[32] R. Trapp, S. Drüe, and G. Hartmann, "Stereo Matching with Implicit Detection of Occlusions," *Proc. European Conf. Computer Vision,* pp. 17-33, 1998.

[33] Univ. of Tsukuba Stereo Pair, obtained from Richard Szeliski at http://www.research.microsoft.com/szeliski/stereo/ Website current as of Sept. 2001.

[34] J. Weng, N. Ahuja, and T.S. Huang, "Two-View Matching," *Proc. Int'l Conf. Computer Vision,* pp. 64-73, 1988.

[35] R.P. Wildes, "Direct Recovery of Three-Dimensional Scene Geometry from Binocular Stereo Disparity," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 13, no. 8, pp. 761-774, 1991.

[36] A.L. Yuille and T. Poggio, "A Generalized Ordering Constraint for Stereo Correspondence," Artificial Intelligence Lab 777, MIT, Cambridge, Mass. 1984.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** http://computer.org/publications/dilb.

Bidirectional Deformable Matching with Application to Handwritten Character Extraction

Kwok-Wai Cheung, *Member,*
IEEE Computer Society,
Dit-Yan Yeung, *Member, IEEE,* and
Roland T. Chin, *Member,*
IEEE Computer Society

Abstract—To achieve integrated segmentation and recognition in complex scenes, the model-based approach has widely been accepted as a promising paradigm. However, the performance is still far from satisfactory when the target object is highly deformed and the level of outlier contamination is high. In this paper, we first describe two Bayesian frameworks, one for classifying input patterns and another for detecting target patterns in complex scenes using deformable models. Then, we show that the two frameworks are similar to the forward-reverse setting of Hausdorff matching and that their matching and discriminating properties are complementary to each other. By properly combining the two frameworks, we propose a new matching scheme called bidirectional matching. This combined approach inherits the advantages of the two Bayesian frameworks. In particular, we have obtained encouraging empirical results on shape-based pattern extraction, using a subset of the CEDAR handwriting database containing handwritten words of highly varying shape.

Index Terms—Model-based segmentation, deformable models, Bayesian inference, bidirectional matching, Hausdorff matching.

——————————— ◆ ———————————

1 INTRODUCTION

To achieve integrated segmentation and recognition in complex scenes, the model-based approach has widely been accepted as a promising paradigm. For example, one can search for the presence of a rigid object in an input image by optimizing some data mismatch measure with respect to the geometric transformation applied to the model. However, if the object of interest is *nonrigid*, the potential shape variations can no longer be described by a compact set of transformation parameters. Instead, more flexible representations, commonly called *deformable models*, are required. Extracting nonrigid shapes using deformable models is known to be highly ill-posed. Very often, regularization techniques are used to alleviate the problem, where some model smoothness criteria are added to the data mismatch measure to form the overall optimization criterion [8], [9].

Even with the introduction of smoothness regularizers, the performance of deformable matching is sometimes still far from satisfactory, especially when the shape of the target object is highly deviated from the reference model and the level of outlier contamination is high. One possible direction to reduce the outlier influence is to enhance the model adequacy. For example, domain-specific constraints obtained via careful design can be imposed on shape variations for the particular application [13], [6]. Also, some model statistics obtained via learning can be incorporated to

———————————————

- *K.-W. Cheung is with the Department of Computer Science, Hong Kong Baptist University, Kowloon Tong, Hong Kong. E-mail: william@comp.hkbu.edu.hk.*
- *D.-Y. Yeung and R.T. Chin are with the Department of Computer Science, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong. E-mail: {dyyeung, roland}@cs.ust.hk.*

enhance the model formulation if sufficient training data are available [11]. In parallel with the use of enhanced models, seach windows with carefully set window size are also commonly used to prevent the outlier data from affecting the matching results.

In this paper, we propose to achieve robust deformable matching using a new approach which has been inspired by the "forward-reverse" idea in Hausdorff matching [7]. The proposed approach is orthogonal to the model enhancing direction and does not require the use of explicit search windows. In the following, two Bayesian frameworks, one for deformable pattern classification and another for deformable pattern detection, are first described together with their relationships to Hausdorff matching. Then, we show that the strengths and shortcomings of the two frameworks regarding their shape matching and discriminating properties are complementary to each other. By combining the two frameworks, a robust matching algorithm called *bidirectional deformable matching* is proposed. To evaluate its performance, we have adopted a particular spline model and have applied the proposed approach to extract characters from the handwritten city name images in the *bb* and *bs* data set of the CEDAR database. To the best of our knowledge, there is no prior work on deformable matching where the forward-reverse idea is used.

2 BAYESIAN FRAMEWORKS FOR MODEL-BASED PATTERN RECOGNITION

In this section, we first review the background of Hausdorff matching. Then, from the perpective of generative models, we draw the relationship between Hausdorff matching and two different Bayesian frameworks, one for *pattern classification* and the other for *pattern detection*.

2.1 Hausdorff Matching: Background

Hausdorff matching [7] is a robust algorithm proposed for matching two sets of points, say, $\mathbf{M} : \{\mathbf{m}_j\}$ (model points) and $\mathbf{D} : \{\mathbf{y}_l\}$ (data points). Its robustness relies on the underlying matching criterion called the *Hausdorff distance*, which is formulated as the maximum of two asymmetric distances, namely the *forward distance* and the *reverse distance*. The forward distance measures the maximum distance from any model point to its closest data point, given as $\mathcal{H}(\mathbf{M}, \mathbf{D}) = \max_j \min_l \|\mathbf{m}_j - \mathbf{y}_l\|$, which is small when *each* model point in $\mathbf{M}$ can find a match to *some* data point in $\mathbf{D}$. The reverse distance $\mathcal{H}(\mathbf{D}, \mathbf{M})$ is defined simply by reversing the role of $\mathbf{M}$ and $\mathbf{D}$ and is small when *each* data point in $\mathbf{D}$ can find a match to *some* model point in $\mathbf{M}$. When both the forward and reverse distances are small, the corresponding Hausdorff distance is small and one can conclude that the two sets of points are close to each other. In the literature, the Hausdorff distance has been shown to be a robust indicator for detecting patterns in complex scenes [12]. But, so far, only rigid object matching with translation and scaling has been considered.

2.2 A Dual View of Generativity

A typical way to apply the generative model approach to deformable model-based pattern recognition [11] is to assume that the input data are generated according to a likelihood function given a model instance, which in turn is generated according to a prior distribution of the model class. Optimal matching is then achieved via Bayesian inference. The notion that *all* the data are generated from the model is in fact analogous to the definition of the reverse distance. We call this Bayesian approach the *reverse framework*, which has been applied to pattern classification [3] and object matching (with the use of search windows) [8]. However, if we consider the scenario of detecting patterns in complex scenes

where many outliers exist,[1] it is unlikely that the model can generate the outliers well and, thus, bad matching is usually resulted. Under such a situation, we argue that it could be more reasonable to exchange the roles of the model and the data and assume that the model instance is generated from a portion of the input data, which resembles the model shape. We call this Bayesian approach the *forward framework*. However, the data-generating-model assumption fails when occlusion occurs. By adopting a similar idea used in Hausdorff matching, these two frameworks can be combined in such a way that the model converges to a spatially localized subset of data with a high probability that the model and the data subset can generate back each other (cf. Helmholtz machine by Dayan et al., [4]). See Section 4 for more details.

2.3 Reverse Framework

Applying the reverse framework to pattern classification has been described in detail in [3]. For the sake of completeness, it is briefly reviewed here. Let H_i denote the shape model of the ith class, D the input image, w the model parameter vector describing shape, α the regularization parameter, and β the noise level parameter. According to [3], the classified output can be computed by finding the ith class that maximizes the *evidence* $p(\mathbf{D}|H_i)$, which is approximated as

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

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

2.3.1 Representation and Criterion Formulation

While the framework is generic and can be applied to different representations, here we model 2D binary patterns using cubic B-splines, each of which is parameterized by a small set of k control points w and affine transform parameters $\{\mathbf{A}, \mathbf{T}\}$. The distribution of the black pixels is represented by a mixture of Gaussians placed along the spline.[2] 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

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

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

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

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

---

1. Outliers in images often appear in the form of cluttered objects or the background.
2. One can extend the data representation to include a feature vector for each pixel to allow also gray-level or color images. An additional generation process is then required to model the per-pixel feature distribution.

TABLE 1
Comparing the Major Formulae of the Two Frameworks

	Reverse Framework	Forward Framework
Model Deformation	$\frac{1}{2}(\mathbf{w}-\mathbf{h})^t \boldsymbol{\Sigma}^{-1}(\mathbf{w}-\mathbf{h})$	
Data Mismatch	$-\sum_{l=1}^{N} \log\left[\frac{1}{N_g}\sum_{j=1}^{N_g}\exp\left(\frac{-\beta\|\mathbf{m}_j-\mathbf{y}_l\|^2}{2}\right)\right]$	$-\sum_{j=1}^{N_g}\log\left[\frac{1}{N}\sum_{l=1}^{N}\exp\left(\frac{-\beta\|\mathbf{m}_j-\mathbf{y}_l\|^2}{2}\right)\right]$
h-function	$\dfrac{\exp(-\beta\frac{\|\mathbf{m}_j(\hat{\mathbf{w}}_n)-\mathbf{y}_l\|^2}{2})}{\sum_{j=1}^{N_g}\exp(-\beta\frac{\|\mathbf{m}_j(\hat{\mathbf{w}}_n)-\mathbf{y}_l\|^2}{2})}$	$\dfrac{\exp(-\beta\frac{\|\mathbf{m}_j(\hat{\mathbf{w}}_n)-\mathbf{y}_l\|^2}{2})}{\sum_{l=1}^{N}\exp(-\beta\frac{\|\mathbf{m}_j(\hat{\mathbf{w}}_n)-\mathbf{y}_l\|^2}{2})}$
α	$\gamma/2E_w(\mathbf{w}^*)$	
β	$\dfrac{2N-\gamma}{2E'_D(\mathbf{w}^*,\mathbf{A}^*,\mathbf{T}^*;\hat{\mathbf{w}},\hat{\mathbf{A}},\hat{\mathbf{T}},\mathbf{D})}$	$\dfrac{2N_g-\gamma}{2E_D^{\S\prime}(\mathbf{w}^*,\mathbf{A}^*,\mathbf{T}^*;\hat{\mathbf{w}},\hat{\mathbf{A}},\hat{\mathbf{T}},\mathbf{D}_i)}$
$\Delta\log\alpha$	$\sqrt{2/\gamma}$	
$\Delta\log\beta$	$\sqrt{2/(2N-\gamma)}$	$\sqrt{2/(2N_g-\gamma)}$

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, Σ is the covariance matrix of $\mathbf{w}$, $\mathbf{S}_j$ is a matrix containing cubic B-spline coefficients, $\mathcal{A}$ and $\mathcal{T}$ are 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 jth Gaussian, N is the number of black pixels, N_g is the number of Gaussians along the spline, α is the regularization parameter, β is the inverse of the Gaussians' variance for modeling the noise level, $\mathbf{y}_l$ 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 Reverse Matching
Model matching is done through MAP estimation of the parameters $\{\mathbf{w},\mathbf{A},\mathbf{T}\}$, where the posterior distribution is given as

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

$E_M(\mathbf{w},\mathbf{A},\mathbf{T};\alpha,\beta,\mathbf{D})=\alpha E_w(\mathbf{w})+E_D(\mathbf{w},\mathbf{A},\mathbf{T};\beta,\mathbf{D})$ and Z_M is the corresponding partition function. The expectation-maximization (EM) algorithm [5] is used. In the E-step, an h-function defined as

$$h_j^l(\hat{\mathbf{w}}_n,\hat{\mathbf{A}}_n,\hat{\mathbf{T}}_n;\beta,\mathbf{y}_l)=\frac{\exp\left(-\beta\frac{\|bfm_j(\hat{\mathbf{w}}_n,\hat{\mathbf{A}}_n,\hat{\mathbf{T}}_n)-\mathbf{y}_l\|^2}{2}\right)}{\sum_{j=1}^{N_g}\exp\left(-\beta\frac{\|\mathbf{m}_j(\hat{\mathbf{w}}_n,\hat{\mathbf{A}}_n,\hat{\mathbf{T}}_n)-\mathbf{y}_l\|^2}{2}\right)}\qquad(7)$$

is involved where $\hat{\mathbf{w}}_n$ and $\{\hat{\mathbf{A}}_n,\hat{\mathbf{T}}_n\}$ are the parameter estimates obtained in the nth EM iteration. Intuitively speaking, h_j^l can be intrepreted as the responsibility of the jth Gaussian to the lth data point and has a high value when the Gaussian center $\mathbf{m}_j$ is close to the data point $\mathbf{y}_l$. In the M-step, the criterion function $E'_M=\alpha E_w(\mathbf{w})+\beta E'_D(\mathbf{w},\mathbf{A},\mathbf{T};\hat{\mathbf{w}},\hat{\mathbf{A}},\hat{\mathbf{T}},\mathbf{D})$ is minimized, where

$$E'_D(\mathbf{w},\mathbf{A},\mathbf{T};\hat{\mathbf{w}},\hat{\mathbf{A}},\hat{\mathbf{T}},\mathbf{D})=$$
$$\sum_{l=1}^{N}\sum_{j=1}^{N_g}h_j^l(\hat{\mathbf{w}}_n,\hat{\mathbf{A}}_n,\hat{\mathbf{T}}_n;\beta,\mathbf{y}_l)\|\mathbf{m}_j(\mathbf{w},\mathbf{A},\mathbf{T})-\mathbf{y}_l\|^2/2.\qquad(8)$$

It should be noted that E'_D can be interpreted as a "soft" version of the reverse distance used in Hausdoff matching (See the Appendix for more detailed explanation). The E-step and the M-step then alternate until convergence. The MAP estimates of α and β are then computed after each EM iteration using the following formula:

$$\alpha^*=\frac{\gamma}{2E_w(\mathbf{w}^*)},\qquad \beta^*=\frac{2N-\gamma}{2E'_D(\mathbf{w}^*,\mathbf{A}^*,\mathbf{T}^*;\hat{\mathbf{w}},\hat{\mathbf{A}},\hat{\mathbf{T}},\mathbf{D})},\qquad(9)$$

where γ is the effective number of parameters computed according to [10].

2.3.3 Classification
Referring to (1), the classification step is based on $p(\mathbf{D}|H_i)$, which requires the matching results and the effective ranges of α and β, given by $\Delta\log\alpha=\sqrt{2/\gamma}$ and $\Delta\log\beta=\sqrt{2/(2N-\gamma)}$.

2.4 Forward Framework
We now proceed to the forward framework which is newly proposed by us for pattern detection. Using the same notations adopted in the reverse framework and given an input image $\mathbf{D}_i$, a shape model H is said to be detected in it when $Pr(H|\mathbf{D}_i)>\phi$, where ϕ is some prespecified threshold parameter. We call $Pr(H|\mathbf{D}_i)$ the *evidence* in the forward framework. Expanding $Pr(H|\mathbf{D}_i)$ using the Bayes rule, it gives

$$Pr(H|\mathbf{D}_i)=\int Pr(H|\alpha,\beta,\mathbf{D}_i)p(\alpha,\beta|\mathbf{D}_i)d\alpha d\beta.\qquad(10)$$

Assuming that the model parameter $\mathbf{w}$ is independent of α when the input $\mathbf{D}_i$ is given[3] and H is independent of $\mathbf{D}_i$ and β when $\mathbf{w}$ is given, $Pr(H|\mathbf{D}_i)$ becomes

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

Further assuming that $Pr(H|\alpha)$ and $p(\mathbf{w}|\alpha)$ are constants, $Pr(H|\mathbf{D}_i)$ can be approximated as

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

In (12), it is noted that the first factor, $p(\mathbf{w}^*|\alpha^*,H)$, is simply the prior distribution in the reverse framework. The second factor, $p(\mathbf{w}^*|\beta^*,\mathbf{D}_i)$, is new and it represents the distribution of the model parameters $\mathbf{w}$ given the data $\mathbf{D}_i$. To formulate this factor using the data-generating-model assumption, we use the same spline model and the mixture of Gaussians formulation as before in the following section.

---

3. This assumption can be 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.
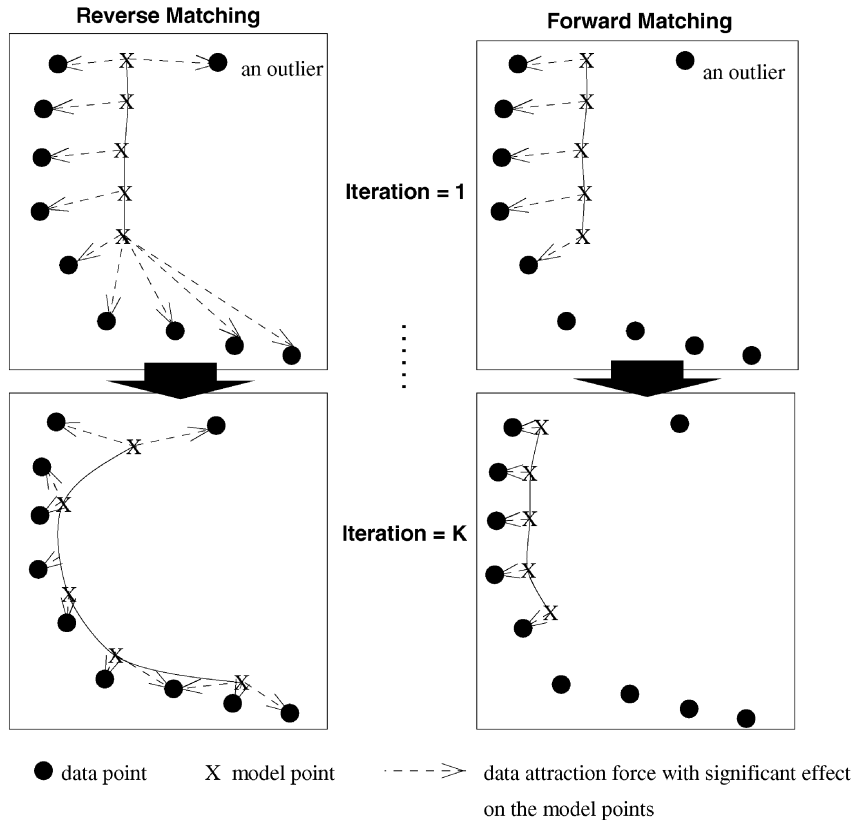
Fig. 1. Illustration of Proposition 1. The arrows indicate the major attracting forces acting on the model points. Note that reverse matching succeeds in extracting the lower portion of the data but is sensitive to the outlier, while forward matching only extracts the data neighboring to the model.

2.4.1   A New Data Mismatch Criterion

To impose the data-generating-model notion, a new data mismatch criterion based on the mixture of Gaussians formulation can be derived by interchanging the two summations in (4), that is

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

and the new "likelihood" becomes

$$p(\mathbf{w}, \mathbf{A}, \mathbf{T} | \beta, \mathbf{D}_i) = \frac{1}{Z_D^\S(\beta)} \exp(-E_D^\S(\mathbf{w}, \mathbf{A}, \mathbf{T}; \beta, \mathbf{D}_i)), \quad (14)$$

where $Z_D^\S(\beta) = \left(\frac{2\pi}{\beta}\right)^{N_g}$. According to (13), each point $\mathbf{m}_j$ along the spline is modeled by a uniformly weighted mixture of Gaussians with their means being the data $\mathbf{D}_i$. In this new definition of data mismatch, β is no longer interpreted as the noise level as in [3], but is related to the effective search area for each model point.

2.4.2   Forward Matching

According to (11), the optimal shape parameters $\{\mathbf{w}^*, \mathbf{A}^*, T^*\}$, can readily be estimated by maximizing $p(\mathbf{w}|\alpha, H)p(\mathbf{w}, \mathbf{A}, \mathbf{T}|\beta, \mathbf{D}_i)$, which is equivalent to minimizing $\alpha E_w(\mathbf{w}) + E_D^\S(\mathbf{w}, \mathbf{A}, \mathbf{T}; \beta, \mathbf{D}_i)$. The optimal α and β, according to (10), are estimated by maximizing $Pr(H|\alpha, \beta, \mathbf{D}_i)p(\alpha, \beta|\mathbf{D}_i)$ with respect to α and β. The EM algorithm is used again for the maximization problem. Due to the page limit, we only tabulate below (Table 1) the major formulae of the forward framework and compare them with those of the reverse framework. More details can be found in [2].

3   PROPERTIES OF THE REVERSE AND FORWARD FRAMEWORKS

For the sake of further discussion, the matching processes of the reverse and forward frameworks are referred to as *reverse matching* and *forward matching*, respectively. Similarly, the evidence measures of the two frameworks are referred to as *reverse evidence* and *forward evidence*. In the following, we compare in particular the matching and discriminating properties of the two frameworks using intuitive examples. Theoretical proofs of the properties can be found in [2].

Proposition 1 (Shape Matching Property). *Reverse matching has good data exploration capability while forward matching is good at finding some localized match.*

As illustrated in Fig. 1, this proposition reveals a dilemma between reverse and forward matching. Reverse matching manages to deform the model to a great extent so as to "explore" and "explain" all the input data, where outliers are also included. This implies that reverse matching is essential for extracting highly deformed shapes but is very sensitive to outliers, even though the outliers are distant from the target object. On the other hand, forward matching only deforms the model to match to some neighboring data points and is thus relatively insensitive to distant outliers. However, the consequence is that its data exploration capability is greatly reduced.

Proposition 2 (Shape Discriminating Property). *The reverse evidence does not penalize partially matched models but the forward evidence does.*

This proposition is illustrated intuitively in Fig. 2. It is noted that the unmatched model points do not contribute much in the overall reverse evidence but they do dominate in the calculation of

Reverse Evidence   **Forward Evidence**

● data point   X model point   - - - - - distance with significant contribution
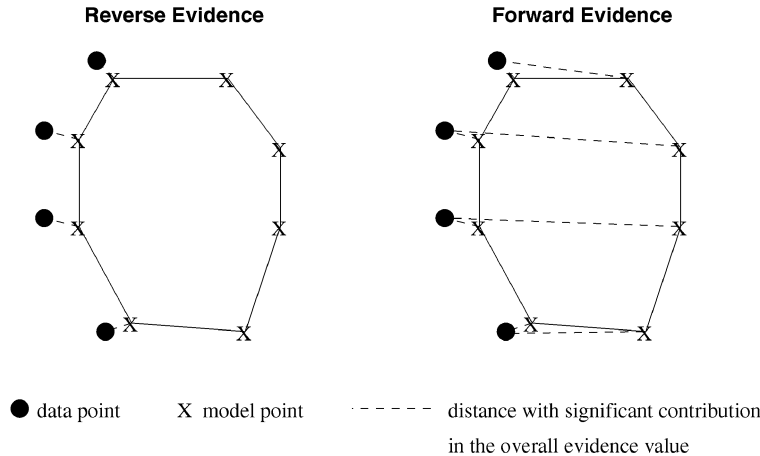in the overall evidence value

Fig. 2. Illustration of Proposition 2. The dashed lines indicate the major distance measures contributed to the overall model evidence value. It is clearly revealed that the unmatched model points are not penalized in the reverse evidence but are heavily penalized in the forward evidence.

---

1. Initialize an input model using chamfer-like matching.

2. Compute the data mismatch, E_D, and the sub-data mismatch, $E_D^\S$,

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 convergence is reached with the difference in $\mathbf{w}$ between two consecutive iterations less than a threshold ϵ'.

Fig. 3. Bidirectional matching algorithm.

the overall forward evidence. In other words, the reverse framework can easily misclassify some input which just happens to resemble a portion of the model shape (e.g., misclassify "C" as "O"), resulting in a high false alarm rate (also called the *subpart* problem in [3]). Nevertheless, this framework is good for extracting patterns containing broken lines or occluded parts. On the contrary, the forward framework *penalizes* models matching to target objects which are occluded but has the advantage that the subpart problem is solved implicitly.

4 BIDIRECTIONAL MATCHING ALGORITHM

Many existing deformable matching algorithms are similar to the approach adopted by the reverse framework and thus suffer from the problem of high sensitivity to outliers. Localization of search is then achieved by confining each model point to a search window of some predetermined size. Based on the duality relationships between the reverse and forward frameworks, we argue that given sufficiently good model initialization, forward matching should be able to ensure that the search is localized while reverse matching can allow the model to explore neighboring data points, especially for those connected to the matched data points. The success of Hausdorff matching gives the cue that taking the maximum of the data mismatch criteria of the two frameworks can be a good way to combine the two frameworks. Based on the idea, we propose an algorithm called *bidirectional matching* outlined in Fig. 3, in which the matching process alternates between the two frameworks according to the values of E_D and $E_D^\S$ until some convergence criterion is satisfied. According to our experimental results, the forward

framework "wins" most of the time during the earlier iterations until a localized shape is identified. During the later iterations, switching between the two frameworks occurs more often. An annealing schedule on the parameter β is required for convergence, the proof of which can be found in [2]. Fig. 4 illustrates a particular case showing the limitations of the two frameworks and the strength of the proposed bidirectional approach.

5 EXPERIMENTAL RESULTS AND DISCUSSION

We have performed experiments to evaluate and compare the matching performance of the forward, reverse[4] and bidirectional matching algorithms. In particular, character extraction from handwritten words is chosen as the test-bed since the shape of handwritings is highly nonrigid. Images sampled from the *bb* and *bs* subsets[5] of the CEDAR database are used as the test set. The data set altogether contains 633 handwritten city name images with stroke anamalies, closely cluttered characters and the printed background. As our focus is to test whether some target character shape can be detected and located in the input image, we assume that the identity of the leftmost character is known[6] and, thus, the

---

4. A search window of fixed size is applied to each of the model points so that the model point will only be attracted by black pixels within the window. Otherwise, reverse matching will try to fit the model to the whole input image.

5. In our experiments, all the images in the data set were preprocessed by simple thresholding and thinning, and rescaled to be of 50-pixel high.

6. 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.
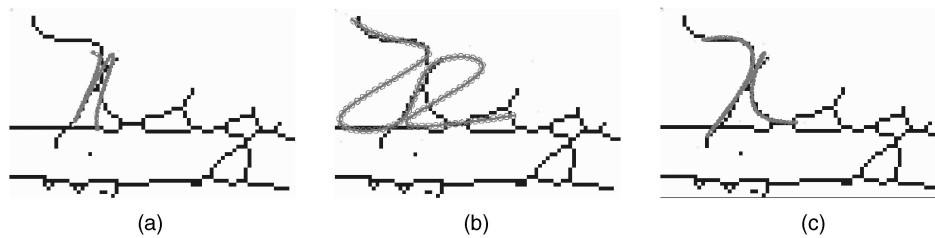
Fig. 4. Comparison of matching performance. (a) For forward matching, it is noted that the model "n" correctly locates the character "n" in the input data but cannot extract the character's fine details. (b) For reverse matching, the model is found to be severely deformed to fit the data as much as possible, including the undesirable outlier data. (c) For the proposed bidirectional matching, the highly nonrigid character "n" can be successfully extracted from the input data.
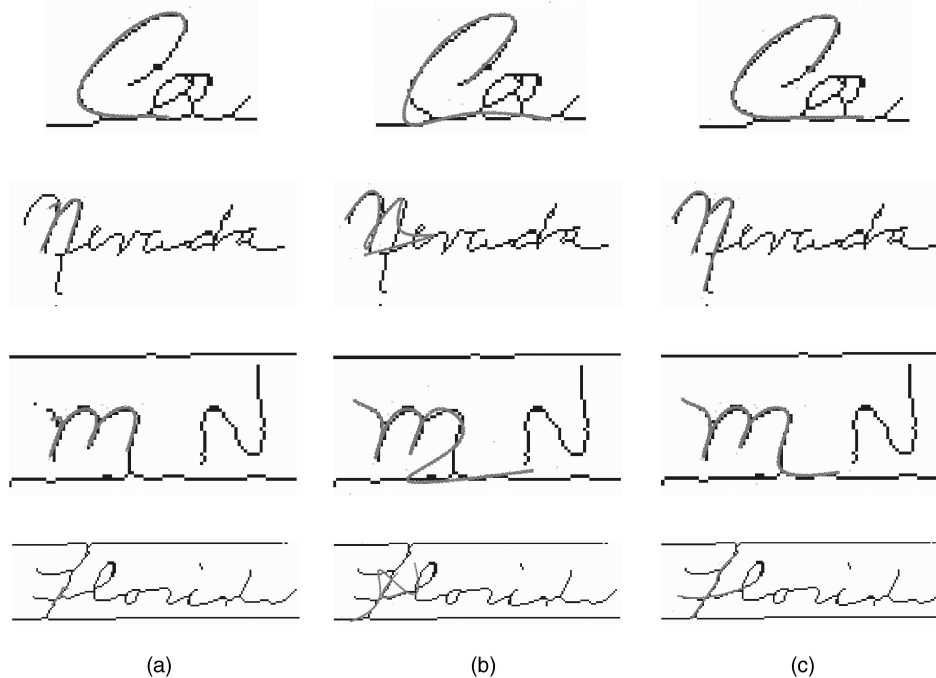

Fig. 5. Some matching results. (a) Forward matching, (b) reverse matching, and (c) bidirectional matching.

corresponding model can be used. Chamfer-like matching was adopted for model initialization and the three matching algorithms were then applied for subsequent matching. A successful match is defined as the case where the majority of the model points lie on the target character shape, which was judged by human visually. Some matching results are shown in Fig. 5. They reveal that reverse matching was highly sensitive to outliers due to cluttered characters and forward matching missed important character strokes for many cases, especially for highly deformed characters. By using bidirectional matching, the successful matching rate given a correct initialization is as high as 92.5 percent (88.5 percent for the *bb* data set and 95.3 percent for the *bs* data set). The overall matching results are tabulated in Table 2. Some failure cases for bidirectional matching are shown in Fig. 6. We believe that some additional smoothness constraints can be introduced to further enhance the matching accuracy.

In our experiments, the correct model initialization rate is 92.5 percent for the *bb* data set and 89.6 percent for the *bs* data set. The sources of errors include handwritings highly distorted from the target character shape, the scene being too complex, and the target character shape being subparts of some other characters (e.g., "C" is a subpart of "O"). Further research effort on model initialization for nonrigid shape is needed. For the computational time, based on our current implementation run on a Pentinum II-MMX (266 MHz) machine, the initialization step and the bidirectional matching step take about 3-7 seconds and 2-3 seconds,

respectively, for each image. The speed variation is mainly due to different degrees of model complexity of the input shapes. We believe that the computational performance can further be enhanced via either some careful data structure design for efficient implementation or some domain-specific heuristics for pruning the search space.

6   CONCLUSION

Two Bayesian frameworks for deformable pattern classification and detection are described in this paper and their matching and discriminating properties have been carefully analyzed. Based on

TABLE 2
Performance Comparison of the Three Matching Algorithms

	no. of images	bidirectional	forward	reverse
Accuracy (bb)	277	88.5%	51.2%	64.8%
Accuracy (bs)	356	95.3%	50.5%	73.2%
Overall	633	92.3%	50.8%	69.5%

(Given correct model initializations.)

Fig. 6. Some failure cases for bidirectional matching.

the relationships between the two frameworks and Hausdorff matching, we integrated the two frameworks using the idea of Hausdorff distance and proposed a new robust matching algorithm called the bidirectional deformable matching. The proposed algorithm can be considered as a deformable version of Hausdorff matching. For the *bb* and *bs* data sets in the CEDAR database with altogether 633 handwritten city name images, we were successful in achieving an overall accuracy of around 92 percent for matching the first characters of the city names.

APPENDIX

SIMILARITY BETWEEN THE MIXTURE-OF-GAUSSIANS FORMULATION AND THE HAUSDORFF DISTANCE

The mixture-of-Gaussians formulation used to describe data mismatch can in fact be considered as a "soft" version of the Hausdorff distance. Recall that, the reverse distance from $\mathbf{D}$ to $\mathbf{M}$ $\mathcal{H}(\mathbf{D}, \mathbf{M})$ is equal to $\max_l \min_j \|\mathbf{m}_j - \mathbf{y}_l\|$. Let $e_j^l \in \{0, 1\}$ denote an indicator variable which takes the value of only when $j = \arg\min_j \|\mathbf{m}_j - \mathbf{y}_l\|$. Then, $\mathcal{H}(\mathbf{D}, \mathbf{M})$ can be rewritten as $\max_l \sum_j e_j^l \|\mathbf{m}_j - \mathbf{y}_l\|$. Furthermore, if we use the L_∞ norm as the distance between $\mathbf{m}_j$ and $\mathbf{y}_l$, $\mathcal{H}(\mathbf{D}, \mathbf{M})$ becomes $\sum_l \sum_j e_j^l \|\mathbf{m}_j - \mathbf{y}_l\|^\infty$. By comparing it with the criterion used in the M-step for the reverse matching (according to (8)) $E_D' = \sum_l \sum_j h_j^l \|\mathbf{m}_j - \mathbf{y}_l\|^2 / 2$, it can be shown that they are very similar to each other. The formulation of h_j^l, according to its definition is in fact a *soft-min* function on $\|\mathbf{m}_j - \mathbf{y}_l\|$. Instead of using the L_∞ norm, the L_2 norm is used instead in E_D'. When the model and the data are close to each other, $\mathcal{H}(\mathbf{D}, \mathbf{M})$ and E_D' also become close to each other. Due to the symmetry of the problems, similar analogy also applies to the forward distance and the criterion used in forward matching.

ACKNOWLEDGMENTS

REFERENCES

[1] K.W. Cheung, D.Y. Yeung, and R.T. Chin, "A Bidirectional Matching Algorithm for Deformable Pattern Detection with Application to Handwritten Word Retrieval," *Proc. Seventh Int'l Conf. Computer Vision*, pp. 1105-1110, Sept. 1999.
[2] K.W. Cheung, "Bayesian Frameworks for Deformable Pattern Recognition and Retrieval.," PhD thesis, Dept. of Computer Science, the Hong Kong Univ. of Science and Technology, (http://www.comp.hkbu.edu.hk/william/pub/thesis.ps.gz). Jan. 2000.
[3] K.W. Cheung, D.Y. Yeung, and R.T. Chin, "A Bayesian Framework for Deformable Pattern Recognition with Application to Handwritten Character Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1382-1388, Dec. 1998.
[4] P. Dayan, G.E. Hinton, R.M. Neal, and R.S. Zemel, "Helmholtz Machine," *Neural Computation*, vol. 7, no. 5, pp. 889-904, Sept. 1995.
[5] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum-Likelihood from Incomplete Data via the EM Algorithm," *J. Royal Statistical Soc., Series B*, vol. 39, pp. 1-38, 1977.
[6] M.P. Dubuisson, S. Lakshmanan, and A.K. Jain, "Vehicle Segmentation and Classification Using Deformable Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 3, pp. 293-308, Mar. 1996.
[7] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge, "Comparing Images Using the Hausdorff Distance under Translation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850-863, Sept. 1993.
[8] A.K. Jain, Y. Zhong, and S. Lakshmanan, "Object Matching Using Deformable Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 3, pp. 267-278, Mar. 1996.
[9] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int'l J. Computer Vision*, vol. 1, no. 4, pp. 321-331, 1987.
[10] D.J.C. MacKay, "Bayesian Interpolation," *Neural Computation*, vol. 4, no. 3, pp. 415-447, 1992.
[11] M. Revow, C.K.I. Williams, and G.E. Hinton, "Using Generative Models for Handwritten Digit Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 6, pp. 592-606, June 1996.
[12] W. Rucklidge, *Efficient Visual Recognition Using the Hausdorff Distance.* Berlin: Springer, 1996.
[13] A.L. Yuille, D.S. Cohen, and P. Hallinan, "Feature Extraction from Faces Using Deformable Templates," *Int'l J. Computer Vision*, vol. 8, no. 2, pp. 99-111, Aug. 1992.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** http://computer.org/publications.