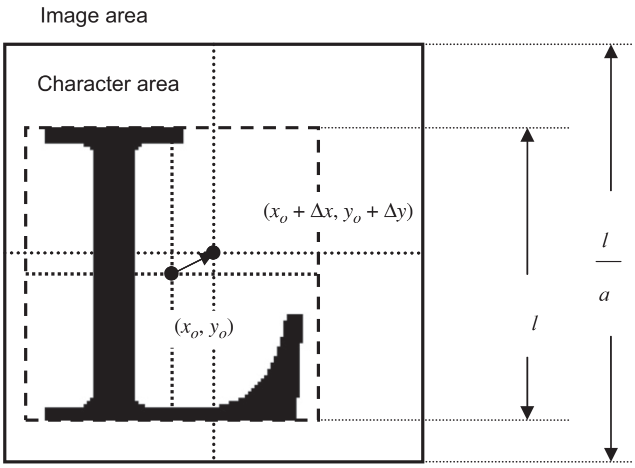

Fig. 3. Areas and parameters for segmentation.

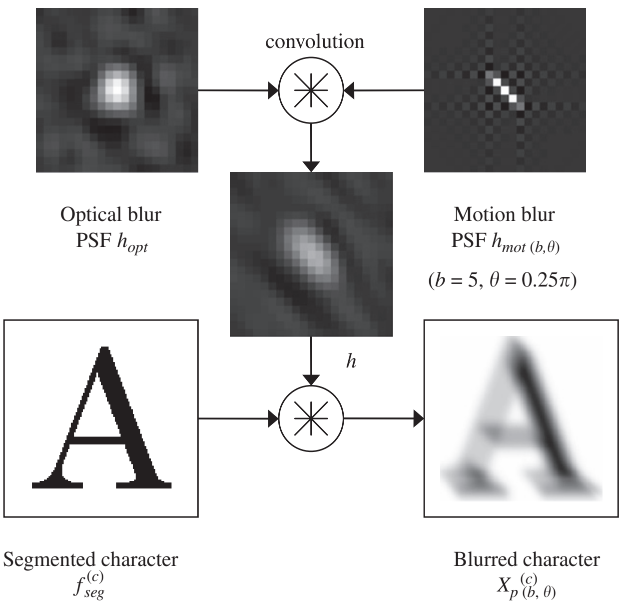

Fig. 4. Generation of a training image using PSFs.

A parameter vector $\boldsymbol{p}$ consisting of the parameters introduced above is defined as

$$\boldsymbol{p} = (b, \theta, \Delta x, \Delta y, a, d). \tag{9}$$

This parameter vector is used for generating training images.

2.2. Generation of training images

Training images are generated using the estimated optical blur PSF $h_{opt}(x, y)$ and the parameter $\boldsymbol{p}$. Fig. 4 shows the generation process by the PSFs. Let $\mathrm{x}_{\boldsymbol{p}}^{(c)}$ be a training image generated from category c's original image $f^{(c)}$ by a parameter vector $\boldsymbol{p}$ as

$$\mathrm{x}_{\boldsymbol{p}}^{(c)}(x, y) = \sum_{i,j} h(i, j) f_{seg}^{(c)}(x - id, y - jd), \tag{10}$$

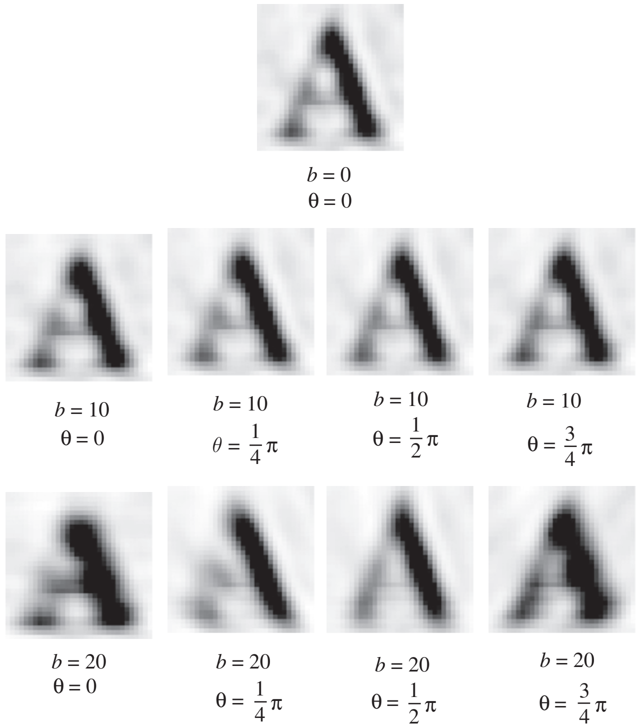

Fig. 5. Examples of the generated images for category "A" ($\Delta x = \Delta y = 0$, $a = 15/16$, $d = 2$).

where

$$h(x, y) = h_{opt}(x, y) * h_{mot(b,\theta)}(x, y) \tag{11}$$

and

$$f_{seg}^{(c)}(x - x_0, y - y_0) = f^{(c)}\left(\frac{x}{a} + \Delta x, \frac{y}{a} + \Delta y\right). \tag{12}$$

Examples of the generated training images are shown with corresponding values of $\boldsymbol{p}$ in Fig. 5.

3. First step of recognition: recognition by the subspace method

3.1. Construction of a subspace

In the training step, a subspace is constructed from various training images for each category. The constructed subspaces have ability to classify low-quality characters robustly, except for some structurally similar categories.

Let $\mathcal{P}$ be a set of N different parameter vectors $\boldsymbol{p}_n$ ($n = 1, 2, \ldots, N$), where N is the number of training images used for constructing a subspace of a category. N training images are generated from parameter vectors $\boldsymbol{p}_n \in \mathcal{P}$. For each training image $\mathrm{x}_{\boldsymbol{p}_n}^{(c)}$, a vector $\boldsymbol{x}_{\boldsymbol{p}_n}^{(c)}$ is constructed from pixel values of the image as described below. First, $\mathrm{x}_{\boldsymbol{p}_n}^{(c)}$ is converted to a vector $\tilde{\boldsymbol{x}}_{\boldsymbol{p}_n}^{(c)}$ such that the mean of its elements becomes 0 by

$$\begin{aligned} \tilde{\boldsymbol{x}}_{\boldsymbol{p}_n}^{(c)} = [&\mathrm{x}_{\boldsymbol{p}_n}^{(c)}(0, 0) - \bar{x}_{\boldsymbol{p}_n}^{(c)} \quad \cdots \quad \mathrm{x}_{\boldsymbol{p}_n}^{(c)}(w - 1, 0) - \bar{x}_{\boldsymbol{p}_n}^{(c)} \\ &\cdots \quad \mathrm{x}_{\boldsymbol{p}_n}^{(c)}(0, h-1) - \bar{x}_{\boldsymbol{p}_n}^{(c)} \quad \cdots \quad \mathrm{x}_{\boldsymbol{p}_n}^{(c)}(w-1, h-1) - \bar{x}_{\boldsymbol{p}_n}^{(c)}]^\top, \end{aligned} \tag{13}$$

Fig. 6. Top four eigenvectors for category "A".

where w and h are the width and the height of the image, respectively, and

$$\bar{x}_{\boldsymbol{p}_n}^{(c)} = \frac{1}{wh} \sum_{x=0}^{w-1} \sum_{y=0}^{h-1} \mathrm{x}_{\boldsymbol{p}_n}^{(c)}(x, y).$$

Second, this vector is normalized to $\boldsymbol{x}_{\boldsymbol{p}_n}^{(c)}$ whose norm is 1 by

$$\boldsymbol{x}_{\boldsymbol{p}_n}^{(c)} = \frac{\widetilde{\boldsymbol{x}}_{\boldsymbol{p}_n}^{(c)}}{\|\widetilde{\boldsymbol{x}}_{\boldsymbol{p}_n}^{(c)}\|}. \tag{14}$$

A matrix $\boldsymbol{Q}_1^{(c)}$ is then calculated as

$$\boldsymbol{Q}_1^{(c)} = \frac{1}{N} \boldsymbol{X}_1^{(c)} (\boldsymbol{X}_1^{(c)})^\top, \tag{15}$$

where matrix $\boldsymbol{X}_1^{(c)}$ is represented by a list of the N vectorized training images $\boldsymbol{x}_{\boldsymbol{p}_n}^{(c)}$ as

$$\boldsymbol{X}_1^{(c)} = [\boldsymbol{x}_{\boldsymbol{p}_1}^{(c)} \ \cdots \ \boldsymbol{x}_{\boldsymbol{p}_N}^{(c)}]. \tag{16}$$

Next, the eigenvalues and corresponding eigenvectors of this matrix $\boldsymbol{Q}_1^{(c)}$ are calculated. The eigenvectors are sorted in order of the magnitude of their corresponding eigenvalues, and the largest R_1 ($R_1 < N$) eigenvectors $\boldsymbol{e}_{r_1}^{(c)}$ ($r_1 = 1, 2, \ldots, R_1$) are used for the recognition. Examples of the eigenvectors are illustrated in Fig. 6.

3.2. Character recognition using multiple frames

Yanadume et al. demonstrated that multiple-frame integration improves the recognition accuracy of low-resolution characters [17]. Given M frames of the same character, and letting $\boldsymbol{z}_m$ denote the vectorized and normalized target image in the mth frame, the recognition result in the first step, $\hat{c}_1$ is determined from the inner product to the R_1 eigenvectors $\boldsymbol{e}_{r_1}^{(c)}$ ($r_1 = 1, \ldots, R_1$), by

$$\hat{c}_1 = \arg\max_{\forall c} \sum_{m=1}^{M} \sum_{r_1=1}^{R_1} (\boldsymbol{e}_{r_1}^{(c)\top} \boldsymbol{z}_m)^2. \tag{17}$$

4. Second step of recognition: reclassification using blur information

The recognition results obtained in the first step tend to involve misclassification within certain groups of structurally similar categories. The second step attempts to reclassify such dubious results to the correct category using the eigenspace method [11]. The blur parameters estimated from camera motion are used for the matching of characters in this step. This attempt is based on the idea that the blur parameters should supply supplementary information for differentiating structurally similar categories.

4.1. Difference between subspace and eigenspace

The difference between the subspace method and the eigenspace method is described here.

Fig. 7 illustrates the recognition schemes of these methods. In the subspace method, an input image is compared with subspaces of all categories and then classified according to similarities to the subspaces. These subspaces are constructed from training images with various degradation parameters. Consequently, the subspace method has a general capability to recognize the degraded images, although the similarities to each training image cannot be evaluated.

Unlike subspace, eigenspace is constructed for a set of categories. All the images, regardless of training or input, are projected onto the same eigenspace. The input image is then classified to the nearest category in the eigenspace. The eigenspace method has the advantage that it can evaluate a distance to individual training images as a dissimilarity. It can use the blur information to select the images for the distance calculation, whereas the subspace method cannot use such information in the recognition step.

4.2. Grouping structurally similar characters

For each category g, characters that are frequently misclassified to category g are grouped and described as $\mathcal{G}^{(g)}$. Such groups can be organized by applying the first step of recognition to a certain amount of samples. Let $\rho(g|c)$ denote the rate at which a character in category c is classified to category g in the first step. The category c is grouped if $\rho(g|c) \geqslant \tau$, where τ is a grouping threshold; and of course, g itself also needs to be a member of $\mathcal{G}^{(g)}$. In brief, $\mathcal{G}^{(g)}$ is organized as

$$\mathcal{G}^{(g)} = \{c | \rho(g|c) \geqslant \tau\} \cup \{g\}. \tag{18}$$

4.3. Construction of an eigenspace in groups

An eigenspace used for this second step of recognition is constructed in each group. Similar to the step in which the subspaces were constructed, a covariance matrix $\boldsymbol{Q}_2^{(g)}$ of group g is calculated as

$$\boldsymbol{Q}_2^{(g)} = \frac{1}{KN} \boldsymbol{X}_2^{(g)} (\boldsymbol{X}_2^{(g)})^\top, \tag{19}$$

except that a matrix $\boldsymbol{X}_2^{(g)}$ is represented by all the training images of categories $c_k \in \mathcal{G}^{(g)}$ ($1 \leqslant k \leqslant K = |\mathcal{G}^{(g)}|$) and their mean $\boldsymbol{\mu}^{(g)}$ as

$$\boldsymbol{X}_2^{(g)} = [\widetilde{\boldsymbol{X}}_1^{(c_1)} \ \cdots \ \widetilde{\boldsymbol{X}}_1^{(c_K)}], \tag{20}$$

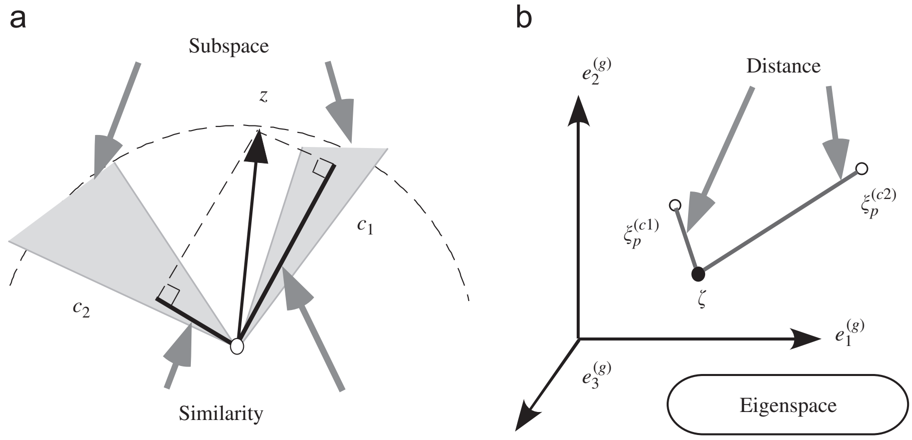

Fig. 7. While the subspace method evaluates similarities to the subspace, the eigenspace method evaluates distances (dissimilarities) to each training image. (a) In the subspace method, the similarity of input image $\boldsymbol{z}$ to the subspace $\{\boldsymbol{e}_r^{(c)}\}$ is defined as $\sum_r (\boldsymbol{e}_r^{(c)\top}\boldsymbol{z})^2$. (b) In the eigenspace method, the distance between input image $\boldsymbol{\zeta}$ and training image $\boldsymbol{\xi}_{\boldsymbol{p}}^{(c)}$, both of which are projected on the eigenspace, is defined as $\|\boldsymbol{\zeta} - \boldsymbol{\xi}_{\boldsymbol{p}}^{(c)}\|$.

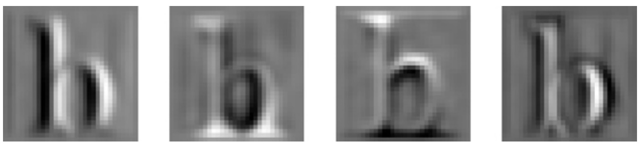

Fig. 8. Top four eigenvectors for group $\mathcal{G}^{(\text{“h”})} = \{\text{“b”}, \text{“h”}\}$.

where

$$\widetilde{\boldsymbol{X}}_1^{(c_k)} = [\boldsymbol{x}_{\boldsymbol{p}_1}^{(c_k)} - \boldsymbol{\mu}^{(g)} \quad \cdots \quad \boldsymbol{x}_{\boldsymbol{p}_N}^{(c_k)} - \boldsymbol{\mu}^{(g)}] \tag{21}$$

and

$$\boldsymbol{\mu}^{(g)} = \frac{1}{KN}\sum_{k=1}^{K}\sum_{n=1}^{N} \boldsymbol{x}_{\boldsymbol{p}_n}^{(c_k)}. \tag{22}$$

Next, the eigenvalues and corresponding eigenvectors of this matrix $\boldsymbol{Q}_2^{(g)}$ are calculated. The eigenvectors are sorted in order of the magnitude of their corresponding eigenvalues, and the largest R_2 ($R_2 < KN$) eigenvectors $\boldsymbol{e}_{r_2}^{(g)}$ ($r_2 = 1, 2, \ldots, R_2$) are used. Examples of the eigenvectors are illustrated in Fig. 8.

4.4. Projection of the training images to the eigenspace

All the training images are projected onto the eigenspace as points. The following operation projects category c's training images $\boldsymbol{x}_{\boldsymbol{p}}^{(c)}$ ($c \in \mathcal{G}^{(g)}, \boldsymbol{p} \in \mathcal{P}$), and thereby the projected points $\boldsymbol{\xi}_{\boldsymbol{p}}^{(c)}$ in the eigenspace is obtained as

$$\boldsymbol{\xi}_{\boldsymbol{p}}^{(c)} = [\boldsymbol{e}_1^{(g)} \quad \cdots \quad \boldsymbol{e}_{R_2}^{(g)}]^\top (\boldsymbol{x}_{\boldsymbol{p}}^{(c)} - \boldsymbol{\mu}^{(g)}). \tag{23}$$

4.5. Character recognition using blur information

This second step of recognition utilizes blur information obtained from camera motion. Given that a recognition result from the first step is g, if $|\mathcal{G}^{(g)}| = 1$, there are no other candidates for consideration as a recognition result. Hence the final recognition result is also g, whereas if $|\mathcal{G}^{(g)}| \geqslant 2$, we dismiss the results from the first step once and compute the final result as described below.

(1) *Projection to the eigenspace*: First, the target image is projected onto the eigenspace of group $\mathcal{G}^{(g)}$. Letting $\boldsymbol{z}_m$ denote the vectorized and normalized target image in the mth frame, a projected point $\boldsymbol{\zeta}_m$ corresponding to the image is obtained by

$$\boldsymbol{\zeta}_m = [\boldsymbol{e}_1^{(g)} \quad \cdots \quad \boldsymbol{e}_{R_2}^{(g)}]^\top (\boldsymbol{z}_m - \boldsymbol{\mu}^{(g)}). \tag{24}$$

(2) *Estimation of blur information*: As for blur information, the extent and angle of the motion blur can be estimated from the camera motion. Let x_m and y_m represent the location of the target character in the image of the mth frame. A blur extent parameter $\hat{b}_m$ and a blur angle parameter $\hat{\theta}_m$ are estimated as follows:

$$\hat{b}_m = \sqrt{(x_m - x_{m-1})^2 + (y_m - y_{m-1})^2}, \tag{25}$$

$$\hat{\theta}_m = \tan^{-1}\frac{y_m - y_{m-1}}{x_m - x_{m-1}}. \tag{26}$$

(3) *Calculation of similarity*: In the eigenspace method, the smaller the distance between the projected points is, the more similar these original images are. Here we need to evaluate the distance between $\boldsymbol{\zeta}_m$ and points $\boldsymbol{\xi}_{\boldsymbol{p}}^{(c)}$ ($b = \hat{b}_m, \theta = \hat{\theta}_m$) for each category $c \in \mathcal{G}^{(g)}$. However, the estimated $\hat{b}_m$ and $\hat{\theta}_m$ generally do not coincide with any parameters $\boldsymbol{p} \in \mathcal{P}$ in the trained set, and more importantly, these estimated values can differ to a certain extent from the actual blur PSF. Accordingly, reference points are selected from training sets with a limited parameter range $\mathcal{B} \subset \mathcal{P}$ defined as

$$\mathcal{B} = \{\boldsymbol{p} \in \mathcal{P} \mid 0 \leqslant b \leqslant \hat{b} + \Delta b, \hat{\theta} - \Delta\theta \leqslant \theta \leqslant \hat{\theta} + \Delta\theta\}, \tag{27}$$

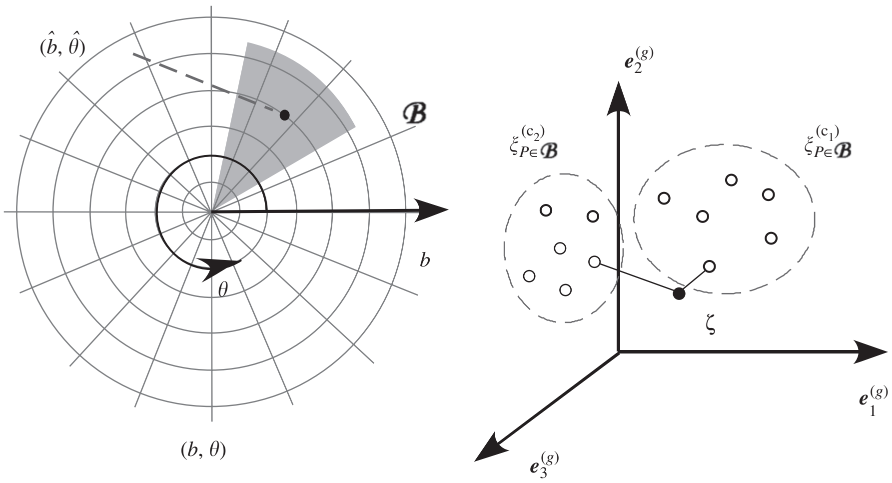

Fig. 9. Reference points used for the classification are selected from parameter range $\mathcal{B}$ that is restricted by estimated blur parameters.

assuming that the actual b is not much greater than the estimated $\hat{b}$, and that the difference between θ and $\hat{\theta}$ is not large. The classification is based on the nearest neighbor rule. Fig. 9 illustrates the classification scheme of the proposed method. For each category, a distance to the nearest reference point is evaluated. The final result $\hat{c}_2$ is computed using

$$\hat{c}_2 = \arg \min_{c\in\mathcal{G}^{(g)}} \sum_{m=1}^{M} \min_{\boldsymbol{p}\in\mathcal{B}} \|\boldsymbol{\zeta}_m - \boldsymbol{\xi}_{\boldsymbol{p}}^{(c)}\|. \tag{28}$$

5. Experiment

5.1. Conditions

The performance of the proposed method was evaluated with a digital camera (Panasonic DMC-FX9) that provides the ability to record video with a spatial resolution of 640×480 pixels and 30 frames per second. As test data, 62 characters (A–Z, a–z, 0–9: Century font) printed on a paper were captured. The average size of the printed characters was $5\,\text{mm}^2$. The distance to the paper was 30 cm, and the focal length of the camera was 5.8 cm; the average character size in the captured images was 11×11 pixels. The segmented area for each character was the minimum square that includes the whole character. The anti-blur function of the digital camera was kept off during the experiment.

5.2. Training step

First, the optical blur PSF of the camera needs to be estimated. Two hundred images were taken for PSF estimation from a distance of 30 cm. In the generation step, we controlled the parameters so that the training images set should vary. The synthetic degradation parameter d was changed in four steps ($d = 0.5, 1.0, 1.5, 2.0$), the blur extent parameter b by 11 steps ($b = 0, 2, \ldots, 20$), the blur angle parameter θ by 12 steps ($\theta = 0, \pi/12, \ldots, 11\pi/12$), the expansion rate parameter a by three steps ($a = 14/16, 15/16, 1$), and the segmentation parameters Δx and Δy individually by three steps ($-a, 0, a$), which resulted in obtaining 14 256 training images (32×32 pixels) per category. The original images were also in Century font. We used the top 10 eigenvectors for the first step of the recognition ($R_1 = 10$). The number of eigenvectors was determined such that the cumulative contributions of all subspaces were over 97.5%.

Table 1
Groups ($g : \mathcal{G}^{(g)}$) within which characters are reclassified

$\tau = 0.01$	$\tau = 0.02$	$\tau = 0.05, 0.10$	$\tau = 0.20$
L: {L, t}	S: {S, 8}	V: {V, v}	V: {V, v}
O: {O, o}	V: {V, v}	W: {W, w}	h: {b, h}
R: {F, R}	W: {W, w}	h: {b, h}	l: {I, i, l,1}
S: {S, 8}	h: {b, h}	l: {I, i, l, 1}	
V: {V, v}	l: {I, i, l, 1}		
W: {W, w}	1: {i, 1}		
h: {b, h}			
l: {I, i, j, l, 1}			
1: {i, 1}			

If recognition result in the first step is character g, final result is determined from category set $\mathcal{G}^{(g)}$.

Next, the groups used for the reclassification were organized. As the recognition samples for the grouping, 300 image sequences composed of 10 successive frames each were taken in the same way as the test data. For the grouping threshold in Eq. (18), five cases ($\tau = 0.01, 0.02, 0.05, 0.10, 0.20$) were tested. Table 1 lists the organized groups in these cases. The eigenspaces were constructed as described in Section 4. As for the value of R_2 in Eq. (28), we used the smallest number of eigenvectors whose cumulative contribution was over 80%.

5.3. Comparison with other methods

The sets of points, which were used for distance evaluation in the second step of recognition, were computed by projecting

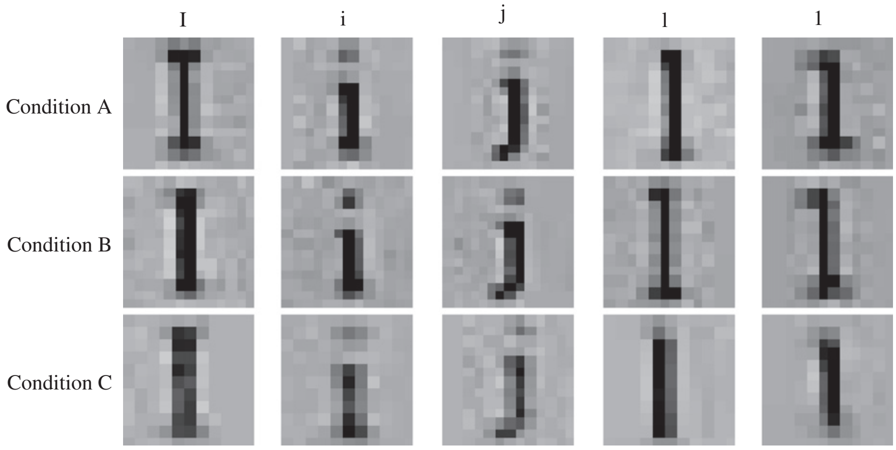

Fig. 10. Examples of the test data.

the training images onto the eigenspaces. The parameter range in Eq. (27) was set as $(\Delta b, \Delta\theta) = (2, \pi/6)$.

5.4. Methods

In order to evaluate the performance of the proposed method, it was compared with the subspace method and with the original eigenspace method.

Recognition results of the subspace method are obtained by Eq. (17). The proposed method is equivalent to the subspace method if the second step of recognition is not employed, namely if $\tau > 1$.

In the original eigenspace method compared here, all the results are obtained by the distance calculation in a single eigenspace, which is also known as an universal eigenspace [11]. Its recognition process is basically the same as described in Section 4, except that the eigenspaces of groups are replaced by the universal eigenspace. The proposed method is equivalent to the original eigenspace method if all categories g have a group $\mathcal{G}^{(g)}$ consisting of all categories, namely if $\tau = 0$.

5.5. Conditions and recognition results

Three photographic conditions were set for this experiment. Image sequences for the test data were taken under:

Condition A: Still camera on a tripod.
Condition B: Camera held as still as possible.
Condition C: Camera held by vibrating hand.

We used image sequences composed of 10 successive frames for the tests. The number of the image sequences for each Conditions A, B, and C were 300, 1736, and 503, respectively. The image sequences for Conditions A and B were taken by six persons. Fig. 10 shows some examples of the test data, while the distribution of the estimated blur extent $\hat{b} = \sum_{m=1}^{M} \hat{b}_m / M$ is given in Fig. 11, where we can see that $\hat{b}$ is not always zero even under Condition B. The results are shown in Table 2.

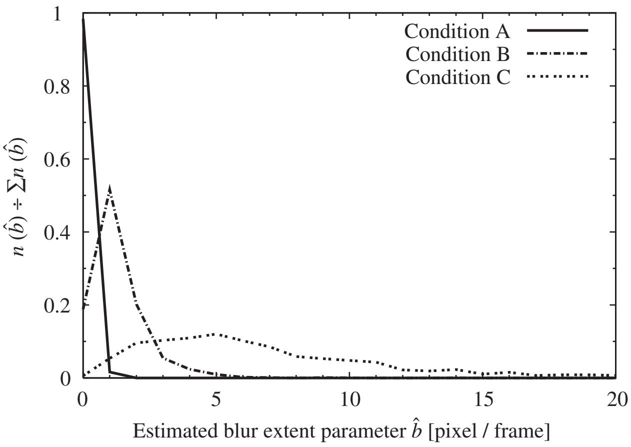

Fig. 11. Distributions of estimated blur extent $\hat{b}$ by conditions. $n(\hat{b})$ is the number of samples whose blur extent is $\hat{b}$.

5.6. Discussion

By comparing the results to those of the subspace method, the availability of the second recognition step is demonstrated. According to the results, the recognition rates were improved by introducing the second step. The proposed method was effective particularly under Conditions B and C, indicating that some blurred characters, which were misclassified in the first step, were correctly reclassified in the second step. It is also worth remarking that Condition B was more appropriate for the recognition than Condition A. This is because integrating time-varying images by Eqs. (17) and (28) was effective for recognizing low-resolution characters.

While the usefulness of the second step using the eigenspace method was shown, the recognition rates from the original eigenspace method were low under any condition. One reason is that the number of categories was large. The original eigenspace method using only an universal eigenspace is not suitable for the purpose such as character recognition. The eigenspace is useful if it is constructed among the small number of structurally similar categories and if the image appearance varies

Table 2
Recognition rates of characters under various conditions

Method	EM	Proposed method				SM
Grouping threshold	($\tau = 0$)	$\tau = 0.01$	$\tau = 0.02$	$\tau = 0.05, 0.10$	$\tau = 0.20$	($\tau > 1$)
Condition A	77.22	97.01	97.37	97.39	97.39	97.30
Condition B	77.49	98.38	98.73	98.69	98.69	98.21
Condition C	75.03	93.58	94.29	94.29	94.30	93.76

The proposed method is compared also with the subspace method (SM) and the original eigenspace method (EM).

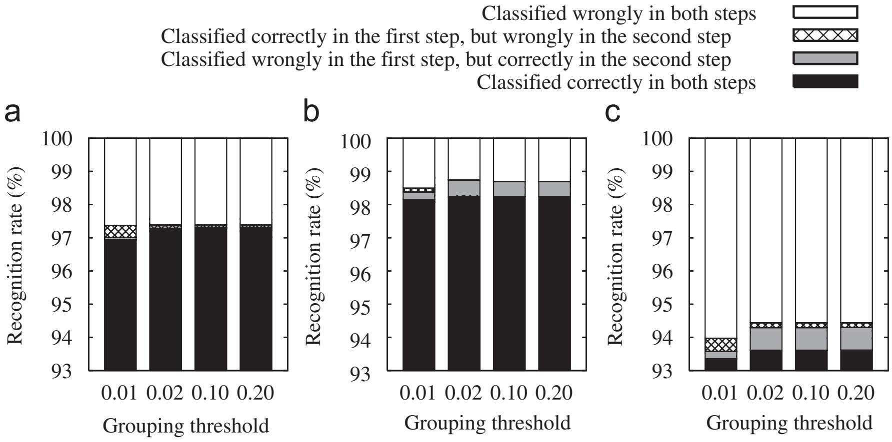

Fig. 12. The number of misclassified characters in each step. Recognition rates for various τ are compared: (a) Condition A; (b) Condition B; (c) Condition C.

depending on parameters. The proposed method uses the eigenspace effectively for the classification within groups of such characters.

A problem connected with grouping is the determination of threshold τ. As discussed above, desirable grouping contributes to a better performance. If we assume that the recognition accuracy of the second step is identical to that of the original eigenspace method, it is about 80% due to the results in Table 2. Accordingly, it can be effective to determine τ such that $\tau \leqslant 1-0.8=0.2$. When $\tau=0.01$, however, the performance under Conditions A and C was lower than the subspace method. Fig. 12 shows how much characters were classified correctly or wrongly in each step. Setting τ too small increased the number of misclassified characters in the second recognition step. As shown in Table 1, the number of groups increases as τ decreases, which can result in over-grouping. These results indicate that the second recognition step is effective for frequently misclassified categories and therefore over-grouping should be avoided.

6. Conclusion

In this paper, we proposed a method for improving the recognition accuracy of camera-captured characters without restoring images. This method uses the blur information estimated from camera motion. A reclassification step using this information was introduced to reduce classification errors. It was experimentally proved that the effective use of motion blur improves the recognition accuracy of blurred characters. Since the training step is based on the generative learning method, our method can be applied easily to characters of any size and font. Evaluating the method's performance under various conditions is a future work, together with improving the reclassification accuracy. Parameters from other models such as rotation model could also be valuable as supplementary information for the classification.

Acknowledgment

Parts of this research were supported by the Grants-In-Aid for Scientific Research (16300054 and 17650050) and the 21st century COE program from the Ministry of Education, Culture, Sports, Science and Technology. This work is implemented based on the MIST library (http://mist.suenaga.m.is.nagoya-u.ac.jp/).

References

[1] D. Doermann, J. Liang, H. Li, Progress in camera-based document image analysis, in: Proceedings of the 5th International Conference on Document Analysis and Recognition, Edinburgh, Scotland, 2003, pp. 606–616.

[2] H. Andrew, B. Hunt, Digital Image Restoration, Prentice-Hall, Englewood Cliffs, NJ, 1977.

[3] J. Hobby, H. Baird, Degraded character image restoration, in: Proceedings of the 5th UNLV Symposium on Document Analysis and Information Retrieval, Las Vegas, USA, 1996, pp. 177–189.

[4] H. Li, D. Doermann, Text enhancement in digital video using multiple frame integration, in: Proceedings of the 7th ACM International Conference on Multimedia, Orlando, Florida, 1999, pp. 19–22.
[5] C. Mancas-Thillou, M. Mirmehdi, Super-resolution text using the Teager filter, in: Proceedings of the 1st International Workshop on Camera-Based Document Analysis and Recognition, Seoul, Korea, 2005, pp. 10–16.
[6] N. Tsunashima, M. Nakajima, Estimation of point spread function using compound method and restoration of blurred images, IEICE Trans. J81-D-II (11) (1998) 2688–2692 (in Japanese).
[7] Y. Yitzhaky, N. Kopeika, Identification of blur parameters from motion blurred images, Graphical Models Image Process. 59 (5) (1997) 310–320.
[8] M. Ben-Ezra, S. Nayar, Motion-based motion deblurring, IEEE Trans. Pattern Anal. Mach. Intell. 26 (6) (2004) 689–698.
[9] H. Ishida, S. Yanadume, T. Takahashi, I. Ide, Y. Mekada, H. Murase, Recognition of low-resolution characters by a generative learning method, in: Proceedings of the 1st International Workshop on Camera-Based Document Analysis and Recognition, Seoul, Korea, 2005, pp. 45–51.
[10] H. Ishida, T. Takahashi, I. Ide, Y. Mekada, H. Murase, A generative learning method for the recognition of blurred characters taken by portable cameras, IEICE Trans. J89-D (9) (2006) 2055–2064 (in Japanese).
[11] H. Murase, S. Nayar, Three-dimensional object recognition from appearance—parametric eigenspace method, Systems Comput. Japan 26 (8) (1995) 45–54.
[12] E. Oja, Subspace Methods of Pattern Recognition, Research Studies, Hertfordshire, UK, 1983.
[13] S. Yanadume, T. Takahashi, I. Ide, Y. Mekada, H. Murase, Recognition of very low-resolution characters from motion images taken by portable digital cameras, IEICE Trans. J89-D (2) (2006) 323–331 (in Japanese).
[14] K. Ishii, Generation of distorted characters and its applications, Systems Comput. Japan 14 (6) (1983) 19–27.
[15] T. Horiuchi, K. Toraichi, L. Yamamoto, H. Yamada, On method of training dictionaries for handwritten character recognition using relaxation matching, in: Proceedings of the 2nd International Conference on Document Analysis and Recognition, Tsukuba, Japan, 1993, pp. 638–641.
[16] J. Sun, Y. Hotta, Y. Katsuyama, Low-resolution character recognition by dual eigenspace and synthetic degraded patterns, in: Proceedings of the 1st ACM Workshop on Handcopy Document Processing, Washington, DC, USA, 2004, pp. 15–22.
[17] S. Yanadume, Y. Mekada, I. Ide, H. Murase, Recognition of very low-resolution characters from motion images, in: Proceedings of the 5th Pacific-Rim Conference on Multimedia, Lecture Notes in Computer Science, vol. 3331, Springer, Berlin, 2004, pp. 247–254.
[18] M. Potmesil, Modeling motion blur in computer-generated images, Comput. Graphics 17 (3) (1983) 389–399.

About the Author—HIROYUKI ISHIDA received his B.S. and M.S. degrees from the Department of Information Engineering and from the Graduate School of Information Science, respectively, at Nagoya University. He is currently pursuing a Ph.D. in Information Science at Nagoya University.

About the Author—TOMOKAZU TAKAHASHI received his B.S. degree from the Department of Information Engineering at Ibaraki University, and his M.S. and Ph.D. from the Graduate School of Science and Engineering at Ibaraki University. His research interests include computer graphics and image recognition.

About the Author—ICHIRO IDE received his B.S. degree from the Department of Electronic Engineering, his M.S. degree from the Department of Information Engineering, and his Ph.D. from the Department of Electrical Engineering at the University of Tokyo. He is currently an Associate Professor in the Graduate School of Information Science at Nagoya University.

About the Author—YOSHITO MEKADA received his B.S., M.S., and Ph.D. degrees from the Graduate School of Information Engineering at Nagoya University. He is currently a Professor in the School of Life System Science and Technology at Chukyo University.

About the Author—HIROSHI MURASE received his B.S., M.S., and Ph.D. degrees from the Graduate School of Electrical Engineering at Nagoya University. He is currently a Professor in the Graduate School of Information Science at Nagoya University. He received the Ministry Award from the Ministry of Education, Culture, Sports, Science and Technology in Japan in 2003. He is a Fellow of the IEEE.