Structurally Based Template Matching of On-line Handwritten Characters

Jakob Sternby
Centre for Mathematical Sciences
Sölvegatan 18, Box 118
S-221 00, Lund, Sweden

Abstract

A large share of the early work on on-line handwriting recognition involved structural and syntactical methods. These approaches were soon abandoned in favor of template matching and statistical methods due to the difficulty in defining reliable rules dealing with the large variability in on-line handwritten characters. However, any method for HWR utilize the structural information implicitly and one could argue that their success depends on how well this is done. This paper presents a novel template matching method, the Frame Deformation Energy (FDE) matching, that utilizes the explicit structure of the samples to model the non-linear global variations by a set of affine transformations through a structural reparameterization. Experiments on a large data set show that for single models the FDE, despite its ad hoc implementation in this paper, outperforms conventionally used template matching schemes such as DTW and Active Shape.

1 Introduction

Although template matching systems for single character recognition still exist in the form of various implementations of the DTW algorithm [15, 18], most of the focus has been shifted towards statistically based systems utilizing HMM and/or Neural Networks [1, 6, 7, 12]. Although dependent on the data sets used for experiments as well as the implementation, this shift by the general research community has probably been triggered by the higher recognition rates reported from such statistical systems during the past decade [2]. There are, however, merits of template matching systems as well. Since they do not require any type of training, they are well suited for incorporation of adaptivity as well as user customization [15]. Furthermore, the minimal distance concept seems to be something comparatively transparent to the human mind and some reports indicate that humans have a more intuitive understanding of errors made by a template matching scheme [10].

The simplest template matching scheme for on-line handwriting is to compare samples with the natural Euclidean distance in the space induced by their dimension as decided by the number of coordinates placed under an arclength parameterization of the handwritten curves. An arclength parameterization is, however, dependent upon the length of different curve segments thus causing a varied number of coordinates to be placed on corresponding curve segments of different samples of the same symbol. The

remedy to this problem is DTW which allows a warping path to connect coordinates to the closest coordinate of the matched curve and thus under some constraints divert from the pairwise sequential matching order. This type of DTW, however, does not attempt to solve issues of the variations of the internal structure of the segments and it has been shown previously that, especially for complex symbols, significant improvements can be made to the conventional DTW strategy [17].

In this paper we present a novel template matching method that we call Frame Deformation Energy (FDE) matching. This method assumes that a structural parameterization identifying a segment decomposition has been conducted on the handwritten samples. The matching is then conducted in three steps by calculating global transformation properties, the affine transformation of the constituting segments and finally obtaining some matching distance value for the remaining intermittent points, thus implicating a coarse to fine strategy for the matching procedure. We have constructed a simple system based on these ideas and tested it on a large data set. Comparisons are made both for single models as well as in a kNN context. Although still in the cradle of its development the new method FDE outperforms DTW for single models on the large data set used in experiments.

2 Structural Reparameterization with Core Points

Extracting points of significant structural importance is a subject that has been studied in detail for the purpose of segmenting cursive script in the past [11, 13]. A comprehensive summation of many of these techniques can be found in paper [9]. It has been realized by many that extreme points in the direction orthogonal to the writing direction contain a large portion of such interesting points for the on-line single character problem. We will therefore use this subset, that we call the **core point frame** of a sample, as a basis for our structural reparameterization. The curve between two successive points in the core point frame will be referred to as a **segment**.
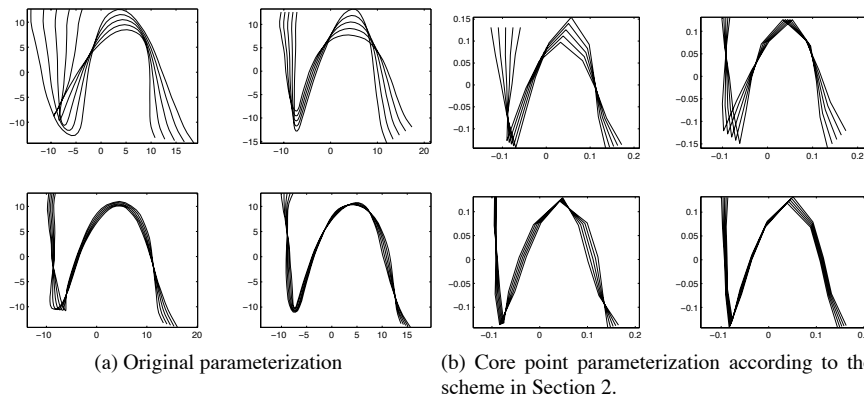

(a) Original parameterization   (b) Core point parameterization according to the scheme in Section 2.

Figure 1: The first four components of principal component analysis of one allograph of n, automatically separated by the allograph separation scheme of Section 2.1.

Before the extraction of the core point frame, every sample is centered and normalized. We also assume that they are roughly rotationally aligned so that the x-axis is aligned with the writing direction. Since we aim at removing redundance from the samples and minimize the number of points between the coordinates chosen in the core point frame we try to make a clever choice of a fixed number of intermittent points. For the English alphabet we have observed that the most complex curve, the *s*, is legible even if we limit the number of intermittent points to three and hence we have chosen to place this number of fixed points on every segment. Instead of just spacing the points evenly on the segment as we would if we were to use conventional arclength parameterization, we first try to find all points that have a significant curvature. This search is done recursively by picking any point with a diversion from the line between the start and end point that exceeds a threshold. If the number of curvature points chosen in this manner is less than our fixed number of points per segment we add points and try to space them as evenly as possible. We call the complete set of extracted points with the core point frame and intermittent points, the *core points* $C(X)$ of a handwritten character sample X. The first modes of some samples of allographs that have been structurally reparameterized in this way are shown in Figure 1. It clearly shows that our reparameterization has managed to model the movement of the core points independently from the statistical variations of the curvatures on each segment.

In the case of severe noise parallel to the *y*-axis there may be false points in the core point frame. Future research will therefore concentrate on the development of methods for assessing the validity of segments in an extracted core point frame. Since this problem may be viewed as a way of extracting a configuration of points of a certain shape from a larger context of points, we could possibly apply the Vector Context Matrix to solve the problem [14]. Further investigation of this issue is, however, beyond the scope of this paper. It is also evident from the experimental results that it is not a crucial issue since the results of automatic training and testing with the crude parameterization strategy presented here are highly competitive.

2.1   Allograph Separation

An important topic for handwriting recognition systems that utilize unsupervised training data is allograph separation. Normally this problem is referred to as clustering in the field of HWR, since the most common method of allograph separation is to perform clustering of the training samples [16]. For *k*NN, clustering is mainly performed to identify and exclude outliers as well as to reduce the data set used for matching. For single model matching the allograph separation and outlier removal plays a crucial role in the appearance of the obtained set of single models. With *single model matching* we mean that only one artificial sample, usually the cluster center, has been chosen as a template for every allograph. In this paper we have chosen a crude allograph separation scheme based on the structural reparameterization of Section 2. We define an allograph cluster sequence as the sequence of labeled core points, with labels $L \in \{N, S, n, s, m\}$. Here the $\{N, S, n, s\}$ points denote the local maxima *n* and minima *s* of the core point frame, further differentiated by capitalization depending on convexity. The intermittent points are labeled *m*. The first modes of two allograph clusters separated by this crude technique are shown in Figure 1.

The strategy explained above is not optimal in the sense that some allograph clusters may be very similar to others. Since this problem is not a topic of this paper the strategy

was sufficient for the experimental section of this paper. Outliers have not been removed in our implementation and this is also reflected in the reported recognition accuracy of the various systems implemented in Section 4. Outliers can be removed by removing clusters with few members as in paper [1] or by a voting strategy depending on internal distance calculations [15]. It could be interesting to evaluate the effect of such a removal on the different methods for single model matching (only cluster removal) as well as kNN matching (also removing individual samples from clusters).

3 The Frame Deformation Energy Model

One of the limitations of template matching techniques such as DTW lies in the fact that the normalization is global. Even though DTW is successful at enabling matching between samples of varying dimension it is still dependent on normalization and thereby also sensitive to non-linear variations. In other words a handwritten sample X is in general not only the result of global transformation of the reference template but also of individual transformations of each segment of the core point frame. One such way to depict the transformations of just the segments is to compute the thin-plate spline transformation of one core point frame to the other. The deformed grids for one in-class and one inter-class example of such transformations are depicted in Figure 2.
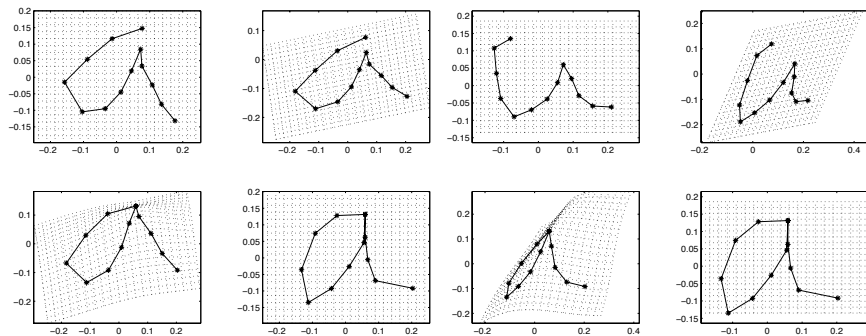
(a) An in-class example of bending one example of an a to another.

(b) An inter-class example of bending a sample of a u to an a

Figure 2: Bending the core point frame while leaving the parameterization fixed. The four figures in both cases display the original sample, the affine approximation to the target sample, the complete thin plate spline of the core point frame and finally the target sample.

These facts motivate the search for a method that tries to find both the local segment transformations as well as calculate the resulting distance to the transformed segments. In short we want to divide the matching process of a sample $X = \{x_i\}$ to a template (prototype) $P = \{p_i\}$ into three stages:

1. Find the best global linear transformation $A_P = \mathrm{argmin}_L \|P - LX\|$

2. Find the frame bending transformation $\mathbf{B}_P$, $p_i = \mathbf{B}_P(x_i)$, $\forall x_i, p_i$ in their respective core point frames

3. Calculate a distance value dependent on the transformations $\mathbf{A}_P$, $\mathbf{B}_P$ and the remaining difference $P - \mathbf{B}_P(\mathbf{A}_P(X))$

Analysis of samples of on-line handwritten characters clearly shows that in-class global transformations of handwritten characters are constrained linear transformations. There are no reflections and only limited rotation and skew. We define the frame bending transformation as the set of affine transformations identifying corresponding core point frames. Since we have left the parametrization of each segment fixed during these transformations we have actually obtained a kind of Bookstein coordinates on each individual segment [4]. An example of two samples being matched to each other in this way is shown in Figure 3.
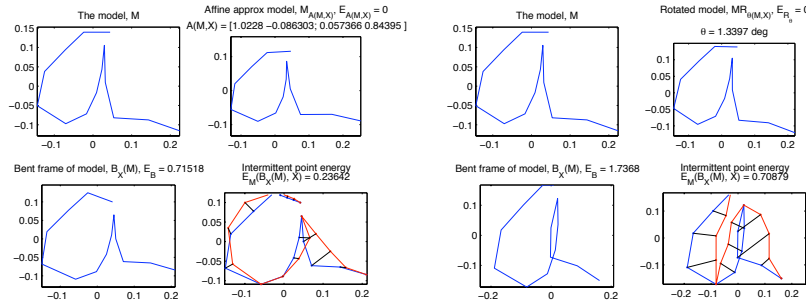
(a) A sample of an *a* matched the mean template of an allograph of *a*

(b) A sample of an *a* matched the mean template of an allograph of *n*

Figure 3: A sample of an *a* being matched to mean templates according to the scheme of Section 3.

3.1 Frame Energy Model Distance Measure

A popular method to achieve exact transformations between templates in pattern recognition is thin-plate splines. Although there have been successful applications of thin-plate splines to the character recognition problem in the past [3] it has obvious shortcomings. The main problem is that common variations in handwritten patterns involves points on the extension of a line being distributed on either side of the line causing folding of the thin-plate spline [5]. To counter this problem we introduce a much simpler energy model for the frame. We let each segment be modeled by a robust spring that is equally sensitive to compression and prolongation and let the connection between each segment be a coiled spring that is equally sensitive to torque in both directions. The most simple distance measure for the bending energy of the frame between a sample X and a template P of m core points, with frames $F_X = (f_X^1, \ldots, f_X^n)$ to $F_P = (f_P^1, \ldots, f_P^n)$ is then given by

$$E_{\mathrm{B}}(X,P) = k_x \sum_{i=1}^{n-1} (\|f_X^{i+1} - f_X^i\| - \|f_P^{i+1} - f_P^i\|)^2 +$$

$$k_a \sum_{i=1}^{n-2} \kappa_i^2 (\arg(f_X^{i+2} - f_X^{i+1}, f_X^{i+1} - f_X^i) - \arg(f_P^{i+2} - f_P^{i+1}, f_P^{i+1} - f_P^i))^2, \quad (1)$$

where k_x, k_a are the spring constants for the segment springs and the inter-segment springs respectively. $\kappa_i = \frac{\min(\|f_X^{i+2} - f_X^{i+1}\|, \|f_X^{i+1} - f_X^i\|)}{\pi}$ is a normalization constant to balance angle and length variations in the core point frame. For notational convenience we will imply a modula 2π for the subtraction operation of angles retrieved by the $\arg : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, 2\pi)$ operator. As described in the previous section the result of the bent frame $\mathrm{B}_{F_P}(F_X)$ is that $\|\mathrm{B}_{F_P}(F_X) - F_P\| = 0$, however, the intermittent points are just Bookstein coordinates in their respective surrounding segment and will generally not be identical.

From an implementation point of view the most simple way to model the energy of transforming points from one curve to the other is to to find a model that corresponds to the Euclidean measure. This is achieved by imagining that each of the intermittent points is attached to the corresponding point in the sample being matched by elastic strings. This induces an energy measure for the intermittent points by

$$E_M^{\mathrm{Euc}}(\mathrm{B}_P(\mathrm{A}_P(X)), P) = \sum_{j=1}^m k_j (\mathrm{B}_P(\mathrm{A}_P(x_j)) - p_j)^2, \quad (2)$$

where k_j is the spring constant for the string attached to core point j in P. Evidently setting $k_j = 1, j = 1, \ldots, m$ gives the square Euclidean distance of the bent frame transformed sample $\|\mathrm{B}_P(\mathrm{A}_P(X)) - P\|^2$.

Table 1: Results of k-NN matching on the MIT database.

k-Distance Measure	Original data	Reparameterized
1-Euclidean	86.3%	89.2%
3-Euclidean	86.5%	**89.4%**
1-D_R^{Euc}	N/A	87.4%

Above we have described methods to account for the two steps of frame bending and curve segment comparison. It is not entirely obvious how to fit a suitable penalization of global transformations into this. On one hand global transformations are natural variations of isolated handwritten character data and on the other some kind of penalization is necessary since the energy $E_{\mathrm{B}}(X,P)$ of (1) is invariant to global rotation. We have tried global transformations of just rotation and of the triple scale, rotation and skew. For the global parameters of scale (λ_x, λ_y), rotation θ and skew η we can use a distance measure similar to that of the bending energy by setting

$$E_{RSS}(X,P) = k_\lambda ((\frac{\lambda_x}{\lambda_y})^{t_\lambda} - 1)^2 + k_\theta (\frac{\mathrm{mod}\,(\theta, \pi)}{\pi})^2 + k_\eta (\frac{\mathrm{mod}\,(\eta, \pi)}{\pi})^2, \quad (3)$$

where t_λ is 1 $\lambda_x < \lambda_y$ and -1 otherwise. We will use the notation $E_R(X,P)$ for the global transformation energy of just the rotational component in (3).

Combining the distance components for global transformation, bending energy and curve segment into a weighted sum produces the following distance functions:

$$
\begin{aligned}
D_R^{Euc}(X,P) &= w_R E_R(X,P) &&+ w_{BE} E_B(A(X),P) \\
&&&+ w_M E_M^{Euc}(A(B(X)),P), &&(4) \\
D_{RSS}^{Euc}(X,P) &= w_{RSS} E_{RSS}(X,P) &&+ w_{BE} E_B(A(X),P) \\
&&&+ w_M E_M^{Euc}(A(B(X)),P). &&(5)
\end{aligned}
$$

We refer to matching with the distance measures in (4) or (5) as Frame Deformation Energy Matching (FDE).
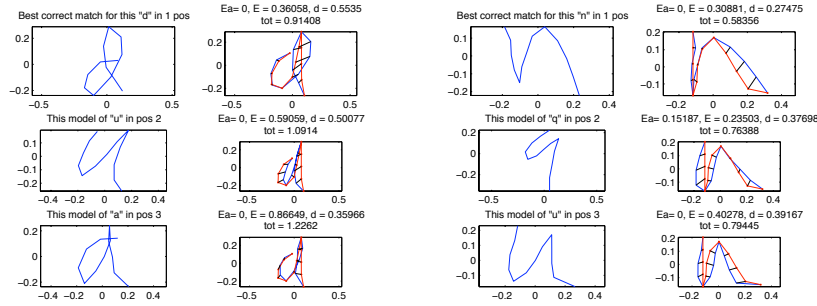
(a) A sample of the three best FDE matches of a sample of a *d*

(b) A sample of the three best FDE matches of a sample of a *n*

Figure 4: The best matches of two samples from the MIT-database

4 Experiments

We have conducted recognition experiments on the MIT single character database [8]. We chose the set of 37161 samples from the *w* section (single characters from words) as the test set and the 2825 samples from the *l* section as the training set. Both the test set and the training set were separated according to the labeling technique of Section 2.1 independently of whether the reparameterization was performed or not. The best matching models for two samples matched with single model FDE are shown in Figure 4.

For single models we compared our new template matching method FDE with DTW as well as a Gaussian Active Shape Model (AS). For the FDE and DTW methods we constructed a single model for each allograph as the mean of the samples belonging to that allograph class. For AS we built one model for each allograph class. As expected from the qualitative plots of the first modes, the recognition rate for AS increases significantly when we use the structural technique instead of conventional arclength parameterization. The FDE was implemented with the D_R^{Euc} measure in the most simple way by setting all

Table 2: Results of single model matching on the MIT database.

Method	Original data	Reparameterized
AS	77.2 %	82.4 %
DTW-mean	79.5%	-
FDE-mean	N/A	**82.9%**

the spring constants in (1), (2) and (3) to 1, as well as all of the weights in (4). Even with this simple ad hoc setting the FDE method outperforms AS and FDE for these settings as is seen in Table 2.

For some reason the FDE with the same parameter settings as in the single model case did not perform well here. It is possible that the FDE method may be more sensitive to outliers, which were not removed in the experiments, and that this has an impact on results when using all the samples in each allograph class for recognition.

5 Discussion and Conclusions

In this paper we have presented a novel template matching method, the Frame Energy Matching (FDE), which is based on a structural parameterization. The method outperforms both DTW and AS for single models even in a very basic ad hoc implementation. The FDE introduces a coarse to fine strategy for matching templates of handwritten characters and it is thus probable that it better handles the non-linearity of global variations present in handwriting data. Surprisingly the setting implemented in this paper gave slightly worse results for classification via the 1-nn rule. Possibly this is an indication that the flexible match of the FDE is more sensitive to outliers in data. We will continue to work with improvements to the new method. A natural next step is to investigate ways of optimizing all the various spring constants and weights. One can also try to use this framework with a statistical method. It might be even more efficient to view the problem in a probabilistic way by determining the class C with a model M_C that has the highest probability $P(C|A(M_C,X),B_X,B_X(M_C)-X)$. Since the novel FDE technique already at this early stage has shown great capacity for computationally efficient single models it will be especially useful in on-line cursive script systems based on segmentation graphs.

References

[1] C. Bahlmann and H. Burkhardt. The writer independent online handwriting recognition system *frog on hand* and cluster generative statistical dynamic time warping. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(3):299–310, March 2004.

[2] E. J. Bellegarda, J. R. Bellegarda, D. Nahamoo, and K. Nathan. A fast statistical mixture algorithm for on-line handwriting recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(12):1227–1233, 1994.

[3] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(24):509–522, 2002.

[4] F. L. Bookstein. Size and shape spaces for landmark data in two dimensions. *Statistical Science*, 1(2):181–242, 1986.

[5] A. P. Erikson and K. Åström. On the bijectivity of thin-plate splines. pages 109–112, Malmö, March 2005. SSBA.

[6] N. Gauthier, T. Artières, B. Dorizzi, and P. Gallinari. Strategies for combining on-line and off-line information in an on-line handwriting recognition system. In *Proc. of the 6th International Conference on Document Analysis and Recognition*, pages 412–416, Seattle, 2001.

[7] J. Hu, S.G. Lim, and M. K. Brown. Writer independent on-line handwriting recognition using an hmm approach. *Pattern Recognition*, (33):133–147, 2000.

[8] R. Kassel. The mit on-line character database. ftp://lightning.lcs.mit.edu/pub/handwriting/mit.tar.Z.

[9] X. Li, M. Parizeau, and R. Plamondon. Segmentation and reconstruction of on-line handwritten scripts. *Pattern Recognition*, 31(6):675–684, 1998.

[10] R. Niels. Dynamic time warping - an intuitive way of handwriting recognition? Master's thesis, Radboud University Nijmegen, 2005.

[11] M. Parizeau and R. Plamondon. A handwriting model for syntactic recognition of cursive script. In *Proc. 11th International Conference on Pattern Recognition*, volume II, pages 308–312, August 31 to September 3 1992.

[12] R. Plamondon and S. Srihari. On-line and off-line handwriting recognition: A comprehensive survey. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(1):63–84, January 2000.

[13] C. De Stefano, M. Garutto, and A. Marcelli. A saliency-based multiscale method for on-line cursive handwriting shape description. In *Proceedings of the Ninth International Workshop on Frontiers in Handwriting Recognition*, pages 124–129, 2004.

[14] J. Sternby. On-line signature verification by explicit solution to the point correspondence problem. In *First International Conference on Biometric Authentication*, pages 569 – 576, july 2004.

[15] V. Vuori, J. Laaksonen, E. Oja, and J. Kangas. On-line adaptation in recognition of handwritten alphanumeric characters. In *Proc. of the 5th International Conference on Document Analysis and Recognition*, pages 792–795, September 1999.

[16] L. Vuurpijl and L. Schomaker. Finding structure in diversity: a hierarchical clustering method for the categorization of allographs in handwriting. In *Proc. of the 4th International Conference on Document Analysis and Recognition*, pages 387–393, 1997.

[17] T. Wakahara. On-line cursive script recognition using local affine transformation. In *Proc. 9th International Conference on Pattern Recognition*, pages 1133–1137, Piscataway, NJ, 1988. IEEE Press.

[18] T. Wakahara and K. Odaka. On-line cursive kanji character recognition using stroke-based affine transformation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(12):1381–1385, 1997.