Predictive DP Matching for On-Line Character Recognition

Daiki Baba, Seiichi Uchida, Hiroaki Sakoe
Kyushu University, Fukuoka-shi, 819–0395 Japan
uchida@is.kyushu-u.ac.jp

Abstract

For on-line character recognition, predictive DP matching is proposed where two physically different features, coordinate features and directional features, are handled in a unified manner. For this unification, the distance of the directional features is converted into a distance of the coordinate features by a feature prediction technique. An experimental result showed that the predictive DP matching could attain a recognition rate comparable to the rate by the conventional DP matching which requires the costly optimization of the weight to balance the two features.

1. Introduction

Dynamic programming (DP) matching is an elastic matching technique and has been widely applied to on-line character recognition. One of its early trails can be found in 1970's [1] and it is still employed in recent trials. It is interesting to note that a DP matching technique showed very good recognition performance in the competition of on-line Tamil character recognition at IWFHR2006 [2].

In on-line character recognition techniques including DP-matching based techniques, each character pattern is often represented as a temporal sequence of coordinate feature and directional feature [3]. Those two features are mutually complementary. The coordinate feature can represent entire character shape and is sensitive to shape distortion. In contrast, the directional feature is robust to shape distortion and ambiguous to represent the complete shape. (For example, "6" and "0" may be represented by the same directional feature sequence.)

On the use of the two features, careful consideration is necessary because they are physically different features. Most DP matching techniques (e.g., [4]) have employed a weighted sum of a coordinate distance and a directional distance on evaluating the matching. Although the weighted sum is a practical way, its theoretical grounding is rather weak. In addition, there is no established way to determine the weight. Thus, the weight was often determined empirically in past trials.

The predictive DP matching proposed in this paper can handle the two features in a unified manner. A new distance, called a predictive distance, is introduced to evaluate the distance of directional features in terms of a distance of coordinate features. By using the predictive distance instead of the conventional directional distance, it is possible to evaluate the matching by only coordinate features and thus to exclude the troublesome weight parameter.

The contributions of the predictive DP matching are summarized as follows. First, the predictive DP matching provides a reasonable strategy to convert the directional feature to the coordinate feature for handling the two features in a unified manner. Second, the predictive DP matching involves a novel *prediction* mechanism to generate reference patterns adaptively according to each input pattern.

2. Related Work

We can find several strategies to handle two physically different features (i.e., coordinate feature and directional feature) other than the simple weighted sum, which will be detailed in Section 3. Separated use will be another simple strategy. For example, Kobayashi et al. [5] have proposed a two-step matching technique where the directional feature is used in the first step for a rough classification and the positional feature is used in the second step.

Stochastic representation will be a more general strategy and can unify physically different features. In this strategy, each reference pattern will be represented as a sequence of three-dimensional probability density functions $p_j(x, y, \theta), j = 1, \ldots, J$ of x-y coordinates and local direction θ, and the matching between the ith point of the input pattern and the jth point of the reference pattern is evaluated by, for example, $-\log p_j(x_i, y_i, \theta_i)$, where (x_i, y_i, θ_i) is the feature vector of the ith point. In this evaluation, all the features are *normalized* by their (co-)variances in the stochastic representation and thus it is not necessary to mind the physical difference between features. This strategy can be found in several DP-based methods (e.g., [6]) as well as HMM-based methods. The weakness of the strategy is a large amount of training patterns for the estimation of reliable variances.

Another unification strategy can be found in [7]. This strategy is somewhat similar to the proposed technique in the sense that the directional distance is evaluated in terms of a set of coordinate distances. Specifically, a sum of four coordinate distances was used for the matching evaluation. Two distances are for main evaluations. The other two distances are for auxiliary evaluations and used conditionally to compensate weakness of the main evaluations on several cases. Our evaluation is defined as a sum of two distances and therefore much simpler than [7]. In addition, due to the weakness, its recognition accuracy was lower than the proposed technique as shown in a later section.

The proposed technique is further distinguishable from the above past strategies because the proposed technique involves a feature prediction mechanism and has a potential to improve its performance by extending its prediction mechanism. Several matching-based recognition techniques with some prediction mechanisms [8, 9, 10] have been proposed in the area of speech recognition. We can consult those techniques to extend our prediction mechanism.

3. Conventional DP Matching Algorithm

3.1. Coordinate feature and directional feature

Let $\boldsymbol{E}$ denote an input pattern,

$$\boldsymbol{E} = \boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_i, \ldots, \boldsymbol{e}_I \tag{1}$$

and $\boldsymbol{R}_c$ denote the reference pattern of the class $c \in [1, C]$,

$$\boldsymbol{R}_c = \boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_j, \ldots, \boldsymbol{r}_J. \tag{2}$$

For simplicity, the notations $\boldsymbol{r}_j$ and J are used here instead of $\boldsymbol{r}_{c,j}$ and J_c. The vectors $\boldsymbol{e}_i$ and $\boldsymbol{r}_j$ are the following 3-dimensional feature vectors,

$$\left\{ \begin{array}{lcl} \boldsymbol{e}_i & = & (x_i, y_i, \theta_i), \\ \boldsymbol{r}_j & = & (X_j, Y_j, \Theta_j), \end{array} \right. \tag{3}$$

where (x_i, y_i) and (X_j, Y_j) are coordinate features, and θ_i and Θ_j are directional features representing the local angles of writing directions. The directional feature is often derived from the coordinate features as follows:

$$\left\{ \begin{array}{lcl} \theta_i & = & \tan^{-1}(y_i - y_{i-1})/(x_i - x_{i-1}), \\ \Theta_j & = & \tan^{-1}(Y_j - Y_{j-1})/(X_j - X_{j-1}). \end{array} \right. \tag{4}$$

3.2. Coordinate distance and directional distance

The evaluation of the matching between $\boldsymbol{E}$ and $\boldsymbol{R}_c$ is based on the distance between feature vectors, $\boldsymbol{e}_i$ and $\boldsymbol{r}_j$. Let $d_{\text{pos+dir}}(\boldsymbol{e}_i, \boldsymbol{r}_j)$ denote the distance between $\boldsymbol{e}_i$ and $\boldsymbol{r}_j$. A naive definition that $d_{\text{pos+dir}}(\boldsymbol{e}_i, \boldsymbol{r}_j) = \|\boldsymbol{e}_i - \boldsymbol{r}_j\|$ does not work because the coordinate feature and the directional feature are physically different quantities and have different ranges. An alternative definition of $d_{\text{pos+dir}}(\boldsymbol{e}_i, \boldsymbol{r}_j)$ is a weighted sum of two different distances (e.g., [4]), i.e.,

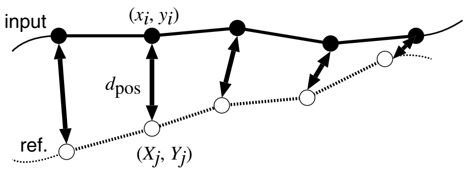

Figure 1. Coordinate distance d_{pos}.

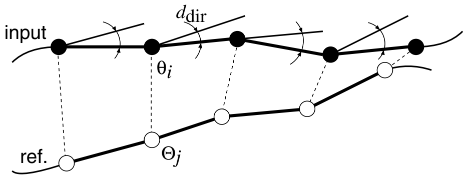

Figure 2. Directional distance d_{dir}.

$$d_{\text{pos+dir}}(\boldsymbol{e}_i, \boldsymbol{r}_j) = (1-\alpha) d_{\text{pos}}(\boldsymbol{e}_i, \boldsymbol{r}_j) + \alpha d_{\text{dir}}(\boldsymbol{e}_i, \boldsymbol{r}_j), \tag{5}$$

where $d_{\text{pos}}(\boldsymbol{e}_i, \boldsymbol{r}_j)$ is the *coordinate distance* defined as

$$d_{\text{pos}}(\boldsymbol{e}_i, \boldsymbol{r}_j) = \|(x_i, y_i) - (X_j, Y_j)\|, \tag{6}$$

and $d_{\text{dir}}(\boldsymbol{e}_i, \boldsymbol{r}_j)$ is the *directional distance* defined as

$$d_{\text{dir}}(\boldsymbol{e}_i, \boldsymbol{r}_j) = |\theta_i - \Theta_j|. \tag{7}$$

The constant α $(0 \leq \alpha \leq 1)$ is the weight to balance those two distances. Note that the directional distance ranges from 0 to π. Figures 1 and 2 illustrate the coordinate and the directional distances, respectively.

3.3. DP matching

The matching between $\boldsymbol{E}$ and $\boldsymbol{R}_c$ are formulated as the following optimization problem:

Minimize

$$\mathcal{J} = \sum_{i=1}^{I} d_{\text{pos+dir}}(\boldsymbol{e}_i, \boldsymbol{r}_{j_i}) \tag{8}$$

with respect to $\{j_i \mid i = 1, \ldots, I\}$

subject to

$$\left\{ \begin{array}{l} j_i \in \{1, 2, \ldots, J\}, \\ j_{i-1} \in \{j_i - 2,\ j_i - 1,\ j_i\}, \\ j_1 = 1, \\ j_I = J, \end{array} \right.$$

where the set of variables $\{j_i\}$ specify the matching (i.e., the point-to-point correspondence) between $\boldsymbol{E}$ and $\boldsymbol{R}_c$.

It is well-known that the above optimal matching problem can be solved efficiently by a DP algorithm. The DP algorithm relies on the following recursive equation:

$$g(i,j) = d_{\text{pos+dir}}(\boldsymbol{e}_i, \boldsymbol{r}_j) + \min_{j' \in \{j-2, j-1, j\}} g(i-1, j'), \tag{9}$$

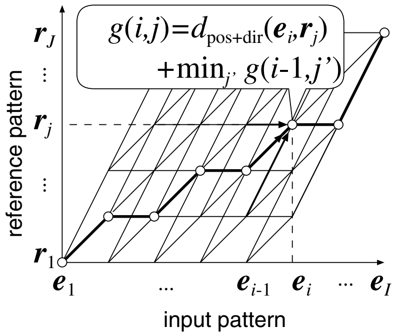

Figure 3. Conventional DP algorithm.

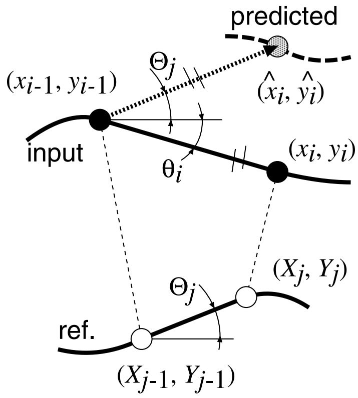

Figure 4. Derivation of coordinate feature $(\hat{x}_i, \hat{y}_i)$.

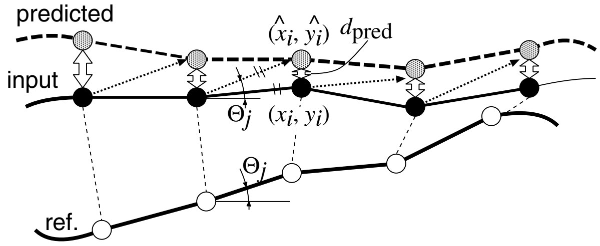

Figure 5. Predictive distance d_{pred}.

where $g(i,j)$ denotes the minimum matching cost between $\boldsymbol{e}_1, \dots, \boldsymbol{e}_i$ and $\boldsymbol{r}_1, \dots, \boldsymbol{r}_j$. Thus, this equation implies the fact that the minimum matching cost can be calculated recursively from $i = 1$ to I. (This fact relies on so-called "the principle of optimality.") The minimum $\mathcal{J}$ is provided as $g(I, J)$ and used for the minimum distance discrimination of $\boldsymbol{E}$.

Figure 3 illustrates the DP algorithm. Any path from $(i,j) = (1,1)$ to (I, J) on the i–j search graph represents a sequence $j_1, \dots, j_i, \dots, j_I$ and thus represents a possible matching between $\boldsymbol{E}$ and $\boldsymbol{R}_c$. Accordingly, the optimal matching problem is reduced to the minimum cost path problem and solved by the DP-recursion (9).

3.4. Optimization of weight α

There is no established way to determine the weight α. This may be because the coordinate feature and the directional feature are physically different quantities and their sum itself is somewhat groundless. As a practical solution, the weight α is often determined empirically in a trial-and-error manner.

4. Predictive DP Matching

4.1. Predictive distance

The key idea of the proposed predictive DP matching is the fact that the directional feature θ_i is derived from the difference between two consecutive coordinate features (x_{i-1}, y_{i-1}) and (x_i, y_i) according to (4). This fact can be interpreted in another way that the coordinate feature (x_i, y_i) can be derived from (x_{i-1}, y_{i-1}) and θ_i, i.e.,

$$(x_i, y_i) = (x_{i-1}, y_{i-1}) + \|(x_i, y_i) - (x_{i-1}, y_{i-1})\| \cdot (\cos\theta_i, \sin\theta_i). \tag{10}$$

Consider the replacement of θ_i by Θ_j in (10). If θ_i is very close to Θ_j, the equality in (10) still holds. In contrast, if θ_i is different from Θ_j, the right-hand side of (10) becomes different from the left-hand side. Thus, the difference between both sides can be used to evaluate the distance between θ_i and Θ_j. Let $(\hat{x}_i, \hat{y}_i)$ denote the right-hand side of (10) after the replacement of θ_i by Θ_j, i.e.,

$$(\hat{x}_i, \hat{y}_i) = (x_{i-1}, y_{i-1}) + \|(x_i, y_i) - (x_{i-1}, y_{i-1})\| \cdot (\cos\Theta_j, \sin\Theta_j). \tag{11}$$

The distance between θ_i and Θ_j can be evaluated as the difference between (x_i, y_i) and $(\hat{x}_i, \hat{y}_i)$, i.e.,

$$d_{\text{pred}}(\boldsymbol{e}_i, \boldsymbol{r}_j) = \|(x_i, y_i) - (\hat{x}_i, \hat{y}_i)\|. \tag{12}$$

Although the distance d_{pred} depends on $\boldsymbol{e}_{i-1}$ in addition to $\boldsymbol{e}_i$ and $\boldsymbol{r}_j$, we will use the notation $d_{\text{pred}}(\boldsymbol{e}_i, \boldsymbol{r}_j)$ for simplicity[1].

Figure 4 illustrates the equation (11), i.e., the derivation of $(\hat{x}_i, \hat{y}_i)$. As shown in this figure, $(\hat{x}_i, \hat{y}_i)$ and (x_i, y_i) are equidistant from (x_{i-1}, y_{i-1}). Figure 5 illustrates the distance $d_{\text{pred}}(\boldsymbol{e}_i, \boldsymbol{r}_j)$.

The properties of d_{pred} can be summarized as follows.

- The distance d_{pred} is the difference between two coordinate features. Thus, d_{pred} has the same physical meaning as d_{pos} and can be added directly with d_{pos}.

[1] If the input pattern is perfectly resampled so that $\|(x_i, y_i) - (x_{i-1}, y_{i-1})\|$ equals to a constant D, the distance d_{pred} no longer depends on $\boldsymbol{e}_{i-1}$ because $(\hat{x}_i, \hat{y}_i) = (x_{i-1}, y_{i-1}) + D \cdot (\cos\Theta_j, \sin\Theta_j)$.

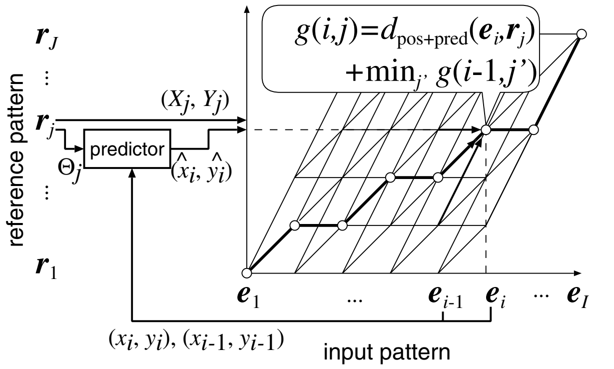

Figure 6. Predictive DP matching.

- The distance d_{pred} evaluates the difference between two directional features, θ_i and Θ_j. In fact, if $\theta_i = \Theta_j$, $d_{\text{pred}} = 0$. If $\theta_i - \Theta_j = \pm\pi$ (i.e., if θ_i and Θ_j differ the most), d_{pred} reaches its maximum value, $2\|(x_i, y_i) - (x_{i-1}, y_{i-1})\|$.

- The distance d_{pred} is independent of the coordinate features of $\boldsymbol{R}_c$.

- The distance d_{pred} can be called a *predictive distance*, because the coordinate feature $(\hat{x}_i, \hat{y}_i)$ can be considered as the prediction of (x_i, y_i) by (x_{i-1}, y_{i-1}) and Θ_j. If $\theta_i = \Theta_j$, we can expect the perfect prediction. Otherwise, we will have a prediction error as a non-zero value of d_{pred}.

We finally have the distance $d_{\text{pos+pred}}$ as an alternative to $d_{\text{pos+dir}}$. The distance $d_{\text{pos+pred}}$ is defined as the sum of two coordinate distances, d_{pos} and d_{pred}, i.e.,

$$d_{\text{pos+pred}}(\boldsymbol{e}_i, \boldsymbol{r}_j) = d_{\text{pos}}(\boldsymbol{e}_i, \boldsymbol{r}_j) + d_{\text{pred}}(\boldsymbol{e}_i, \boldsymbol{r}_j). \quad (13)$$

Again, since d_{pos} and d_{pred} are based on the same physical quantities, the weight α is not necessary to balance them.

4.2. Incorporation of predictive distance into DP matching

Formally, the incorporation of the predictive distance d_{pred} into DP matching is very simple; it is done just by replacing $d_{\text{pos+dir}}$ by $d_{\text{pos+pred}}$ in the objective function $\mathcal{J}$ of (8). Hereafter, we call the DP matching based on $d_{\text{pos+pred}}$ *predictive DP matching*.

Figure 6 illustrates the predictive DP matching. The current and the last input feature vectors $\boldsymbol{e}_i$ and $\boldsymbol{e}_{i-1}$ are sent to a predictor to obtain the predicted coordinate feature vector $(\hat{x}_i, \hat{y}_i)$. The detailed mechanism of the predictor is already described as (11). The directional feature Θ_j (to be more accurate, Θ_{j_i}) acts as the inner parameter of the predictor. Then, the distance between the predicted feature $(\hat{x}_i, \hat{y}_i)$ and the current feature (x_i, y_i) is calculated as d_{pred} and added to d_{pos} to have $d_{\text{pos+pred}}$.

Figures 5 and 6 illustrate the important property of the predictive DP matching; the predictive DP matching can adapt the reference vector $\boldsymbol{r}_j$ to the input vectors $\boldsymbol{e}_i$ as $(\hat{x}_i, \hat{y}_i)$. In other words, the prediction mechanism involved in the predictive DP matching generates an adapted reference pattern, $(\hat{x}_1, \hat{y}_1), \ldots, (\hat{x}_i, \hat{y}_i), \ldots, (\hat{x}_I, \hat{y}_I)$. The predictive distance d_{pred} can be considered as the distance between the input pattern and this adapted reference pattern.

Table 1. Recognition rates (%).

$d_{\text{pos+dir}}$	d_{pos}	d_{dir}	d_{Burr}	$d_{\text{pos+pred}}$	d_{pred}
95.9	92.8	87.3	94.8	95.2	87.0

5. Experimental Results

5.1. Data preparation

Isolated numeral patterns from Ethem Alpaydin Digit, which is an on-line character database in the Unipen format, were used. The database contains 7,494 training patterns by 30 writers and 3,498 test patterns by other 14 writers. As preprocessing, each character pattern with multiple strokes was converted into a single-stroke pattern by connecting pen-up parts linearly. Then, a linear size normalization was performed to be 128$\times$ 128 while keeping the original aspect ratio. Re-sampling was also performed to make every pair of consecutive points equidistant.

Then, 21 reference patterns were created from the training patterns by the clustering method called CLUSTER [4]. Four references were assigned to the class “8”, three were to “5”, “7”, and “9”, two were “1” and “4”, and one was “0”, “2”, “3”, and “6.”

5.2. Recognition result by conventional DP matching

First, the recognition performance of the conventional DP matching based on (8) was evaluated by using the above test data set. The weight α was optimized by observing the recognition rates for the training data at every $\alpha \in \{0, 0.01, 0.02, \ldots, 0.99, 1\}$. Thus, the recognition experiment was repeated 100 times. This trial-and-error optimization revealed that the optimal $\alpha = 0.41$ could attain the rate 98.2% for training data.

The recognition rates of the test data were summarized in Table 1. The recognition rate by the conventional technique ($d_{\text{pos+dir}}$) was 95.9% at $\alpha = 0.41$. This table also shows the rate 92.8% attained by d_{pos} alone and the rate 87.3% attained by d_{dir} alone. Those two rates were equivalent to the rates attained by $d_{\text{pos+dir}}$ at $\alpha = 0$ and 1. This fact indicates that a thoughtless selection of α may degrade the recognition performance by $d_{\text{pos+dir}}$ drastically.

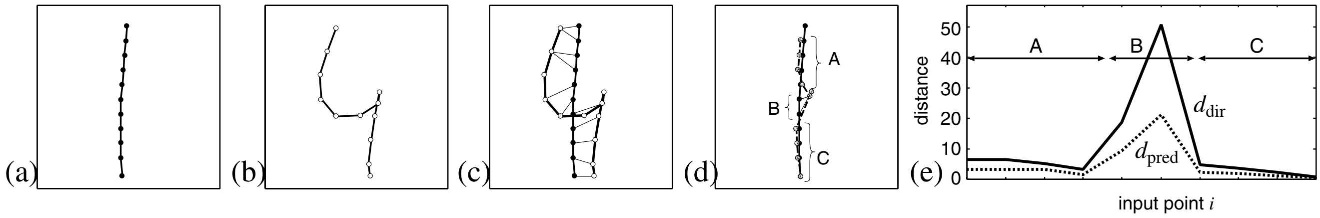

Figure 7. The effect of $d_{\rm pred}$. (See text.)

Table 1 also shows the result by $d_{\rm Burr}$, which is the distance proposed in [7]. As noted in Section 2, this distance evaluates a directional distance by a coordinate distance, and therefore has a close relation to the proposed technique. Its recognition rate, however, was rather insufficient as shown in the table.

5.3. Recognition result by predictive DP matching

As shown in Table 1, the recognition rate achieved by the proposed technique ($d_{\rm pos+pred}$) was 95.2% and thus comparable to 95.9% by the conventional technique ($d_{\rm pos+dir}$). This result should be appreciated by considering that the performance of the conventional technique is degraded by a thoughtless α. Furthermore, the proposed technique is free from the costly optimization of α. Those results and facts will show the usefulness of the proposed technique.

Figure 7 shows how $d_{\rm pred}$ plays the role of $d_{\rm dir}$. In this figure, (a) is an input pattern $\boldsymbol{E}$ of the class "1", (b) is a reference pattern $\boldsymbol{R}_c$ of the class "4", (c) is a result of matching between $\boldsymbol{E}$ and $\boldsymbol{R}_c$, and (d) is the sequence of the predicted coordinate features (, or the adapted reference pattern) $(\hat{x}_1, \hat{y}_1), \ldots, (\hat{x}_i, \hat{y}_i), \ldots, (\hat{x}_I, \hat{y}_I)$. Since the characters "1" and "4" have the similar directional features in their beginning and end, the predicted coordinate features $(\hat{x}_i, \hat{y}_i)$ are very close to the input pattern (x_i, y_i) around the beginning (A) and the ending (C) parts in (d). In contrast, the characters have different directional features in their middle parts, and therefore we can observe some difference between the predicted coordinate features and the input pattern around the middle (B) part. The graph (e) compares $d_{\rm pred}$ with $d_{\rm dir}$ at each point i. Those two distances show very similar curves. Thus, the distance $d_{\rm pred}$ is a promising alternative to $d_{\rm dir}$.

6. Conclusion

For on-line character recognition, predictive DP matching has been proposed. First, it could handle two physically different features, coordinate features and directional features, in a unified manner. Second, it involved a feature prediction mechanism for the unification. Therefore predictive DP matching has a potential to improve its performance by extending its prediction mechanism. The unification has been done by replacing the conventional directional distance by a new coordinate distance, called a predictive distance. In other words, the directional distance is converted into a new coordinate distance by the feature prediction. The results of a recognition experiment have shown that predictive DP matching can achieve a reasonable recognition accuracy and that the predictive distance can take the place of the directional distance.

Future work will focus on the development of a more sophisticated prediction mechanism to improve the overall performance of the predictive DP matching.

References

[1] K. Ikeda, et al., "On-line recognition of hand-written characters using positional and stroke vector sequences," Proc. 4th IJCPR, pp. 813–815, 1978.

[2] S. Madhvanath and S. M. Lucas, "IWFHR 2006 online Tamil handwritten character recognition competition," Proc. 10th IWFHR, CD-ROM, 2006.

[3] C. C. Tappert, C. Y. Suen, and T. Wakahara, "The state of art in on-line handwriting recognition," IEEE Trans. PAMI, vol. 12, no. 8, pp. 787–808, 1990.

[4] S. D. Connell and A. K. Jain, "Template-based online character recognition," Pattern Recognit., vol. 34, no. 1, pp. 1–14, 2001.

[5] M. Kobayashi, et al., "RAV (reparameterized angle variations) algorithm for online handwriting recognition, IJDAR, vol. 3, pp. 181-191, 2001.

[6] C. Bahlmann and H. Burkhardt, "The writer independent online handwriting recognition system *frog on hand* and cluster generative statistical dynamic time warping," IEEE Trans. PAMI, vol. 26, no. 3, pp. 299–310, 2004.

[7] D. J. Burr, "Designing a handwriting reader," IEEE Trans. PAMI, vol. PAMI-5, no. 5, pp. 554–559, 1983.

[8] K. Iso and T. Watanabe, "Speaker-independent word recognition using a neural prediction model," Proc. ICASSP, vol. 1, pp. 441-444, 1990.

[9] S. Garcia-Salicetti, P. Galiinari, B. Dorizzi, A. Mellouk, and D. Fanchon, "A neural predictive approach for online cursive script recognition," Proc. IEEE ICASSP, vol. 5, pp. 3463-3466, 1995.

[10] P. Kenny, M. Lennig, and P. Mermelstein, "A linear predictive HMM for vector-valued observations with applications to speech recognition," IEEE Trans. ASSP, vol. 38, no. 2, pp. 220-225, 1990.