Handwritten Chinese Radical Recognition Using Nonlinear Active Shape Models

D. Shi, S.R. Gunn, and
R.I. Damper, *Senior Member*, *IEEE*

Abstract—Handwritten Chinese characters can be recognized by first extracting the basic shapes (radicals) of which they are composed. Radicals are described by nonlinear active shape models and optimal parameters found using the chamfer distance transform and a dynamic tunneling algorithm. The radical recognition rate is 96.5 percent correct (writer-independent) on 280,000 characters containing 98 radical classes.

Index Terms—Handwritten Chinese character recognition, active shape model, kernel principal component analysis, chamfer distance transform, dynamic tunneling algorithm.

1 Introduction

OFFLINE handwritten Chinese character recognition is an important, but very difficult pattern recognition problem because it involves complex structures, serious interconnection among the components, numerous pattern variations, absence of dynamic (pen movement) information, and a large number of characters. Traditionally, the problem has been approached through stroke extraction, but each character typically has a large number of strokes in a complex spatial relationship with one another. Fortunately, the many thousands of Chinese characters can be composed from a relatively small number of basic shapes, or *radicals* [1], in a specific spatial relationship. In this paper, we consider the use of nonlinear active shape models (ASMs) to decompose handwritten Chinese characters into their (position-dependent) radicals so as to simplify the problem of character recognition. After decomposition, recognition can be achieved by optimizing the combination of the component radicals. Fig. 1a illustrates an example character which can be decomposed into two radicals. The upper radical (which looks like a 2×2 grid), Fig. 1b, means *land*; the lower radical (which looks like a bending person working), Fig. 1c, means *labor*. This exposes how our ancestors described a man: as *land labor*!

Recently, Chung and Ip [2] applied snakes [3] to partition handwritten Chinese characters into radicals. The external energy in their work consists of two different functionals: displacement and intersection. The displacement functional avoids the snake deviating too much from the original template; the intersection functional avoids the intersection of strokes with the template. However, snakes are forced to partition the images by smoothness and some salient features. They did not discuss how to deal with the important problems of false salient features resulting from broken strokes and thinning algorithms.

Radical recognition can be treated as model-based shape extraction from a given image, avoiding the complexities of stroke extraction. However, deformable modeling is required to capture handwriting variability. Jain and Zongker [4] investigated the application of deformable templates to handwritten digit recognition. Their system represents a binary image in terms of its contour and, then, iteratively computes parameters of a continuous displacement function to map the contour template as closely as possible onto the edges of the target image. They reported a 99.25 percent recognition rate on a 2,000 character subset of NIST Special Database 1.

Cootes et al. [5] proposed active shape models to capture shape variations in an iterative search procedure, capable of locating the modeled structures in noisy, cluttered images—even if partially occluded. ASMs have similarities to snakes, in which a contour is fitted to the image evidence by minimizing an energy functional. However, a snake only has generic prior knowledge, such as smoothness. A much greater amount of prior information can be recovered from training sets and encoded within an ASM.

Active shape modeling extracts the eigenvectors, $\mathbf{U}$, of the training examples by principal component analysis (PCA). Thereafter, deformable models can be generated by adjusting shape parameters, $\mathbf{b}$, corresponding to the principal modes of variation:

$$\mathbf{\Gamma} = \mathbf{\Psi} + \mathbf{U}\mathbf{b}, \tag{1}$$

where $\mathbf{\Psi}$ is the mean vector of the training examples.

From (1), it is seen that the original ASMs are only suitable for representing linear variations. However, nonlinear shape variations are common in handwriting, such as different writing styles from person-to-person and time-varying distortion. To generalize ASMs to the nonlinear case, Sozou et al. [6] introduced polynomial regression. The basis for their method is to further reduce the residuals once a linear mode has been extracted by fitting a polynomial along the direction of the principal components. This requires, however, that the second eigenvector can be modeled as a function of the first; otherwise, implausible shapes may be generated. To solve this problem, they also applied a multilayer perceptron (MLP) [7] to find the nonlinear functionals among the shape parameters, $\mathbf{b}$. The many well-known disadvantages of the MLP method include the possibilities of over or underfitting the training data and the sensitivity to the choice of perceptron architecture and to the initial start point(s) for training.

Heap and Hogg [8] described applications where the training examples form a highly nonlinear space, which can be successfully represented by piecewise linear submodels. They adopted a two-level hierarchical approach: An initial global PCA is carried out to produce a lower dimensional space and, then, the linear subregions are constructed in this new lower space. They used the simplest way to constrain the valid shape region, i.e., Euclidean-distance-based clustering, which, however, cannot properly address the probabilistic nature of the problem.

Cootes and Taylor [9] represented shape variations with a probability density function which can be used to determine if a generated shape is plausible. They introduced a kernel-based density estimation technique, constructed from a large number of kernels. To reduce computational cost, they used a mixture of a small number of Gaussians, which can be fitted to the kernel estimate using a modification of the expectation-maximization algorithm. One problem is how to determine the optimal number of mixture components.

Romdhani et al. [10] were the first to apply kernel PCA to active shape models. Original input shapes are mapped to a feature space through kernel functions and, then, active shape modeling in feature space is employed as in the linear case. The expected ASMs are pre-images of the generated ASMs in feature space. By this means, they modeled nonlinear 2D shapes of nonrigid 3D objects, simultaneously recovering object pose at multiple views and across the view sphere. To address the problem of determination of valid shape regions (by an upper bound on the modulus in feature space), Twining and Taylor [11] designed a "proximity to data" functional and demonstrated its performance on both artificial and real-world nonlinear examples. However, such a function may not be continuous, which calls for special treatment in searching shape parameters.

The remainder of this paper is structured as follows: Section 2 describes how ASMs are improved by kernel PCA to capture nonlinear handwriting variations. Section 3 details the use of

- *D. Shi is with the School of Computer Engineering, Nanyang Technological University, Nanyang Avenue, Singapore 639798. E-mail: ASDMShi@ntu.edu.sg.*
- *S.R. Gunn is with the Image, Speech, and Intelligent Systems (ISIS) Group, Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, United Kingdom. E-mail: srg@ecs.soton.ac.uk.*
- *R.I. Damper is with the Image, Speech and Intelligent Systems (ISIS) Research Group, Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, England. E-mail: rid@ecs.soton.ac.uk.*

Manuscript received 27 July 2001; revised 7 May 2002; accepted 23 Aug. 2002. Recommended for acceptance by A. Khotanzad.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 114616.

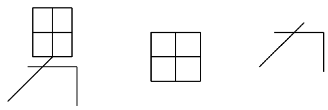

Fig. 1. Illustration of a Chinese character decomposed into its radicals. (a) The Chinese character for *man* is composed of (b) the upper radical for *land* and (c) the lower radical for *labor*.

nonlinear ASMs in handwritten character recognition. The chamfer distance transform is applied to get a satisfactory basin of attraction and the dynamic tunneling algorithm is employed to overcome problems of local minima when searching for the optimal shape parameters. In Section 4, we describe some experiments to determine the performance of these techniques on handwritten Chinese radical recognition, before concluding in Section 5.

2 Shape Representation by Kernel PCA

Principal component analysis is a technique for extracting structure from possibly high-dimensional datasets. It is readily performed by solving an eigenvalue problem or by using iterative algorithms. PCA is useful when the original pattern space can be accurately described by a subspace spanned by the first several principal eigenvectors. Often, the data lie in a subspace and, if this is linear, then a small number of principal components is sufficient to account for most of the variation.

Kernel PCA [12] extends linear PCA to nonlinear subspaces. Here, linear PCA is performed in a high-dimensional feature space $\mathcal{F}$ related to the input space by a nonlinear map $\Phi : \mathbb{R}^{2N} \rightarrow \mathbb{R}^{M}$. The number of nonlinear components obtained by kernel PCA can be greater than the original input dimension; the dimensionality of the feature space equals the number of training examples. However, the method confers no advantage if the data lie in a linear subspace. The main challenge is to choose an appropriate nonlinear transformation.

We start with a training set of examples for each radical $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_M\}$, represented by N landmark points with 10 points used for each stroke. A line is drawn between the (manually determined) start and termination points of a stroke, and intervening points are automatically determined. Hence: $\mathbf{e}_k = (x_{k0}, y_{k0}, \ \ldots, x_{k(N-1)}, y_{k(N-1)})^T$. The mean vector of the set is then defined by $\mathbf{\Psi} = \frac{1}{M}\sum_{k=1}^{M} \mathbf{e}_k$. In the original active shape models, $\mathbf{\Psi}$ is calculated after the training examples are aligned and, then, each training example is rotated and scaled into the tangent space so as to minimize the distance between them. This process iterates until convergence [5], [9]. However, such an alignment strategy does not work well with handwritten Chinese radicals. One of the reasons is that the local rotation may lead to different radicals. So:

1. Characters are deskewed based on the document layout analysis;
2. Each segmented character is normalized to a (64×64) dot matrix;
3. In recognition, the unknown normalized character image is rotated to a few positions and all the radical classes are matched with the rotated images.

The details of kernel PCA applied to ASMs are given in [10]. The projections of a data point $\mathbf{e}$ onto the eigenvectors $\mathbf{V}^k$ in the feature space $\mathcal{F}$ are defined as:

$$\beta_k(\mathbf{e}) = \mathbf{V}^{k^T}\widetilde{\mathbf{\Phi}}(\mathbf{e}) = \sum_{i=1}^{M'} \alpha_i^k \widetilde{\mathbf{\Phi}}(\mathbf{e}_i)^T \widetilde{\mathbf{\Phi}}(\mathbf{e}) = \sum_{i=1}^{M'} \alpha_i^k \widetilde{K}(\mathbf{e}_i, \mathbf{e}),$$

where M' is the number of principal components (modes). Any example in the training set can be approximated using the mean vector and a weighted sum of these deviations obtained from the first M' modes. Here, $\widetilde{\mathbf{\Phi}}(\mathbf{e}_i)$ is the centralized point in the feature space $\mathcal{F}$ corresponding to the vector $\mathbf{\Psi}$. However, we have no centralized data in $\mathcal{F}$, so we cannot compute $\widetilde{K}_{ij} = \widetilde{\mathbf{\Phi}}(\mathbf{e}_i) \cdot \widetilde{\mathbf{\Phi}}(\mathbf{e}_j)$ directly, but we can do so relying on the noncentered counterpart $\mathbf{K}$. Defining the notations $1_{ij} = 1$ for all i, j, and $(1_M)_{ij} = 1/M$, we compute [13]: $\widetilde{K}_{ij} = (\mathbf{K} - \mathbf{1}_M\mathbf{K} - \mathbf{K}\mathbf{1}_M + \mathbf{1}_M\mathbf{K}\mathbf{1}_M)_{ij}$.

Our purpose is to build up models for each class, which requires approximate representations of the data in input space rather than in feature space. To this end, by introducing shape parameters we can also generate ASMs on the basis of kernel PCA with the following two steps:

1. Generate active shape models in feature space with the mean vectors $\mathbf{\Psi}$. We define an operator $P_{M',b}$ by:

$$P_{M',b}\widetilde{\mathbf{\Phi}}(\mathbf{\Psi}) = \sum_{k=1}^{M'} \beta_k(\mathbf{\Psi}) b_k \mathbf{V}^k.$$

2. Find the active model $\mathbf{\Gamma}$ which is a pre-image in the feature space so as to minimize:

$$\begin{aligned}\rho(\mathbf{\Gamma}) &= ||P_{M',b}\widetilde{\mathbf{\Phi}}(\mathbf{\Psi}) - \widetilde{\mathbf{\Phi}}(\mathbf{\Gamma})||^2 \\ &= \widetilde{\mathbf{K}}(\mathbf{\Gamma}, \mathbf{\Gamma}) - 2\sum_{k=1}^{M'} b_k \beta_k(\mathbf{\Psi}) \sum_{i=1}^{M} \alpha_i^k \widetilde{\mathbf{K}}(\mathbf{e}_i, \mathbf{\Gamma}) + \widetilde{\mathbf{K}}(\mathbf{\Psi}, \mathbf{\Psi}).\end{aligned} \tag{2}$$

Points in the input space will be mapped to a hyperplane in the feature space (Fig. 2). A point in the feature space will move away from this hyperplane under linear operations (e.g., component truncation and shape parameter). Hence, its pre-image will be approximated by that corresponding to its nearest hyperplane point in the feature space.

3 Radical Recognition with Nonlinear ASMs

With models generated by adjusting shape parameters via (2), radicals can be matched to the observed character by minimizing the distance between the radical model and the character skeleton.

3.1 Chamfer Distance Transform

When adjusting shape parameters, a satisfactory basin of attraction is needed to find the optimal shape parameters efficiently: We use the chamfer distance transform [14] for this. An important property of the transform is its ability to handle noisy and distorted data as the edge points of one image undergo a set of parametric transformations which describe how the images can be geometrically distorted in relation to one another. We have previously applied the chamfer distance transform to character recognition with some success [15], but only for the linear PCA case.

A shape model, denoted $\mathbf{\Gamma}(\mathbf{b})$, can be generated by changing parameters $\mathbf{b}$. In the matching phase, the shape models for all the classes are superimposed upon the target image. Treating the target image as a functional of shape models, i.e., $I(\gamma_j(\mathbf{b}))$, the energy of a shape model is given by:

$$E(\mathbf{\Gamma}(\mathbf{b})) = \sum_{j=1}^{N} D_{\text{chamfer}}(I(\gamma_j(\mathbf{b}))), \tag{3}$$

where $\gamma_j()$ is the jth point of $\mathbf{\Gamma}()$ located in two-dimensional position (x_j, y_j) and $D_{\text{chamfer}}()$ denotes the chamfer distance transform. Hence, searching for the optimal shape is equivalent to minimizing the energy $E()$.

3.2 Gradient Descent with Dynamic Tunneling Algorithm

One way to minimize the energy $E()$ in (3) is to use gradient descent. Assuming initial shape parameters $\mathbf{b}$, an initial model $\mathbf{\Gamma}()$ is obtained. Given a test character image, it is matched against each model and the corresponding minimum energy and the test image classified into the class with the overall minimum energy.

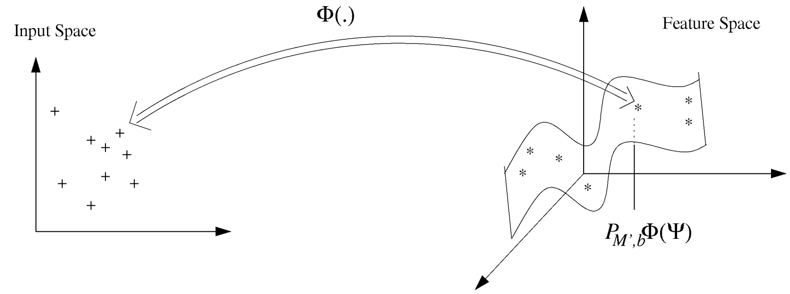

Fig. 2. Illustration of pre-image in input space.

To overcome problems with local minima, we use a dynamic tunneling algorithm [16] based on a physical analogy to the quantum-mechanical tunneling of a particle through a potential barrier. The degree of tunneling allowed is a function of time, t, increasing during iterative search. The search procedure can jump to another basin of attraction where the new, initial search point is even lower in energy. From this new starting point, gradient descent can again be used to find a lower minimum. RoyChodhury et al. [17] show that a sensible step size is $(\frac{2t}{3\rho})^{3/2}$, where ρ is the strength of repeller/attractor. To prevent implausible deformations, the shape parameters are only allowed to vary by $\pm 3\sqrt{\lambda_k}$, where λ_k is the eigenvalue of the kth principal component. The algorithm for searching for the kth shape parameter, b_k, is given by [15] as:

Step 1 $b_k^* = 0$, $b_k^+ = 0$.
Step 2 Search for local optimum with gradient descent from point b_k^* to obtain new current global minimum b_k^*.
Step 3 Dynamic tunneling phase begins, setting the tunneling time to $t = 0$.
Step 4 Update: $t = t + 1$.
Step 5 In the case of positive direction, select the point

$$b_k = b_k^* + \sqrt{\left(\frac{2t}{3\rho}\right)^3};$$

In the case of negative direction, select the point $b_k = b_k^* - \left(b_k^+ + \sqrt{\left(\frac{2t}{3\rho}\right)^3}\right)$, i.e., treating b_k^+ as an offset. ($\rho = 7,000$ in our current experiments.)
Step 6 In the case of positive direction, if $b_k > 3\sqrt{\lambda_k}$, then $b_k^+ = b_k^*$, and go to **Step 3**;
In the case of negative direction, if $b_k < -3\sqrt{\lambda_k}$, go to **Step 10**.
Step 7 Calculate the energy function E_Γ in b_k.
Step 8 If $E_\Gamma(b_k) > E_\Gamma(b_k^*)$ go to **Step 3** to continue dynamic tunneling phase.
Step 9 Otherwise, go to **Step 2** to start a gradient descent phase.
Step 10 End.

4 Experiments and Results

Radical recognition experiments have used a subset of the database collected by Harbin Institute of Technology and Hong Kong Polytechnic University, which comprises a total of 751,000 loosely constrained handwritten Chinese characters. There are 3,755 categories written by 200 different writers [15]. Experiments were carried out with Gaussian kernels, $k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma^2}\right)$, which gave better results than polynomial or spline kernels. Fig. 3a shows an example character skeleton and its chamfer transform, while Fig. 3b shows a typical mean model for characters from one of the classes. Fig. 3c shows the ASMs generated as the number of principal components varies from 1 to 5, whereupon the chamfer distance between the models and the input skeleton is 198, 104, 43, 32, and 26, respectively. Since our purpose is not accurate reconstruction of the image but pattern classification, the number of principal components can be truncated such that

$$\frac{\sum_{i=1}^{M'} \lambda_i}{\sum_{j=1}^{M} \lambda_j} > 90 \text{ percent.}$$

We now compare our proposed nonlinear ASM method (Method 1 below) with some representative other works on radical recognition.

Method 1. Using the *nonlinear* PCA approach described above, experiments were conducted on 98 radicals covering 1,400 loosely constrained Chinese character categories written by 200 different writers, i.e., 280,000 characters. The total number of radicals is approximately 590,0000. There are 60 randomly selected character examples with semiautomatic landmark labeling for training each radical model. Although $60 \times 98 = 5,880$ radicals have been seen in training, this is small relative to the total of 590,000 (less than 1 percent) so that we consider all examples are unseen. This also prevents us to having to exclude from the test set a large number of complete characters for which only one (out of up to four) of its radicals has been seen in training.

Method 2. Active shape models with *linear* PCA [15]. The experiments were conducted on the same test set as for Method 1.

Method 3. The stroke-based approach of Wang and Fan [18]. Their experiments for radical extraction were conducted on just 32 radical classes in 1,856 test characters from six writers.

Method 4. The snake-partitioning approach of Chung and Ip [2]. Their experiments were conducted on 100 character categories written by 10 people (i.e., 1,000 test examples only). They considered and reported results on the six most common radical combination schemes (vertical, left-down, surrounding, horizontal, up-left, and cover) only.

For ASM-based Methods 1 and 2, a correct recognition is defined as the correct radical identified at the correct position in the image. Table 1 shows that these two methods are easily the best among the existing radical approaches. They deal with the largest

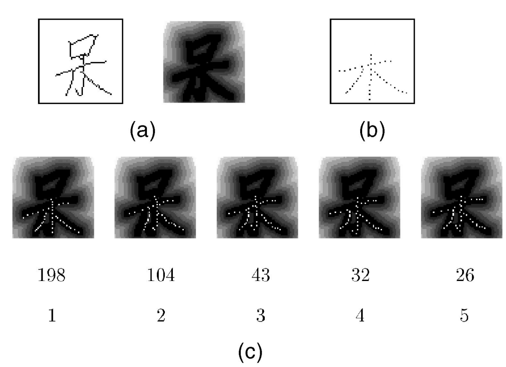

Fig. 3. Active model generation by adjusting shape parameters. (a) An example character skeleton and its chamfer transformed image. (b) Mean model of lower radical. (c) Active models generated for the lower radical with the number of principal components varying from 1 to 5.

TABLE 1
Performance of Different Approaches to Chinese Radical Recognition

	Test set size (characters)	Number of radicals trained	Number of writers	% radicals correct
Method 1 (Nonlinear ASM)	280,000	98	200	96.5
Method 2 (ASM, linear PCA)	280,000	98	200	94.2
Method 3 (stroke-based)	1856	32	6	92.5
Method 4 (snake partitioning)	1000	100	10	68.0–95.0

number of radicals on a test set which is significantly larger than other workers have used, yet still achieve better matching rates. (The range of values for Method 4 relates to Chung and Ip's report of separate figures for each of their six radical combination schemes.) If we assume a binomial distribution of errors with error rate p, then the standard deviation of the error distribution in our experiments is $\sqrt{np(1-p}$ with $n = 590,000$. Thus, the 99 percent confidence interval (corresponding to plus or minus three standard deviations) is 96.4 to 96.6 percent and 94.1 to 94.3 percent for Methods 1 and 2, respectively. It follows that Method 1 gives significantly better results than Method 2, indicating the importance of modeling nonlinear handwriting variations.

The advantage of our proposed ASM approach is the capability to handle nonlinear writer variations with only a small number of shape parameters. Our models also avoid stroke extraction, which is a source of considerable difficulty in handwriting recognition. Although the method requires relatively long matching times (0.4 seconds per character on a Pentium III PC, 455 MHz, 128 MB RAM) caused by working at the pixel level and shape-parameter searching, given the pace of development of high-performance computing, this is unlikely to remain a problem in the future. We also require landmark labels, which at present are obtained semi-automatically.

5 Conclusions and Future Work

The paper has introduced a novel method of handwritten Chinese radical recognition based on nonlinear active shape models. Kernel principal component analysis is employed to capture the main nonlinear variation of the training examples around the mean vector. In matching, the dynamic tunneling algorithm is incorporated with gradient descent to search for the optimal shape parameters in terms of chamfer distance minimization. The experimental results show that our method achieves superior performance compared with existing similar representative works.

We have concentrated on radical recognition as the first stage of two-stage Chinese character recognition. The second stage requires us to find the optimal radical combination, which can be done with standard methods based on a precompiled lexicon and dynamic programming (Viterbi decoding). Errors in radical recognition are unlikely to propagate to a large degree since the prior structural knowledge embodied in the lexicon allows only legal (semantically meaningful) combinations.

Our future work also includes (fully) automatic landmark point labeling. One possible approach is stroke extraction associated with labeled reference radicals. A reference radical is created for each class, in which all the strokes involved are labeled. An alternative is to solve reparameterization, alignment, and clustering problems simultaneously [19], although further investigation is required to deal with touching/broken strokes in our case of Chinese character recognition.

References

[1] S.K. Chang, "An Interactive System for Chinese Character Generation and Retrieval," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 3, no. 3, pp. 257-265, 1973.

[2] F. Chung and W.W.S. Ip, "Complex Character Decomposition Using Deformable Model," *IEEE Trans. Systems, Man, and Cybernetics—Part C: Applications and Reviews,* vol. 31, no. 1, pp. 126-132, 2001.

[3] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int'l J. Computer Vision,* vol. 1, no. 4, pp. 321-331, 1987.

[4] A.K. Jain and D. Zongker, "Representation and Recognition of Handwritten Digits Using Deformable Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 12, pp. 1386-1391, Dec. 1997.

[5] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Garaham, "Active Shape Models—Their Training and Application," *Computer Vision and Image Understanding,* vol. 61, no. 1, pp. 38-59, 1995.

[6] P.D. Sozou, T.F. Cootes, C.J. Taylor, and E.C. Di Mauro, "Non-Linear Generalization of Distribution Models Using Polynomial Regression," *Image and Vision Computing,* vol. 13, no. 5, pp. 451-457, 1995.

[7] P.D. Sozou, T.F. Cootes, C.J. Taylor, and E.C. Di Mauro, "Non-Linear Point Distribution Modelling Using a Multi-Layer Perceptron," *Proc. British Machine Vision Conf.,* pp. 107-116, 1995.

[8] T. Heap and D. Hogg, "Improving Specificity in PDMs Using a Hierarchical Approach," *Proc. British Machine Vision Conf.,* pp. 80-89, 1997.

[9] T.F. Cootes and C.J. Taylor, "A Mixture Model for Representing Shape Variation," *Image and Vision Computing,* vol. 17, no. 8, pp. 567-574, 1999.

[10] S. Romdhani, S. Gong, and A. Psarrou, "A Multi-View Nonlinear Active Shape Model Using Kernel PCA," *Proc. British Machine Vision Conf.,* pp. 483-492, 1999.

[11] C.J. Twining and C.J. Taylor, "Kernel Principal Component Analysis and the Construction of Non-Linear Active Shape Models," *Proc. British Machine Vision Conf.,* vol. 1, pp. 23-32, 2001.

[12] B. Schölkopf, A.J. Smola, and K. Müller, "Kernel Principal Component Analysis," *Advances in Kernel Methods,* B. Schölkopf, C.J.C. Burges, and A.J. Smola, eds., pp. 327-352, MIT Press, 1998.

[13] B. Schölkopf, S. Mika, A. Smola, G. Rätsch, and K. Müller, "Kernel PCA Pattern Reconstruction via Approximate Pre-Images," *Proc. Eighth Int'l Conf. Artificial Neural Networks,* L. Niklasson, M. Bodén, and T. Ziemke, eds., pp. 147-152, 1998.

[14] G. Borgefors, "Hierarchical Chamfer Matching: A Parametric Edge Matching Algorithm," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 10, no. 6, pp. 849-865, June 1988.

[15] D. Shi, S.R. Gunn, and R.I. Damper, "A Radical Approach to Handwritten Chinese Character Recognition Using Active Handwriting Models," *Proc. Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 670-675, 2001.

[16] Y. Yao, "Dynamic Tunneling Algorithm for Global Optimization," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 19, no. 5, pp. 1222-1230, 1989.

[17] P. RoyChodhury, Y.P. Singh, and R.A. Chansarkar, "Hybridization of Gradient Descent Algorithms with Dynamic Tunneling Methods for Global Optimization," *IEEE Trans. Systems, Man, and Cybernetics—Part A: Systems and Humans,* vol. 30, no. 3, pp. 384-390, 2000.

[18] A.B. Wang and K.C. Fan, "Optical Recognition of Handwritten Chinese Characters by Hierarchical Radical Matching Method," *Pattern Recognition,* vol. 34, no. 1, pp. 15-35, 2001.

[19] N. Duta, A.K. Jain, and M.P. Dubuisson-Jolly, "Automatic Construction of 2D Shape Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 5, pp. 433-446, May 2001.