

Hierarchical Active Shape Models, Using the Wavelet Transform

Christos Davatzikos*, Xiaodong Tao, and Dinggang Shen

Abstract—Active shape models (ASMs) are often limited by the inability of relatively few eigenvectors to capture the full range of biological shape variability. This paper presents a method that overcomes this limitation, by using a hierarchical formulation of active shape models, using the wavelet transform. The statistical properties of the wavelet transform of a deformable contour are analyzed via principal component analysis, and used as priors in the contour's deformation. Some of these priors reflect relatively global shape characteristics of the object boundaries, whereas, some of them capture local and high-frequency shape characteristics and, thus, serve as local smoothness constraints. This formulation achieves two objectives. First, it is robust when only a limited number of training samples is available. Second, by using local statistics as smoothness constraints, it eliminates the need for adopting ad hoc physical models, such as elasticity or other smoothness models, which do not necessarily reflect true biological variability. Examples on magnetic resonance images of the corpus callosum and hand contours demonstrate that good and fully automated segmentations can be achieved, even with as few as five training samples.

Index Terms—Active shape model, deformable contours, the wavelet transform.

I. INTRODUCTION

DEFORMABLE shape models have played an important role in the medical image analysis literature during the past 15 years. They have been used for segmentation and deformable registration of biomedical images. They overcame some of the limitations of earlier “bottom-up” approaches by imposing higher level constraints on the kinds of shapes that are likely to be encountered in biomedical images. They typically impose some form of local smoothness constraint [1] via elastic forces or some combination of local and global constraints [2]. Computationally efficient finite-element formulations of elastic deformable models have also been proposed [3]. Extensive application of these models soon revealed one of their limitations: elastic models are often too flexible, and they can be trapped by spurious edges or by edges adjacent to the structure of interest, thus converging to a suboptimal, and often very poor, solution. Therefore, a very careful initialization is

typically necessary, rendering these models subjective, as they depend on human interaction, and less practical for routine use.

Some of the limitations of deformable shape models were overcome by models that impose constraints not pertaining to some physical model (e.g., elastic), but rather to a statistical shape model derived from a training set. In [4] and [5], a frequency-based model was proposed, and in [6] a model using principal component analysis (PCA) was used. Other investigators subsequently used these models with some success [7]–[12]. Active shape models (ASMs) certainly mitigated the problem of local minima, by restricting the possible configurations a deformable shape can assume. However, they themselves suffer from a fundamental limitation: they often restrict the deformable shape too much, particularly if they have been trained on a relatively small number of samples. This is because the number of eigenvectors that can be used to represent any shape is at most equal to the number of training samples minus one, whereas the number of points comprising the shape might be two or more orders of magnitude higher. In other words, it is difficult to estimate a high-dimensional probability distribution of a shape from a relatively small number of samples. As a result, the subspace of “allowable shapes” spanned by the relatively few eigenvectors limits the ability of an ASM to follow the finer details of a shape. Combinations of physical and statistical shape models have also been proposed [6], [12], but they suffer from the limitations of each individual component of the model, depending on the relative weights assigned to the physical versus the statistical prior.

Our premise is that physical models, although often offer a convenient way to impose shape priors, do not necessarily reflect the variation of real biological shapes. Accordingly, we present a model that is based entirely on statistical shape priors. A hierarchical formulation of ASMs allows our model to capture both the global shape characteristics and the finer local details, thus overcoming current limitations of ASMs. Experimental results demonstrate that the proposed hierarchical model performs substantially better than the standard ASM, if the ASM is trained on relatively small number of training samples.

A. Overview

Deformable contours and deformable surfaces are sometimes represented in parametric forms by ordered lists of landmark points, or structured meshes. Classic active shape model is built from the statistics of the positions of landmark points on the contours in a set of training samples. The model is iteratively fitted to an image from a rough starting approximation [13]. A multiresolution implementation for image segmentation using active shape model was proposed [14] to improve the robustness, accuracy and speed. In that method, the image to be segmented is smoothed using Gaussian kernels with different sizes so that a

Manuscript received October 15, 2001; revised September 10, 2002. This work was supported in part by the National Institutes of Health (NIH) under Grant R01 AG14971–03. *Asterisk indicates corresponding author.*

*C. Davatzikos is with the Section for Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: hristos@rad.jhu.edu).

X. Tao is with the Section for Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104 USA and also with Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218 USA.

D. Shen is with the Section for Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104 USA.

Digital Object Identifier 10.1109/TMI.2003.809688

sequence of multiscale versions of the testing image is obtained. The active shape model is applied first to the coarse, low resolution version of the image, and then refined on higher resolution versions. In the process of image segmentation, the shape model is the same for different scales. Therefore, the model space for different scales remains the same.

In this paper, we propose new methods for building the statistical shape model for planar shapes based on multiresolution representation of the deformable contours. Multiresolution representation of curves and surfaces has been studied in computer graphics community for more than two decades. Methods based on subdivision [15], scale space representation [16] and hierarchical splines [17] have been used extensively in geometric modeling and computer vision. Recently, wavelet based multiresolution representation of curves and surfaces has received much attention [18], [19], and [20], because wavelet coefficients provide a convenient way to describe parametric curves and surfaces in different resolutions. In this paper, we use wavelet decomposition of the coordinates of the landmark points of training shapes to build hierarchical active shape models and apply the models for segmentation of callosal boundaries in the midsagittal sections of magnetic resonance images (MRIs) of human brains and contours of hand images.

The remainder of the paper is organized as follows: Section II provides a detailed description of the hierarchical active shape models. In this section, two methods are described. One method uses heuristic partition of the shape to achieve shape hierarchies, and the other one uses wavelet decomposition. Section III shows some experimental results. Finally Section IV discusses those results and gives a brief summary.

II. HIERARCHICAL ACTIVE SHAPE MODELS

Assume that N training samples are available in the form of sets of K corresponding landmarks in the D -dimensional space. Assume, also, that standard Procrustes alignment [21] has been performed on the training samples to eliminate the variations introduced by translation, scaling, and rotation. Let the vectors \mathbf{x}_n , $n = 1, \dots, N$, be formed by concatenating the coordinates of the K landmark points of the n th sample. Let $\mathbf{e}_1, \dots, \mathbf{e}_{N-1}$ be the eigenvectors corresponding to the nonzero eigenvalues of the covariance matrix, \mathbf{C} , calculated from these vectors. (There could, in principle, be fewer than $N - 1$ eigenvectors with nonzero eigenvalues, but this is rarely the case when only a limited number of samples is available, which is the focus of our paper.) If, as is typically the case, $N \ll D \cdot K$, it is likely that the space, \mathcal{S} , spanned by $\mathbf{e}_1, \dots, \mathbf{e}_{N-1}$ does not represent the full range of shape variation one expected to find in biomedical images. Therefore, in a standard ASM, which includes a step in which tentative shapes are projected onto \mathcal{S} , final shape reconstructions often appear to lack fine details. Clearly, the more samples are available, the larger \mathcal{S} is, and the more faithful the shape reconstruction is.

Assume, now, that we are only interested in representing a segment of a shape, comprising K' points, rather than the whole shape (see Fig. 1). If K' is small enough, then N samples will be adequate to represent the variability of that segment, thus allowing the model to locally capture the fine details of the segment. This would be achieved, however, at the expense of dis-

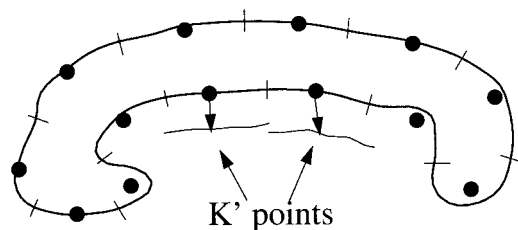


Fig. 1. A contour can be partitioned into smaller segments. The global shape characteristics can be captured by a statistical prior relating the centers of mass (solid circles) of all segments. Finer and local shape characteristics can be captured by statistical priors for each segment individually. When the global prior is used to reposition the center of mass points, the respective segments are repositioned rigidly by the same displacement vector. This might lead to discontinuities. In order to remedy this situation, we use overlapping segments, with a continuous blending.

regarding spatial relationships between points of that segment and points of other segments of the shape. Such relationships can be captured by a statistical prior relating the positions of the segments relative to each other. For example, one can apply a standard ASM on the centers of mass of these segments. This is one of the simplest forms of a hierarchical ASM, and is referred to as *Method 1* next. A more formal framework is the wavelet decomposition of a contour or surface representing the shape, followed by estimation of the statistics of the resulting expansion coefficients. This is referred to as *Method 2* below. Although Method 1 is somewhat heuristic, it is presented along with Method 2, because it provides a more intuitive explanation of the framework of hierarchical ASM.

A. Hierarchical ASM Based on Heuristic Partition of the Shape

We define a deformable contour as a collection of M segments, each comprising K' points. Let \mathbf{y}_m , $m = 1, \dots, M$, be a vector corresponding to the m th segment, and formed by concatenating the coordinates of all points in that segment. Let, also, \mathbf{c}_m be the center of mass of the points on the m th segment. In one of its simplest forms, a hierarchical ASM can be represented by two levels. At the bottom level, the covariance matrix of \mathbf{y}_m is used as prior for the deformation of the m th segment, thus imposing some sort of local smoothness constraint. At the top level, statistics of the vector $\mathbf{c} = [\mathbf{c}_1^T, \dots, \mathbf{c}_M^T]^T$ are collected from the training set, and are used as prior for more global shape properties. The algorithm is as follows.

- Step 1) Move each point of the contour toward a nearby target location (e.g., toward a nearby edge);
- Step 2) Calculate the center of mass of each segment of the contour, forming the vector \mathbf{c} ;
- Step 3) Project \mathbf{c} to the space formed by the respective eigenvectors of the top level of the statistical prior. Contour segments are translated according to their respective centers of mass (see Fig. 1);
- Step 4) Project the coordinate vector of each individual segment to its respective subspace spanned by the eigenvectors for that segment;
- Step 5) Go to Step 1.

For simplicity, we have omitted standard steps of the ASMs that include mapping of a shape into and out of a model space, by accounting for overall pose and size differences.

In practice, we implement this algorithm using overlapping segments, in order to avoid discontinuities that can be introduced at connecting points (see Fig. 1). Although we used two levels of statistical priors in our description above, the method can be generalized to include any number of levels, so that at each level, the length of the vector on which the statistical prior is calculated is small enough for the number of available samples. “Small enough” depends not only on the number of training samples, but on the compactness of the underlying distributions, as well. Therefore, it will vary from one application to another.

B. Hierarchical ASM Based on Wavelet Decomposition

Method 1 effectively divides the spectrum of the parametric representation of the deformable contour into two levels, in a rather crude way. The first level is a coarse sampling of the contour. It corresponds to the contour represented by the center of mass of the segments and it can be thought as a subsampled contour following spatial smoothing by a filter with “box” type impulse response. The second level contains the details of the contour. It is represented by the local shape properties of the contour segments. Below Method 2 formalizes and generalizes this idea, by using a wavelet transform.

1) *Wavelets and Wavelet Transform:* Wavelets, $\psi_{a,b}$, $a, b \in Z$, are L^2 functions generated from a *mother wavelet* ψ by dilations and translations [22]. In this paper, we only consider the discrete wavelets generated using dilations by powers of two and integer translations

$$\psi_{a,b} = \sqrt{2^{-a}}\psi(2^{-a}t - b), a, b \in Z. \quad (1)$$

Because wavelets are well localized in both space and frequency domains, they are suitable for representing local frequency components in a given signal. The wavelet transform is a tool for applying wavelets to perform localized analysis of signals [23]. It has been applied with great success to a wide range of signal/image processing and analysis problems, such as signal compression [24], denoising [25], and pattern recognition [26].

Instead of the heuristic partition of the spatial domain described in Section II-A, the wavelet transform provides an elegant way to perform a scale-space decomposition. In the context of this paper, we use a logarithm tree 2-band wavelet packet to divide the space-frequency domain. For a p -level wavelet packet, we divide the space-frequency domain into $B = 2^p$ bands (see Fig. 2).

2) *Analysis Steps Using Wavelet Transform:* Now, let us consider the wavelet transform of the deformable contours. As stated before, we assume that N training samples are available in the form of sets of K corresponding landmarks. Assume that the x -, y -coordinates of the k th landmark of the n th training sample after Procrustes alignment are $(u_{n,k}, v_{n,k})$, $n = 1, \dots, N$ and $k = 1, \dots, K$. Then each training sample can be represented by its x -, y -coordinate vectors

$$\mathbf{u}_n = [u_{n1}, u_{n2}, \dots, u_{nK}]^T \quad (2)$$

$$\mathbf{v}_n = [v_{n1}, v_{n2}, \dots, v_{nK}]^T, n = 1, \dots, N. \quad (3)$$

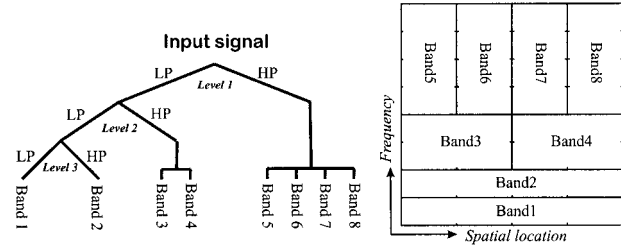


Fig. 2. Three-level logarithm tree 2-band wavelet packet. The wavelet transform is iterated for the low-pass branches (marked as LP in the figure). The outputs of the high-pass branches (marked as HP in the figure) are divided into a number of bands. The number of bands for each high-pass branch is determined by the level at which the high-pass filtering is performed. In this way, for a p level wavelet decomposition the frequency domain is divided into $B = 2^p$ bands. The space-frequency domain partition of the signal is shown on the right.

By applying wavelet transform to \mathbf{u}_n and \mathbf{v}_n , we get their wavelet coefficients

$$\mathbf{c}_n = \mathcal{W}(\mathbf{u}_n) = [c_{n,i}, i = 1, \dots, L] \quad (4)$$

$$\mathbf{d}_n = \mathcal{W}(\mathbf{v}_n) = [d_{n,i}, i = 1, \dots, L] \quad (5)$$

where $\mathcal{W}(\bullet)$ denotes the wavelet decomposition. Note that \mathbf{c}_n and \mathbf{d}_n are not necessarily K -dimensional vectors. Their dimension, L , depends on what scaling/wavelet functions we use in the transform. In this paper, we use a p -level logarithm tree 2-band wavelet packet to divide the digital frequency domain into $B = 2^p$ bands (see Fig. 2). In particular, $p = 6$ and $B = 64$. By using the wavelet transform, large i s in (4) and (5) correspond to high-frequency information (local shape information), while small i s carry relatively lower frequency information (global shape information). Moreover, the index i is related to the spatial location along the contour, around which shape information is collected.

Let $\mathbf{w}_n = [c_{n1}, d_{n1}, c_{n2}, d_{n2}, \dots, c_{nL}, d_{nL}]^T$ be the $2L \times 1$ vector formed by concatenating the wavelet coefficients of the n th subject. We divide the vectors $\mathbf{w}_1, \dots, \mathbf{w}_N$ into $B = 64$ bands each. Let $\mathbf{w}_n^{(b)}$ be the wavelet coefficient vector for band b , $n = 1, \dots, N$, $b = 1, \dots, B$. We calculate the inter-individual variation of the wavelet coefficients within each band (see Fig. 3) from vectors $\mathbf{w}_n^{(b)}$. The complete statistical shape model includes the mean and covariance matrix of wavelet coefficients for each band

$$\mu^{(b)} = \frac{1}{N} \sum_{n=1}^N \mathbf{w}_n^{(b)} \quad (6)$$

$$\mathbf{C}^{(b)} = \frac{1}{N-1} \sum_{n=1}^N (\mathbf{w}_n^{(b)} - \mu^{(b)}) (\mathbf{w}_n^{(b)} - \mu^{(b)})^T. \quad (7)$$

We then perform PCA in each band individually, thus obtaining 64 sets of eigenvectors and eigenvalues. Let $\Phi^{(b)}$ be the matrix formed by the eigenvectors for the wavelet coefficients of the b th band and $\Lambda^{(b)}$ be the vector of the eigenvalues of the wavelet coefficients of the b th band. The eigenvectors and eigenvalues of the first few bands represent relatively global aspects of shape variability, whereas bands with higher indices represent higher frequency and more localized aspects of shape variability. Each band has a relatively small number of variables,

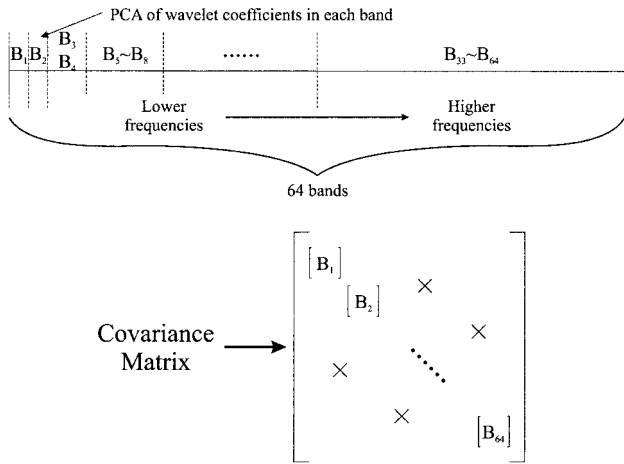


Fig. 3. A wavelet transform is applied to the parametric functions representing a deformable contour. The resulting wavelet coefficients are then grouped into 64 bands (see Fig. 2), and the joint distribution of each band is estimated from the available training samples, via its mean and covariance matrix. This effectively transforms the covariance matrix of the full joint distribution into a matrix that is close to, but not necessarily exactly, a block diagonal matrix. The submatrix corresponding to band B_1 reflects global shape characteristics, whereas the submatrix corresponding to B_{64} reflects local shape characteristics at a particular segment. B_{33} reflects local shape characteristics of a neighboring segment.

compared to the total number of points. Moreover, these variables form a fairly compact distribution, since they tend to be correlated. Therefore, its eigenvectors span a relatively large subspace. In this way, the overconstraining effect of standard ASMs is overcome.

3) *Reconstruction of the Contour From the Wavelet Coefficients:* With the eigenvectors and eigenvalues for each of the B bands, $\Phi^{(b)}$, and $\Lambda^{(b)}$, $b = 1, \dots, B$, we can approximate any new shapes by projecting the wavelet coefficients of each band to the corresponding eigen space and, thus, reconstruct the plausible shapes. Suppose a new shape, \mathbf{a} , is represented by a list of landmark points (u_k, v_k) , $k = 1, \dots, K$. We approximate \mathbf{a} using the statistical shape model as follows.

- Step 1) Align \mathbf{a} with the mean shape in the model using appropriate translation, scaling, and rotation to get $\mathcal{T}(\mathbf{a})$.
- Step 2) Apply the wavelet transform $\mathcal{W}(\bullet)$ to the x - and y -coordinate vectors of $\mathcal{T}(\mathbf{a})$ to get the wavelet coefficients \mathbf{c} and \mathbf{d} [(4) and (5)].
- Step 3) Divide \mathbf{c} and \mathbf{d} into B bands each using the same partition of the space-frequency domain used in building the model. The coefficient vector for each band is $\mathbf{s}^{(b)}$, $b = 1, \dots, B$.
- Step 4) Approximate $\mathbf{s}^{(b)}$ using the statistical shape model as

$$\hat{\mathbf{s}}^{(b)} = \mu^{(b)} + \Phi^{(b)} \mathbf{q}_b \quad (8)$$

where \mathbf{q}_b is a parameter vector computed by truncating each component of $\Phi^{(b)\top} \cdot (\mathbf{s}^{(b)} - \mu^{(b)})$ with two standard deviations of corresponding eigen variation for the b th band.

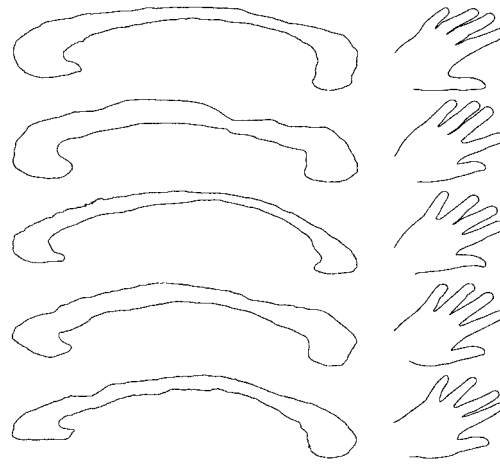


Fig. 4. 5 representative training samples used in the experiments of this paper.

- Step 5) Group $\hat{\mathbf{s}}^{(b)}$, $b = 1, \dots, B$ into wavelet coefficients $\hat{\mathbf{c}}$ and $\hat{\mathbf{d}}$, and use inverse wavelet transform to reconstruct the shape $\hat{\mathbf{a}}$.
- Step 6) Apply the inverse of the transform in Step 1) to get $\hat{\mathbf{a}} = \mathcal{T}^{-1}(\hat{\mathbf{a}}) = (\hat{u}_k, \hat{v}_k)$, $k = 1, \dots, K$, which is the approximation of \mathbf{a} in the image space.

C. Justification in Terms of the Underlying Prior Distribution

In calculating a statistical shape prior, and under the assumption of a Gaussian distribution, one needs to invert the covariance matrix of a very high-dimensional distribution

$$P(\mathbf{x}) = \frac{1}{z} \cdot \exp \{ -(\mathbf{x} - \mu)^T \mathbf{C}^{-1} (\mathbf{x} - \mu) \} \quad (9)$$

where z is a normalization constant. However, \mathbf{C} is a ‘‘highly singular’’ matrix, since it is estimated from a very small number of training samples. Therefore, an estimate of $(\mathbf{x} - \mu)^T \mathbf{C}^{-1} (\mathbf{x} - \mu)$, which is found via projections of $(\mathbf{x} - \mu)$ on the eigenvectors of \mathbf{C} by the PCA, is often a very rough approximation of the true value of $(\mathbf{x} - \mu)^T \mathbf{C}^{-1} (\mathbf{x} - \mu)$. Accurately estimating the value of $(\mathbf{x} - \mu)^T \mathbf{C}^{-1} (\mathbf{x} - \mu)$ and, therefore, the probability of a particular shape, from few samples is not possible. *However, it is possible to accurately estimate the inverse of certain submatrices of \mathbf{C} and, therefore, of marginal distributions that pertain to local and finer shape details.* Such estimation can be achieved if the dimensionality of these submatrices is ‘‘small enough.’’ ‘‘Small enough’’ here is relative to the degree of spatial correlation among neighboring points, and to the number of available training samples. Low degree of spatial correlation would require either larger number of training samples or smaller bands (submatrices), or both. *These marginal distributions replace local smoothness constraints, which are often borrowed in an ad hoc way from physical systems, such as elastic objects.*

It is worth noting that the goal of our hierarchical scheme is to utilize marginal distributions that can be estimated accurately. Equivalently, it utilizes marginal distributions whose covariance matrix has eigenvectors that can span a relatively large subspace. This goal is achieved simultaneously in two ways. First, smaller sets of highly correlated variables are examined; these are the small contour segments and correspond to the high-index bands in the wavelet transform ($B_{33} \sim B_{64}$). Second, the joint

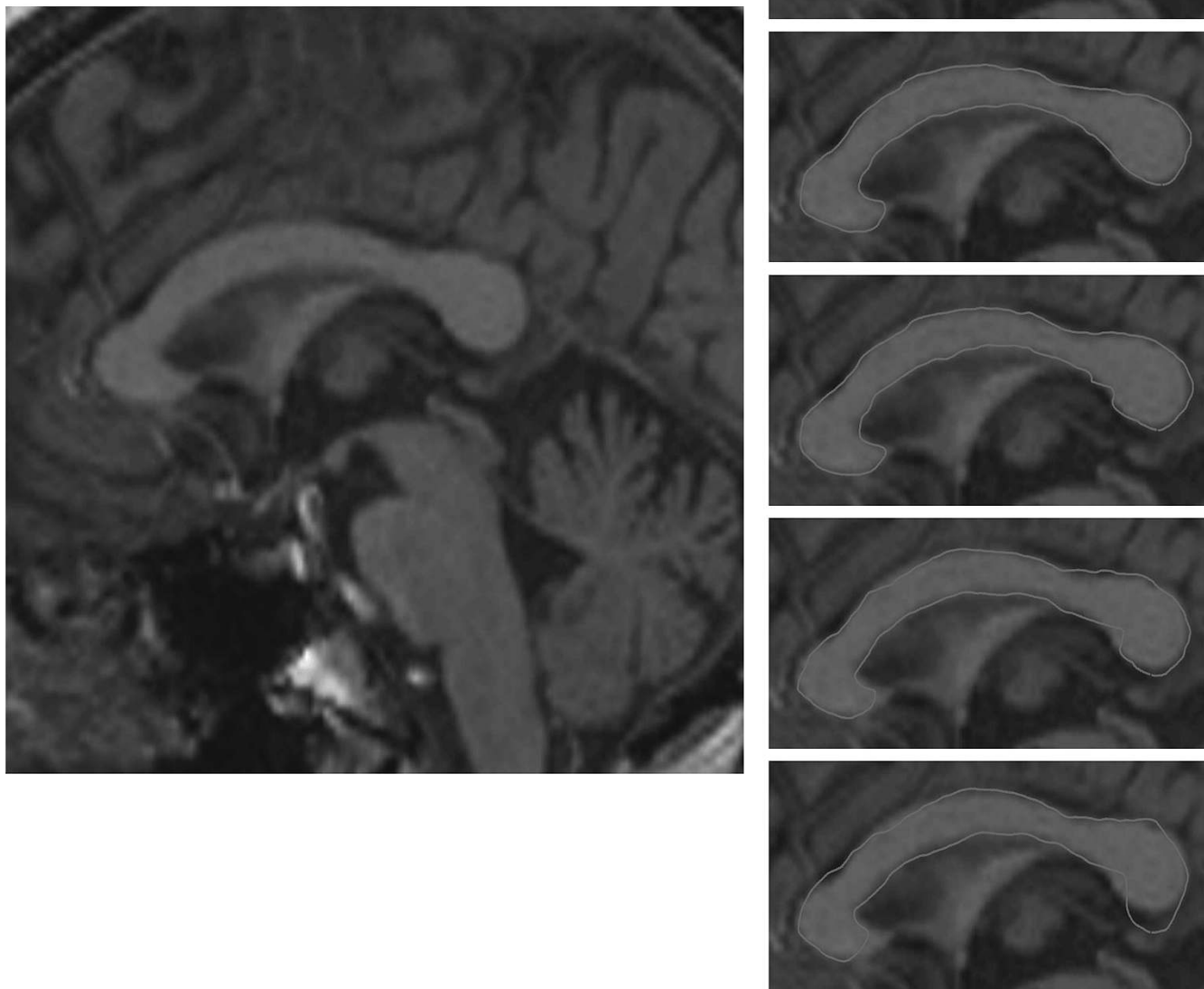


Fig. 5. (Left) Midsagittal MR image. (Right) Automated segmentation obtained using the standard active shape model (ASM) trained on 99, 50, 20, 10, and 5 samples (top to bottom, respectively). The ASM performs well when enough training samples are available, relative to the variability of the structure, but starts to fail when relatively few training samples are available.

distribution of a large number of variables is considered after smoothing, which effectively makes the distribution more compact. For example, band B_1 in Fig. 3 examines the joint distribution of a number of linear combinations of all points along the contour, since the wavelet transform is a linear operator.

III. EXPERIMENTS

In this section, we provide experimental results and quantitative analysis of the performance of different models as we change the number of training samples.

We applied the standard active shape model, Method 1, and Method 2 to the midsagittal sections of magnetic resonance images of human brains to find the callosal boundaries and to the hand images to find the hand contours. For the corpus callosum, we randomly selected midsagittal sections from our database of normal elderly subjects participating in the Baltimore Longitudinal Study of Aging [27]. The training set was built by having an expert outline the callosal boundary, and specify two landmark points, namely the anterior most and posterior-most tips of the callosum along its axis of symmetry. Training contours were then parameterized by piece-wise constant speed parameteriza-

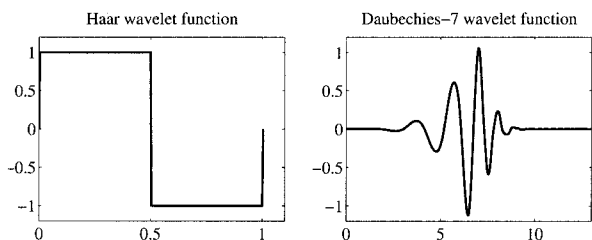


Fig. 6. Wavelet functions for Haar (left) and Daubechies-7 (right) wavelets. Haar wavelet has only one vanishing moment, while Daubechies-7 wavelet has seven vanishing moments. Therefore, Daubechies-7 wavelet provides much smoother results.

tions in-between landmarks. The hand contours used as training samples were manually outlined from the gray level hand images with the points corresponding to the tips of the fingers and the points in between fingers specified. The contours were then parameterized by piece-wise constant speed parameterizations in-between these points. Both the callosal boundaries and the hand contours were represented by 512 landmark points. Fig. 4 shows some examples for callosal boundaries and hand contours we used to train our model.

In order to compare the performance of different models, we used the same initialization and the same deformation scheme for all models. During the deformation, each landmark point was moved to its nearby feature in the image. And this displacement was propagated to its neighbors according to a Gaussian fall-off function in order to achieve a smooth deformation. After all landmark points were moved according to the image features, the tentative contour was mapped to the shape space (constrained by two standard deviations from the mean in the direction of each eigenvector) to initialize the next iteration. The iterative deformation was terminated when the difference between two successive iterations was small enough.

A. Corpus Callosum

We first tested the standard ASM for different numbers of training samples. The images are 256×256 pixels in size. Fig. 5 shows a test image on the left, and the results obtained using ASMs trained on 99, 50, 20, 10, and 5 samples. Although for an adequately large number of samples a good segmentation was obtained, the segmentation error increased rapidly when less than 50 samples were used.

We then used the five samples of Fig. 4 to train our hierarchical ASM. For Method 2, we tested two different wavelets: the Haar and the Daubechies-7, whose wavelet functions are shown in Fig. 6. Figs. 7 and 8 show the principal modes of variation for the lowest frequency component of callosal boundaries derived for each of these wavelets. Since Daubechies-7 wavelet has wider support and more vanishing moments, it gives smoother contours. Therefore, we used the Daubechies-7 wavelet for all the experiments in the following.

Fig. 9 shows three examples of segmentations obtained using ASM and the hierarchical methods proposed in this paper. It is clear that ASM needs to be trained on a sufficiently large number of samples in order to capture the finer details of individual shapes, which is not the case with the hierarchical methods.

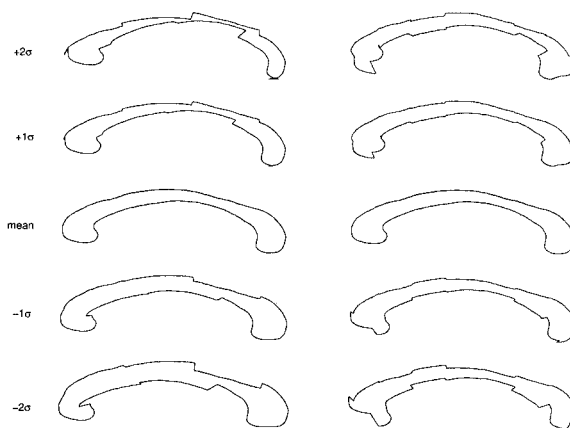


Fig. 7. The first (left) and second (right) mode of variation of the lowest frequency component using the limited support Haar wavelet. Choppiness of the resulting shapes is due to the small support of the basis functions.

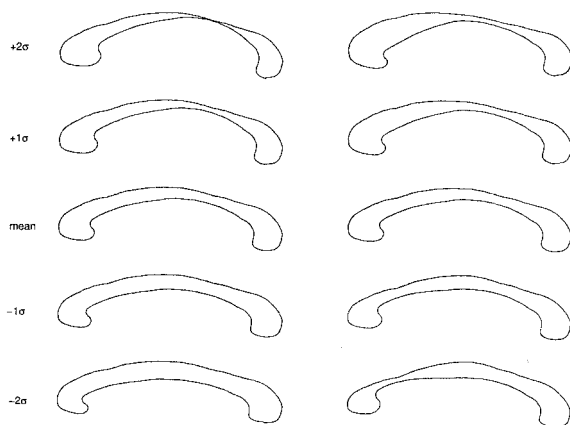


Fig. 8. The first two principal modes of variation of the lowest frequency component using the Daubechies-7 wavelet. The broader support of the basis functions provides a smooth representation of shape variation in terms of the eigenvectors of the wavelet coefficients of different bands.

B. Hand Contours

We also conducted the above experiment with hand images. Fig. 10 shows the model for the hand contour trained on five samples. The first and second columns show the first two eigen variations of the traditional ASMs. The last three columns show a few eigen variations obtained using Method 2, for example, the third and the fourth columns show the first two eigen variations for Band-1; and the last column shows the most significant eigen variation for Band-3. From the figure, we can see that Band-1 can be used to describe global shape variations such as the movements of figures, while Band-3 encodes some more local shape variations.

Fig. 11 shows three hand images we used to test the methods. Fig. 12 shows the image searching results using different methods. The first and the second rows show the results using standard ASMs trained on 5 and 50 samples. The third and fourth rows show the results using Method 1 and Method 2 trained on five samples. When enough samples (50) were available, the standard ASM can achieve satisfying results. But when only a very limited number (5) of training samples were available, the standard ASM failed to find the desired contours, because the ASM trained on five samples is too restrictive.

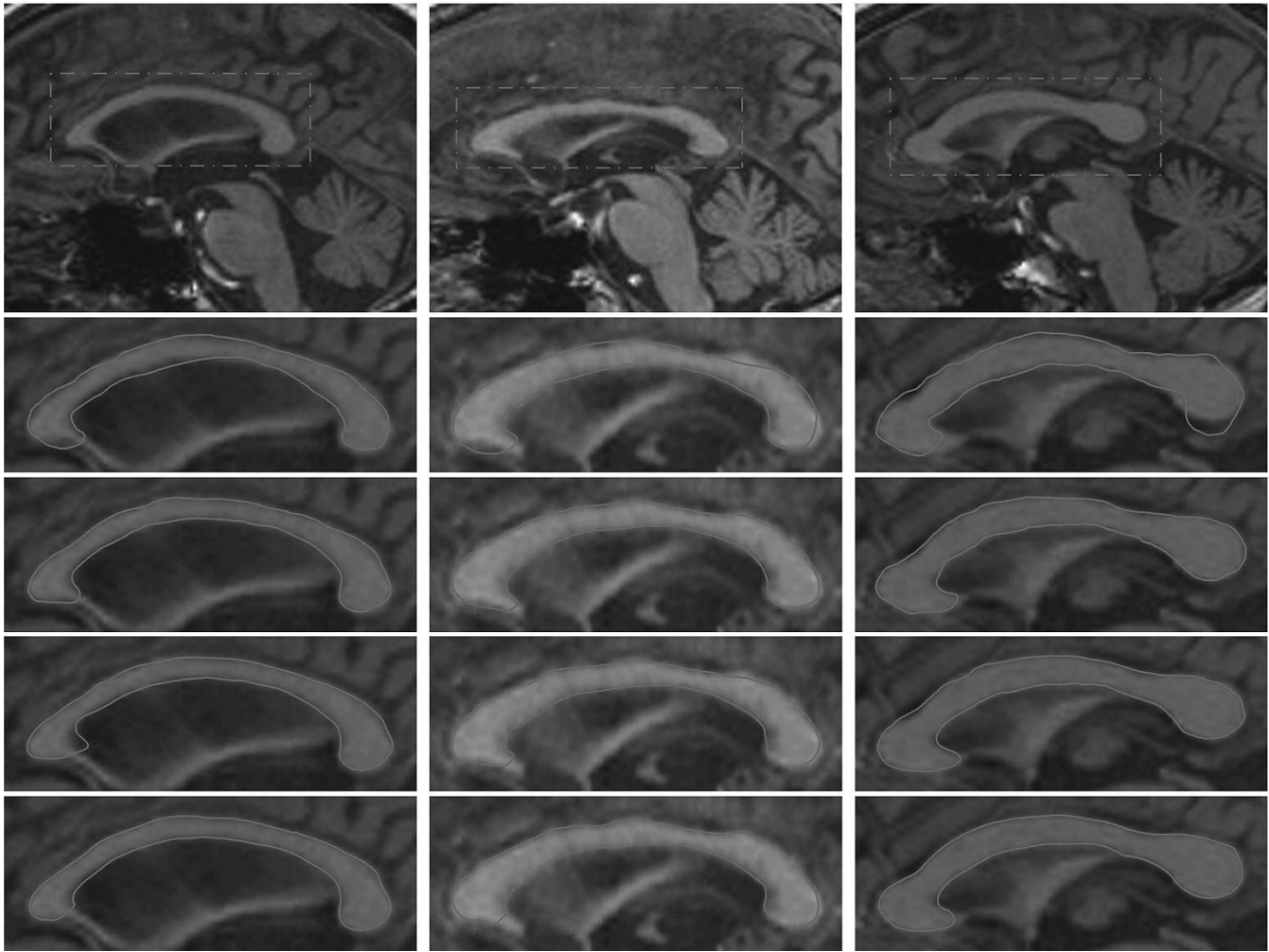


Fig. 9. Top row: Midsagittal images of the corpus callosum of three subjects. Second-fifth rows: Segmentations obtained using ASM trained on five samples, ASM trained on 99 samples, Method 1, and Method 2 trained on five samples. It is clear that ASM needs to be trained on a sufficiently large number of samples in order to capture the finer details of individual shapes, which is not the case with the hierarchical methods.

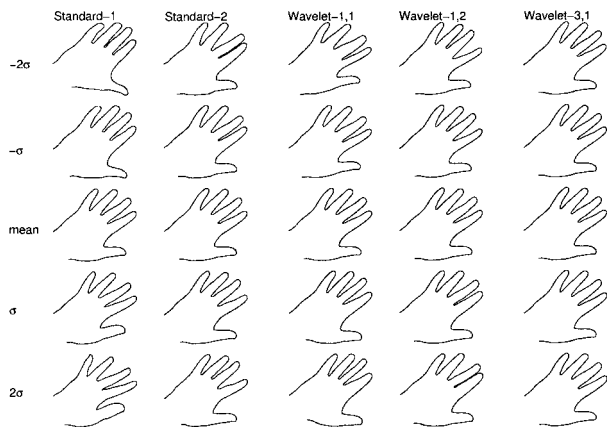


Fig. 10. Eigen variations of hand contour for standard shape model and wavelet-based hierarchical shape model. The first two columns show the first two eigen variations of the standard shape model. The third and fourth columns show the first two eigen variations of the lowest frequency band (Band-1) of the hierarchical shape model. These two columns show the global variations of the hand contour. The last column shows the first eigen variation of Band-3 of the hierarchical shape model. It shows relatively local variation of the hand contour.

During the deformation, image features (edge points) try to attract the landmark points. When the tentative contours are

projected to the shape space, only a very small portion of tentative deformation is kept for the future iterations, which can result in failure because the projection may move the deformable contour far away from its desired location.

For Methods 1 and 2, five samples are adequate to achieve good results since these methods provide efficient ways to describe both global and local information presented in the training samples. However, for the standard ASMs, details in the training samples are often overwhelmed by the global shape variations and are not reflected by the statistical model.

C. Quantitative Analysis of the Methods

In this experiment, we applied standard ASMs, Method 1, and Method 2, trained on different number of training samples, to the midsagittal sections of MR images of eight human brains and eight hand images, and computed the error between the results using different shape models and the corresponding contours manually outlined by the human expert. The error is calculated by averaging the distance from the points on a contour to the closest point on the contours outlined manually by the human expert. Table I shows the average errors of the eight callosal boundaries. And Table II shows the average errors of the

TABLE I

AVERAGE ERRORS BETWEEN THE RESULTS OF DEFORMABLE SEGMENTATION OF EIGHT CALLOSAL BOUNDARIES USING DIFFERENT MODELS TRAINED ON DIFFERENT NUMBER OF TRAINING SAMPLES. THE RESULTS OF DEFORMABLE SEGMENTATION ARE COMPARED WITH THE CORRESPONDING BOUNDARIES MANUALLY DRAWN BY THE HUMAN EXPERT. NUMBERS IN THE TABLE ARE IN PIXELS

Method	Number of training samples									
	5	10	15	20	30	40	50	65	80	99
ASM	2.26	2.11	2.07	1.99	1.99	2.03	1.90	1.95	1.91	1.90
Method 1	2.01	1.98	1.98	1.96	1.97	1.95	1.92	1.93	1.91	1.92
Method 2	1.97	1.95	1.93	1.94	1.94	1.94	1.91	1.93	1.92	1.91

TABLE II

AVERAGE ERRORS BETWEEN THE RESULTS OF DEFORMABLE SEGMENTATION OF EIGHT HAND CONTOURS USING DIFFERENT MODELS TRAINED ON DIFFERENT NUMBER OF TRAINING SAMPLES. THE RESULTS OF DEFORMABLE SEGMENTATION ARE COMPARED WITH THE CORRESPONDING BOUNDARIES MANUALLY DRAWN BY THE HUMAN EXPERT. NUMBERS IN THE TABLE ARE IN PIXELS

Method	Number of training samples									
	5	10	15	20	25	30	35	40	45	50
ASM	5.27	2.81	2.61	2.16	1.98	1.99	2.03	2.02	1.95	1.84
Method 1	1.78	1.50	1.39	1.31	1.27	1.30	1.27	1.29	1.30	1.28
Method 2	1.68	1.70	1.62	1.56	1.40	1.42	1.36	1.35	1.37	1.34



Fig. 11. Hand images used to test the methods.

eight hand contours. Fig. 13 shows the average errors. From the tables and the figure, we can see that for the standard ASMs, the errors drop quickly as we augment the training set when the model is trained on relatively small number of samples; the errors drop slowly as we increase the training sample number if we already have enough (more than 25) training samples. For Methods 1 and 2, the errors drop relatively slowly as we increase the training sample number. Also, from the table, we can see that for the hierarchical shape models, the errors are substantially smaller than the error for standard ASMs when only a small number of training examples is available. From Table I, we also observe that once we have enough training samples (more than 50), the errors for the three methods are comparable to each other. This is because when we have enough training samples, the shape spaces generated using three methods do not differ much from each other. They all reflect the true biological variation of the shape.

IV. SUMMARY AND DISCUSSION

We presented a hierarchical formulation of active shape models, which allows a statistical shape prior to be able to capture fine as well as coarse shape characteristics, and which overcomes limitations of previous models based on PCA of the covariance matrix. A hierarchical representation of a shape in terms of its wavelet transform was followed by a PCA on the wavelet coefficients. The premise of our method is that, although a relatively small number of training samples cannot necessarily capture the high-dimensional probability density

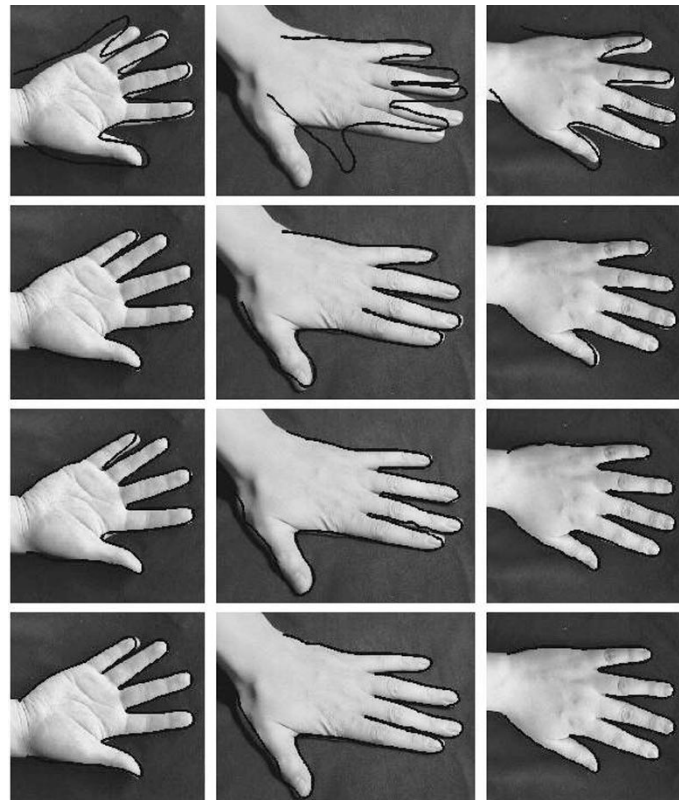


Fig. 12. Final segmentation results of different methods. Top row shows the result using the standard active shape model trained on five samples. The method failed for all three testing images because extremely small training set results in a very restrictive shape model. The second row shows the result of using standard active shape model trained on 50 samples. In this case, the model can follow the detail of the shape fairly well. The third and the fourth rows show the result using Method 1 and Method 2 trained on five samples, respectively. Again, as we observed in Fig. 9, Methods 1 and 2 give satisfying results when very small number of training samples are used to train the model.

function of a shape, it can be used to estimate patterns of covariation of smaller numbers of variables, thus allowing the hierarchical model to capture local and fine shape details, without the need for an ad hoc physically-based smoothness

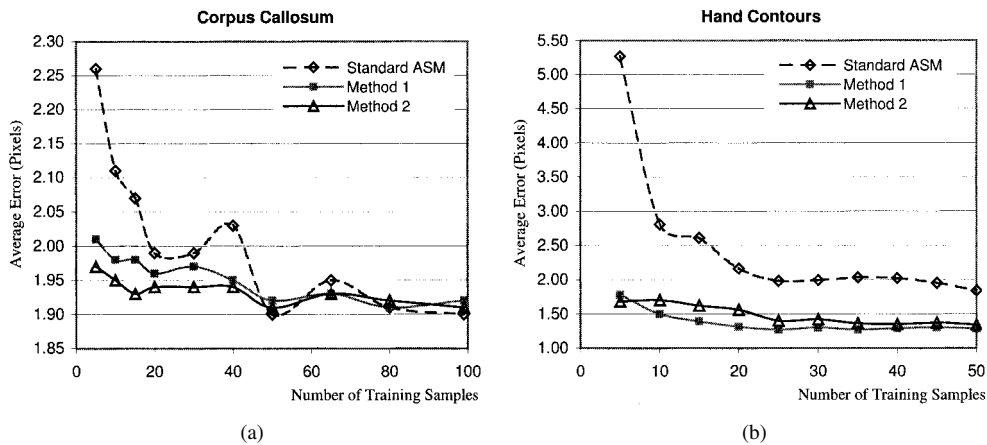


Fig. 13. Average errors between the results of deformable segmentation for corpus callosum (left) and hand contour (right) using different models trained on different number of training samples.

constraint (e.g., elasticity). It is true that if certain variation of the shape is not presented in the training samples. Then no matter how many samples we use to train the model, there is no way for the model to recover this variation. This is true for not only the standard active shape model, but also the methods proposed in this paper. However, the methods proposed can provide ways to better describe the variations that occur in the training samples.

In the paper, we also described a more heuristic Method 1, which is based on a crude multiscale representation of the shapes. It provides some intuition of multiscale curve representation and a rather simplistic way to build hierarchical active shape model. Method 2 is based on the wavelet decomposition. It has a rigorous mathematical background, and the multiresolution hierarchy is more meaningful than the one used by Method 1. Since Method 2 requires wavelet decomposition and reconstruction when mapping a shape into and out of a model space, it is more computationally demanding than Method 1.

Experiments performed on midsagittal sections of the corpus callosum and hand images showed that the proposed hierarchical models perform better than standard active shape models when a limited number of samples is available. We note that, for the case of callosal boundaries and hand contours parameterized with 512 landmark points, when 50 or more training samples were available, then standard ASM performed reliably for callosal boundaries and hand contours. However, we anticipate that on more complex and three-dimensional (3-D) structures, such as cortical structures, a prohibitively large number of training samples would be necessary, if standard ASMs were to be used, whereas our experiments herein indicate that remarkably low numbers of training samples might be sufficient when the hierarchical model is used.

Method 1 can be extended to 3-D objects easily since it involves only an overlapping partition of the object and standard ASM procedures for each partition and for the centers of mass of partitions. Method 2 can be extended into 3-D using some existing wavelet based multiresolution representation of surfaces, such as the one described in [18]. An extensive validation will be necessary, in order to better understand strengths and weaknesses of this model. In particular, we need to better understand how performance of our algorithm depends on parameters such

as the number of levels used in the wavelet transform, and the number and size of bands used in the statistical analysis, and how performance is affected by different wavelets.

REFERENCES

- [1] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1987.
- [2] D. Terzopoulos and D. Metaxas, "Dynamic 3D models with local and global deformations: Deformable superquadrics," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 703–714, July 1991.
- [3] A. Pentland and S. Sclaroff, "Closed-form solutions for physically based shape modeling and recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 715–729, July 1991.
- [4] L. H. Staib and J. S. Duncan, "Boundary finding with parametrically deformable models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 1061–1075, Nov. 1992.
- [5] —, "Model-based deformable surface finding for medical images," *IEEE Trans. Med. Imag.*, vol. 15, no. 5, pp. 720–731, 1996.
- [6] T. F. Cootes and C. J. Taylor, "Combining point distribution models with shape models based on finite element analysis," *Image Vis. Computing*, vol. 13, no. 5, pp. 403–9, 1995.
- [7] G. Szekely, A. Kelemen, C. Brechbuhler, and G. Gerig, "Segmentation of 2-D and 3-D objects from MRI volume data using constrained deformations of flexible Fourier contour and surface models," *Med. Image Anal.*, vol. 1, no. 1, pp. 19–34, 1996.
- [8] A. Kelemen, G. Szekely, and G. Gerig, "Elastic model-based segmentation of 3-D neuroradiological data sets," *IEEE Trans. Med. Imag.*, vol. 18, pp. 828–839, Oct. 1999.
- [9] D. Shen, E. Herskovits, and C. Davatzikos, "Segmentation and shape modeling of 3D MR structures using an adaptive focus statistical shape model," *IEEE Trans. Med. Imag.*, vol. 20, pp. 257–270, Apr. 2001.
- [10] D. Shen and C. Davatzikos, "An adaptive focus deformable model using statistical and geometric information," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 906–913, Aug. 2000.
- [11] N. Duta and M. Sonka, "Segmentation and interpretation of MR brain images using an improved knowledge-based active shape model," in *Information Processing in Medical Imaging*. Berlin, Germany: Springer-Verlag, 1997, pp. 375–380.
- [12] Y. Wang and L. H. Staib, "Boundary finding with prior shape and smoothness models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 738–743, July 2000.
- [13] T. F. Cootes, C. J. Taylor, A. Hill, and J. Haslam, "The use of active shape models for locating structures," in *Proc. 13th Int. Conf. Information Processing in Medical Imaging*, 1993, pp. 33–47.
- [14] T. F. Cootes, C. J. Taylor, and A. Lanitis, "Multi-resolution search with active shape models," in *Proc. Int. Conf. Pattern Recognition*, vol. I, Oct. 1994, pp. 610–612.
- [15] D. Ballard, "Strip-trees: A hierarchical representation for curves," *Commun., Assoc. Computing Machinery*, vol. 24, no. 5, pp. 310–321, 1981.

- [16] A. K. Mackworth and F. Mokhtarian, "The renormalized curvature scale space and the evolution properties of planar curves," *Proc. IEEE Computer Vision Pattern Recognition '88*, pp. 318–326, 1988.
- [17] D. R. Forshey and R. H. Bartels, "Hierarchical B-spline refinement," *Comput. Graph.*, vol. 22, no. 4, pp. 205–212, 1988.
- [18] L. M. Reissell, "Wavelet multiresolution representation of curves and surfaces," *Graphical Models and Image Processing: GMIP*, vol. 58, no. 3, pp. 198–217, 1996.
- [19] A. Finkelstein and D. H. Salesin, "Multiresolution curves," *Comput. Graph.*, vol. 28, pp. 261–268, 1994.
- [20] Q. M. Tieng and W. W. Boles, "Recognition of 2-D object contours using the wavelet transform zero-crossing representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 910–916, Aug. 1997.
- [21] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*. New York: Wiley, 1998.
- [22] I. Daubechies, *Ten lectures on Wavelets*. Philadelphia, PA: SIAM (Soc. Ind. Appl. Math), 1992.
- [23] —, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, vol. 36, pp. 961–1005, Sept. 1990.
- [24] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.
- [25] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Processing Lett.*, vol. 3, pp. 10–12, Jan. 1996.
- [26] D. Shen and H. H. S. Ip, "Discriminative wavelet shape descriptors for invariant recognition of 2-D patterns," *Pattern Recogn.*, vol. 32, no. 2, pp. 151–165, Feb. 1999.
- [27] S. M. Resnick, A. Goldszal, C. Davatzikos, S. Golski, M. A. Kraut, E. J. Metter, R. N. Bryan, and A. B. Zonderman, "One-year age changes in MRI brain volumes in older adults," *Cereb. Cortex*, vol. 10, pp. 464–472, 2000.