

# A Bidirectional Matching Algorithm for Deformable Pattern Detection with Application to Handwritten Word Retrieval

Kwok-Wai Cheung, Dit-Yan Yeung and Roland T. Chin

Department of Computer Science

The Hong Kong University of Science and Technology

Clear Water Bay, Hong Kong

Tel: +852-23588771 Fax: +852-23581477

{william, dyyeung, roland}@cs.ust.hk

## Abstract

A Bayesian framework for deformable pattern classification has been proposed in [1] with promising results for isolated handwritten character recognition. Its performance, however, degrades significantly when it is applied to detect deformable patterns in complex scenes, where the amount of outliers due to other neighboring objects or the background is usually large. Also, the fact that the associated evidence measure does not penalize models resting on white space results in a high false alarm rate. In this paper, another Bayesian framework for deformable pattern detection is proposed. The framework possesses the intrinsic property of matching with only part of an image (segmentation) and its associated evidence measure can penalize white space implicitly. However, limited data exploration capability is the major trade-off. By properly combining the two frameworks, a new matching algorithm called bidirectional matching is proposed. This combined approach possesses the advantages of the two frameworks and gives robust results for non-rigid shape extraction. To evaluate the performance of the proposed approach, we have applied it to shape-based handwritten word retrieval. Using a subset of the *bb* dataset in the CEDAR database, we can achieve a recall rate of 59% and a precision rate of 43%.

## 1. Introduction

An important advantage of model-based object recognition is its potential to achieve integrated segmentation and recognition in complex scenes. If the object of interest is non-rigid, the corresponding shape model is required to be deformable for more flexible matching. This leads to the need for distinguishing between the *inlier* data, which is important for controlling model deformation, and the *outlier* data, which should be suppressed because of the distraction

it causes to the model. For example, to extract a character from a handwritten word, the outliers may appear in the form of stroke anomalies, closely cluttered characters or the printed background.

The Bayesian framework proposed in [1] for deformable matching, though with promising results for isolated handwritten character recognition, does suffer from the outlier problem mentioned above. Besides, it has been pointed out that the evidence used for classification in the adopted framework does not penalize models resting on white space, which in turn leads to a high false alarm rate. These two intrinsic deficiencies have been addressed in [4], where the proposed solution for the outlier problem is to adopt a uniform noise model, which unfortunately has not been rigorously tested.<sup>1</sup> For the white space problem, the proposed solution in [4] is to model also the pixels of the white space, which as discussed in [4], will be too computationally expensive. In this paper, a new Bayesian framework for deformable pattern detection is proposed, which, when compared with the framework proposed in [1], is proved to have the nice implicit properties of a) localizing a model to match with only part of an image and b) penalizing white space. However, limited data exploration capability is the major trade-off. By combining the two frameworks in a way similar to the “forward-backward” idea of Hausdorff matching, we propose a new deformable matching algorithm called *bidirectional matching*. It possesses the advantages of the two Bayesian frameworks while its computational complexity is still of the same order as the two individual frameworks. To evaluate the effectiveness of the proposed bidirectional matching algorithm, we have applied it to the retrieval of handwritten word images based on input shape queries, using the *bb* dataset of the CEDAR database containing handwritten city name images as the test set.

---

<sup>1</sup>Our preliminary experimental results show that the noise modelling approach deteriorates quickly as the amount of outliers increases.

## 2. Bayesian Frameworks for Deformable Pattern Recognition

### 2.1. A Dual View of Generativity

In [4], there are some discussions about the perspective of generative models, on which the Bayesian framework for deformable pattern classification proposed in [1] is based. To summarize, an important assumption of the framework is that the input data (including both the inliers and outliers) is generated from the shape model. While lacking a good probabilistic model for the outliers, it is almost impossible for the shape model to generate back some data resembling its shape together with the outliers.

However, the roles of the model and the data can be interchanged and we then assume that the model shape is “generated” from the data. The question of interest now becomes whether the model can be generated from the data, or in other words, whether the model can be *detected* in the data. An input dataset, with an instance of the model embedded in it, normally contains a subset which resembles the model shape (except for the occlusion cases) and this makes the generation process possible. This changes the underlying assumption of the Bayesian framework proposed in [1] and suggests a new Bayesian framework for deformable pattern detection, which will be described in the rest of this section.

### 2.2. Analogy with the Forward and Reverse Distances in Hausdorff Matching

To cite from [5]: “*The Hausdorff distance is actually composed of two asymmetric distances: the forward distance, which is the distance from the model to the image, the reverse distance, the distance from the image to the model. The forward distance is small when every point in the model is close to some point in the image, and the reverse distance is small when every point in the image is close to some point in the model.*” (cf. Helmholtz machine by Dayan *et al.* [3]).

From the above citation, one could see that the notion of model generating data is analogous to that of the reverse distance, while the notion of data generating model is analogous to that of the forward distance. Due to their similarity and for the sake of subsequent discussions, we denote the framework proposed in [1], which is restated here again for completeness, as the *reverse framework* while the newly proposed framework, to be described in the following section, as the *forward framework*.

### 2.3. Reverse Framework

Let  $H_i$  denote the shape model of the  $i$ -th character class,  $\mathbf{D}$  the input image,  $\mathbf{w}$  the model parameter vector describing character shape,  $\alpha$  the regularization parameter, and  $\beta$

the character stroke width. The parameters  $\alpha$  and  $\beta$  are referred to as hyperparameters. Assuming equal prior probabilities for different character classes, the classified output is computed by finding the  $i$ -th class that maximizes  $p(\mathbf{D}|H_i)$ , which is approximated in [1] as

$$Pr(\mathbf{D}|H_i) \propto \frac{p(\mathbf{D}|\mathbf{w}, \beta^*, H_i)p(\mathbf{w}|\alpha^*, H_i)}{p(\mathbf{w}|\mathbf{D}, \alpha^*, \beta^*, H_i)} p(\alpha^*, \beta^*|H_i) \Delta \log \alpha \Delta \log \beta \quad (1)$$

where  $p(\mathbf{w}|\alpha, H_i)$  is the prior parameter distribution,  $p(\mathbf{D}|\mathbf{w}, \beta, H_i)$  the likelihood function,  $p(\mathbf{w}|\mathbf{D}, \alpha, \beta, H_i)$  the posterior parameter distribution given the data  $\mathbf{D}$  and  $\Delta \log \alpha$  and  $\Delta \log \beta$  are the *effective* ranges of  $\alpha$  and  $\beta$ , respectively. Parameters with the superscript “\*” denote their MAP estimates.

#### 2.3.1 Representation and Criterion Formulation

Handwritten digits are represented as cubic B-splines, each of which is parameterized by a small set of  $k$  control points  $\mathbf{w}$  and affine transform parameters  $\{\mathbf{A}, \mathbf{T}\}$ . The distribution of the black pixels is represented by a mixture of Gaussians. The prior parameter distribution and the likelihood function are defined by two criterion functions - model deformation  $E_w$  and data mismatch  $E_D$ , respectively. They are given as

$$E_w(\mathbf{w}) = \frac{1}{2}(\mathbf{w} - \mathbf{h})^t \boldsymbol{\Sigma}^{-1}(\mathbf{w} - \mathbf{h}) \quad (2)$$

$$p(\mathbf{w}|\alpha, H_i) = \frac{1}{Z_w(\alpha)} \exp(-\alpha E_w(\mathbf{w})) \quad (3)$$

$$E_D(\mathbf{w}, \mathbf{A}, \mathbf{T}; \mathbf{D}) = - \sum_{l=1}^N \log \left[ \frac{1}{N_g} \sum_{j=1}^{N_g} \exp \left( -\beta \frac{\|\mathbf{m}_j(\mathbf{w}, \mathbf{A}, \mathbf{T}) - \mathbf{y}_l\|^2}{2} \right) \right] \quad (4)$$

$$p(\mathbf{D}|\mathbf{w}, \mathbf{A}, \mathbf{T}, \beta, H_i) = \frac{1}{Z_D(\beta)} \exp(-E_D(\mathbf{w}, \mathbf{A}, \mathbf{T}; \mathbf{D})) \quad (5)$$

where  $Z_w(\alpha) = \left(\frac{2\pi}{\alpha}\right)^k |\boldsymbol{\Sigma}|^{1/2}$ ,  $Z_D(\beta) = \left(\frac{2\pi}{\beta}\right)^N$ ,  $\mathbf{h}$  is the mean control point,  $\boldsymbol{\Sigma}$  is the covariance matrix of  $\mathbf{w}$ ,  $\mathbf{S}_j$  is a matrix containing cubic B-spline coefficients,  $\mathcal{A}$  and  $\mathcal{T}$  composed of  $\mathbf{A}$  submatrices and  $\mathbf{T}$  subvectors respectively,  $\mathbf{m}_j(\mathbf{w}, \mathbf{A}, \mathbf{T}) = \mathbf{S}_j^t(\mathcal{A}\mathbf{w} + \mathcal{T})$  is the mean of the  $j$ -th Gaussian,  $N$  is the number of black pixels,  $N_g$  is the number of Gaussians along the spline,  $\alpha$  is the regularization parameter,  $\beta$  is the inverse of the Gaussians’ variance for modeling the character stroke width,  $\mathbf{y}_l$  is the location vector of an individual black pixel and  $\mathbf{D}$  denotes the set  $\{\mathbf{y}_l | 1 \leq l \leq N\}$ .

### 2.3.2 Matching

The matching is done by maximum a posteriori (MAP) estimation of  $\{\mathbf{w}, \mathbf{A}, \mathbf{T}\}$ , where the posterior distribution  $p(\mathbf{w}, \mathbf{A}, \mathbf{T} | \mathbf{D}, \alpha, \beta, H_i)$  is defined as

$$p(\mathbf{w}, \mathbf{A}, \mathbf{T} | \mathbf{D}, \alpha, \beta, H_i) = \frac{1}{Z_M} \exp(-E_M(\mathbf{w}, \mathbf{A}, \mathbf{T}; \mathbf{D})) \quad (6)$$

where  $E_M(\mathbf{w}, \mathbf{A}, \mathbf{T}; \mathbf{D}) = \alpha E_w(\mathbf{w}) + E_D(\mathbf{w}, \mathbf{A}, \mathbf{T}; \mathbf{D})$  and  $Z_M$  is the corresponding partition function. The expectation-maximization (EM) algorithm is used where an  $h$ -function is typically involved and is here defined as

$$h_j^l(\hat{\mathbf{w}}_n, \hat{\mathbf{A}}_n, \hat{\mathbf{T}}_n; \mathbf{y}_l) = \frac{\exp(-\beta \frac{\|\mathbf{m}_j(\hat{\mathbf{w}}_n) - \mathbf{y}_l\|^2}{2})}{\sum_p \exp(-\beta \frac{\|\mathbf{m}_p(\hat{\mathbf{w}}_n) - \mathbf{y}_l\|^2}{2})} \quad (7)$$

where  $\hat{\mathbf{w}}_n$  and  $\{\hat{\mathbf{A}}_n, \hat{\mathbf{T}}_n\}$  are the estimates of the control point vector and the affine transform obtained in the  $n$ -th EM iteration. The MAP estimates of  $\alpha$  and  $\beta$  are also computed after each EM iteration using the formula given as,

$$\alpha^* = \frac{\gamma}{2E_w(\mathbf{w}^*)} \quad \beta^* = \frac{2N - \gamma}{2E_D'(\mathbf{w}^*, \mathbf{A}^*, \mathbf{T}^*; \hat{\mathbf{w}}, \hat{\mathbf{A}}, \hat{\mathbf{T}}, \mathbf{D})} \quad (8)$$

where  $\gamma$  can be interpreted as the effective numbers of parameters and  $E_D'$  is derived from  $E_D$  in the E-step of the EM algorithm. Readers are referred to [1] for more details.

### 2.3.3 Classification

Referring to EQ.(1), the classification step is based on  $p(\mathbf{D} | H_i)$ , which requires the matching results together with the effective ranges of  $\alpha$  and  $\beta$ , given by

$$\Delta \log \alpha = \sqrt{2/\gamma} \quad \Delta \log \beta = \sqrt{2/(2N - \gamma)}. \quad (9)$$

## 2.4. Forward Framework

Let  $H$  denote a shape model to be detected,  $\mathcal{D}$  a set of images, and  $\mathbf{D}_i$  the  $i$ -th image in  $\mathcal{D}$ . Given  $\mathbf{D}_i$ , a shape model  $H$  is said to be detected in it when  $Pr(H | \mathbf{D}_i) > \phi$  where  $\phi$  is a threshold parameter.

Expanding  $Pr(H | \mathbf{D}_i)$  according to the Bayes rule and assuming that  $\mathbf{w}$  is independent of  $\alpha$  when  $\mathbf{D}_i$  is given,<sup>2</sup> it gives

$$\begin{aligned} & Pr(H | \mathbf{D}_i) \\ &= \int Pr(H | \alpha, \beta, \mathbf{D}_i) p(\alpha, \beta | \mathbf{D}_i) d\alpha d\beta \\ &= \int Pr(H | \alpha, \beta, \mathbf{w}, \mathbf{D}_i) p(\mathbf{w} | \beta, \mathbf{D}_i) d\mathbf{w} p(\alpha, \beta | \mathbf{D}_i) d\alpha d\beta. \end{aligned} \quad (10)$$

<sup>2</sup>This assumption is justified by the argument that  $\mathbf{w}$  is generated by  $\mathbf{D}_i$  and without knowing  $H$ , there is no reference for measuring the deviation of  $\mathbf{w}$  and thus no information about the degree of regularization.

We further assume that  $H$  is independent of  $\mathbf{D}_i$  and  $\beta$  when  $\mathbf{w}$  is given and  $Pr(H | \mathbf{D}_i)$  becomes

$$Pr(H | \mathbf{D}_i) = \int \frac{p(\mathbf{w} | \alpha, H) Pr(H | \alpha)}{p(\mathbf{w} | \alpha)} p(\mathbf{w} | \beta, \mathbf{D}_i) d\mathbf{w} p(\alpha, \beta | \mathbf{D}_i) d\alpha d\beta. \quad (11)$$

Assuming that  $Pr(H | \alpha)$  and  $p(\mathbf{w} | \alpha)$  are constants and applying Laplacian approximation,  $Pr(H | \mathbf{D}_i)$  becomes

$$Pr(H | \mathbf{D}_i) \propto p(\mathbf{w}^* | \alpha^*, H) p(\mathbf{w}^* | \beta^*, \mathbf{D}_i) p(\alpha^*, \beta^* | \mathbf{D}_i) \Delta \mathbf{w} \Delta \log \alpha \Delta \log \beta. \quad (12)$$

If we compare EQ.(12) of this forward framework with EQ.(1) of the reverse framework, it can be noted that the first factor,  $p(\mathbf{w}^* | \alpha^*, H)$ , is the prior distribution of  $\mathbf{w}$  and remains unchanged. The second factor,  $p(\mathbf{w}^* | \beta^*, \mathbf{D}_i)$ , is now the probability distribution of the model parameters given the data, instead of the probability distribution of the data given the model parameters.

### 2.4.1 A New Mismatch Criterion

The mismatch criterion for the reverse framework is called a *data* mismatch criterion which measures the discrepancy of *all* the data from the model. To impose the data-generating-model notion, we define a new criterion for the forward framework called *sub-data* mismatch criterion  $E_{D_s}$ , defined as

$$E_{D_s}(\mathbf{w}, \mathbf{A}, \mathbf{T}; \beta, \mathbf{D}_i) = - \sum_{j=1}^{N_g} \log \left[ \frac{1}{N} \sum_{l=1}^N \exp \left( -\beta \frac{\|\mathbf{m}_j(\mathbf{w}, \mathbf{A}, \mathbf{T}) - \mathbf{y}_l\|^2}{2} \right) \right] \quad (13)$$

and we define

$$p(\mathbf{w}, \mathbf{A}, \mathbf{T} | \beta, \mathbf{D}_i) = \frac{1}{Z_{D_s}(\beta)} \exp(-E_{D_s}(\mathbf{w}, \mathbf{A}, \mathbf{T}; \mathbf{D}_i)) \quad (14)$$

where  $Z_{D_s}(\beta) = \left( \frac{2\pi}{\beta} \right)^{N_g}$ . According to EQ.(13) and EQ.(14), each point  $\mathbf{m}_j$  along the spline is modeled by a uniformly weighted mixture of Gaussians with their means being the data  $\mathbf{D}_i$ . This coheres with the data-generating-model assumption. In the new definition, however, we lose the interpretation that  $\beta$  is related to the character stroke width as in [1]. Instead, it is related to the size of the localized search region to be controlled for each model point to explore.

### 3. Implementation Differences between the Forward and Reverse Frameworks

#### 3.1. Matching

##### 3.1.1 Estimation of Shape Parameters

The optimal shape parameters  $\{\mathbf{w}^*, \mathbf{A}^*, T^*\}$  are estimated by maximizing  $p(\mathbf{w}, \mathbf{A}, \mathbf{T}|\alpha, \beta, \mathbf{D}_i, H)$ , which is equivalent to minimizing  $\alpha E_w(\mathbf{w}) + E_{D_s}(\mathbf{w}, \mathbf{A}, \mathbf{T}; \beta, \mathbf{D}_i)$ . The EM algorithm can be used again for the maximization problem, with just the  $h$ -function modified as follows

$$h_j^l(\hat{\mathbf{w}}_n, \hat{\mathbf{A}}_n, \hat{\mathbf{T}}_n; \mathbf{y}_l) = \frac{\exp(-\beta \frac{\|\mathbf{m}_j(\hat{\mathbf{w}}_n) - \mathbf{y}_l\|^2}{2})}{\sum_{p=1}^N \exp(-\beta \frac{\|\mathbf{m}_j(\hat{\mathbf{w}}_n) - \mathbf{y}_p\|^2}{2})}. \quad (15)$$

##### 3.1.2 Estimation of Hyperparameters

The optimal hyperparameters, according to EQ.(10), are estimated by maximizing  $Pr(H|\alpha, \beta, \mathbf{D}_i)p(\alpha, \beta|\mathbf{D}_i)$  with respect to  $\alpha$  and  $\beta$ . To derive the updating formula for  $\alpha$  and  $\beta$ , assuming  $p(\alpha, \beta|\mathbf{D}_i)$  to be constant, we can further expand  $Pr(H|\alpha, \beta, \mathbf{D}_i)$  as

$$Pr(H|\alpha, \beta, \mathbf{D}_i) \propto p(\mathbf{w}^*|\alpha, H)p(\mathbf{w}^*, \mathbf{A}^*, \mathbf{T}^*|\beta, \mathbf{D}_i)\Delta \mathbf{w}. \quad (16)$$

When compared with EQ.(8) of the reverse framework, the updating formula for  $\alpha$  is unchanged but that for  $\beta$  has to be modified as

$$\beta^* = \frac{2N_g - \gamma}{2E'_{D_s}(\mathbf{w}^*, \mathbf{A}^*, \mathbf{T}^*; \hat{\mathbf{w}}, \hat{\mathbf{A}}, \hat{\mathbf{T}}, \mathbf{D}_i)}. \quad (17)$$

#### 3.2. Detection

A shape model  $H$  is considered to be detected in the input data  $\mathbf{D}_i$  if  $Pr(H|\mathbf{D}_i)$  (which will later be referred to as the *evidence* of the forward framework) is found to be greater than a threshold  $\phi$ . The value of  $Pr(H|\mathbf{D}_i)$  can be computed by substituting the prior distribution of  $\mathbf{w}$  of the reverse framework together with the new sub-data mismatch criteria (given by EQ.(13)) into EQ.(12). Also, the computation of  $\Delta \mathbf{w}$  and  $\Delta \log \alpha$  remains the same as that in the reverse framework, while that of  $\Delta \log \beta$  is modified as

$$\Delta \log \beta = \sqrt{2/(2N_g - \gamma)}. \quad (18)$$

#### 4. Comparison of the Two Frameworks

For the sake of further discussions, the matching processes of the reverse and forward frameworks are referred to as *reverse matching* and *forward matching* respectively.

Similarly, the evidences of the two frameworks are referred to as *reverse* and *forward evidences*. It can be theoretically shown that both frameworks have their own strengths and shortcomings [2]. Due to the page limit, we only summarize and discuss the result of the analysis.

#### 4.1. Shape Discriminating Properties

Considering the shape discriminating properties of the two frameworks, it can be shown that:

**Proposition 1** *The reverse evidence does not penalize models resting on white space.*

**Proposition 2** *The forward evidence does penalize models resting on white space.*

These two properties provide a better understanding of the relationship between the two evidences and the reverse and forward distances in Hausdorff matching. The reverse distance is small when every point in the image is close to some point in the model, i.e., white space is allowed for the matched model. This corresponds closely to our reverse evidence. Similar analogy is also true for the forward distance and our forward evidence. From an application point of view, the reverse evidence, which does not incorporate white space penalty, is good for detecting patterns with broken lines or occluded patterns. However, it suffers from the sub-part problem described in [1]. The forward framework, which penalizes white space, is good for minimizing false alarms and solves the sub-part problem implicitly.

#### 4.2. Shape Matching Properties

Considering the shape matching properties of the two frameworks, it can be shown that:

**Proposition 3** *Reverse matching has good data exploration characteristics.*

**Proposition 4** *Forward matching has good localization characteristics.*

These two propositions reveal the dilemma that forward matching is insensitive to outlier influence but lacks good data exploration capability, while reverse matching has good data exploration characteristics but is heavily influenced by outliers. The dilemma is also discussed in Section 6.1.

#### 5. Bidirectional Matching Algorithm and its Convergence Properties

The duality observed from the two frameworks suggests the idea of combining the two frameworks to get the best

of both worlds. The success of Hausdorff matching gives the cue that taking the maximum of the two frameworks' data mismatch related criteria can be a good choice. Based on the idea, we propose an algorithm called *bidirectional matching* outlined in Figure 1, where the matching process switches between the two frameworks according to the values of  $E_D$  and  $E_{D_s}$  until some convergence criterion is satisfied. Figure 2 illustrates the limitation of the individual frameworks and the strength of the proposed *bidirectional matching*. The convergence proof of the algorithm can be found in [2].

1. Initialize an input model using chamfer-like matching.
2. Compute the data mismatch,  $E_D$ , and the sub-data mismatch,  $E_{D_s}$ , for the two frameworks.
3. do
  - (a) If  $E_D > E_{D_s}$ ,  
Perform reverse matching.  
else  
Perform forward matching.
  - (b)  $\beta := (1 + \epsilon)\beta$ ;
  - (c) if  $\beta > 4$ ,  $\beta := 4$ ;  
/\* equivalent to a Gaussian width = 0.5 \*/
4. until a convergence is reached with the difference in  $w$  for two consecutive iterations less than a threshold  $\epsilon'$ .

**Figure 1. The matching algorithm.**

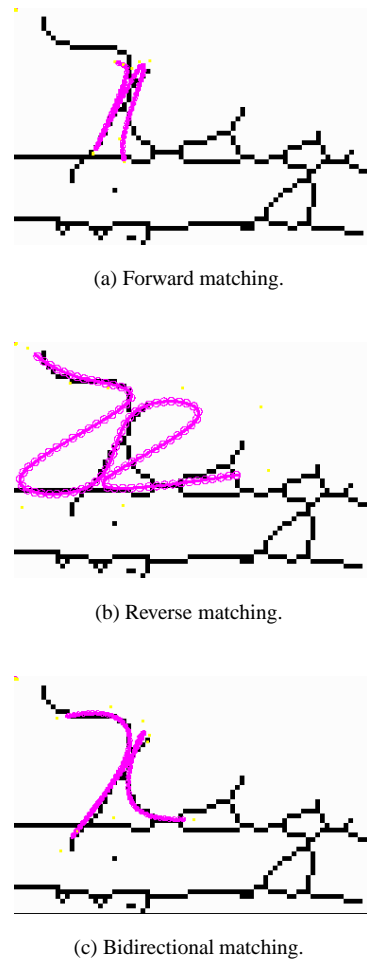
## 6. Experimental Results

In order to evaluate the effectiveness of our bidirectional matching algorithm, we have applied it for recognizing handwritten word images. Our experiment contains two parts. The first part is concentrated on evaluating the matching performance of the bidirectional matching algorithm and the second part is concentrated on the performance of a retrieval application based on the algorithm, which tests both the shape matching and discriminating performance of the system. The *bb* subset<sup>3</sup> of the CEDAR database, which contains around 300 handwritten city name images, is used as the test set. For both experiments, we take the covariance matrix of the model parameter to be an identity matrix and thus no model training is adopted.

### 6.1. Handwritten Character Extraction

To test the matching performance of the algorithm, we applied it to extract the leftmost character from each binary

<sup>3</sup>This subset contains handwritten cursive scripts of city names and was created by the CEDAR group for testing character segmentation algorithms.



**Figure 2. Comparison of matching performance.**

handwritten word image in the *bb* subset of the CEDAR database.<sup>4</sup> We assume that the identity of the leftmost character is known<sup>5</sup> but the goal is to locate it accurately through deformable matching. Being provided with a hand-drafted character model with an equivalent shape topology as the leftmost character, the algorithm first initializes the model and then starts matching. The matching performance is visually checked. While the reverse matching process fails for almost all the test cases (due to outliers), the success rate for bidirectional matching is found to be 85.5%. The error rates

<sup>4</sup>All the images in the CEDAR database are gray-level ones. So, they are preprocessed by first performing a simple intensity thresholding, and followed by thinning. The threshold for the thresholding is computed by detecting the valley in the image histogram.

<sup>5</sup>Note that the handwriting of a particular alphabet can have very different shapes or even topologies, e.g., "M" and "m" and their identities are considered to be different here.

due to bad initialization, bad matching and others are 9.7%, 3.4% and 1.4%, respectively. This shows the effectiveness of the bidirectional matching algorithm and the importance of good model initialization to good matching performance. Figure 3 depicts some of the matching results.



Figure 3. Some matching results.

## 6.2. Handwritten Word Retrieval based on Character Shape

To apply the algorithm to a non-rigid shape retrieval problem, we match a set of character shape models for the candidate characters, one by one, with the first 100 handwritten word images extracted from the *bb* dataset. We use the forward evidence as the final discriminating measure, which can penalize models resting on white space. Figure 4 shows the model set (7 models) we used for the input queries. To quantify the retrieval results, we first sort the list of word images according to their forward evidence values, cut off the list at a certain point to form the output list and compute the corresponding recall and precision rates. To determine the cut-off point, two different schemes are tested, one using the best- $N$  approach and the other using the evidence thresholding approach. The overall precision rate and recall rate are found to be 43% and 59%, respectively, when the best 10 word images are selected for the output, and 45% and 65%, respectively, when the images with the negative logarithm of the forward evidence values larger than -5.5 are selected.<sup>6</sup>

To compare with the text retrieval problem where false alarms can be resulted from the fact that an input string is part of some other irrelevant strings, our retrieval system's false alarms are caused by the input shape being part of some other irrelevant portions of the handwritten cursive script. The irrelevant portion can be one single character (e.g., a "C" model can find a good match with part of an "O") or a composite of two adjacent characters (e.g., a "U" model can find a good match somewhere between the consecutive characters "dl").

## 7. Conclusion

A Bayesian framework for deformable pattern detection is proposed in this paper which can solve both the out-

<sup>6</sup>As a comparison, the average number of candidates is 12.7.

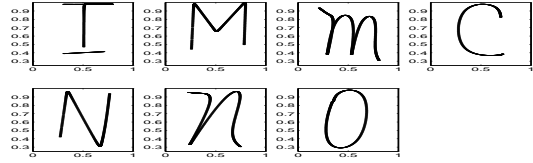


Figure 4. The input character shapes used for testing in our retrieval experiments.

lier and the white space problems implicitly. By comparing it with the framework proposed in [1] for deformable pattern classification, a reverse-and-forward duality is observed. Using only one of the two frameworks for extracting non-rigid shapes will suffer from being either too sensitive towards outliers or too poor in data exploration. By properly combining the two frameworks, a new matching algorithm called bidirectional matching is proposed. The combined framework bears a close theoretical analogy with Hausdorff matching and has been proved to converge. The proposed system has been applied to shape-based handwritten word retrieval. For the *bb* dataset in the CEDAR database with 300 handwritten city name images, we can achieve a correct matching rate of 85.5%. For a particular subset of the dataset containing 100 handwritten city name images, we achieve at best a 59% recall rate with a corresponding 43% precision rate based on the best 10 approach for selecting the candidates.

## References

- [1] K. W. Cheung, D. Y. Yeung, and R. T. Chin. A Bayesian framework for deformable pattern recognition with application to handwritten character recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1382–1388, Dec. 1998.
- [2] K. W. Cheung. *Bayesian Frameworks for Deformable Pattern Classification and Retrieval: Application to Handwriting Recognition*. PhD thesis, The Hong Kong University of Science and Technology, Jan. 1999.
- [3] P. Dayan, G. E. Hinton, R. M. Neal, and R. S. Zemel. Helmholtz machine. *Neural Computation*, 7(5):889–904, Sept. 1995.
- [4] M. Revow, C. K. I. Williams, and G. E. Hinton. Using generative models for handwritten digit recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(6):592–606, June 1996.
- [5] W. Rucklidge. *Efficient Visual Recognition Using the Hausdorff Distance*. Springer, Berlin, 1996.