Affine-invariant Recognition of Handwritten Characters via Accelerated KL Divergence Minimization

Toru Wakahara
Faculty of Computer and Information Sciences
Hosei University
3-7-2 Kajino-cho, Koganei-shi, Tokyo, 184-8584 Japan
E-mail: wakahara@hosei.ac.jp

Yukihiko Yamashita
Graduate School of Engineering and Science
Tokyo Institute of Technology
2-12-1 O-okayama, Meguro-ku, Tokyo, 152-8550 Japan
E-mail: yamasita@ide.titech.ac.jp

Abstract—This paper proposes a new, affine-invariant image matching technique via accelerated KL (Kullback-Leibler) divergence minimization. First, we represent an image as a probability distribution by setting the sum of pixel values at one. Second, we introduce affine parameters into either of the two images' probability distributions using the Gaussian kernel density estimation. Finally, we determine optimal affine parameters that minimize KL divergence via an iterative method. In particular, without using such conventional nonlinear optimization techniques as the Levenberg-Marquardt method we devise an accelerated iterative method adapted to the KL divergence minimization problem through effective linear approximation. Recognition experiments using the handwritten numeral database IPTP CDROM1B show that the proposed method achieves a much higher recognition rate of 91.5% at suppressed computational cost than that of 83.7% obtained by a simple image matching method based on a normal KL divergence.

Keywords-affine-invariant image matching; Gaussian kernel density estimation; KL divergence; character recognition;

I. INTRODUCTION

Most successful OCR systems adopt statistical or probabilistic pattern recognition techniques using a large amount of training data [1]. However, the problem of how to improve the recognition accuracy when we have only a limited quantity of data remains unsolved. To resolve this problem, several promising matching techniques based on deformable models have been proposed. Revow et al. [2] and Jain et al. [3] reinforced their deformable models via probabilistic viewpoints. On the other hand, DP-based 2D warping by Ronee et al. [4], the tangent distance by Simard et al. [5], and GAT correlation by Wakahara et al. [6] belong to enhanced techniques of distortion-tolerant template matching. Moreover, Wakahara et al. [7] extended GAT correlation to PAT correlation to deal with nonlinear distortion.

From the viewpoint of image matching measure the above-mentioned techniques adopted either simple graylevel difference at each pixel or normalized cross-correlation.

On the other hand, Viola et al. [8] introduced the technique of mutual information maximization in an affine-invariant image alignment problem. They adopted the mutual information as a matching measure. However, the nonlinear optimization problem thus introduced has been solved by such general-purpose and/or time-consuming techniques as the gradient descent method and the Levenberg-Marquardt method [9].

In this paper we propose a new, affine-invariant image matching technique via accelerated KL divergence minimization. The KL divergence [10] is an asymmetric measure of the difference between two probability distributions. Also, the mutual information is expressed by the KL divergence. Our proposed method consists of three steps. The first step is representation of an image as a probability distribution by setting the sum of pixel values at one. The second step is introduction of affine parameters into either of the two images' probability distributions using the Gaussian kernel density estimation. The third step is determination of optimal affine parameters that minimize KL divergence via an iterative method. In particular, we devise an accelerated iterative method specially adapted to the KL divergence minimization problem through effective linear approximation.

Experimental results made on the handwritten numeral database IPTP CDROM1B demonstrate that the proposed method achieves a decided improvement in recognition accuracy at suppressed computational cost compared to a simple image matching method based on a normal KL divergence.

II. REPRESENTATION OF PROBABILITY DISTRIBUTIONS OF IMAGES AND KL DIVERGENCE

In this section we explain how to represent probability distributions of two grayscale images to be matched. Then, we introduce the KL divergence of two probability distributions as a matching measure between two grayscale images.

First, we specify two grayscale images to be matched by a reference image, $\boldsymbol{f} = \{\, f_{ij} \,\}$, and an input image, $\boldsymbol{g} = \{\, g_{ij} \,\}, (0 \le i < m, 0 \le j < n)$, where f_{ij} and g_{ij} take integers in $[\,0,\,255\,]$.

In advance, we transform their grayscale values linearly so that the brightest pixels have the value of 255 and the darkest pixels have the value of 0. Also, we assume that

"figure" and "background" are represented by darker pixels and brighter pixels, respectively.

Next, we define a probability distribution of an image as one that each pixel has a probability proportional to its individual grayscale value and the sum of those probabilities for $m \times n$ pixels totals up to one.

We denote a probability distribution of the reference image and a probability distribution of the input image by $\boldsymbol{p} = \{\, p_{ij} \,\}$ and $\boldsymbol{q} = \{\, q_{ij} \,\}, (0 \leq i < m, 0 \leq j < n)$, respectively, and calculate them by

$$
\begin{aligned}
p_{ij} &= \frac{255 - f_{ij} + \alpha}{\sum_{i'=0}^{m-1} \sum_{j'=0}^{n-1} (255 - f_{i'j'} + \alpha)}, \\
q_{ij} &= \frac{255 - g_{ij} + \alpha}{\sum_{i'=0}^{m-1} \sum_{j'=0}^{n-1} (255 - g_{i'j'} + \alpha)},
\end{aligned} \quad (1)
$$

where a positive constant α is introduced so that p_{ij} and q_{ij} always have positive values. Actually, we set the value of α at one. Also, it is clear that $\sum_{i,j} p_{ij} = 1$ and $\sum_{i,j} q_{ij} = 1$ are satisfied.

KL divergence [10] between the probability distributions $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$ is the average additional amount of information required to specify the value of $\boldsymbol{x}$ as a result of using an approximating distribution $q(\boldsymbol{x})$ instead of the true distribution $p(\boldsymbol{x})$ given by

$$
\begin{aligned}
\mathrm{KL}(p \,\|\, q) &= -\int p(\boldsymbol{x}) \ln q(\boldsymbol{x}) d\boldsymbol{x} + \int p(\boldsymbol{x}) \ln p(\boldsymbol{x}) d\boldsymbol{x} \\
&= \int p(\boldsymbol{x}) \ln \left\{ \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} \right\} d\boldsymbol{x}.
\end{aligned} \quad (2)
$$

The KL divergence is not a symmetrical quantity, i.e., $\mathrm{KL}(p \,\|\, q) \neq \mathrm{KL}(q \,\|\, p)$, and satisfies $\mathrm{KL}(p\|q) \geq 0$ with equality if and only if $p(\boldsymbol{x}) = q(\boldsymbol{x})$. Thus we can interpret the KL divergence as a measure of the dissimilarity of the two probability distributions $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$.

Therefore, we propose to use the discretized KL divergence using $\boldsymbol{p} = \{\, p_{ij} \,\}$ and $\boldsymbol{q} = \{\, q_{ij} \,\}$ as the matching or dissimilarity measure between the reference image $\boldsymbol{f}$ and the input image $\boldsymbol{g}$ given by

$$
\mathrm{KL}(\boldsymbol{p} \,\|\, \boldsymbol{q}) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} p_{ij} \ln \left\{ \frac{p_{ij}}{q_{ij}} \right\}. \quad (3)
$$

III. Embedding of affine parameters in KL divergence via kernel density estimation

In this section we propose to affine-transform either of reference and input images and estimate its probability distribution thus transformed via the Gaussian kernel density estimation technique. As a result, we obtain the KL divergence in which affine parameters are embedded.

Following the Gaussian kernel density estimation [10], when observations $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N$ are being drawn from some unknown probability density $p(\boldsymbol{x})$ in some D-dimensional space, we estimate the value of $p(\boldsymbol{x})$ by

$$
p(\boldsymbol{x}) = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{(2\pi h^2)^{D/2}} \exp \left\{ -\frac{\| \boldsymbol{x} - \boldsymbol{x}_n \|^2}{2h^2} \right\}, \quad (4)
$$

where h represents the standard deviation of the Gaussian components. Thus our density model is obtained by placing a Gaussian over each data point and then adding up the contributions over the whole data set, and then dividing by N so that the density is correctly normalized.

For our 2-dimensional image matching problem, we deal with the probability distribution $\boldsymbol{q}$ of the input image $\boldsymbol{g}$ and we can consider that each data point $\boldsymbol{x}_{ij} = (i, j)^T$ has its individual weight q_{ij}. Therefore, using the Gaussian kernel density estimation of (4) we can estimate the probability density $\tilde{\boldsymbol{q}} = \{\, \tilde{q}_{ij} \,\}, (0 \leq i < m, 0 \leq j < n)$ by

$$
\tilde{q}_{ij} = C \sum_{i'=0}^{m-1} \sum_{j'=0}^{n-1} q_{i'j'} \exp \left\{ -\frac{\| \boldsymbol{x}_{ij} - \boldsymbol{x}_{i'j'} \|^2}{2h^2} \right\}, \quad (5)
$$

where C is a normalizing constant by which $\sum_i \sum_j \tilde{q}_{ij} = 1$ is satisfied.

Then, we apply two-dimensional affine transformation expressed by

$$
\mathrm{A} = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}, \quad (6)
$$

to the input image $\boldsymbol{g}$. This transformation moves each data point $\boldsymbol{x}_{ij} = (i, j)^T$ of the input image to $\boldsymbol{x}_{ij}^* = (i^*, j^*)^T$ together with its individual weight q_{ij} by

$$
\boldsymbol{x}_{ij}^* = \mathrm{A} \boldsymbol{x}_{ij} + \boldsymbol{b}, \quad (7)
$$

where $i^* = a_{00} i + a_{01} j + b_0, \quad j^* = a_{10} i + a_{11} j + b_1$.

According to (5) and (7), the probability density $\tilde{\boldsymbol{q}^*} = \{\, \tilde{q}_{ij}^* \,\}, (0 \leq i < m, 0 \leq j < n)$ of the affine-transformed input image $\boldsymbol{g}^*$ is estimated by

$$
\tilde{q}_{ij}^* = C' \sum_{i'=0}^{m-1} \sum_{j'=0}^{n-1} q_{i'j'} \exp \left\{ -\frac{\| \boldsymbol{x}_{ij} - \boldsymbol{x}_{i'j'}^* \|^2}{2h^2} \right\}, \quad (8)
$$

where C' is a normalizing constant by which $\sum_i \sum_j \tilde{q}_{ij}^* = 1$ is satisfied.

Finally, we can obtain the KL divergence between two probability distributions of the reference image $\boldsymbol{f}$ and the affine-transformed input image $\boldsymbol{g}^*$ in which affine parameters A and $\boldsymbol{b}$ of (6) are embedded given by

$$
\mathrm{KL}(\boldsymbol{p} \,\|\, \tilde{\boldsymbol{q}}^*) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} p_{ij} \ln \left\{ \frac{p_{ij}}{\tilde{q}_{ij}^*} \right\}. \quad (9)
$$

IV. Accelerated minimization of KL divergence for affine parameters

In this section we propose an efficient computational model of affine parameters that minimizes the KL divergence of (9). The value of KL divergence thus minimized is considered as an affine-invariant image matching measure.

First, as necessary conditions for minimizing the KL divergence of (9), we set the derivatives of (9) with respect to affine parameters at 0 as follows.

$$\frac{\partial \, \text{KL}(\boldsymbol{p} \,\|\, \tilde{\boldsymbol{q}}^*)}{\partial \, \text{A}} = \text{O}, \qquad \frac{\partial \, \text{KL}(\boldsymbol{p} \,\|\, \tilde{\boldsymbol{q}}^*)}{\partial \, \boldsymbol{b}} = \boldsymbol{0}. \tag{10}$$

Substituting (7), (8), and (9) into (10) then gives the following simultaneous equations of affine parameters.

$$
\begin{aligned}
0 &= \sum_{i,j} \frac{p_{ij}}{\tilde{q}_{ij}^*} \sum_{i',j'} q_{i'j'} \, i' \, (a_{00} \, i' + a_{01} \, j' + b_0 - i) \, Q, \\
0 &= \sum_{i,j} \frac{p_{ij}}{\tilde{q}_{ij}^*} \sum_{i',j'} q_{i'j'} \, j' \, (a_{00} \, i' + a_{01} \, j' + b_0 - i) \, Q, \\
0 &= \sum_{i,j} \frac{p_{ij}}{\tilde{q}_{ij}^*} \sum_{i',j'} q_{i'j'} \, i' \, (a_{10} \, i' + a_{11} \, j' + b_1 - j) \, Q, \\
0 &= \sum_{i,j} \frac{p_{ij}}{\tilde{q}_{ij}^*} \sum_{i',j'} q_{i'j'} \, j' \, (a_{10} \, i' + a_{11} \, j' + b_1 - j) \, Q, \\
0 &= \sum_{i,j} \frac{p_{ij}}{\tilde{q}_{ij}^*} \sum_{i',j'} q_{i'j'} \, (a_{00} \, i' + a_{01} \, j' + b_0 - i) \, Q, \\
0 &= \sum_{i,j} \frac{p_{ij}}{\tilde{q}_{ij}^*} \sum_{i',j'} q_{i'j'} \, (a_{10} \, i' + a_{11} \, j' + b_1 - j) \, Q, \\
Q &\equiv \exp\left\{ -\frac{\| \boldsymbol{x}_{ij} - \boldsymbol{x}_{i'j'}^* \|^2}{2h^2} \right\}.
\end{aligned}
\tag{11}
$$

Because affine parameters appear as arguments of exponential functions in $\tilde{q}_{ij}^*$ and Q, these simultaneous equations are nonlinear and cannot be solved analytically in a closed form.

Such general-purpose techniques as the gradient descent method, the Gauss-Newton method, and the Levenberg-Marquardt method have been used to solve nonlinear optimization problems. However, a straightforward application of those techniques to a particular nonlinear optimization problem is not so effective in general and is usually very time-consuming. Actually, in preliminary experiments we applied the Levenberg-Marquardt method to the KL divergence minimization of (9), but the obtained results were unsatisfactory in terms of both the optimization ability and the computational cost.

Hence, we devise a new accelerated iterative method specially adapted to our problem of the KL divergence minimization of (9).

The key idea is an effective linear approximation of (11). Namely, we set affine parameters appearing in the exponential functions of $\tilde{q}_{ij}^*$ and Q at $\text{A} = \text{I}$ and $\boldsymbol{b} = \boldsymbol{0}$ as a zeroth order approximation. As a result, we obtain the following simultaneous linear equations of affine parameters.

$$
\begin{aligned}
0 &= \sum_{i,j} \frac{p_{ij}}{\tilde{q}_{ij}} \sum_{i',j'} q_{i'j'} \, i' \, (a_{00} \, i' + a_{01} \, j' + b_0 - i) \, Q', \\
0 &= \sum_{i,j} \frac{p_{ij}}{\tilde{q}_{ij}} \sum_{i',j'} q_{i'j'} \, j' \, (a_{00} \, i' + a_{01} \, j' + b_0 - i) \, Q', \\
0 &= \sum_{i,j} \frac{p_{ij}}{\tilde{q}_{ij}} \sum_{i',j'} q_{i'j'} \, i' \, (a_{10} \, i' + a_{11} \, j' + b_1 - j) \, Q', \\
0 &= \sum_{i,j} \frac{p_{ij}}{\tilde{q}_{ij}} \sum_{i',j'} q_{i'j'} \, j' \, (a_{10} \, i' + a_{11} \, j' + b_1 - j) \, Q', \\
0 &= \sum_{i,j} \frac{p_{ij}}{\tilde{q}_{ij}} \sum_{i',j'} q_{i'j'} \, (a_{00} \, i' + a_{01} \, j' + b_0 - i) \, Q', \\
0 &= \sum_{i,j} \frac{p_{ij}}{\tilde{q}_{ij}} \sum_{i',j'} q_{i'j'} \, (a_{10} \, i' + a_{11} \, j' + b_1 - j) \, Q', \\
Q' &\equiv \exp\left\{ -\frac{\| \boldsymbol{x}_{ij} - \boldsymbol{x}_{i'j'} \|^2}{2h^2} \right\}.
\end{aligned}
\tag{12}
$$

We can solve these simultaneous linear equations easily by conventional techniques like the Gaussian elimination [11]. However, because of the above-mentioned linear approximation the obtained affine parameters are not the optimal solution but a sub-optimal solution of (11).

Therefore, we adopt the successive iteration method [11] and iteratively affine-transform the input image by the sub-optimal solution until the KL divergence of (9) arrives at a minimum. The procedure of the successive iteration method used here is as follows.

$Step \, 1$: By using the initial probability distributions $\boldsymbol{p} = \{ p_{ij} \}$ and $\boldsymbol{q} = \{ q_{ij} \}$ of (1), we calculate the initial value $\text{KL}^{(\tau=0)}(\boldsymbol{p} \,\|\, \boldsymbol{q})$ of (3).

$Step \, 2$: We solve (12) to obtain A and $\boldsymbol{b}$ as a sub-optimal solution of (11). Then, we affine-transform the input image $\boldsymbol{g}$ into $\boldsymbol{g}^*$ by A and $\boldsymbol{b}$, and substitute a new $\boldsymbol{g}^*$ for the old $\boldsymbol{g}$.

$Step \, 3$: After updating the probability distribution $\boldsymbol{q} = \{ q_{ij} \}$ and setting $\tau = \tau + 1$, we calculate the updated value $\text{KL}^{(\tau)}(\boldsymbol{p} \,\|\, \boldsymbol{q})$. If there is no further decrease in the KL divergence value, we output the present value as the minimized KL divergence value and stop the iteration. Otherwise, we go to $Step \, 2$.

Finally, it is to be noted that the proposed method includes only one model parameter h of (4) representing the standard deviation of the Gaussian components.

V. Experimental results

In this section we apply the proposed affine-invariant image matching technique via KL divergence minimization to handwritten character recognition.

We use the handwritten numeral database IPTP CDROM1B [12], although this is not a case of insufficient

training data that we are aiming to deal with. This database contains binary images of handwritten digits divided into two groups of 17,985 samples for training and 17,916 samples for test. Actually, the highest recognition rate ever reported for this database is 99.49% obtained via sophisticated discriminant functions in high-dimensional feature space [13].

First, position and size normalization by moments [14] is applied to each binary image so that the center of gravity of black pixels is located at the center of the image and the average distance of black pixels from the center of the image is set at the predetermined value of ρ ($= 12.5$). Then, we transform all of binary images into grayscale images by Gaussian filtering and set the image size at 40×60 pixels. Hence, we have $m = 40$ and $n = 60$ in the notation used in Section II.

Second, we generate a single reference image per digit by averaging each category's training samples.

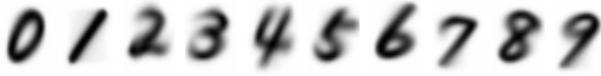Figure 1 shows reference images generated for ten digits.

Figure 1.   Reference images for ten digits.

Third, we apply the proposed affine-invariant image matching technique to each of 17,916 test samples against ten reference images. From preliminary experiments, the value of h of (4) was decided to be set at 1.5.

Furthermore, we obtain two kinds of matching results by applying affine transformation to a test sample in one case and applying affine transformation to reference images in the other case. Then, we adopt the average of two kinds of minimized KL divergence values as the final KL divergence value.

Finally, we denote the initial KL divergence values and the final KL divergence values by $\mathrm{KL}^{org}(\boldsymbol{p}_\omega \| \boldsymbol{q})$ and $\mathrm{KL}^{final}(\boldsymbol{p}_\omega \| \boldsymbol{q})$ ($\omega = 0, 1, \ldots 9$), respectively. Here, ω specifies a digit category.

Now, our first concern is to investigate to what extent affine transformation can reduce the KL divergence values between test samples and their correct category's reference images.

On the other hand, suppression of excessive matching is crucial to distortion-tolerant image matching. Hence, our second concern is to examine how much the proposed affine-invariant image matching technique can improve the recognition accuracy.

Figure 2 shows occurrence rates of the initial and final KL divergence values at intervals of 0.02 being divided into two cases: KL divergence values against correct categoris and incorrect but the most similar or "rival" categories.

From Fig. 2, it is found that the proposed affine-invariant image matching technique achieves a marked decrease in
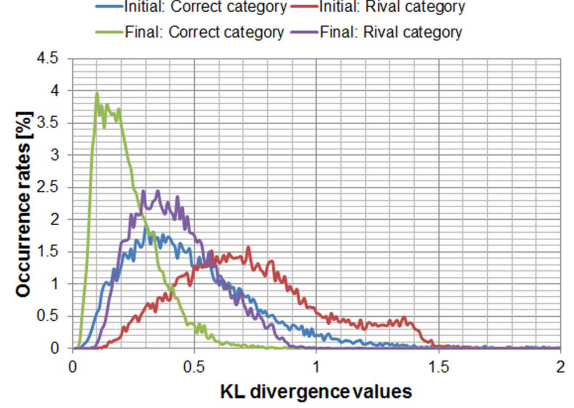

Figure 2.   Occurrence rates of the KL divergence values.

the KL divergence values against not only correct but also rival categories. This fact shows that most of handwriting distortion can be expressed by affine transformation, which has been already utilized by Bunke et al. [15] via the perturbation method based on affine transformation.

In recognition experiments, we assign each test sample with a probability distribution $\boldsymbol{q}$ of (1) to the digit category with the minimum KL divergence value given by

$$\omega^{final} = \operatorname*{argmin}_{\omega} \mathrm{KL}^{final}(\boldsymbol{p}_\omega \| \boldsymbol{q}). \qquad (13)$$

Similarly, a simple image matching method using the initial KL divergence values as a matching measure assigns each test sample to the digit category given by

$$\omega^{org} = \operatorname*{argmin}_{\omega} \mathrm{KL}^{org}(\boldsymbol{p}_\omega \| \boldsymbol{q}). \qquad (14)$$

Figure 3 shows recognition rates obtained by using the initial and final KL divergence values.
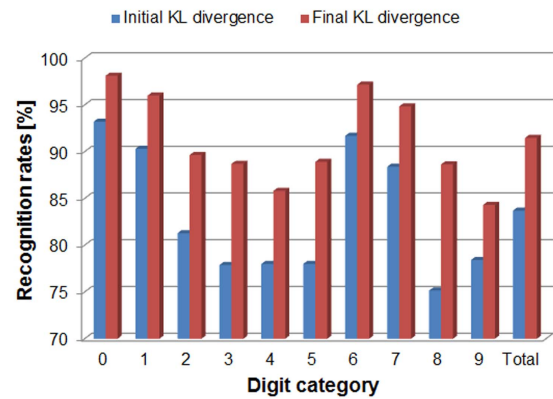

Figure 3.   Recognition rates via the initial and final KL divergence values.

From Fig. 3, it is clear that the proposed affine-invariant image matching technique via KL divergence minimization achieves a substantial increase in discrimination ability as compared to a simple image matching technique using the initial KL divergence values. Actually, the proposed method achieved a much higher recognition rate of 91.5% than that of 83.7% obtained by using the initial KL divergence values.

Finally, we compare the proposed method with the Levenberg-Marquardt method as applied to the same image matching problem explained in this section in terms of both the ability of minimization and the computational cost.

Regarding the computational complexity, both the proposed method and the Levenberg-Marquardt method have a time complexity of $O(m^2n^2)$, where an image has a total of $m \times n$ pixels. However, the Levenberg-Marquardt method needs to evaluate the first and second derivatives of (9), which imposes a heavy computational burden on the method.

Table I shows comparisons between the proposed method and the Levenberg-Marquardt method.

Table I
COMPARISONS BETWEEN THE PROPOSED METHOD AND THE
LEVENBERG-MARQUARDT METHOD.

Comparison items	Proposed method	LM method
Ave. of final KL divergence: Correct	0.223	0.412
Ave. of final KL divergence: Rival	0.416	0.630
Computational cost	1.00	3.17

From Table I, it is first found that the average of the final KL divergence values of the proposed method is much smaller than that of the Levenberg-Marquardt method, which demonstrates the superiority of the proposed method over the Levenberg-Marquardt method in KL divergence minimization ability. Also, it is found that the proposed method can be performed at fairly less computational cost than the Levenberg-Marquardt method.

VI. CONCLUSION

This paper proposed a new, promising technique of affine-invariant image matching via KL divergence minimization. In particular, we devised an accelerated iterative method specially adapted to the KL divergence minimization problem through effective linear approximation.

Recognition experiments using the handwritten numeral database IPTP CDROM1B showed that the proposed method achieved a recognition rate of 91.5% at suppressed computational cost, while the general-purpose Levenberg-Marquardt method took much more time but failed to gain a sufficient decrease in KL divergence values.

Future work is to greatly reduce the computational cost of the proposed method in order to apply this technique to an actual, small sample-size recognition task such as recognition of camera-based character images in natural scenes, where we have only a limited quantity of training data and cannot utilize statistical techniques.

REFERENCES

[1] C.-L. Liu, K. Nakashima, H. Sako, and H. Fujisawa. "Handwritten digit recognition: benchmarking of state-of-the-art techniques". *Pattern Recognition*, 36:2271–2285, 2003.

[2] M. Rebow, C. K. I. Williams, and G. E. Hinton. "Using generative models for handwritten digit recognition". *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-18:592–606, 1996.

[3] A. K. Jain and D. Zongker. "Representation and recognition of handwritten digits using deformable templates". *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-19:1386–1390, 1997.

[4] M. 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*, pages 39–43, Seattle, Sept. 2001.

[5] P. Simard, Y. LeCun, and J. Denker. "Efficient pattern recognition using a new transformation distance". *Advances in Neural Information Processing Systems*, 5:50–58, 1993.

[6] 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.*, PAMI-23:384–395, 2001.

[7] T. Wakahara and Y. Yamashita. "Multi-template GAT/PAT correlation for character recognition with a limited quantity of data". *Proc. of Twentieth Int. Conf. on Pattern Recognition*, pages 2873–2876, Istanbul, Aug. 2010.

[8] P. Viola and W. A. Wells. "Alignment by maximization of mutual information". *International Journal of Computer Vision*, 24:137–154, 1997.

[9] B. Zitová and J. Flusser. "Image registration methods: a survey". *Image and Vision Computing*, 21:977–1000, 2003.

[10] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[11] Mathematical Society of Japan and K. Ito. *Encyclopedic Dictionary of Mathematics*. Second Edition, the MIT Press, 1987.

[12] 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.

[13] M. Shi, Y. Fujisawa, T. Wakabayashi, and F. Kimura. "Handwritten numeral recognition using gradient and curvature of gray scale image". *Pattern Recognition*, 35:2051–2059, 2002.

[14] R. G. Casey. "Moment normalization of handprinted characters". *IBM J. Res. Develop.*, 14:548–557, 1970.

[15] T. M. Ha and H. Bunke. "Off-line, handwritten numeral recognition by perturbation method". *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-19:535–539, 1997.