

# Shape Matching Using GAT Correlation against Nonlinear Distortion and its Application to Handwritten Numeral Recognition

Toru Wakahara

*Faculty of Computer and Information Sciences*

*Hosei University*

*3-7-2 Kajino-cho, Koganei-shi, Tokyo, 184-8584 Japan*

*E-mail: wakahara@k.hosei.ac.jp*

## Abstract

*This paper addresses the problem of to what extent linear transformation can alleviate nonlinear distortion. We investigate a technique of global affine transformation (GAT) correlation to absorb linear distortion between gray-scale images. Features used in GAT correlation are occurrence probabilities of black pixels or gradients. Experiments using the handwritten numeral database IPTP CDROM1B show that the entropy of GAT-superimposed images decreases by around 15%. Furthermore, gray-level-based GAT correlation improves the recognition rate from 85.78% to 91.01%, while gradient-based GAT correlation improves the recognition rate from 91.80% to 94.02%. These results show that GAT correlation has a marked effect of improving both shape matching and discrimination abilities by extracting linear distortion from nonlinear one.*

## 1. Introduction

Most current OCR systems adopt statistical or probabilistic pattern recognition techniques, including neural networks, hidden Markov models, and support vector machines, in high-dimensional feature space. Recently, direct recognition of gray-scale characters instead of binary ones is intensively investigated with emphasis on feature extraction [1]. The success of these techniques totally depends on their high representation and prediction abilities of shape variability. In other words, statistical or probabilistic methods try successfully to “learn by examples.”

On the other hand, structural or model-based methods just try to understand what shape distortion is in a rather qualitative manner. This kind of approach is likely to depend on heuristic shape models and fails to achieve sufficient recognition accuracy against a wide range of handwriting variation. Of course, several challenging shape deformation models have been proposed by employing probabilistic or deterministic techniques, for example, deformable templates [2], [3], [4], the tangent distance [5], and a dynamic programming-based 2D

warping [6]. However, the more probabilistic techniques we utilize in order to improve the recognition accuracy, the less intuitive understanding of what shape distortion is we obtain.

In our previous paper [7], we introduced the concept of global affine transformation (GAT) correlation that achieves both noise tolerance and affine-invariance. We applied this method successfully to matching of input gray-scale images subject to uniform affine transformation together with additive random Gaussian noise against rigid templates. Our principal aim was to show how to extract and absorb linear transformation components embedded in input images.

This paper describes an enhanced GAT correlation method as applied to shape matching against nonlinear distortion occurring in handwritten characters. The key ideas are in two ways. First, we devise matching features that represent occurrence probabilities of black pixels or gradients. Namely, we introduce the probabilistic viewpoint into the GAT correlation method. Second, we determine optimal linear transformation in the 2D plane that maximizes the normalized cross-correlation value in the feature space. From experimental results using the handwritten numeral database IPTP CDROM1B we demonstrate improvements in both shape matching and discrimination abilities of the enhanced GAT correlation method capable of extracting linear distortion from nonlinear ones in gray-scale images.

## 2. IPTP CDROM1B character database

The handwritten numeral database IPTP CDROM1B provided by Institute for Posts and Telecommunications Policy of Japan [8] is used in our experiments. The IPTP CDROM1B contains binary images of handwritten digits. These binary digit images were manually segmented through binarization from 8-bit gray-scale images of three digit ZIP codes optically scanned from real Japanese New Year greeting cards. The size of each binary image is 120 dots  $\times$  80 dots in height and width. This database consists of two groups of 17,985 samples used for training and 17,916 samples used for test.

## 3. Enhanced GAT correlation method

An enhanced GAT correlation method maximizes the value of normalized cross-correlation between target and GAT-superimposed input images in the feature space. Hence, representation of input and target features and appropriate matching procedures are crucial to the success of GAT application.

### 3.1. Feature extraction

An input gray-scale image is generated from an original binary image, and the gradient is calculated. We generate two kinds of feature vectors using gray levels and gradients. The procedure is described below.

(1) Position and size normalization is applied to an original binary image by using moments. Namely, the center of gravity of black pixels is shifted to the center of the image, and the second moment around the center of gravity is set at the predetermined value. We get the binary image  $b(i, j)$ , ( $0 \leq i < 80$ ,  $0 \leq j < 120$ ), where each pixel has the value of 1 (black) or 0 (white).

(2) Neighbor averaging filter of size  $3 \times 3$  is repeatedly applied ten times to the binary image  $b(i, j)$  to generate a gray-scale image  $g(i, j)$ , ( $0 \leq i < 80$ ,  $0 \leq j < 120$ ), where  $g(i, j)$  takes a real value in the range of  $[0, 1]$ .

(3) Roberts cross-gradient operator given by (1) is applied to each pixel  $g(i, j)$  to calculate the gradient direction  $\theta(i, j)$  at the pixel as follows.

$$\begin{aligned} \Delta u &= g(i+1, j+1) - g(i, j), \\ \Delta v &= g(i+1, j) - g(i, j+1), \\ \theta(i, j) &= \tan^{-1} \left( \frac{\Delta v}{\Delta u} \right) - \frac{\pi}{4}. \end{aligned} \quad (1)$$

The gradient direction  $\theta(i, j)$  is quantized to 16 levels with  $\pi/8$  interval.

(4) We generate two kinds of feature vectors of the input image using the information of gray level  $g(i, j)$  and the gradient direction  $\theta(i, j)$ . The first feature vector  $f_1$  represents the two-dimensional distribution  $G(i, j)$  of gray levels obtained simply by down sampling every two horizontal and vertical pixels of  $g(i, j)$  blurred by  $2 \times 2$  neighbor averaging. As a result, we obtain a 2,400-dimensional feature vector:

$$f_1 \equiv (G(i, j)), \quad 0 \leq i < 40, \quad 0 \leq j < 60. \quad (2)$$

The second feature vector  $f_2$  represents the two-dimensional distributions  $\{ G_k(i, j) \}$  ( $0 \leq k \leq 7$ ) of pixels whose gradient direction has either of eight directions with  $\pi/4$  interval. To calculate  $f_2$  we first register each pixel of the original  $g(i, j)$  in either of 16 directional planes according to its gradient direction. Second, the spatial resolution is reduced from  $120 \times 80$  to  $15 \times 10$  by down sampling with both Gaussian and neighbor averaging filters. Similarly, the directional resolution is

reduced from 16 to 8 by weighted down sampling. Finally, we obtain a 1,200-dimensional feature vector:

$$\begin{aligned} f_2 \equiv & (G_0(i, j), G_1(i, j), G_2(i, j), G_3(i, j), \\ & G_4(i, j), G_5(i, j), G_6(i, j), G_7(i, j)), \\ & 0 \leq i < 10, \quad 0 \leq j < 15. \end{aligned} \quad (3)$$

(5) According to definite canonicalization [9] two kinds of feature vectors are linearly transformed to satisfy the following normalization requirements.

$$\begin{aligned} \bar{f}_1 \equiv & \sum_j \sum_i G(i, j) = 0, \quad \|f_1\| = 1, \\ \bar{f}_2 \equiv & \sum_k \sum_j \sum_i G_k(i, j) = 0, \quad \|f_2\| = 1. \end{aligned} \quad (4)$$

### 3.2. Target representation

In our previous paper [7], we used 2D rigid templates as targets and matched them with artificial input images generated by applying uniform affine transformation and additive random Gaussian noise to those templates.

When dealing with handwritten characters it is not so straightforward to prepare appropriate 2D templates as targets. We need a technique of 2D point-wise registration between two images of handwritten characters for getting an average image. However, we have not developed a perfect 2D registration technique as yet.

To resolve the above-mentioned problem, we develop targets that represent occurrence probabilities of 2D distributions of matching features. The concrete procedure for generating targets is as follows.

(1) As targets to be matched with the first feature vector  $f_1$  we introduce the occurrence probabilities  $P_B(i, j)$  of the value of 1 (black) at each pixel  $(i, j)$  calculated by accumulating all training samples' binary images  $b(i, j)$ .

Figure 1 shows targets expressed by gray-scale images where the darker pixel represents the higher occurrence probability  $P_B(i, j)$  of black at the pixel.



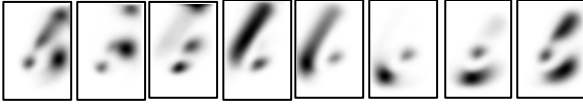
**Figure 1. Targets representing occurrence probabilities of black pixels.**

Finally, we obtain the target feature vector  $r_1 = (P_B(i, j))$ ,  $0 \leq i < 40$ ,  $0 \leq j < 60$ , for each category of ten digits through reduction in the spatial resolution. Also, the procedure of definite canonicalization is applied to  $r_1$  in order that the following conditions hold:

$$\bar{r}_1 \equiv \sum_j \sum_i P_B(i, j) = 0, \quad \|r_1\| = 1. \quad (5)$$

(2) As targets for the second feature vector  $\mathbf{f}_2$  we introduce the occurrence probabilities  $\{P_k(i, j)\}$  ( $0 \leq k \leq 7$ ) of eight gradient directions with  $\pi/4$  interval that are calculated by accumulating all training samples' gray-scale images  $g(i, j)$  in individual gradient direction. Figure 2 shows the target of digit "6" expressed by gray-scale images where the darker pixel represents the higher occurrence probability of each of eight directions. According to the same procedure in dimension reduction as that applied to input feature vectors we obtain the target feature vector  $\mathbf{r}_2 = (P_0(i, j), P_1(i, j), P_2(i, j), P_3(i, j), P_4(i, j), P_5(i, j), P_6(i, j), P_7(i, j))$ ,  $0 \leq i < 10, 0 \leq j < 15$ , for each category of ten digits, where  $\mathbf{r}_2$  satisfies the conditions subject to definite canonicalization given by

$$\bar{\mathbf{r}}_2 \equiv \sum_k \sum_j \sum_i P_k(i, j) = 0, \quad \|\mathbf{r}_2\| = 1. \quad (6)$$



**Figure 2. Target of digit "6" representing occurrence probabilities of each of eight directions with  $\pi/4$  interval.**

### 3.3. Optimal GAT determination

By using vectors,  $\mathbf{f}$  and  $\mathbf{r}$ , subject to definite canonicalization we can calculate the normalized cross-correlation value between  $\mathbf{f}$  and  $\mathbf{r}$  by means of their simple inner product  $(\mathbf{f}, \mathbf{r})$ . The normalized cross-correlation value is robust against blurring and random noise [9].

The aim of GAT correlation is to determine the optimal affine transformation as applied to the input 2D image that achieves the maximum correlation value between target and GAT-superimposed input in the feature space, that is, the value of  $(\mathbf{r}, \text{Affine}[\mathbf{f}])$ . Here,  $\text{Affine}[\cdot]$  denotes the operation of affine transformation in the 2D space, and is defined by a  $2 \times 2$  matrix  $\mathbf{A}$  representing rotation, scale-change, and shearing, and a translation vector  $\mathbf{b}$ :

$$\mathbf{A} = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}. \quad (7)$$

To maximize the value of  $(\mathbf{r}, \text{Affine}[\mathbf{f}])$  we introduce the following objective function  $\Phi$  given by

$$\begin{aligned} \Phi &\equiv (\mathbf{r}, \text{Affine}[\mathbf{f}]) = \sum_j \sum_i r(i, j) \times \text{Affine}[\mathbf{f}](i, j) \\ &= \sum_j \sum_i r(\tilde{i}, \tilde{j}) f(i, j) \rightarrow \max \text{ for } \mathbf{A} \text{ and } \mathbf{b}. \\ \tilde{i} &= a_{00} \times i + a_{01} \times j + b_0, \quad \tilde{j} = a_{10} \times i + a_{11} \times j + b_1. \end{aligned} \quad (8)$$

Then, to eliminate exhaustive trial and error for determining optimal  $\mathbf{A}$  and  $\mathbf{b}$  we substitute  $\Phi$  of (8) with another objective function  $\Psi$  employing the Gaussian kernel as follows.

$$\begin{aligned} \Psi &\equiv \sum_j \sum_i \sum_{j'} \sum_{i'} f(i, j) r(i', j') G(\mathbf{A}, \mathbf{b}, i, j, i', j') \\ &\rightarrow \max \text{ for } \mathbf{A} \text{ and } \mathbf{b}. \end{aligned}$$

$$G(\mathbf{A}, \mathbf{b}, i, j, i', j') = \exp\left(-\frac{(\tilde{i}-i)^2 + (\tilde{j}-j)^2}{D}\right),$$

$$\tilde{i} = a_{00} \times i' + a_{01} \times j' + b_0,$$

$$\tilde{j} = a_{10} \times i' + a_{11} \times j' + b_1, \quad (9)$$

where  $D$  controls the spread of the Gaussian kernel [7]. To solve (9), we derive a set of nonlinear equations by setting the derivatives of  $\Psi$  with respect to each of six unknown parameters,  $a_{00}$ ,  $a_{01}$ ,  $a_{10}$ ,  $a_{11}$ ,  $b_0$ , and  $b_1$ , equal to zero, respectively. Next, we obtain a set of simultaneous linear equations by using the 0<sup>th</sup> order approximation that sets  $\mathbf{A} = \mathbf{I}$  and  $\mathbf{b} = \mathbf{0}$  in the Gaussian kernel. Finally, we solve these simultaneous linear equations by conventional techniques to determine sub-optimal  $\mathbf{A}$  and  $\mathbf{b}$ .

We adopt the successive iteration method in order to obtain the optimal GAT of (8) by iteratively updating the input feature vector by sub-optimal GAT components of (9) until the value of  $\Phi$  arrives at a maximum.

The individual objective functions of (8) corresponding to each of the two kinds of feature extraction are given as follows:

$$\Phi_1 \equiv (\mathbf{r}_1, \text{Affine}[\mathbf{f}_1]) = \sum_{j=0}^{59} \sum_{i=0}^{39} P_B(\tilde{i}, \tilde{j}) \times G(i, j),$$

$$\Phi_2 \equiv (\mathbf{r}_2, \text{Affine}[\mathbf{f}_2]) = \sum_{k=0}^7 \sum_{j=0}^{14} \sum_{i=0}^9 P_k(\tilde{i}, \tilde{j}) \times G_k(i, j),$$

$$\tilde{i} = a_{00} \times i + a_{01} \times j + b_0, \quad \tilde{j} = a_{10} \times i + a_{11} \times j + b_1. \quad (10)$$

We denote optimal GAT determinations using  $\Phi_1$  and  $\Phi_2$  by gray-level GAT and gradient GAT, respectively.

## 4. Experimental results

Nonlinear handwriting distortion cannot be absorbed in principle only by linear transformation. Hence, our primary concern is to investigate to what extent linear transformation can alleviate nonlinear distortion.

In this chapter we demonstrate two kinds of experimental results. First, we prove the high shape matching ability of the enhanced GAT correlation method as applied to each test sample against its correct target. Second, from recognition experiments on IPTP CDROM1B character database we demonstrate distinct improvements in discrimination ability using the maximized normalized cross-correlation values obtained by the proposed method.

#### 4.1. Shape matching ability

We determine the optimal GAT between the feature vector  $f$ , extracted from each test sample, and its correct category's target vector  $r_{\text{correct}}$ . By applying the optimal GAT to the original binary input image  $b(i, j)$ , ( $0 \leq i < 80$ ,  $0 \leq j < 120$ ), we can generate the GAT-superimposed binary image  $b^{\text{GAT}}(i, j)$ , ( $0 \leq i < 80$ ,  $0 \leq j < 120$ ).

Then, we compare the entropy value of a set of original input images with that of a set of GAT-superimposed images. Here, the entropy value  $\varepsilon$  for a set of  $N$  binary images  $\{b_n(i, j)\}$  ( $n = 1, 2, \dots, N$ ) is defined by

$$\varepsilon = -\frac{1}{m_{av}} \sum_j \sum_i p(i, j) \ln p(i, j),$$

$$p(i, j) = \frac{1}{N} \sum_n b_n(i, j), \quad (11)$$

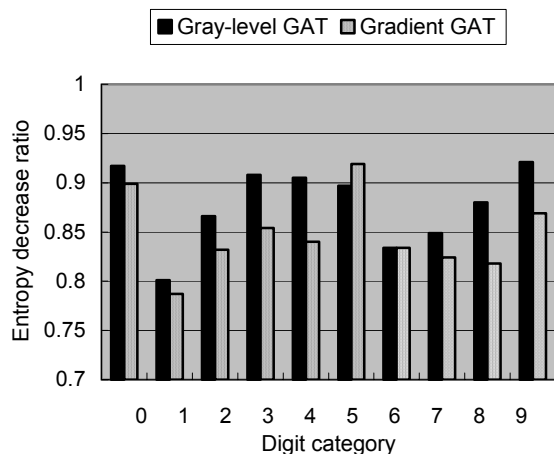
where  $m_{av}$  denotes the mean of the total number of black pixels in the binary input image.

Finally, we introduce the entropy decrease ratio  $\eta$ , as an evaluation criterion, defined by

$$\eta = \varepsilon_{\text{GAT}} / \varepsilon_0, \quad (12)$$

where  $\varepsilon_0$  and  $\varepsilon_{\text{GAT}}$  denote the original entropy value and one obtained by GAT application, respectively.

Figure 3 shows the relation between the  $\eta$  values and categories of ten digits in shape matching experiments using all test samples and their correct targets. The mean values of  $\eta$  obtained by gray-level GAT and gradient GAT were 0.879 and 0.848, respectively.



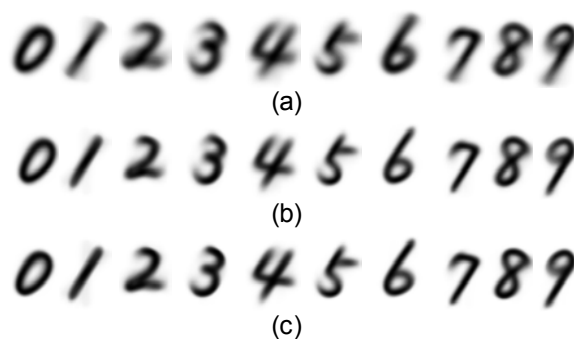
**Figure 3. Relation between entropy decrease ratios and digit categories.**

From Figure 3, it is first found that the GAT correlation technique can alleviate the shape variation substantially, although the degree of the entropy decrease varies with digit category. Also, it is found that gradient GAT has superior shape matching ability over gray-level GAT for

not all but most of digit categories. This result points out that the gradient information is effective in enhancing the shape matching ability.

Figure 4 shows three kinds of overlapped images of original test samples, gray-level-GAT-superimposed and gradient-GAT-superimposed ones.

From Figure 4, it is found that the degree of original shape variation shown in Figure 4(a) depends heavily on digit category. Actually, the entropy values  $\varepsilon$  of “4”, “2”, “5”, and “3” are much higher than those of the remainder. Also, from Figure 4(b) and Figure 4(c), it is seen that the variety of shape distortion of “4” and “2” cannot be absorbed by simple linear transformation to a satisfying extent.



**Figure 4. Overlapped images of test samples: (a) Original, (b) gray-level-GAT-superimposed, and (c) gradient-GAT-superimposed ones.**

#### 4.2. Discrimination ability

Recognition experiments are made on a test data set in ITPC CDR0M1B character database using the maximized normalized cross-correlation values obtained by either gray-level GAT or gradient GAT.

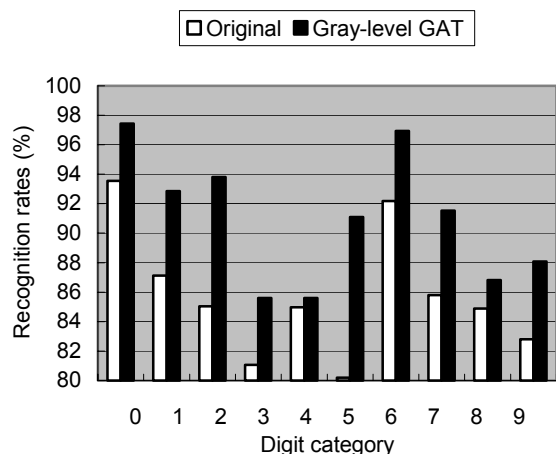
Figure 5 shows the relation between recognition rates and digit categories in case of gray-level GAT. The average recognition rates before and after gray-level GAT application are 85.78% and 91.01%, respectively. In other words, gray-level GAT realized an about 37% reduction in the number of errors. This result means that the very linear transformation is helpful for enhancing naive correlation-based recognition of gray-scale images.

Figure 6 shows the relation between recognition rates and digit categories in case of gradient GAT. The average recognition rate jumped from 91.80% to 94.02%. This is equivalent to an about 27% reduction in the number of errors. However, it is to be noted that gradient GAT produced the contrary effect in case of “4”.

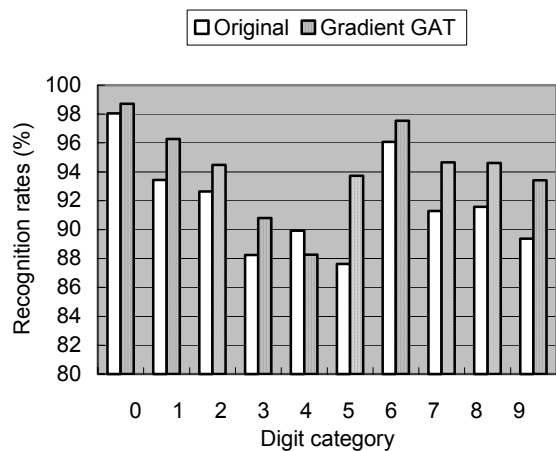
From Figure 5 and Figure 6, it is clear that the original recognition rates of gradient features are much higher than those of gray-level features. This superiority of

gradient features over gray-level features is maintained through GAT correlation application. These results strongly suggest that the GAT correlation method can work in powerful cooperation with sophistication of feature extraction.

Incidentally, from-binary-to-gray-scale conversion of input binary images was found to be very effective in stabilizing matching against targets that represent occurrence probabilities of 2D distributions of features. Direct use of input binary images in GAT correlation deteriorated the recognition rates by about 2%.



**Figure 5. Relation between recognition rates and digit categories by gray-level GAT application.**



**Figure 6. Relation between recognition rates and digit categories by gradient GAT application.**

Finally, generation of multiple templates based on clustering within each digit category is very promising to improve recognition rates. Moreover, it is indispensable

to reinforce a 2D shape model to resolve residual nonlinear distortion after GAT application.

## 5. Conclusion

It is still challenging to tackle the problem of describing handwriting distortion not in a probabilistic manner but in a deterministic manner. From this viewpoint this paper addressed an interesting, essential sub-problem of to what extent linear transformation can alleviate nonlinear distortion.

The enhanced GAT correlation method using features of gray levels and gradients was found to be highly effective in increasing shape matching ability. It is also shown from recognition experiments that the very application of linear transformation to nonlinear distortion achieved notable improvements in discrimination ability.

Future work is modeling of residual, nonlinear shape distortion in a deterministic, controllable manner that is helpful to our understanding of what shape distortion is.

## References

- [1] M. Shi, Y. Fujisawa, T. Wakabayashi, and F. Kimura, "Handwritten numeral recognition using gradient and curvature of gray scale image," *Pattern Recognition*, vol. 35, pp. 2051-2059, 2002.
- [2] M. Revow, C.K.I. Williams, and G.E. Hinton, "Using generative models for handwritten digit recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-18, pp. 592-606, 1996.
- [3] A.K. Jain and D. Zongker, "Representation and recognition of handwritten digits using deformable templates," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-19, pp. 1386-1390, 1997.
- [4] K.W. Cheung, D.Y. Yeung, and R.T. Chin, "A Bayesian framework for deformable pattern recognition with application to handwritten character recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-20, pp. 1382-1388, 1998.
- [5] 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, Morgan Kaufmann, 1993.
- [6] M.A. Ronee, S. Uchida, and H. Sakoe, "Handwritten character recognition using piecewise linear two-dimensional warping," *Proc. of Sixth Int. Conf. on Document Analysis and Recognition*, pp. 39-43, Seattle, Sept. 2001.
- [7] T. Wakahara, Y. Kimura, and A. Tomono, "Affine-invariant recognition of gray-scale characters using global affine transformation correlation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-23, pp. 384-395, 2001.
- [8] K. Osuga, T. Tsutsumida, S. Yamaguchi, and K. Nagata, "IPTP survey on handwritten numeral recognition," *IPTP Research and Survey Report*, R-96-V-02, 1996.
- [9] T. Iijima, *Pattern Recognition*, chapter 6, Tokyo: Corona,

1973 (in Japanese).