Hierarchical Decomposition of Handwriting Deformation Vector Field for Improving Recognition Accuracy

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

Seiichi Uchida
Faculty of Information Science and Electrical Engineering
Kyushu University
744 Motooka, Nishi-ku, Fukuoka-shi, 819-0395 Japan
E-mail: uchida@ait.kyushu-u.ac.jp

***Abstract*—This paper addresses the problem of how to extract, describe, and evaluate handwriting deformation from the deterministic viewpoint for improving recognition accuracy. The key ideas are threefold. The first is to extract handwriting deformation vector field (DVF) between a pair of input and target images by 2D warping. The second is to hierarchically decompose the DVF by a parametric deformation model of global/local affine transformation, where local affine transformation is iteratively applied to the DVF by decreasing window sizes. The third is to accept only low-order deformation components as natural, within-class handwriting deformation. Experiments using the handwritten numeral database IPTP CDROM1B show that correlation-based matching absorbing components of global affine transformation and local affine transformation up to the 3rd order achieved a higher recognition rate of 92.1% than that of 87.0% obtained by original 2D warping.**

***Keywords*-handwriting deformation vector field; 2D warping; global/local affine transformation; character recognition;**

I. Introduction

To accomplish the aim of most accurate handwriting recognition we will adopt statistical or probabilistic pattern recognition techniques, including sophisticated discriminant functions, neural networks, and support vector machines or kernel methods [1], [2]. However, we can say that the problem of what is handwriting deformation remains unsolved in the sense that the statistical description of handwriting deformation in a high-dimensional feature space cannot deepen our real understanding of handwriting deformation.

From this viewpoint, 2D image elastic matching based on deformation models have been proposed [3]. The tangent distance [4] and global affine transformation (GAT) correlation method [5] tried to extract handwriting deformation by means of parametric linear transformation, but, could not deal with nonlinear distortion. On the other hand, 2D warping methods via dynamic programming (DP) [6] realized pointwise correspondence between input and target images using nonparametric, rather loose matching constraints, but, suffered from excessive, unnatural matching.

This paper proposes a new, promising technique to extract, describe, and evaluate linear/nonlinear handwriting deformation in a deterministic, parametric manner to improve recognition accuracy. The key ideas are threefold; generation of handwriting deformation vector field (DVF) using a 2D warping technique between a pair of input and target images, hierarchical decomposition of the DVF by a parametric deformation model of global/local affine transformation, and acceptance of only low-order deformation components as natural, within-class handwriting deformation. Experimental results using the handwritten numeral database IPTP CDROM1B show that the proposed method substantially improves recognition accuracy by discriminating natural handwriting deformation from unnatural one via hierarchical decomposition of the DVF.

II. Generation of handwriting DVF by 2D warping

We deal with the problem of optimal matching between a pair of input and target images in grayscale.

Here, we denote input and target images by $\boldsymbol{F} = \{\, f(\boldsymbol{r})\,\}$ and $\boldsymbol{G} = \{\, g(\boldsymbol{r})\,\}$, respectively, where $\boldsymbol{r} = (x, y)^t, (1 \leq x \leq M, 1 \leq y \leq N)$, is a loci vector in a 2D image plane, and $f(\boldsymbol{r})$ and $g(\boldsymbol{r})$ denote grayscale values at $\boldsymbol{r}$.

To generate handwriting deformation vector field (DVF) between $\boldsymbol{F}$ and $\boldsymbol{G}$ we have to solve the problem of determining a vector mapping function $\tau(\boldsymbol{r})$ that specifies a pointwise correspondence between $f(\boldsymbol{r})$ and $g(\tau(\boldsymbol{r}))$. As a result, a deformation vector $\boldsymbol{d}(\boldsymbol{r})$ for $\boldsymbol{F}$ at $\boldsymbol{r}$ is defined by

$$\boldsymbol{d}(\boldsymbol{r}) \equiv \tau(\boldsymbol{r}) - \boldsymbol{r}. \tag{1}$$

The objective function $\Psi(\tau \mid \boldsymbol{F}, \boldsymbol{G})$ for determining an optimal mapping function $\tau(\boldsymbol{r})$ is as follows.

$$\Psi(\tau \mid \boldsymbol{F}, \boldsymbol{G}) = \sum\nolimits_{\boldsymbol{r}} \{|f(\boldsymbol{r}) - g(\tau(\boldsymbol{r}))| \quad + \\ \|\nabla f(\boldsymbol{r}) - \nabla g(\tau(\boldsymbol{r}))\|\} \;\longrightarrow\; \text{min for } \tau. \tag{2}$$

To solve this optimization problem we adopt DP-based piecewise linear 2D warping (PL2DW) method [6]. It is to be noted that dynamic programming (DP) guarantees global optimization of 2D warping.

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.459

Fig. 1 shows two typical versions of 2D warping. Fig. 1(b) corresponds to the above-mentioned PL2DW.

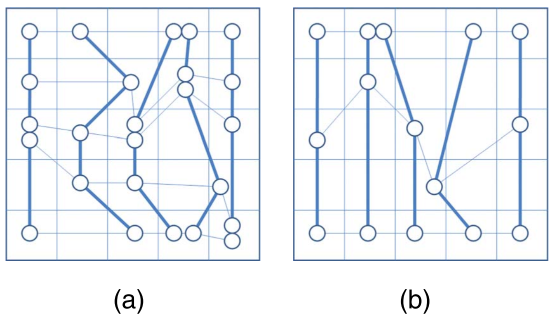

Figure 1. Typical versions of 2D warping. (a) Approximate "rubber sheet". (b) PL2DW.

In Fig. 1(a), the mappings of 4-adjacent pixels are mutually constrained by monotonicity and continuity that approximates ideal "rubber-sheet" matching. However, the optimization of this version is an NP-hard problem. On the other hand, Fig. 1(b) has been proposed as computationally tractable version of Fig. 1(a) by piecewise linear approximation. Namely, each column of $\boldsymbol{F}$ is fitted to $\boldsymbol{G}$ as a broken line with one corner, called pivot. As a result, the pointwise correspondence between $\boldsymbol{F}$ and $\boldsymbol{G}$ can be determined by linear interpolation except for the pivot and boundary.

PL2DW can compensate for fully two-dimensional deformation although it adopts piecewise linear approximation. In this sense, PL2DW is considered a very powerful tool for generating the handwriting DVF. Conversely speaking, PL2DW is likely to suffer from excessive, unnatural matching.

III. Hierarchical decomposition of DVF by global/local affine transformation

In this section we introduce hierarchical decomposition of the handwriting DVF by a parametric deformation model of global/local affine transformation and propose to accept only low-order deformation components as natural, within-class handwriting deformation.

A. Extraction of global affine transformation component from DVF

The aim of extracting global affine transformation component from DVF is to approximate the DVF by optimal affine parameters as closely as possible. Here, we denote global affine transformation by a 2×2 matrix, A, representing rotation, scale-change, and shearing, and a 2D translation vector $\boldsymbol{b}$:

$$\mathrm{A} = \begin{pmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}. \tag{3}$$

Next, we introduce the objective function $\Phi(\mathrm{A}, \boldsymbol{b})$ to determine optimal affine parameters given by

$$\begin{aligned} \Phi(\mathrm{A}, \boldsymbol{b}) &= \sum\nolimits_{\boldsymbol{r}} \| \boldsymbol{d}(\boldsymbol{r}) - (\mathrm{A}\ \boldsymbol{r} + \boldsymbol{b} - \boldsymbol{r}) \|^2 \\ &= \sum\nolimits_{\boldsymbol{r}} \| \tau(\boldsymbol{r}) - (\mathrm{A}\ \boldsymbol{r} + \boldsymbol{b}) \|^2 \\ &\longrightarrow \ \text{min for A and } \boldsymbol{b}. \end{aligned} \tag{4}$$

By setting the derivatives of $\Phi(\mathrm{A}, \boldsymbol{b})$ with respect to each of six unknown parameters, $a_{00}, a_{01}, a_{10}, a_{11}, b_0$, and b_1, equal to zero, we obtain six simultaneous linear equations. We can solve these simultaneous linear equations by conventional techniques such as Gaussian elimination [7].

We know that the optimization problem of Eq. (4) is exactly equivalent to the least-squares criterion and its solution is easy-to-follow.

Here, we denote the determined optimal affine transformation by $\mathrm{A}^{\mathrm{GAT}}$ and $\boldsymbol{b}^{\mathrm{GAT}}$.

Then, the decomposition of the DVF is given by

$$\begin{aligned} \boldsymbol{d}(\boldsymbol{r}) &= (\mathrm{A}^{\mathrm{GAT}}\ \boldsymbol{r} + \boldsymbol{b}^{\mathrm{GAT}} - \boldsymbol{r}) \ + \\ & \quad (\tau(\boldsymbol{r}) - (\mathrm{A}^{\mathrm{GAT}}\ \boldsymbol{r} + \boldsymbol{b}^{\mathrm{GAT}})) \\ &= \boldsymbol{d}^{\mathrm{GAT}}(\boldsymbol{r}) \ + \ \tilde{\boldsymbol{d}}_0(\boldsymbol{r}), \end{aligned} \tag{5}$$

where $\tilde{\boldsymbol{d}}_0(\boldsymbol{r})$ represents the residual component of the DVF after the optimal global affine transformation.

B. Expansion of residual DVF by local affine transformation

The expansion of the residual DVF by local affine transformation is given by

$$\begin{aligned} \boldsymbol{d}(\boldsymbol{r}) &= \boldsymbol{d}^{\mathrm{GAT}}(\boldsymbol{r}) + \sum_{k=1}^{K} \boldsymbol{d}_k^{\mathrm{LAT}}(\boldsymbol{r}) + \tilde{\boldsymbol{d}}_K(\boldsymbol{r}), \\ \boldsymbol{d}_k^{\mathrm{LAT}}(\boldsymbol{r}) &= \boldsymbol{s}_k(\boldsymbol{r}) - \boldsymbol{s}_{k-1}(\boldsymbol{r}), \\ \boldsymbol{s}_k(\boldsymbol{r}) &= \mathrm{A}_k(\boldsymbol{r})\ \boldsymbol{s}_{k-1}(\boldsymbol{r}) + \boldsymbol{b}_k(\boldsymbol{r}), \\ \boldsymbol{s}_0(\boldsymbol{r}) &= \mathrm{A}^{\mathrm{GAT}}\ \boldsymbol{r} + \boldsymbol{b}^{\mathrm{GAT}}, \end{aligned} \tag{6}$$

where $\tilde{\boldsymbol{d}}_K(\boldsymbol{r})$ denotes the residual component of the DVF after global affine transformation and a set of the subsequent K local affine transformations.

The objective function for determining the kth local affine transformation is given as follows.

$$\begin{aligned} &\Phi(\mathrm{A}_k(\boldsymbol{r}), \boldsymbol{b}_k(\boldsymbol{r}) \mid \theta_k) = \sum\nolimits_{\boldsymbol{r}'} w_{k-1}(\boldsymbol{r}'|\boldsymbol{r}) \ \times \\ &\left\| \tilde{\boldsymbol{d}}_{k-1}(\boldsymbol{r}') - (\mathrm{A}_k(\boldsymbol{r})\boldsymbol{s}_{k-1}(\boldsymbol{r}') + \boldsymbol{b}_k(\boldsymbol{r}) - \boldsymbol{s}_{k-1}(\boldsymbol{r}')) \right\|^2 \\ &\qquad \longrightarrow \text{min for } \mathrm{A}_k(\boldsymbol{r}), \boldsymbol{b}_k(\boldsymbol{r}) \quad \text{for } 1 \le k \le K, \\ &w_{k-1}(\boldsymbol{r}'|\boldsymbol{r}) = \exp\left(- \frac{\| \boldsymbol{s}_{k-1}(\boldsymbol{r}') - \boldsymbol{s}_{k-1}(\boldsymbol{r}) \|^2}{2\theta_k^2} \right). \end{aligned} \tag{7}$$

By introducing a Gaussian window function $w_{k-1}(\boldsymbol{r}'|\boldsymbol{r})$ around the point of $\boldsymbol{s}_{k-1}(\boldsymbol{r})$, optimal $\mathrm{A}_k(\boldsymbol{r})$ and $\boldsymbol{b}_k(\boldsymbol{r})$

are determined to approximate the residual DVF in the neighborhood of $\boldsymbol{s}_{k-1}(\boldsymbol{r})$ as closely as possible.

In particular, the parameter θ_k specifies the spread of the Gaussian window function around $\boldsymbol{s}_{k-1}(\boldsymbol{r})$, and controls the stiffness of matching by local affine transformation. Namely, the smaller the value of θ_k is, the softer the matching by local affine transformation is.

Now, we propose to decompose the DVF into a series of deformation components using successive local affine transformations with decreasing window sizes given by

$$\theta_k = \frac{\theta_1}{2^{k-1}} \qquad \text{for } k = 1, 2, \ldots, K. \tag{8}$$

Actually, we consider that the window size is no less than one. Hence, by setting the value of θ_1 at 2^p, we obtain a total of $(p+1)$ $(=K)$ window sizes.

Finally, it is to be noted that the optimization problem of Eq. (7) can also be easily solved according to the weighted least-squares method.

Fig. 2 illustrates a hierarchical decomposition of a deformation vector.

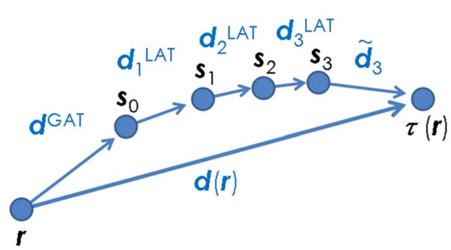

Figure 2. Hierarchical decomposition of a deformation vector.

C. Determination of natural handwriting deformation via DVF decomposition

Pointwise correspondence by 2D warping is subject to excessive matching due to rather loose constraints and a kind of too straightforward optimization. As a result, a naive 2D warping absorbs not only within-class but also between-class handwriting deformation, and deteriorates recognition accuracy.

From the above viewpoint, we propose to accept only low-order components of DVF decomposition as natural, within-class handwriting deformation.

According to the DVF decomposition formula of Eq. (6) we can generate the kth-order deformation-absorbed input image, $\boldsymbol{F}^*{}_k = \{ f_k^*(\boldsymbol{r}) \}$, given by

$$f_k^*(\boldsymbol{s}_k(\boldsymbol{r})) \equiv f(\boldsymbol{r}) \quad (0 \le k \le K). \tag{9}$$

Then, we can determine the optimum order $\hat{k}$ that absorbs only natural, within-class handwriting deformation for maximizing the recognition accuracy.

IV. Experimental Results

We use the handwritten numeral database IPTP CDROM1B [8]. This database contains binary images of handwritten digits divided into two groups of 17,985 samples for training and 17,916 samples for test.

As preprocessing, position and size normalization is conducted using 1st and 2nd moments. After being transformed into grayscale images by mean filtering, all images are downsized to 40×60 pixels. Hence, we have $M = 40$ and $N = 60$ in the notation used in Section II.

On the other hand, we generate a single target image per digit by averaging each category's training samples.

Then, we generate the DVF between each of 17,916 test samples and each target image by PL2DW featuring one pivot on each column with a warp range of 15 pixels.

Each DVF is decomposed into the series of deformation components by hierarchical application of global/local affine transformation. Here, we set the value of θ_1 at 2^5, that is, $K = 6$ in Eq. (8). As a result, we have a total of seven deformation components extracted from the DVF: one global affine transformation and six subsequent local affine transformations.

Fig. 3 shows examples of hierarchical decomposition of DVF.

Figure 3. Examples of hierarchical decomposition of DVF. (a) Input images. (b) Global affine transformation. (c) - (e) Successive local affine transformations with θ = 16, 8, and 4. (f) PL2DW. (g) Target images.

From Fig. 3, it is first found that PL2DW-superimposed input images of Fig. 3(f) are likely to suffer from unnatural, excessive matching between input and target images. This is mainly because matching constraints are rather loose although the DP-based PL2DW itself guarantees global optimization subject to those constraints. Also, it is clear that global affine transformation cannot compensate for nonlinear handwriting deformation as shown in Fig. 3(b). Furthermore, it is important to note that the successive local affine transformations should be interrupted at a moderately large local window size, e.g. eight, in order to avoid excessive matching.

As explained in Section III-C, we generate a total of seven kinds of deformation-absorbed input images, $\{\boldsymbol{F^*}_k\}(0 \leq k \leq 6)$ against each target image $\boldsymbol{G}$.

Now, we conduct recognition experiments using normalized cross-correlation values between GAT/LAT deformation-absorbed input images and target images.

Fig. 4 shows relations between recognition rates and absorption of global/local affine transformation components, where LATk $(1 \leq k \leq 6)$ specifies absorption of global affine transformation and subsequent local affine transformation up to the kth order. Also, ORG and PL2DW stand for correlation-based matching of original input images and PL2DW-superimposed input images against target images, respectively.

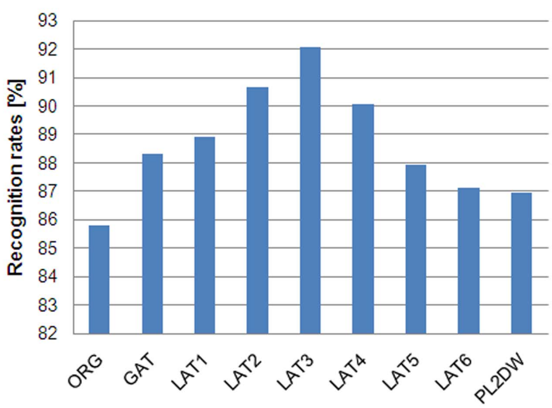

Figure 4. Relations between recognition rates and absorption of global/local affine transformation components.

From Fig. 4, it is first found that the use of GAT/LAT deformation-absorbed input images improves the recognition accuracy by comparing GAT, LATk $(1 \leq k \leq 6)$ with ORG. It is secondly found that LAT3 maximizes the recognition rate while PL2DW achieves no substantial improvement in recognition accuracy due to excessive matching. This fact means that the DVF components of global affine transformation and subsequent local affine transformation up to the 3rd order represent natural, within-class handwriting deformation and should be absorbed in matching to improve recognition accuracy.

Also, it is to be noted that the optimum order $\hat{k}$ that absorbs only natural, within-class handwriting deformation for maximizing the recognition accuracy is data-dependent.

From these results, we can say that the proposed method provides a most effective, powerful means for improving recognition accuracy by discriminating between natural, within-class handwriting deformation and unnatural one.

V. Conclusion

It is very interesting and still challenging to extract, describe, and evaluate handwriting deformation from not the statistical but the deterministic viewpoint.

This paper proposed one powerful solution; DP-based 2D warping generates the DVF between a pair of input and target images as a global optimization problem, and the successive global/local affine transformations decompose the DVF into linear/nonlinear deformation components in a truly parametric manner.

Experiments using the handwritten numeral database IPTP CDROM1B showed that absorption of the DVF components of global affine transformation and subsequent local affine transformation up to the 3rd order substantially improved recognition accuracy. We can say that the proposed method provides a promising clue for discriminating natural deformation components from unnatural ones.

Future work is to combine this deterministic method with statistical techniques in a cooperative manner.

References

[1] 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.

[2] 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.

[3] S. Uchida and H. Sakoe. "A survey of elastic matching techniques for handwritten character recognition". *IEICE Trans. Inf. & Syst.*, E88-D:1781–1798, 2005.

[4] 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.

[5] 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.

[6] 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.

[7] Mathematical Society of Japan. *Encyclopedic Dictionary of Mathematics*. MIT Press, Cambridge, MA., 1997.

[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.