Handwritten character recognition using elastic matching based on a class-dependent deformation model

Seiichi Uchida and Hiroaki Sakoe
Faculty of Information Science and Electrical Engineering, Kyushu University
6-10-1 Hakozaki, Higashi-ku, Fukuoka-shi, 812-8581 Japan
{uchida,sakoe}@is.kyushu-u.ac.jp

Abstract

For handwritten character recognition, a new elastic image matching (EM) technique based on a class-dependent deformation model is proposed. In the deformation model, any deformation of a class is described by a linear combination of eigen-deformations, which are intrinsic deformation directions of the class. The eigen-deformations can be estimated statistically from the actual deformations of handwritten characters. Experimental results show that the proposed technique can attain higher recognition rates than conventional EM techniques based on class-independent deformation models. The results also show the superiority of the proposed technique over those conventional EM techniques in computational efficiency.

1. Introduction

Elastic image matching (EM) techniques have been employed in handwritten character recognition tasks. The purpose of EM is to find the optimal fitting between a pair of character image patterns. The distance between the patterns under the optimal fitting will be invariant to their deformations and thus favorable for deformation-tolerant recognition systems.

The fitting ability (i.e., the range of compensable deformations) depends on the *deformation model* of the EM. The deformation models of conventional EM techniques are class-independent. Namely, all classes are assumed to have the same deformation tendency. It seems, however, more natural to assume that each class has a intrinsic deformation tendency. For example, the two vertical strokes of "M" are often tilted to be closer, whereas those of "H" are not (Fig. 1). This simple example reveals the limitation of the conventional class-independent models.

In this paper, a new EM technique based on a class-dependent deformation model is proposed. In this model any deformation of a class is described by a weighted linear combination of intrinsic deformation directions, called *eigen-deformations*, of the class. The eigen-deformations can be estimated statistically and automatically from the actual deformations of handwritten characters.

The fitting problem of the proposed EM technique is a nonlinear optimization problem and hard to solve directly. This is due to the weights at the linear combination, which are the control variables to be optimized, are contained in an image function. Thus, the approximation scheme employed in the tangent distance (TD) method by Simard et al. [4]-[6] are exploited to solve the problem.

The original idea of the eigen-deformations can be found in the Point Distribution Models (PDM) proposed by Cootes et al. [1]. Uchida et al. [7] have extended the PDM to deal with fully 2D deformations and have used it in an EM-based handwritten character recognition system. In their system, the eigen-deformations are used indirectly, that is, they are used in a posterior evaluation of the results of the EM based on a class-*in*dependent model. Contrary to this, in the proposed technique the eigen-deformations are directly embedded into a deformation model. Experimental results will show the advantageous characteristics of this novel and straightforward use of the eigen-deformations.

2. The Proposed Elastic Matching Technique

2.1. Problem Formulation

In general, *parametric* EM techniques are formulated as the following minimization problem

$$d(\boldsymbol{P}_c, \boldsymbol{E}) = \min_{\alpha} \|s(\alpha, \boldsymbol{P}_c) - \boldsymbol{E}\| \tag{1}$$

where $\boldsymbol{P}_c = \{P_c(x,y)\}$ is the reference (i.e., prototype) character image of the class c and $\boldsymbol{E} = \{E(x,y)\}$ is an input character image, where $x, y = 1, 2, \ldots, I$. The function $s(\alpha, \boldsymbol{P}_c)$ is a deformation model and provides deformed images of $\boldsymbol{P}_c$ according to M control parameters $\alpha = \{\alpha_1, \ldots, \alpha_m, \ldots, \alpha_M\}$. If this deformation model covers all deformations of the class c, the minimized distance $d(\boldsymbol{P}_c, \boldsymbol{E})$ will be a deformation-invariant distance between $\boldsymbol{P}_c$ and $\boldsymbol{E}$ and thus favorable for deformation-tolerant recognition systems.

Figure 1. Class-dependent deformation tendency. The counter-tilting of two vertical strokes can be easily found in "M" (left) but not in "H" (right).

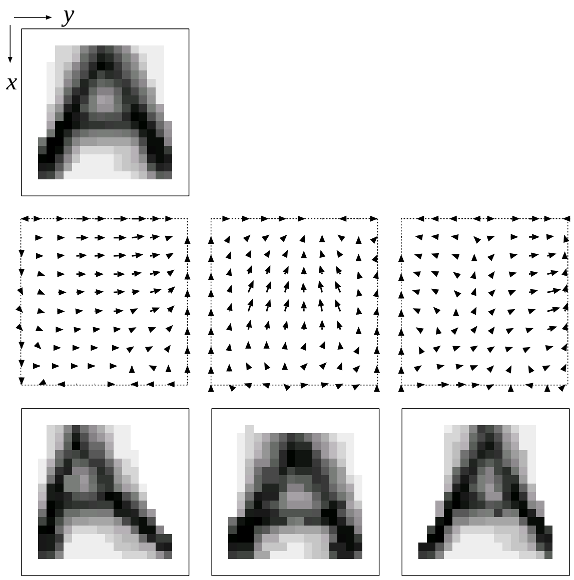


Figure 2. The reference pattern P_c of the class "A" (top), the first three eigen-deformations $\boldsymbol{u}_{c,1} = \{(X_{c,1}(x,y), Y_{c,1}(x,y))\}$, $\boldsymbol{u}_{c,2}$, and $\boldsymbol{u}_{c,3}$ of the class (middle), and the reference patterns deformed by these eigen-deformations, i.e., $\{P_c(x + X_{c,m}(x,y), y + Y_{c,m}(x,y))\}$ (bottom).

In this paper, a new parametric EM technique based on the following *class-dependent* deformation model:

$$
\begin{aligned}
& s(\alpha, P_c(x,y)) \\
& \quad = \quad P_c\Big(x + \sum_{m=1}^{M} \alpha_m X_{c,m}(x,y), \\
& \qquad\qquad y + \sum_{m=1}^{M} \alpha_m Y_{c,m}(x,y)\Big)\Big), \qquad (2)
\end{aligned}
$$

where $(X_{c,m}(x,y), Y_{c,m}(x,y))$ is the (x,y)th element of the mth *eigen-deformation* $\boldsymbol{u}_{c,m}$ of the class c. The eigen-deformation $\boldsymbol{u}_{c,m}$ is a $2I^2$-dimensional vector and represents the mth intrinsic deformation direction of the class. The element $X_{c,m}(x,y)$ and $Y_{c,m}(x,y)$ are the vertical and the horizontal displacements at the position (x,y), respectively. In the deformation model (2), any deformation of the reference $\boldsymbol{P}_c$ is expressed by a weighted linear combination of the M eigen-deformations. As we will see in Section 2.4, the eigen-deformations can be estimated automatically from actual handwritten character samples.

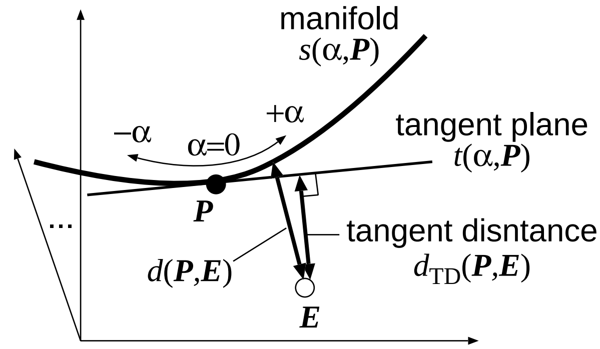


Figure 3. Manifold $s(\alpha, \boldsymbol{P})$, its tangent plane $t(\alpha, \boldsymbol{P})$, and tangent distance [4]-[6].

Figure 2 shows the first three eigen-deformations of the class "A". (The displacement $(X_{c,m}(x,y), Y_{c,m}(x,y))$ is represented by an arrow at each pixel (x,y).) The eigen-deformations are ordered by their frequencies, and thus this figure shows that the global slant transformation appeared as $\boldsymbol{u}_{c,1}$ is the most frequent deformation of the class "A".

2.2. Optimization in Tangent Distance Framework

The optimization problem (1) with the deformation model (2) is hard to solve directly, because the parameters α to be optimized are contained in the nonlinear 2D-1D function $\boldsymbol{P}_c$. Thus, some approximation is required to solve the optimization problem. In the following, the subscript c is dropped for notational simplicity, whenever there is no confusion.

Here, we employ the approximation scheme used in the tangent distance (TD) method [4]-[6], where it is considered that an M-dimensional manifold is formed by the the model $s(\alpha, \boldsymbol{P})$ in an I^2-dimensional pattern space. Figure 3 illustrates the manifold at $M = 1$. In this case, the manifold is reduced to a curve controlled by α_1. Note that the manifold passes the original reference pattern $\boldsymbol{P}$ at $\alpha = 0$.

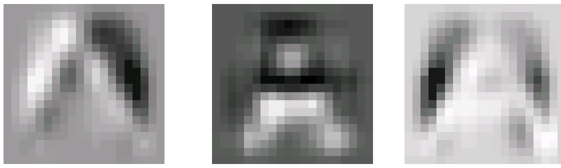
As shown in Fig. 3, the distance $d(\boldsymbol{P}, \boldsymbol{E})$ can be approximated by the following TD

$$
d_{\mathrm{TD}}(\boldsymbol{P}, \boldsymbol{E}) = \min_{\alpha} \| t(\alpha, \boldsymbol{P}) - \boldsymbol{E} \| , \qquad (3)
$$

where $t(\alpha, \boldsymbol{P})$ is the tangent plane of the manifold $s(\alpha, \boldsymbol{P})$ at $\alpha = 0$. This tangent plane is obtained by the first-order Taylor expansion of $s(\alpha, \boldsymbol{P})$ around $\alpha = 0$, i.e.,

$$
t(\alpha, \boldsymbol{P}) = \boldsymbol{P} + \sum_{m=1}^{M} \alpha_m \left[\frac{\partial s(\alpha, \boldsymbol{P})}{\partial \alpha_m} \right]_{\alpha=0} . \qquad (4)
$$

With the proposed deformation model (2), the tangent plane



φ1 φ2 φ3

Figure 4. The first three tangent vectors of the class "A". (Their values were normalized to be displayed as gray-scale images.)

(4) becomes

$$t(\alpha, \boldsymbol{P}) = \boldsymbol{P} + \sum_{m=1}^{M} \alpha_m \phi_m, \tag{5}$$

where ϕ_m is the tangent vector obtained as

$$\phi_m(x, y) = P_x(x, y)X_m(x, y) + P_y(x, y)Y_m(x, y), \tag{6}$$

where $P_x(x, y)$ and $P_y(x, y)$ are x- and y-derivatives of $P(x, y)$, respectively. Since the tangent vector ϕ_m can also be considered as an $I \times I$ image, Equation (5) indicates that any deformed reference pattern can be approximated by a simple linear combination of M images $\phi_1, \ldots, \phi_M$ and its original image $\boldsymbol{P}$. Figure 4 shows the first three tangent vectors of the class "A" obtained from the eigen-deformations of Fig.2.

Substituting (5) into (3), the approximated solution α can be obtained as

$$\alpha = \Phi^{-1}\lambda \tag{7}$$

where Φ and λ are an $M \times M$ matrix and an M-dimensional vector respectively. Their elements $\Phi_{m,m'}$ and λ_m $(m, m' = 1, \ldots, M)$ are

$$\Phi_{m,m'} = \sum_{x,y} \phi_m(x, y)\phi_{m'}(x, y),$$

and

$$\lambda_m = \sum_{x,y} \phi_m(x, y)(E(x, y) - P(x, y)).$$

The above TD approximation has the advantage that a closed-form solution can be obtained as (7). The TD approximation also has the disadvantage that the approximation may be accurate only around $\alpha = 0$, i.e., accurate for relatively small deformations.

2.3. Implementation Techniques and Computational Complexity

A sophisticated way for obtaining the derivatives P_x and P_y is to use the blurred image $P(x, y) * G_\sigma(x, y)$ instead of $P(x, y)$ [4]-[6], where $G_\sigma(x, y)$ is the Gaussian function. Then,

$$P_x(x, y) \sim \frac{\partial P(x, y) * G_\sigma(x, y)}{\partial x} = P(x, y) * \frac{\partial G_\sigma(x, y)}{\partial x}.$$

In our experiments, the standard deviation σ was fixed at 1.25 by a preliminary experiment. (As remarked in [5], this value was not critical. In fact, similar results were obtained with $\sigma \in [0.75, 1.5]$ in our experiment.)

The computational complexity to obtain $d_{\mathrm{TD}}(\boldsymbol{P}, \boldsymbol{E})$ is about $O(MI^2 + M^2)$. In practice, the number of parameters M can be a very small value (~ 3) as we shall see in Section 3. Consequently, the complexity becomes comparative to or far less than that of conventional EM techniques. Note that the matrix Φ^{-1} and $\phi_m(x, y)$ are independent of the input character $\boldsymbol{E}$ and therefore they can be calculated and stored in advance.

2.4. Estimation of Eigen-Deformations

The remaining problem is how to estimate the eigen-deformations $\{\boldsymbol{u}_{c,m}\}$ for each class c. As proposed in [7], this estimation can be done automatically using some conventional (i.e., class-independent model-based) EM technique and principal component analysis (PCA). The two-step procedure for the estimation is outlined as follows. (For more details, see [7].)

Step 1: Collection of actual deformations
Using some conventional EM, $\boldsymbol{P}_c$ is fitted to each of N training samples $\boldsymbol{T}_{c,n}(n = 1, \ldots, N)$. Then the deformation of $\boldsymbol{T}_{c,n}$ is detected from the optimized fitting (i.e., pixel-to-pixel correspondence) between two image patterns as a $2I^2$ (or less)-dimensional displacement vector.

Step 2: Estimation using PCA
The eigen-deformations $\{\boldsymbol{u}_{c,m}\}$ can be estimated by applying PCA to the collected displacement vectors of the class c. Specifically, the eigen-deformations are obtained as the eigen-vectors of the covariance matrix of the N displacement vectors. The eigen-vector which has the largest eigenvalue becomes the first eigen-deformation $\boldsymbol{u}_{c,1}$.

The conventional EM technique employed in Step 1 should have accuracy and flexibility enough to detect the exact deformations of $\boldsymbol{T}_{c,n}$. Such EM technique often has two drawbacks for its use in recognition systems, that is, (i) it requires numerous computations, and (ii) it has the high risk of over-fitting, which is the phenomenon that an image pattern of a class is closely fitted to another image pattern of a different class. These drawbacks, however, do not matter in the present framework, because the conventional EM is only used to match of two patterns of the same class in training process and not used in recognition process. In the experiment of Section 3, a non-parametric EM technique called piecewise-linear two-dimensional warping (PL2DW)

[8] was employed, where the pixel-correspondence is directly optimized by a combinatorial optimization scheme.

3. Experimental Results

3.1. Database

Recognition experiments was conducted on 26 (classes) $\times$ 1100 isolated handwritten English uppercase character images from the standard character database ETL6, which is commonly used mainly in Japan. Each sample was normalized linearly so that its character region became 16×16. The margin of 2 pixels was then added around the normalized sample; thus $I = 20$. After this preprocessing, the first 100 samples of each class were simply averaged to create one reference pattern $\boldsymbol{P}_c$. The next 500 samples were used as training samples $\boldsymbol{T}_{c,n}$ in the estimation of the eigen-deformations. The remaining 500 samples ($13000 = 26 \times 500$ samples in total) were used as test samples $\boldsymbol{E}$ in the following recognition experiment.

The pixel feature ($P_c(x, y)$ and $E(x, y)$) was a five dimensional vector which consists of four-dimensional directional features [2] and one-dimensional gray-level feature. The modification of the proposed technique for the use of this high-dimensional pixel feature is easy and omitted here.

3.2. Recognition Results

Figure 5 shows the recognition rates attained by the proposed EM as a function of the number of the eigen-deformations M. The highest recognition rate is 99.21% and it is 1.12% (145 samples) improvement from the simple rigid matching.

The recognition rates are saturated fast at $M = 3$. This fact shows that the proposed deformation model (2) can cover most of actual deformations with a few parameters. Furthermore, since the amount of the computations required by the proposed EM is relative to M, this fact also shows the high computational efficiency of the proposed technique.

3.3. Comparison to EM Techniques Based on Class-Independent Deformation Models

In order to ensure the effectiveness of the class-dependent deformation model of (2), the proposed technique was experimentally compared to two conventional EM techniques based on class-independent deformation models. The comparison is done on their recognition scores and computational efficiency.

Comparison 1:
The first class-independent EM technique is the original TD-based EM [4]-[6], where the following affine deformation model is employed in all classes

$$s(\alpha, P_c(x, y)) \\ = P_c(x + \alpha_1 x + \alpha_2 y + \alpha_3,\ y + \alpha_4 x + \alpha_5 y + \alpha_6) \quad (8)$$

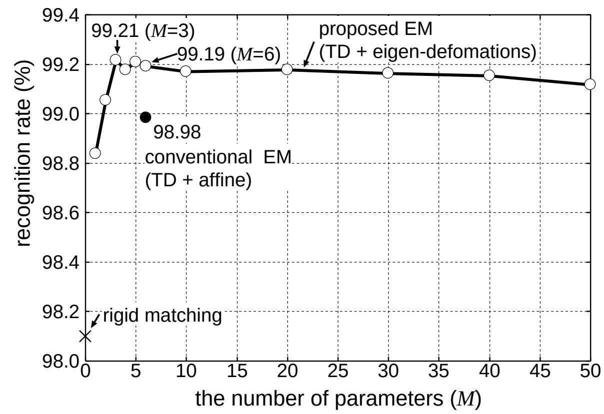


Figure 5. Recognition rates attained by the proposed EM and the affine-based EM.

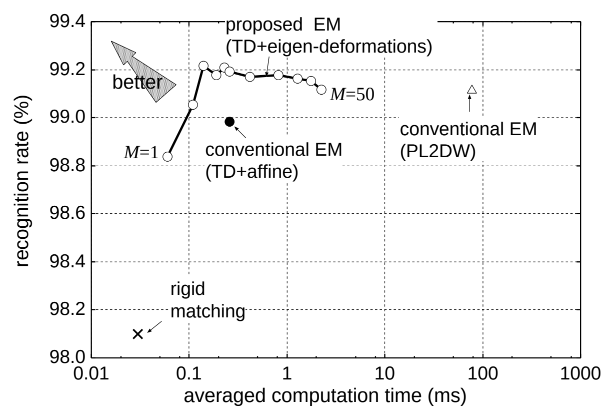


Figure 6. Relation between computation time (ms) and recognition rate (%).

instead of the class-dependent model (2). This model has six parameters (i.e., $M = 6$).

As shown in Fig. 5, its recognition rate was 98.98% and lower than 99.21% of the proposed EM at $M = 3$. This fact shows that the deformation model (2) can cover the actual deformations more compactly than the commonly used affine deformation model. In addition, it is shown that the proposed deformation model can provide more computationally efficient EM than the affine deformation model.

Comparison 2:
The second class-independent EM technique is PL2DW [8], which was used for the estimation of the eigen-deformations as noted in Section 2.4. PL2DW is a non-parametric EM technique, that is, the displacement of each pixel in $\boldsymbol{P}_c$ is directly optimized as a variable in a costly 2D combinatorial optimization scheme.

The best recognition rate of PL2DW was 99.12% and slightly lower than 99.22% of the proposed EM. As suggested in 2.4, this inferiority will be due to over-fitting which results from the high and class-independent flexibility of PL2DW.

On computational efficiency, their difference becomes more clear. In fact, the proposed EM technique is far more computationally efficient than PL2DW. Figure 6 plots the relation of their recognition rates and computation times for one matching required by a PC (CPU: Xeon 2.0GHz). This result shows that the proposed EM only requires about 1/500∼ 1/1000 computation time of PL2DW for almost the same (or higher) recognition rates.

4. Discussions and Future Work

4.1. Another Usage of Eigen-Deformations in EM

Another usage of the eigen-deformations has been proposed in [7], where the eigen-deformations are used *indirectly* in a EM-based recognition system. Specifically, the eigen-deformations are used in a posterior evaluation of the result of some conventional (i.e., class-independent) EM. From the viewpoint of recognition accuracy, this approach will be good enough as reported in [7]. The approach, however, is not computationally efficient because highly-flexible conventional EM (such as PL2DW) should be performed with large computational burden in advance to the posterior evaluation. Furthermore, the conventional EM will waste the computations for useless deformation candidates (e.g., the counter-tilting deformation of "H" illustrated in Fig. 1). Such useless candidates are totally excluded in the proposed EM by embedding the eigen-deformations directly into the deformation model.

4.2. Another Optimization Strategy

The linear approximation is a good strategy for solving the nonlinear optimization problem (1) with less computations, but has the side effect that the deformations compensable by the model (2) are limited to relatively small ones. Thus, iterative algorithms can be an alternative strategies for solving the problem because the iterative algorithms are free from the linear approximation. Wakahara et al. [9] have proposed such an iterative algorithm for EM based on the affine deformation model (8). This algorithm can be extended to deal with the class-dependent deformation tendencies by replacing its affine deformation model by our class-dependent deformation model (2).

4.3. Relation with Subspace Methods

The tangent approximation (5) is similar to the subspace methods [3] in the sense that in the both methods a deformed pattern is approximated by a linear combination of basis images derived from training patterns using PCA (,or KLT). However, there is an essential difference between these two methods; the basis images of the subspace methods are orthogonal in the original pattern space, whereas those of the proposed EM, i.e., ϕ_m are not. Their experimental comparison is left as future work.

5. Conclusion

For handwritten character recognition, a novel elastic image matching (EM) technique based on a class-dependent deformation model was proposed. In the model, any deformation of the class is represented by a linear combination of the eigen-deformations, which are intrinsic deformation directions of the class. In order to solve the nonlinear optimization problem with respect to the model, tangent distance approximation was employed. Experimental results showed that the recognition system using the pattern distance provided by the proposed EM technique can attain higher recognition rates than conventional EM techniques based on class-independent deformation models. The results also showed the superiority of the proposed technique over those conventional techniques in computational efficiency.

Acknowledgement: The authors thank the Electrotechnical Laboratory for providing us ETL6. This work was supported in part by The Ministry of Education, Culture, Sports, Science and Technology in Japan under a Grant-in-Aid for Scientific Research No.14780293.

References

[1] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models - their training and application," *Comput. Vis. Image Und.*, 61(1):38–59, 1995.

[2] S. Mori, K. Yamamoto, and M. Yasuda, "Research on machine recognition of handprinted characters," *IEEE Trans. PAMI*, 6(4):386–405, 1984.

[3] E. Oja, Subspace methods of pattern recognition, Research Studies Press Ltd., 1983.

[4] P. Simard, Y. Le Cun, J. Denker, and B. Victorri, "An efficient algorithm for learning invariances in adaptive classifier," *Proc. 11th ICPR*, 2:651–655, 1992.

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

[6] P. Simard, Y. Le Cun, and J. Denker, "Memory-based character recognition using a transformation invariant metric,"*Proc. 12th ICPR*, 2:262–267, 1994.

[7] S. Uchida and H. Sakoe, " Eigen-deformations for elastic matching based handwritten character recognition," *Pattern Recognition*, to appear, 2003.

[8] S. Uchida, and H. Sakoe, "Piecewise linear two-dimensional warping," *Proc. 15th ICPR*, 3:538–541, 2000.

[9] T. Wakahara, Y. Kimura, and A. Tomono, "Affine-invariant recognition of gray-scale characters using global affine transformation correlation,"