Affine-Invariant Recognition of Gray-Scale Characters Using Global Affine Transformation Correlation

Toru Wakahara, *Member*, *IEEE*, Yoshimasa Kimura, *Member*, *IEEE*, and Akira Tomono

Abstract—This paper describes a new, promising technique of gray-scale character recognition that offers both noise tolerance and affine-invariance. The key ideas are twofold. First is the use of normalized cross-correlation as a matching measure to realize noise tolerance. Second is the application of global affine transformation (GAT) to the input image so as to achieve affine-invariant correlation with the target image. In particular, optimal GAT is efficiently determined by the successive iteration method using topographic features of gray-scale images as matching constraints. We demonstrate the high matching ability of the proposed GAT correlation method using gray-scale images of numerals subjected to random Gaussian noise and a wide range of affine transformation. Moreover, extensive recognition experiments show that the achieved recognition rate of 94.3 percent against rotation within 30 degrees, scale change within 30 percent, and translation within 20 percent of the character width along with random Gaussian noise is sufficiently high compared to the 42.8 percent offered by simple correlation.

Index Terms—Gray-scale character recognition, normalized cross-correlation, global affine transformation, noise-tolerant and affine-invariant image matching, successive iteration method.

1 Introduction

Most current OCR systems binarize the input image as the preprocessing operation because documents are assumed to be "black" and "white," i.e., binary images. However, most documents contain gray or different color backgrounds, textured backgrounds, different types of ink, and often suffer from severe gray-scale degradation introduced by the imaging process [1], [2]. In such cases, the conventional binarization process loses a significant amount of information conducive to segmentation and recognition. This is a real problem because demand is increasing for the direct recognition of gray-scale characters not only in degraded text but also in video frames and WWW images.

There are two major approaches to the direct recognition of gray-scale characters.

The first approach uses topographic features [1], [2]. If we consider the gray-scale image to be a surface, then its topographic features correspond to shape features of the original image. For example, a thin stroke will produce a ridge, a narrow gap will produce a saddle point, etc. The topographic features of gray-scale images were demonstrated to be effective for the segmentation and recognition of touching and overlapping characters [3]. However, topographic features using gray-scale gradients and curvatures on the image surface are sensitive to image defects and noise.

The second approach relates to matched filters or correlation-based matching [4], [5]. Representative of this kind of method is the use of "normalized cross-correlation," which offers robustness against gray-scale degradation [6], [7]. In particular, Iijima [8] founded the theory of definite canonicalization to provide a theoretical proof that the normalized cross-correlation is robust against image blurring and degradation. However, the correlation in itself is weak against geometric image distortion-like affine transformation.

The "perturbation method" [9], [10], [11] and the "tangent-distance" [12] were proposed with the express aim of distortion-tolerant image matching. The perturbation method tries to "reverse" the distortion; the input image is returned to one of the "standard templates" by using a preselected set of geometrical transformations like rotation and slant and/or morphological operations of dilation and erosion. The tangent distance measure absorbs any local image distortion that is expressible by linear transformation in each local area. However, both the perturbation method and tangent distance technique cannot deal with unexpected or nonsmall affine transformation.

In our previous paper [13], we introduced the concept of global affine transformation (GAT) as a general deformation model to realize distortion-tolerant shape matching as applied to binary images of characters. By extending the concept of GAT for application to the matching of gray-scale images, we proposed the promising technique of affine-invariant correlation of gray-scale characters using GAT iteration [14]. Conventional correlation-based

- *T. Wakahara is with the NTT Cyber Solutions Laboratories, Nippon Telegraph and Telephone Corporation, 1-1 Hikari-no-oka, Yokosuka-shi, Kanagawa, 239-0847 Japan. E-mail: waka@marsh.hil.ntt.co.jp.*
- *Y. Kimura is with the NTT Service Integration Laboratories, Nippon Telegraph and Telephone Corporation, 3-9-11 Midori-cho, Musashino-shi, Tokyo, 180-8585 Japan. E-mail: kimura.yoshimasa@lab.ntt.co.jp.*
- *A. Tomono is with the Department of Electronics, Faculty of Engineering, Tokai University, 1117 Kitakaname, Hiratsuka-shi, Kanagawa, 259-1292 Japan. E-mail: tomono@keyaki.cc.u-tokai.ac.jp.*

Manuscript received 23 Feb. 2000; revised 25 Oct. 2000; accepted 28 Nov. 2000.
Recommended for acceptance by T.K. Ho.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 111545.

matching was greatly reinforced in two ways: first, the use of normalized cross-correlation as a noise-tolerant matching measure.

Second is the application of GAT to the input gray-scale image so as to realize affine-invariant correlation with the target gray-scale image.

This paper demonstrates the high matching ability of the proposed method using gray-scale images of numerals subjected to random Gaussian noise and a wide range of affine transformation including rotation, scale change, shearing, and translation. Moreover, extensive recognition experiments show that the proposed GAT correlation method achieves far superior recognition rates to the normalized cross-correlation without GAT and the conventional normalized inner product of the two images.

Section 2 explains the normalized cross-correlation based on definite canonicalization. Section 3 introduces the formulation of affine-invariant correlation by GAT. In Section 4, the successive iteration method for optimal GAT determination is described in detail. Section 5 shows the results of matching and recognition experiments using gray-scale images of numerals. Section 6 discusses the effect of topographic constraints on stabilizing GAT convergence, the computational complexity of the proposed method, and the problems of GAT in real-world applications.

2 Normalized Cross-Correlation Using Definite Canonicalization

This paper deals with the direct matching of two gray-scale images. We denote the two images as $\mathbf{F}$, the input gray-scale image, and $\mathbf{G}$, the target gray-scale image or template, and represent $\mathbf{F}$ and $\mathbf{G}$ by gray level functions $f(\boldsymbol{r})$ and $g(\boldsymbol{r})$, respectively, as follows:

$$\mathbf{F} = \{f(\boldsymbol{r})\}, \mathbf{G} = \{g(\boldsymbol{r})\}, \quad \boldsymbol{r} \in K, \tag{1}$$

where $\boldsymbol{r}$ denotes a 2D loci vector defined in the bounded 2D domain of K. Of course, gray level functions $f(\boldsymbol{r})$ and $g(\boldsymbol{r})$ take on only nonnegative values. Here, we consider $f(\boldsymbol{r})$ and $g(\boldsymbol{r})$ as two vectors in the image space with infinite dimensions by assuming that a variable of $\boldsymbol{r}$ is continuous in the 2D domain of K.

The conventional and most popular matching measure for the two images is the normalized inner product or "simple similarity measure" $S(f,g)$ defined by

$$\begin{aligned} S(f,g) &\equiv (f,g)/\parallel f \parallel \cdot \parallel g \parallel \\ &= \int_K f(\boldsymbol{r})\, g(\boldsymbol{r}) d\boldsymbol{r} / \left[\int_K f(\boldsymbol{r})^2 d\boldsymbol{r} \int_K g(\boldsymbol{r})^2 d\boldsymbol{r} \right]^{1/2}, \end{aligned} \tag{2}$$

where (f,g) denotes the inner product of the two vectors f and g, and $\parallel \bullet \parallel$ denotes Euclidian norm. It is well-known that if we denote the angle between the two vectors of f and g by θ, $S(f,g)$ is represented simply by $\cos\theta$. Although the simple similarity measure $S(f,g)$ is very easy to calculate, its discrimination ability deteriorates considerably in the presence of image defects and noise [7].

To resolve the above-mentioned problem, Iijima [8] founded the theory of definite canonicalization. According to definite canonicalization, all image vectors are projected

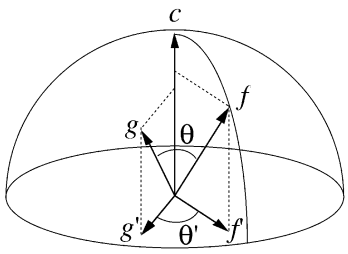

Fig. 1. Definite canonicalization.

to the complementary space perpendicular to the uniform vector c with the unit norm as follows:

$$\begin{aligned} &f'(\boldsymbol{r}) = f(\boldsymbol{r}) - (c,f)\, c(\boldsymbol{r}), \quad g'(\boldsymbol{r}) = g(\boldsymbol{r}) - (c,g)\, c(\boldsymbol{r}), \\ &c(\boldsymbol{r}) = 1/\left[\int_K 1\, d\boldsymbol{r} \right]^{1/2}. \end{aligned} \tag{3}$$

Accordingly, we adopt the simple similarity measure using $f'(\boldsymbol{r})$ and $g'(\boldsymbol{r})$ given by

$$S(f',g') = (f',g')/\parallel f' \parallel\parallel \cdot \parallel g' \parallel \equiv \cos\theta', \tag{4}$$

where θ' represents the angle between the two new vectors f' and g' obtained by definite canonicalization. Most importantly, Iijima proved theoretically that $S(f',g')$ or $\cos\theta'$ provides robustness against image blurring and image degradation.

Fig. 1 shows the meaning of definite canonicalization. In general, when the image blurring operation is applied to the input image $\mathbf{F}$, its image vector f moves toward the uniform vector c in the image space. Because the target image vector g remains the same, the angle θ between f and g increases. As a result, the value of $S(f,g)$ or $\cos\theta$ of (2) gradually decreases as the degree of image blurring increases. On the other hand, it is clear that the value of $S(f',g')$ or $\cos\theta'$ remains unchanged under the image blurring operation, as shown in Fig. 1. This is an intuitive explanation of the robustness of definite canonicalization against image blurring.

By the way, we often use the normalized cross-correlation [6], [7] defined by

$$\begin{aligned} &C(f,g) = \\ &\int_K (f(\boldsymbol{r}) - \mu)(g(\boldsymbol{r}) - \nu) d\boldsymbol{r} / \\ &\left[\int_K (f(\boldsymbol{r}) - \mu)^2 d\boldsymbol{r} \int_K (g(\boldsymbol{r}) - \nu)^2 d\boldsymbol{r} \right]^{1/2}, \\ &\mu \equiv (c,f)c(\boldsymbol{r}), \nu \equiv (c,g)c(\boldsymbol{r}), \end{aligned} \tag{5}$$

where μ and ν specify the mean gray levels of $f(\boldsymbol{r})$ and $g(\boldsymbol{r})$ over K, respectively.

Interestingly enough, we can easily show that the simple similarity measure after definite canonicalization, $S(f',g')$ of (3) and (4), and the normalized cross-correlation, $C(f,g)$ of (5), are exactly the same. Therefore, according to the above-mentioned effect of definite canonicalization, the normalized cross-correlation of (5) as a matching measure guarantees robustness against image blurring and image degradation.

Moreover, we can assume that input and target images have zero mean and unit norm through the following simple transformation:

$$\begin{array}{ll} \int_K f(\boldsymbol{r})d\boldsymbol{r} = 0, & \int_K g(\boldsymbol{r})d\boldsymbol{r} = 0, \\ \int_K f(\boldsymbol{r})^2 d\boldsymbol{r} = 1, & \int_K g(\boldsymbol{r})^2 d\boldsymbol{r} = 1. \end{array} \tag{6}$$

As a result, we obtain the matching measure of normalized cross-correlation simply given by

$$C(f,g) = \int_K f(\boldsymbol{r})g(\boldsymbol{r})d\boldsymbol{r}. \tag{7}$$

However, we still have the problem that the correlation measure in itself cannot compensate for geometric image distortion such as affine transformation.

3 Formulation of Affine-Invariant Correlation by GAT

This section introduces an objective function of affine-invariant correlation using global affine transformation (GAT).

Our previous paper [13] proposed the concept of adaptive or category-dependent normalization using global/local affine transformation (GAT/LAT) as applied to binary images of handwritten characters. GAT was a uniform affine transformation as applied to the input binary image to generate the normalized input image. In particular, the criterion of optimal GAT determination was minimization of the mean of nearest-neighbor interpoint distances between "black" points of the normalized input image and the template. However, because the present paper deals with gray-scale images, not binary ones, we cannot adopt the above-mentioned criterion of optimal GAT determination as applied to the matching of binary images. Therefore, we propose a new computational model of optimal GAT determination for matching gray-scale images [14]. In particular, in order to realize both noise-tolerant and distortion-tolerant image matching, we introduce the new criterion of maximizing the normalized cross-correlation of input and target images through GAT application.

First, we define global affine transformation (GAT). GAT is a uniform affine transformation as applied to an input gray-scale image $\mathbf{F}$ to generate a GAT-superimposed input gray-scale image $\mathbf{F}^* = \{f^*(\boldsymbol{r})\}$ and a loci vector $\boldsymbol{r}$ with gray level of $f(\boldsymbol{r})$ is transformed into a new loci vector $\boldsymbol{r}^*$ with the same gray level of $f(\boldsymbol{r})$ as follows:

$$\boldsymbol{r}^* = \mathrm{A}\boldsymbol{r} + \boldsymbol{b}, \quad \boldsymbol{r} \in K \tag{8}$$

$$f^*(\boldsymbol{r}^*) = f^*(\mathrm{A}\boldsymbol{r} + \boldsymbol{b}) = f(\boldsymbol{r}), \tag{9}$$

where $\mathrm{A} = (\boldsymbol{a}_1 \boldsymbol{a}_2)$ is a 2×2 matrix representing rotation, scale change, and shearing; and $\boldsymbol{b} = (\mathrm{b_x}, \mathrm{b_y})^t$ is a 2D translation vector. Here, $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ represent two basis vectors of the affine-transformed space.

Fig. 2 illustrates the meaning of affine transformation.

Second, we define a fundamental objective function Φ of optimal GAT for affine-invariant correlation using the matching measure of normalized cross-correlation of (7) by

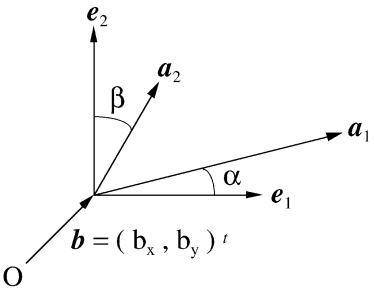

Fig. 2. The meaning of affine transformation. e_1 and e_2 are original unit basis vectors α and β are the arguments of new basis vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ with respect to e_1 and e_2, respectively.

$$\Phi = \int_K f^*(\boldsymbol{r})g(\boldsymbol{r})d\boldsymbol{r} = \int_K f(\boldsymbol{r})g(\mathrm{A}\boldsymbol{r} + \boldsymbol{b})d\boldsymbol{r} \;\rightarrow\; \max \text{ for } \mathrm{A}, \boldsymbol{b}, \tag{10}$$

which, however, requires exhaustive trial and error to determine optimal A and $\boldsymbol{b}$ because A and $\boldsymbol{b}$ are directly embedded in the variable of the specific gray-level function g.

To eliminate this trial and error step, we adopt an equivalent objective function Ψ by introducing a convolution operation using the Gaussian kernel of A and $\boldsymbol{b}$ into as

$$\begin{aligned} \Psi = & \iint_K f(\boldsymbol{r})\; g(\boldsymbol{r}') \exp(- \parallel \mathrm{A}\boldsymbol{r} + \boldsymbol{b} - \boldsymbol{r}' \parallel^2 / D)\; d\boldsymbol{r}d\boldsymbol{r}' \\ & \rightarrow \max \text{ for } \mathrm{A}, \boldsymbol{b}, \end{aligned} \tag{11}$$

where D controls the spread of the Gaussian kernel. Here, it is to be noted that unknown A and $\boldsymbol{b}$ appear only in the differentiable Gaussian kernel, not in the specific gray-level function g. As a result, the procedure of Ψ maximization in (11) becomes much easier than that of Φ maximization in (10). Moreover, it can be shown that Ψ is equivalent to Φ in the limit of $D \rightarrow 0$ by noting that the Gaussian kernel behaves like Dirac's delta function $\delta(\mathrm{A}\boldsymbol{r} + \boldsymbol{b} - \boldsymbol{r}')$ for sufficiently small D [15], which yields

$$\begin{aligned} \Psi &= \iint_K f(\boldsymbol{r})\; g(\boldsymbol{r}') \exp(- \parallel \mathrm{A}\boldsymbol{r} + \boldsymbol{b} - \boldsymbol{r}' \parallel^2 / D)\; d\boldsymbol{r}d\boldsymbol{r}' \\ &\rightarrow \iint_K f(\boldsymbol{r})g(\boldsymbol{r}')\; \delta(\mathrm{A}\boldsymbol{r} + \boldsymbol{b} - \boldsymbol{r}')d\boldsymbol{r}d\boldsymbol{r}' \\ &= \int_K f(\boldsymbol{r})g(\mathrm{A}\boldsymbol{r} + \boldsymbol{b})d\boldsymbol{r} = \Phi \text{ as } D \rightarrow 0. \end{aligned}$$

Therefore, the appropriate control of the decrease in the D value is crucial to the success of this objective function Ψ. Here, we give the D value in a deterministic way as follows:

$$\begin{aligned} D = 1/2 \Bigg\{ & \mathrm{mean}_{\boldsymbol{r} \in \mathbf{F}} \left[\min_{\boldsymbol{r}' \in \mathbf{G}} \parallel \boldsymbol{r} - \boldsymbol{r}' \parallel^2 \text{ and } f(\boldsymbol{r}) = g(\boldsymbol{r}') \right] + \\ & \mathrm{mean}_{\boldsymbol{r}' \in \mathbf{G}} \left[\min_{\boldsymbol{r} \in \mathbf{F}} \parallel \boldsymbol{r}' - \boldsymbol{r} \parallel^2 \text{ and } g(\boldsymbol{r}') = f(\boldsymbol{r}) \right] \Bigg\}. \end{aligned} \tag{12}$$

Hence, D stands for the mean value of Euclidian distances between nearest-neighbor points, one in $\mathbf{F}$ and the other in $\mathbf{G}$, with the same gray level. Also, as is explained in the next section, the successive iteration method guarantees that the D value decreases monotonically in the process of optimal

GAT determination. Moreover, because the D value is controlled in a deterministic manner, we can say that the proposed objective function Ψ of (11) for affine-invariant correlation contains no free parameter. This fact is extremely beneficial from the practical viewpoint.

4 Successive Iteration Method for Optimal GAT Determination

This section describes a strict computational model for optimal GAT determination. The key idea is the successive iteration method as applied to the maximization problem. Moreover, we introduce matching constraints using the topographic features of gray-scale images into the successive iteration method to stabilize and accelerate its convergence.

4.1 Linear Equations for Optimal GAT Determination

First of all, following the necessary condition of Ψ maximization yields a set of simultaneous equations by setting both derivatives of Ψ with respect to each element of A and $\boldsymbol{b}$ equal to zero as

$$\begin{aligned} \mathrm{O} &= \partial\Psi/\partial\mathrm{A} \\ &\propto \iint_K f(\boldsymbol{r})g(\boldsymbol{r}')\ \boldsymbol{r}(\mathrm{A}\boldsymbol{r}+\boldsymbol{b}-\boldsymbol{r}')^t \\ &\quad \exp(-\parallel \mathrm{A}\boldsymbol{r}+\boldsymbol{b}-\boldsymbol{r}'\parallel^2/D)d\boldsymbol{r}d\boldsymbol{r}', \end{aligned} \tag{13}$$

$$\begin{aligned} \boldsymbol{0} &= \partial\Psi/\partial\boldsymbol{b} \\ &\propto \iint_K f(\boldsymbol{r})\ g(\boldsymbol{r}')\ (\mathrm{A}\boldsymbol{r}+\boldsymbol{b}-\boldsymbol{r}') \\ &\quad \exp(-\parallel \mathrm{A}\boldsymbol{r}+\boldsymbol{b}-\boldsymbol{r}'\parallel^2/D)d\boldsymbol{r}d\boldsymbol{r}', \end{aligned} \tag{14}$$

where O denotes a 2×2 zero matrix and (13) and (14) constitute simultaneous equations for the six unknown scalars of A and $\boldsymbol{b}$.

However, it is clear that we cannot solve (13) and (14) analytically because these equations are nonlinear with regard to unknown A and $\boldsymbol{b}$. It is well-known, compared to directly solving a set of nonlinear equations, that it is much easier and more effective to derive an equivalent set of linear equations from the nonlinear equations and then apply the successive iteration method to the obtained linear equations.

Hence, as the zeroth order approximation, we substitute identity matrix I for A and zero vector $\boldsymbol{0}$ for $\boldsymbol{b}$ in the Gaussian kernel in (13) and (14). This allows us to obtain the following set of simultaneous linear equations:

$$\mathrm{O} = \iint_K f(\boldsymbol{r})\ g(\boldsymbol{r}')\boldsymbol{r}(\mathrm{A}\boldsymbol{r}+\boldsymbol{b}-\boldsymbol{r}')^t \exp(-\parallel \boldsymbol{r}-\boldsymbol{r}'\parallel^2/D)d\boldsymbol{r}d\boldsymbol{r}', \tag{15}$$

$$\boldsymbol{0} = \iint_K f(\boldsymbol{r})\ g(\boldsymbol{r}')\ (\mathrm{A}\boldsymbol{r}+\boldsymbol{b}-\boldsymbol{r}')\exp(-\parallel \boldsymbol{r}-\boldsymbol{r}'\parallel^2/D)d\boldsymbol{r}d\boldsymbol{r}', \tag{16}$$

where these simultaneous linear equations are easily solved by conventional techniques such as Gaussian elimination [15].

4.2 Use of Topographic Features as Matching Constraints

To start with, by examining the structure of the simultaneous linear equations of (15) and (16), we can extract the following two characteristics:

1. The product of $f(\boldsymbol{r})$ and $g(\boldsymbol{r}')$ is integrated for correlation by assuming that each point $\boldsymbol{r}$ in $\mathbf{F}$ corresponds to a point $\boldsymbol{r}'$ in $\mathbf{G}$ through GAT operation. Also, there is an assumption that the loci vector of $\mathrm{A}\boldsymbol{r}+\boldsymbol{b}$ is approximately equal to $\boldsymbol{r}'$.
2. The probability that each point $\boldsymbol{r}$ in $\mathbf{F}$ corresponds to a point $\boldsymbol{r}'$ in $\mathbf{G}$ is estimated as weighting factor by the Gaussian kernel, $\exp(-\parallel \boldsymbol{r}-\boldsymbol{r}'\parallel^2/D)$. This weighting factor utilizes only position information of $\boldsymbol{r}$ and $\boldsymbol{r}'$ via Euclidian distance.

The first characteristic is the goal of affine-invariant correlation by GAT in itself. However, regarding the second characteristic, we can say that the position information of $\boldsymbol{r}$ and $\boldsymbol{r}'$ alone is insufficient for validating their correspondence. Hence, we reinforce the matching constraints on $\boldsymbol{r}$ and $\boldsymbol{r}'$ as follows.

The key idea is the use of not only position information but also topographic features as matching constraints. This is because the similarity of shape characteristics is well-estimated in terms of topographic features. Concretely, we introduce an enhanced weighting function $\gamma(\boldsymbol{r},\boldsymbol{r}')$ that estimates the similarity in gray-scale gradients using the topographic features of two points $(\boldsymbol{r})$ in $\mathbf{F}$ and $(\boldsymbol{r}')$ in $\mathbf{G}$:

$$\gamma(\boldsymbol{r},\boldsymbol{r}') = \max\{(\nabla f(\boldsymbol{r}),\nabla g(\boldsymbol{r}')),0\}\cdot\exp(-\parallel \boldsymbol{r}-\boldsymbol{r}'\parallel^2/D), \tag{17}$$

where $(\nabla f(\boldsymbol{r}),\nabla g(\boldsymbol{r}'))$ is the inner product of two gradient vectors, $\nabla f(\boldsymbol{r})$ and $\nabla g(\boldsymbol{r}')$. Using the factor $\gamma(\boldsymbol{r},\boldsymbol{r}')$ has two major advantages. First, $\gamma(\boldsymbol{r},\boldsymbol{r}')$ equals zero when $(\nabla f(\boldsymbol{r}),\nabla g(\boldsymbol{r}'))\le 0$, i.e., the angle between two gradient vectors is more than 90°. This satisfies the similarity in directions of two gradient vectors. Second, when the angle between two gradient vectors is less than 90°, the greater the norms of $\nabla f(\boldsymbol{r})$ and $\nabla g(\boldsymbol{r}')$ are, the greater the value of the weighting function of (17) is. As a result, $\gamma(\boldsymbol{r},\boldsymbol{r}')$ reinforces the contribution of steep regions or edges to correlation matching while suppressing or neglecting the contribution of flat regions or grounds.

Finally, we obtain the following set of simultaneous linear equations using $\gamma(\boldsymbol{r},\boldsymbol{r}')$:

$$\mathrm{O} = \iint_K \gamma(\boldsymbol{r},\boldsymbol{r}')f(\boldsymbol{r})\ g(\boldsymbol{r}')\boldsymbol{r}(\mathrm{A}\boldsymbol{r}+\boldsymbol{b}-\boldsymbol{r}')^t d\boldsymbol{r}d\boldsymbol{r}', \tag{18}$$

$$\boldsymbol{0} = \iint_K \gamma(\boldsymbol{r},\boldsymbol{r}')f(\boldsymbol{r})\ g(\boldsymbol{r}')\ (\mathrm{A}\boldsymbol{r}+\boldsymbol{b}-\boldsymbol{r}')d\boldsymbol{r}d\boldsymbol{r}'. \tag{19}$$

4.3 Successive Iteration Method for Affine-Invariant Correlation Determination

Instead of solving the nonlinear equations of (13) and (14) directly, we apply the successive iteration method to the linearized equations of (18) and (19) reinforced by topographic constraints. The successive iteration method [15] provides the iterative procedure for optimal GAT

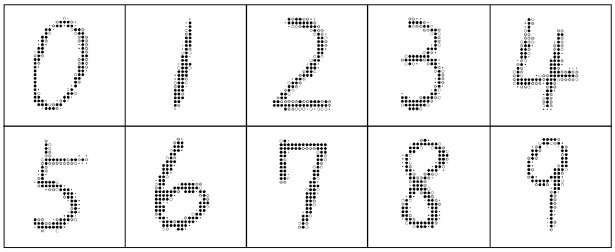

Fig. 3. Target gray-scale images of numerals.

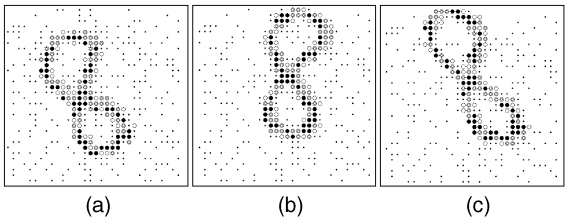

Fig. 4. Examples of artificial input gray-scale images of "eight." (a) Pure rotation and noise. (b) Pure translation and noise. (c) More general affine transformation and noise.

determination and calculation of affine-invariant correlation as follows:

Start: Calculate the initial value of $C_0 = C(f, g)$ of (7) between the original input and target gray-scale images, $\mathbf{F}$ and $\mathbf{G}$.

Loop: Determine the GAT components of A and $\boldsymbol{b}$ by solving the set of simultaneous linear equations of (18) and (19). Next, generate the GAT-superimposed input gray-scale image $\mathbf{F}^* = \{f^*(\boldsymbol{r})\}$ by (9) and substitute $\mathbf{F}^*$ for the input image $\mathbf{F}$. At the same time, update the value of D of (12) and $\gamma(\boldsymbol{r}, \boldsymbol{r}')$ of (17).

Pause: Calculate the updated value of $C_1 = C(f, g)$ of (7) between the renewed $\mathbf{F}$ and $\mathbf{G}$, and compare C_1 with C_0. If the correlation value increases, i.e., $C_1 > C_0$, substitute C_1 for C_0, and then go back to *Loop*; otherwise, output the maximal correlation value of C_0 as the final result of affine-invariant correlation and stop.

In particular, the value of D of (12) decreases monotonically through the above-mentioned iteration process. Therefore, we can expect that the converged correlation value and GAT components obtained by the successive iteration method are approximately equivalent to those obtained by the original fundamental objective function Φ of (10).

5 Experimental Results

In this section, we present results that confirm the high matching and discrimination ability of the proposed method; matching and recognition experiments were conducted using gray-scale images of numerals subjected to random Gaussian noise and a wide range of affine transformation.

5.1 Target and Input Images

In the experiments, we used gray-scale images of numerals as target images. For each of the 10 digits, a single character image free of noise was selected as a target image. Target images originally had 16 discrete gray levels (nonnegative integers) and were transformed so as to have zero mean and unit norm according to (3) and (6). Hence, gray level functions of the images take real values $\in (-1.0, +1.0)$. The spatial resolution of the character image was 32×32. Also, as we deal with digital images, the integral of any image function $h(\boldsymbol{r})$ over the continuous 2D image domain K reduces to the summation of the digital image function $h(\boldsymbol{i})$ defined at integer loci vectors of $\boldsymbol{i} = (\mathrm{i_x}, \mathrm{i_y})^t$ over the digitized 2D domain of $1 \leq \mathrm{i_x}, \mathrm{i_y} \leq 32$.

Fig. 3 shows the target gray-scale images of numerals. To simplify the display, each point was denoted by either a black circle or a double circle or a white circle or a dot or a blank according to temporary quantization into five discrete levels.

The input gray-scale images were artificially generated by applying a wide range of affine transformation and random Gaussian noise to the target images. The concrete procedure for generating input images is as follows.

5.1.1 Affine Transformation Application

Arbitrary affine transformation, specified by $\mathrm{A} = (\boldsymbol{a}_1\ \boldsymbol{a}_2)$ and $\boldsymbol{b} = (\mathrm{b_x}, \mathrm{b_y})^t$ in Fig. 2, was applied to each target image $\mathbf{G} = \{g(\boldsymbol{j})\}$ to generate an input image $\mathbf{F} = \{f(\boldsymbol{i})\}$. In order to assign gray levels to all discrete points of the input image $\mathbf{F}$, we need to apply the affine transformation "backwards" and make full use of interpolation. We used the well-known technique of "bilinear interpolation" [5] as the simplest such method.

5.1.2 Random Gaussian Noise Addition

Random Gaussian noise n with zero mean and unit variance was added to each point of the input image $\mathbf{F}$ to generate the noise-added input image $\mathbf{F}' = \{\ f'(\boldsymbol{i})\ \}$ as follows:

$$f'(\boldsymbol{i}) = f(\boldsymbol{i}) + \kappa \cdot \delta \cdot n(\boldsymbol{i}), \kappa \in (0, 1], \tag{20}$$

where κ is a proportional constant and δ denotes the dynamic range of f.

The input gray-scale image $\mathbf{F}'$ of (20), transformed again to have zero mean and unit norm, is denoted by $\mathbf{F}$.

Fig. 4 shows examples of input gray-scale images of "eight." Figs. 4a, 4b, and 4c show artificial input images generated by a pure rotation by $\alpha = -\beta = +30^\circ$ and noise, a pure translation by $\boldsymbol{b} = (3, 3)^t$ and noise, and more general affine transformation and noise, respectively, as applied to the target image of "eight" in Fig. 3. The proportional constant of κ controlling the magnitude of random Gaussian noise was set at 0.7.

By the way, for the weighting function $\gamma(\boldsymbol{r}, \boldsymbol{r}')$ of (17), we need to calculate gray-scale gradients of the digital images. We used the *Roberts* operator [5] to estimate the norm and direction of gray-scale gradients. Moreover, regarding the direction of gray-scale gradients, we simply used the popular 8-directional quantization in 45° steps.

5.2 Results of Matching Experiments

Fig. 5 shows an example of the process of affine-invariant correlation using GAT iteration between the target image of

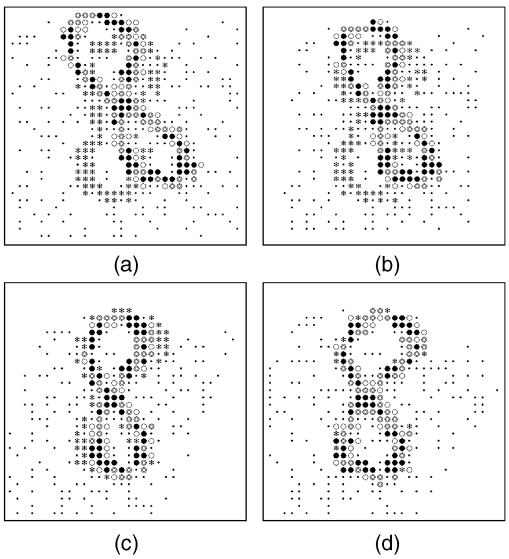

Fig. 5. Example of the GAT matching process. (a) Initial input and target images of "eight." (b), (c), and (d) GAT iterations of 1, 10, and 25.

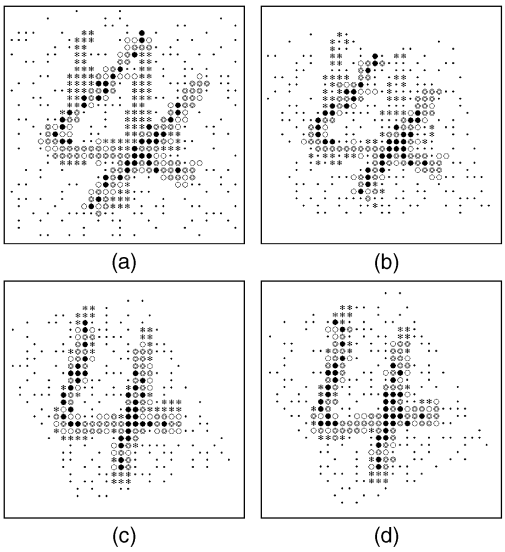

Fig. 6. Example of the GAT matching process. (a) Initial input and target images of "four." (b), (c), and (d) GAT iterations of 3, 9, and 19.

"eight" in Fig. 3 and the input image of Fig. 4c generated by the following affine parameters:

$$\| \boldsymbol{a}_1 \| = 0.90, \| \boldsymbol{a}_2 \| = 1.20, \alpha = 15^\circ, \\ \beta = -30^\circ, \mathrm{b_x} = 2, \mathrm{b_y} = 3, \tag{21}$$

and random Gaussian noise of $\kappa = 0.7$. Fig. 5a shows the initial overlapping of the target and input images, where the points belonging to the target image are denoted by asterisks. Figs. 5b, 5c, and 5d show GAT matching results after iterations of 1, 10, and 25, respectively. The GAT iteration for this example converged at iteration 25. As shown in Fig. 5, the matching degree increases gradually with the GAT iteration number. Actually, we obtained the series of monotonically increasing correlation values of 0.180, 0.251, 0.627, and 0.850 in this order in Figs. 5a, 5b, 5c, and 5d. Regarding the converged result of Fig. 5d, we can estimate the equivalent affine parameters as follows:

$$\| \boldsymbol{a}'_1 \| = 0.962, \| \boldsymbol{a}'_2 \| = 1.18, \alpha' = 14.1^\circ, \\ \beta' = -29.8^\circ, \mathrm{b}'_\mathrm{x} = 1.94, \mathrm{b}'_\mathrm{y} = 3.06,$$

which is very satisfactory compared to the correct values of (21). Also, the converged correlation value of 0.850 at the 25th iteration is sufficiently high compared to the ideal correlation value of 0.884 obtained by the exact inverse transformation of (21).

Fig. 6 shows another example of the process of affine-invariant correlation using GAT iteration between the target image of "four" in Fig. 3 and the input image generated by the following affine parameters:

$$\| \boldsymbol{a}_1 \| = 1.20, \| \boldsymbol{a}_2 \| = 1.10, \alpha = -20^\circ, \\ \beta = 30^\circ, \mathrm{b_x} = 2, \mathrm{b_y} = -2, \tag{22}$$

and random Gaussian noise of $\kappa = 0.7$. Fig. 6a shows the initial overlapping of the target and input images. Figs. 6b, 6c, and 6d show GAT matching results at iterations of 3, 9, and 19, respectively. We obtained the series of monotonically increasing correlation values of 0.282, 0.394, 0.795, and 0.827 in this order in Figs. 6a, 6b, 6c, and 6d. The estimated affine parameters for the result of Fig. 6d are satisfactory as compared to the correct values of (23) as follows:

$$\| \boldsymbol{a}'_1 \| = 1.27, \| \boldsymbol{a}'_2 \| = 1.30, \alpha' = -18.8^\circ, \\ \beta' = 32.9^\circ, \mathrm{b}'_\mathrm{x} = 2.48, \mathrm{b}'_\mathrm{y} = -1.48,$$

Also, the converged correlation value of 0.827 after 19 GAT iterations is sufficiently high compared to the ideal correlation value of 0.895 obtained by the exact inverse transformation of the affine parameters of (22).

Next, we show three kinds of quantitative results gained by applying the proposed GAT correlation method to the matching of the target images in Fig. 3 and artificial input images corresponding to a wide variety of either pure rotation or pure scale change or pure translation and random Gaussian noise.

First, input images were artificially generated by applying pure rotation and random Gaussian noise to each target image in Fig. 3. The rotation angle, $\alpha = -\beta$ in Fig. 2, was varied from -45 degrees to $+45$ degrees in five degree steps; $\boldsymbol{b}$ was taken as a zero vector. Also, random Gaussian noise of $\kappa = 0.7$ was added to each point of each input image. Hence, the total number of input images was 10 digits times 19 rotation angles. GAT matching was then applied to each input image against its correct target image.

Fig. 7 shows the relation between the mean of normalized cross-correlation values and the rotation angle along with random Gaussian noise.

From Fig. 7, it is clear that, even if the initial correlation value was less than 0.2, the proposed method achieved correlation values that exceeded 0.7 through GAT iteration.

Second, input images were artificially generated by applying pure scale change and random Gaussian noise to each target image in Fig. 3. The corresponding affine

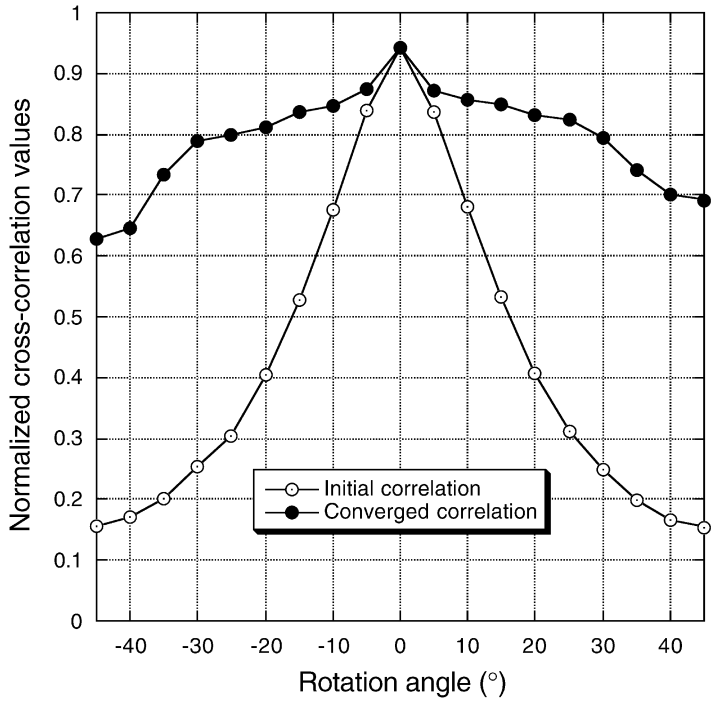

Fig. 7. Relation between the mean of normalized cross-correlation values and the rotation (random Gaussian noise present).

transformation was as follows: $\mathrm{A} = \gamma(\boldsymbol{e}_1\ \boldsymbol{e}_2)$ and $\boldsymbol{b} = (0,0)^t$, where $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ represent the original unit basis vectors. Here, the scale change parameter, λ, was varied from 0.5 to 1.5 in 0.1 steps. Also, random Gaussian noise of $\kappa = 0.7$ was added to each point of each input image. Hence, the total number of input images was 10 digits times 11 scale changes. GAT matching was then applied to each input image against its correct target image.

Fig. 8 shows the relation between the mean of normalized cross-correlation values and the scale change along with random Gaussian noise.

From Fig. 8, it is to be noted that the GAT correlation method is more robust against image expansion than image shrinking. This asymmetry is due to the introduction of the Gaussian kernel, which brings about the effect of image blurring in the weighting function $\gamma(\boldsymbol{r}, \boldsymbol{r}')$ of (17).

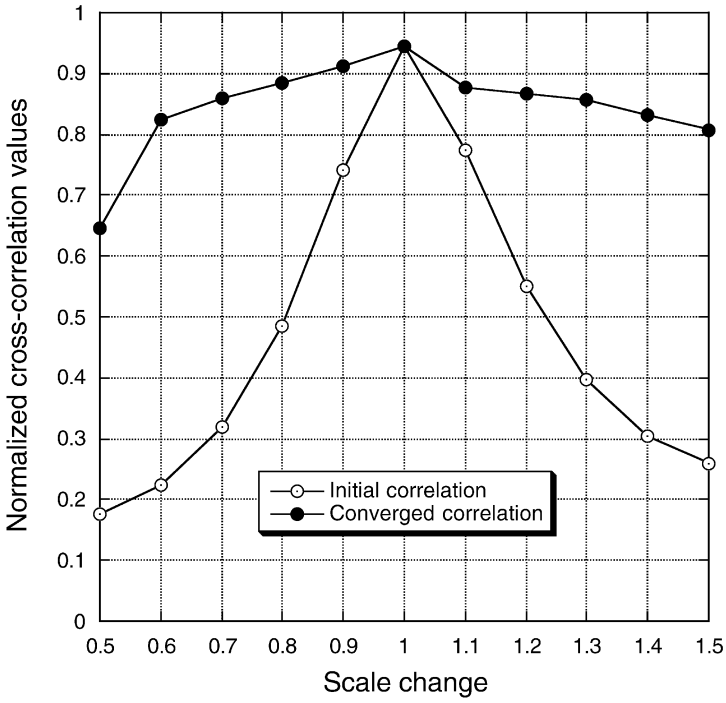

Fig. 8. Relation between the mean of normalized cross-correlation values and the scale change (random Gaussian noise present).

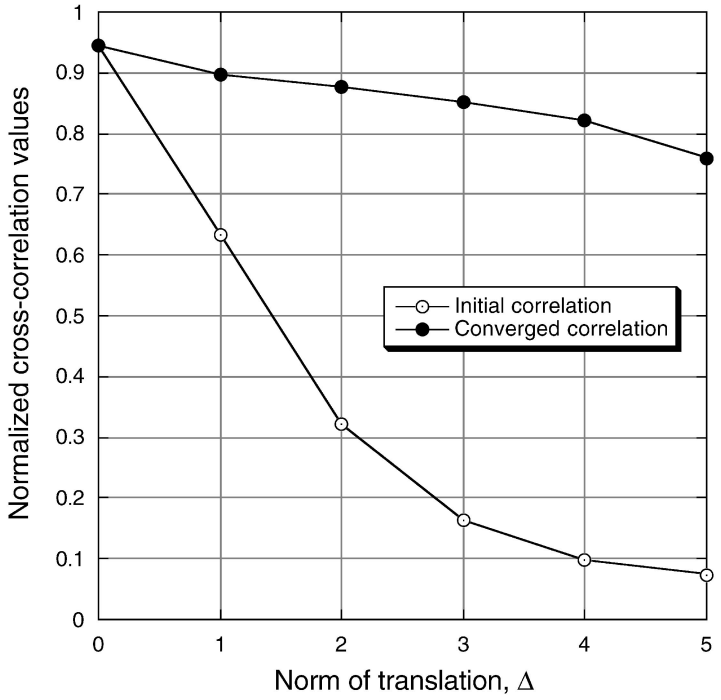

Fig. 9. Relation between the mean of normalized cross-correlation values and the norm of translation (random Gaussian noise present).

Third, we show experimental results for pure translation and noise. Input images were artificially generated by applying translation and random Gaussian noise to the target images. This time, the corresponding affine transformation was as follows: $\mathrm{A} = \mathrm{I}$ and $\boldsymbol{b} = (\mathrm{b_x}, \mathrm{b_y})^t$, where I is an identity matrix and $\mathrm{b_x}$ and $\mathrm{b_y}$ took integer values between -5 and $+5$. Also, random Gaussian noise of $\kappa = 0.7$ was added to each point of each input image. Hence, the total number of input images was 10 digits times 121 translations. GAT matching was then applied to each input image against its correct target image.

Fig. 9 shows the relation between the mean of normalized cross-correlation values and the norm of translation, Δ, along with random Gaussian noise. The norm of translation, Δ, was defined by $\Delta = \max(|\mathrm{b_x}|, |\mathrm{b_y}|)$. Therefore, the value of Δ ranged from zero to five.

From Fig. 9, it is clearly seen that the proposed method is robust against a fairly large amount of translation. Actually, even if the initial correlation value was less than 0.2, the proposed method achieved correlation values in excess of 0.8 through GAT iteration.

Finally, from Figs. 7, 8, and 9, it is found that the proposed GAT correlation method is more robust against translation than against rotation or scale change. This is because the matching stabilizing factor $\gamma(\boldsymbol{r}, \boldsymbol{r}')$ of (17) requires high similarity in the directions of gray-scale gradients of matched points.

5.3 Results of Recognition Experiments

In Section 5.2, we demonstrated the high matching ability of the proposed method as applied to artificially generated input images and their correct targets. However, it is clear that the high matching ability itself does not necessarily guarantee the high discrimination ability of the GAT correlation method as applied to input and target images belonging to not only the same but also different categories.

This section shows the results of extensive recognition experiments conducted using artificial input images subjected to combinations of rotation, scale change, and translation, along with random Gaussian noise against target images in Fig. 3. That is, GAT matching was applied to each input image against all target images of the 10 digits and the category of the target image with which the maximal correlation value was achieved through GAT application was output as the recognition result of the input image.

From Fig. 7, the range of rotation was set at within 30 degrees in 5 degree steps. Also, from Fig. 8, the range of scale change was set at between 0.7 and 1.3 in 0.1 steps. Regarding translation, from Fig. 9, we set $\mathrm{b_x}$ and $\mathrm{b_y}$ to take integer values between -4 and $+4$. Hence, the total number of artificially generated input images per digit was $4,459 = 13 \times 7 \times 49$. Moreover, random Gaussian noise of $\kappa = 0.7$ was added to each point of each input image.

Fig. 10 shows the relation between the recognition rates and the norm of translation, Δ, along with random Gaussian noise. The corresponding recognition rates obtained by using the normalized cross-correlation without GAT and the normalized inner product of (2) are also plotted. Here, we call these two conventional methods "simple correlation" in contrast to the proposed method of GAT correlation.

Fig. 10 clearly shows that the discrimination ability of GAT correlation is far superior to that of simple correlation. Actually, the achieved recognition rate of 94.3 percent against rotation within 30 degrees, scale change within 30 percent, and translation of $\Delta \leq 3$ or within 20 percent of

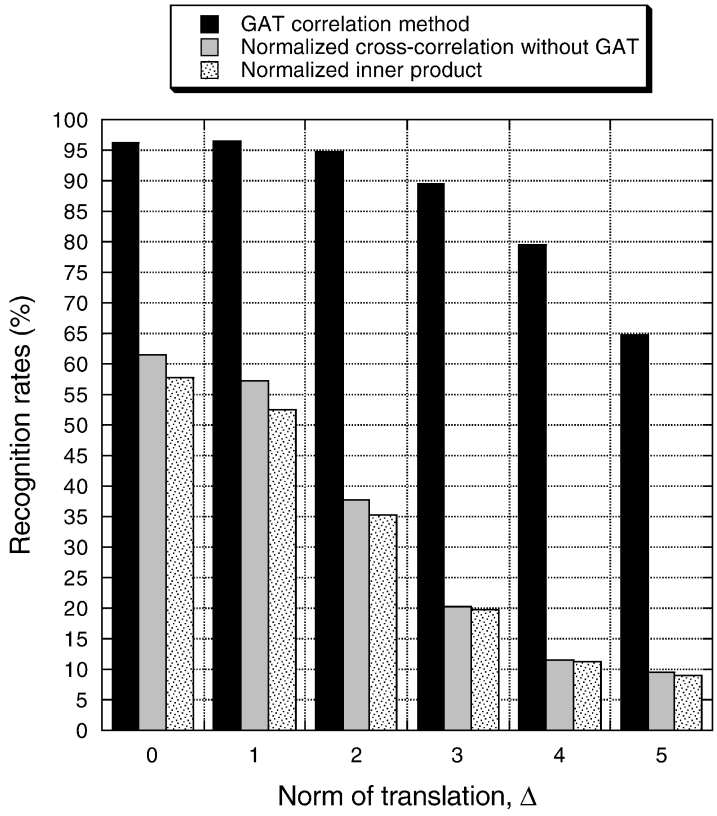

Fig. 10. Relation between the recognition rates and the norm of translation under rotation, scale change, and random Gaussian noise.

the character width is very high compared to the 42.8 percent offered by simple correlation.

Three major causes for low GAT correlation are as follows: The first is the well-known "local optimum" problem in determining optimal GAT by the successive iteration method. The second is the excessive image contraction that occurs when the values of D of (12) are too large. The third is the limit of topographic constraints using gray-scale gradients in $\gamma(\boldsymbol{r}, \boldsymbol{r}')$ of (17) when image distortion and degradation spoils such gradient information.

6 Discussion

In this section, we discuss three problems: how to stabilize GAT convergence so as to avoid "local optimum," how to suppress the computational complexity, and, finally, how to confirm the GAT correlation method in real-world applications.

6.1 GAT Stabilization

The successive iteration method described in Section 4 does not guarantee that the GAT iteration process necessarily gives a global optimal solution for maximizing the original objective function Φ of (10). This is the well-known "local optimum" problem that is inherent in any kind of "hill-climbing" iteration method based on local constraints and remains unsolved. However, as long as the exhaustive combinatorial approach is intractable from the viewpoint of computational cost, it is better to tackle the reinforcement of the iteration methods themselves by introducing efficient constraints.

From this viewpoint, the $\gamma(\boldsymbol{r}, \boldsymbol{r}')$ of (17) were added to (18) and (19) as topographic constraints to realize optimal GAT determination. In this section, we verify the effectiveness of topographic constraints in the GAT correlation method.

In a comparison to the matching results for the case of pure rotation along with random Gaussian noise described in Section 5.2, we conducted GAT correlation experiments using the same conditions but without using the topographic constraints. That is, we solved (15) and (16) instead of (18) and (19).

Fig. 11 shows the relation between the mean of normalized cross-correlation values and the rotation angle without employing topographic constraints. The corresponding results obtained by employing topographic constraints shown in Fig. 7 are also plotted.

From Fig. 11, it is clearly found that the topographic constraints substantially improve the converged correlation values, especially when initial correlation values are low. Moreover, we investigated the average number of GAT iterations required for convergence in the abovementioned experiments. The resulting average number of GAT iterations without topographic constraints was 15.0. On the other hand, the average number of GAT iterations with topographic constraints was 19.2. This fact means that using topographic constraints in GAT iteration is very effective in avoiding "local optima" and continual increases in the correlation values.

Generally speaking, we can say that the effective matching constraints using appropriate shape characteristics are indispensable for stabilizing and accelerating the iterative shape matching process. In particular, such shape characteristics must be invariant, even if only

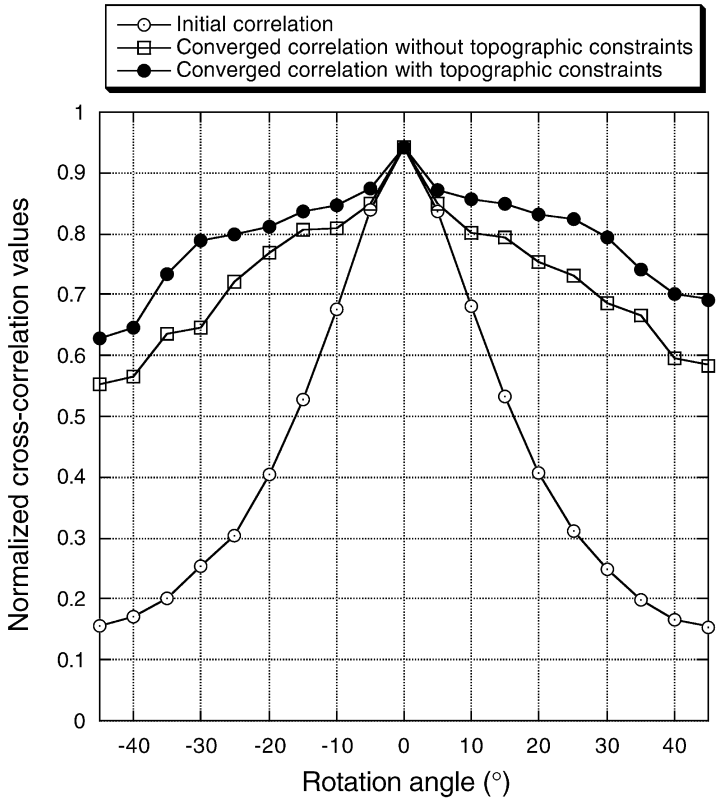

Fig. 11. Relation between the mean of normalized cross-correlation values and the rotation angle without employing topographic constraints.

approximately, against the image distortion to be considered. In this sense, the present use of gray-scale gradients as shape characteristics is one of the successful challenges, but further reinforcement is needed. For example, more global shape characteristics like line densities [16] must be taken into consideration.

6.2 Computational Complexity

First, we present a computational complexity analysis of the GAT correlation method. As described in Section 4, the single matching cycle determines A and $\boldsymbol{b}$ by solving the following simultaneous linear equations:

$$\mathrm{O} = \iint_K \gamma(\boldsymbol{r}, \boldsymbol{r}') f(\boldsymbol{r})\ g(\boldsymbol{r}')\boldsymbol{r}(\mathrm{A}\boldsymbol{r} + \boldsymbol{b} - \boldsymbol{r}')^t d\boldsymbol{r} d\boldsymbol{r}', \tag{18}$$

$$\boldsymbol{0} = \iint_K \gamma(\boldsymbol{r}, \boldsymbol{r}') f(\boldsymbol{r})\ g(\boldsymbol{r}')\ (\mathrm{A}\boldsymbol{r} + \boldsymbol{b} - \boldsymbol{r}') d\boldsymbol{r} d\boldsymbol{r}'. \tag{19}$$

When dealing with digital images of size $N \times N$, calculation of $\gamma(\boldsymbol{r}, \boldsymbol{r}')$ takes $\mathrm{O}(N^4)$ time. Thus, calculating the coefficients for unknown A and $\boldsymbol{b}$ by the multiplication and addition of $\gamma(\boldsymbol{r}, \boldsymbol{r}')$, $f(\boldsymbol{r})$, $g(\boldsymbol{r}')$, and $\boldsymbol{r}$ takes $\mathrm{O}(N^4)$ time in like manner. Last, the simultaneous linear equations with the six unknown scalars of A and $\boldsymbol{b}$ are solved by conventional techniques like Gaussian elimination [15]. Hence, the overall time required for the single matching cycle is only proportional to N^4 and poses no problem to actual implementation. Of course, as the total matching time is equal to the product of the single matching time and the number of GAT iterations, the acceleration of GAT convergence is desired.

Next, we discuss the trade-off between recognition time and recognition ability of GAT correlation as compared to simple correlation.

If we adopt simple correlation by preparing multiple templates against rotation within 30 degrees in 5 degree steps, scale change between 0.7 and 1.3 in 0.1 steps, and translation of $\Delta \leq 3$ in integer steps, the total number of templates is $4,459 = 13 \times 7 \times 49$. The present recognition time of GAT correlation is about 10^4 times larger than that of simple correlation using a single template. In this situation, simple correlation using $4,459$ templates per digit is about two times faster than GAT correlation. As is clear, if we try to prepare templates against a wide range of affine transformation, including shearing, the number of templates needed for simple correlation would be enormous. From this consideration, we can say that GAT correlation provides a very powerful approach to the problem of affine-invariant gray-scale character recognition.

6.3 Problems of GAT in Real-World Applications

We successfully tested the GAT correlation method using artificial image data subjected to random Gaussian noise and a variety of affine transformations with the correct solution of exact affine parameters. Against this background, it is necessary and interesting to discuss how practical the proposed method is for real-world applications.

First of all, when dealing with real images, we should clarify what kind of transformation or deformation must be considered. The GAT correlation method suffices for any application if only a single, particular font is subjected to a specified range of affine transformation and the degradation common to real images. However, if the amount of

residual deformation after GAT operation is not negligible from the viewpoints of matching and discriminating abilities, we have to tackle the nonlinear or nonrigid deformation on real images in a definite manner using more general deformation models. In our approach, we adopt the local affine transformation (LAT) method [13] if it is extended to gray-scale image matching after GAT operation in a hierarchical way. Therefore, it is essential to specify the deformation involved.

On the other hand, even if we can assume that the range of deformation being considered in an actual application is well-approximated by global affine transformation, we must then address the following inherent problems in order to make a success of the GAT correlation method: appropriate generation or selection of templates, location and isolation of input images to be matched with templates, and rough normalization with regard to position and size as applied to the input image as preprocessing. In particular, the biggest challenge is to test the proposed GAT correlation method to see whether it can realize both the isolation and the discrimination simultaneously of a wanted character in a messy background.

7 Conclusion

Direct recognition of gray-scale characters without binarization is very promising. However, no published technique offers sufficient robustness against both gray-scale degradation due to defects and noise and geometrical distortion like affine transformation. To resolve this problem, we proposed a new approach to GAT correlation. First, we adopted normalized cross-correlation as a noise-tolerant matching measure. Second, we introduced the concept of global affine transformation (GAT) as applied to the input image to absorb image distortion that is expressible by uniform affine transformation. Finally, we formulated a computational model for optimal GAT determination that maximizes the normalized cross-correlation by the successive iteration method. In particular, adopting matching constraints that use gray-scale gradients greatly stabilizes and accelerates GAT convergence.

We have demonstrated successful experiments in which the GAT correlation method was applied to the matching and recognition of gray-scale images of numerals subjected to random Gaussian noise and a wide variety of affine transformation. From these results, we can conclude that the proposed affine-invariant correlation method using GAT iteration is a very promising tool for the direct matching of degraded and distorted gray-scale character images. Future work is to apply our GAT correlation method to gray-scale character recognition of not only in real-life paper documents but also the characters in video frames and WWW images.

Acknowledgments

The authors would like to thank Dr. Noboru Sonehara and Dr. Mutsuo Sano for their sincere encouragement and support. They are also indebted to Dr. Yozo Tamura for many fruitful discussions regarding this work. A preliminary version of this paper was presented at the 15th International Conference on Pattern Recognition, 3-8 September, 2000, Barcelona, Spain.

References

[1] L. Wang and T. Pavlidis, "Direct Gray-Scale Extraction of Features for Character Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 10, pp. 1053-1067, Oct. 1993.

[2] S.-W. Lee and Y.J. Kim, "Direct Extraction of Topographic Features for Gray-Scale Character Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, no. 7, pp. 724-729, July 1995.

[3] S.-W. Lee, D.-J. Lee, and H.-S. Park, "A New Methodology for Gray-Scale Character Segmentation and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 10, pp. 1045-1050, Oct. 1996.

[4] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis,* chapter 7. New York: Wiley, 1973.

[5] A. Rosenfeld and A.C. Kak, *Digital Picture Processing,* chapters 4 and 9, second ed. San Diego, Calif.: Academic Press, 1982.

[6] M. Uenohara and T. Kanade, "Use of Fourier and Karhunen-Loéve Decomposition for Fast Pattern Matching with a Large Set of Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 8, pp. 891-898, Aug. 1997.

[7] M. Sawaki and N. Hagita, "Recognition of Degraded Machine-Printed Characters Using a Complementary Similarity Measure and Error-Correction Learning," *IEICE Trans. Information & Systems,* vol. E79-D, pp. 491-497, May 1996.

[8] T. Iijima, *Pattern Recognition,* chapter 6. Tokyo: Corona 1973. (in Japanese).

[9] Y. Yasuda, K. Yamamoto, and H. Yamada, "Effect of the Perturbed Correlation Method for Optical Character Recognition," *Proc. Second Int'l Conf. Document Analysis and Recognition,* pp. 830-833, Nov. 1993.

[10] M. Suzuki, N. Kato, H. Aso, and Y. Nemoto, "A Handprinted Character Recognition System Using Image Transformation Based on Partial Inclination Detection," *IEICE Trans. Information & Systems,* vol. E79-D, pp. 504-509, May 1996.

[11] T.M. Ha and H. Bunke, "Off-Line, Handwritten Numeral Recognition by Perturbation Method," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 5, pp. 535-539, May 1997.

[12] P. Simard, Y. LeCun, and J. Denker, "Efficient Pattern Recognition Using a New Transformation Distance," *Advances in Neural Information Processing Systems,* vol. 5, pp. 50-58, Jan. 1993.

[13] T. Wakahara and K. Odaka, "Adaptive Normalization of Handwritten Characters Using Global/Local Affine Transformation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 12, pp. 1332-1341, Dec. 1998.

[14] T. Wakahara and Y. Kimura, "Affine-Invariant Correlation of Gray-Scale Characters Using GAT Iteration," *Proc. Fifth Int'l Conf. Document Analysis and Recognition,* pp. 613-616, Sept. 1999.

[15] Math. Soc. of Japan, *Encyclopedic Dictionary of Mathematics.* Cambridge, Mass.: MIT Press, 1977.

[16] H. Yamada, K. Yamamoto, and T. Saito, "A Nonlinear Normalization Method for Handprinted Kanji Character Recognition—Line Density Equalization," *Pattern Recognition,* vol. 23, pp. 1023-1029, Sept. 1990.

Toru Wakahara received the BE and ME degrees in applied physics and the PhD degree in mathematical engineering and information physics from the University of Tokyo, Tokyo, Japan, in 1975, 1977, and 1986, respectively. Since 1977, he has been with Nippon Telegraph and Telephone Corporation (NTT), where he has been engaged in basic research on online and offline handwritten character recognition, document image analysis, and biometric person authentication. His research interests include learning and generalizations in pattern recognition, human visual perception, and human-machine interaction. From 1991 to 1993, he was posted to the Institute for Posts and Telecommunications Policy (IPTP), Ministry of Posts and Telecommunications, Japan, where his group conducted IPTP character recognition competitions and studies on multiexpert system for 3-digit handwritten postcode recognition. He is presently a senior research engineer, supervisor, at NTT Cyber Solutions Laboratories. Dr. Wakahara is a member of the Institute of Electronics, Information and Communication Engineers of Japan (IEICE) and the IEEE.

Yoshimasa Kimura received the BE and ME degrees in electrical engineering and the PhD degree in systems engineering from the University of Tokushima, Tokushima, Japan, in 1977, 1979, and 1995, respectively. Since 1979, he has been with the Nippon Telegraph and Telephone Corporation (NTT), where he has been engaged in the research and development of printed/handwritten Kanji pattern recognition systems, online handwriting recognition, and handy-type pen-input interface. His current research interests include character recognition and neural networks. He is presently a senior research engineer at NTT Service Integration Laboratories. Dr. Kimura is a member of the Institute of Electronics, Information and Communication Engineers of Japan (IEICE), the Japanese Neural Network Society (JNNS), the Audio Visual Information Research Group of Japan (AVIRG), and the IEEE.

Akira Tomono received the BE and ME degrees in electronics engineering form Yamanashi University, Japan, in 1974 and 1976, respectively, and the DrEng degree in electrical engineering from the Tokyo Institute of Technology, Japan, in 1994. In 1976, he joined the Electrical Communication Laboratories of NTT, Japan. From 1987 to 1991, he was at the Advanced Telecommunication Research Institute International (ATR), Kyoto, Japan. In 1991, he returned to NTT Laboratories, where he was engaged in research on pattern recognition, intelligent telemonitoring systems, telemedicine systems, and biometrics personal identification. He has been a professor since April 2000 at Tokai University, Kanagawa, Japan. His research interests include human interface, ergonomics, biometrics, and virtual reality. Dr. Tomono is a member of the Institute of Electronics, Information and Communication Engineers of Japan (IEICE). He received the 1994 Best Paper Award from the Institute of Television Engineers.