Active Shape Model Segmentation With Optimal Features

Bram van Ginneken*, Alejandro F. Frangi, Joes J. Staal, Bart M. ter Haar Romeny, and Max A. Viergever

Abstract—An active shape model segmentation scheme is presented that is steered by optimal local features, contrary to normalized first order derivative profiles, as in the original formulation [Cootes and Taylor, 1995, 1999, and 2001]. A nonlinear kNN-classifier is used, instead of the linear Mahalanobis distance, to find optimal displacements for landmarks. For each of the landmarks that describe the shape, at each resolution level taken into account during the segmentation optimization procedure, a distinct set of optimal features is determined. The selection of features is automatic, using the training images and sequential feature forward and backward selection. The new approach is tested on synthetic data and in four medical segmentation tasks: segmenting the right and left lung fields in a database of 230 chest radiographs, and segmenting the cerebellum and corpus callosum in a database of 90 slices from MRI brain images. In all cases, the new method produces significantly better results in terms of an overlap error measure ($p < 0.001$ using a paired T-test) than the original active shape model scheme.

Index Terms—Active shape models, medical image segmentation, model-based segmentation.

I. Introduction

SEGMENTATION is one of the key areas in computer vision. During the 1970s and 1980s, many researchers approached the segmentation problem in a *bottom-up* fashion: emphasis was on the analysis and design of filters for the detection of local structures such as edges, ridges, corners and T-junctions. The structure of an image can be described as a collection of such syntactical elements and their (spatial) relations, and such descriptions can be used as input for generic segmentation schemes. Unfortunately, these segmentations are often not very meaningful. On the other hand, *top-down* strategies (also referred to as *model-based* or *active* approaches) for segmentation were used successfully in highly constrained environments, e.g., in industrial inspection tasks. Often these methods are based

on template matching. Templates incorporate knowledge about both the shape of the object to be segmented and its gray-level appearance in the image, and are matched for instance by correlation or with generalized Hough transform techniques. But template matching, or related techniques, are likely to fail if the object and/or background exhibit a large variability in shape or gray-level appearance, as is often the case in real-life images and medical data.

Active contours or snakes [4], [5] and wave propagation methods such as level sets [6], have been heralded as a new paradigms for segmentation. It was their ability to deform freely instead of rigidly that spurred this enthusiasm. Nevertheless, such methods have two inherent limitations which make them unsuited for many medical segmentation tasks. First, little *a priori* knowledge about the shape to be segmented can be incorporated, except for adjusting certain parameters. Second, the image structure at object boundaries is prescribed by letting the snakes attract to edges or ridges in the image, or by termination conditions for propagating waves. In practice, object boundaries do not necessarily coincide with edges or ridges.

To overcome these limitations, researchers experimented with hand-crafted parametric models. An illustrative example is the work of Yuille *et al.* [7] where a deformable model of an eye is constructed from circles and parabolic patches and a heuristic cost function is proposed for the gray-level appearance of the image inside and on the border of these patches. There are two problems with parametric models. First of all they are *dedicated*, that is, limited to a single application. Second, there is no proof that the shape model and cost function proposed by the designer of the model are the optimal choice for the given application.

Consequently, there is a need for generic segmentation schemes that can be trained with examples as to acquire a model of the shape of the object to be segmented (with its variability) and the gray-level appearance of the object in the image (with its variability). Such methods are prototype-based which makes it easy to adapt them to new applications by replacing the prototypes; they use statistical techniques to extract the major variations from the prototypes in a principled manner.

Several of such schemes have been proposed. For an overview see the book of Blake and Isard [8] and the review by Jain *et al.* [9]. In this paper, we focus on active shape models (ASMs) put forward by Cootes and Taylor [1], [3]. We have implemented the method based on the description of the ASM segmentation method detailed in [10]. The shape model in ASMs is given by the principal components of vectors of landmark points. The gray-level appearance model is limited to the border of the object and consists of the normalized first derivative of profiles

*B. van Ginneken is with the Image Sciences Institute, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands (e-mail: bram@isi.uu.nl).

A. F. Frangi is with the Grupo de Tecnología de las Comunicaciones, Departamento de Ingeniería Electrónica y Comunicaciones, Universidad de Zaragoza, 50015 Zaragoza, Spain.

J. J. Staal and M. A. Viergever are with the Image Sciences Institute, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands.

B. M. ter Haar Romeny is with the Faculty of Biomedical Engineering, Medical and Biomedical Imaging, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands.

centered at each landmark that run perpendicular to the object contour. The cost (or energy) function to be minimized is the Mahalanobis distance of these first derivative profiles. The fitting procedure is an alternation of landmark displacements and model fitting in a multiresolution framework.

Several comparable approaches are found in the literature. Shapes and objects have been modeled by landmarks, finite-element methods, Fourier descriptors and by expansion in spherical harmonics (especially for surfaces in three dimensions [11], [12]). Jain *et al.* [13] have presented a Bayesian framework in which templates are deformed and more probable deformations are more likely to occur. They use a coarse-to-fine search algorithm. Ronfard [14] has used statistics of object and background appearance in the energy function of a snake. Brejl and Sonka [15] have described a scheme similar to ASMs but with a nonlinear shape and appearance model that is optimized with an energy function after an exhaustive search to find a suitable initialization. Pizer *et al.* [16] describe an object model that consists of linked primitives which can be fitted to images using methods similar to ASMs. Cootes and Taylor have explored active appearance models (AAMs) [2], [17], [18] as an alternative to ASMs. In AAMs, a combined principal component analysis of the landmarks and pixel values inside the object is made which allows one to generate plausible instances of both geometry and texture. The iterative steps in the optimization of the segmentation are steered by the difference between the true pixel values and the modeled pixel values within the object. Sclaroff and co-workers [19], [20] have proposed a comparable method in which the object is modeled as a finite-element model.

While there are differences, the general layout of these schemes is similar in that there are: 1) a shape model that ensures that the segmentation can only produce plausible shapes; 2) a gray-level appearance model that ensures that the segmentation places the object at a location where the image structure around the border or within the object is similar to what is expected from the training images; and 3) an algorithm for fitting the model by minimizing some cost function. Usually, the algorithm is implemented in a multiresolution fashion to provide long-range capture capabilities.

ASMs have been used for several segmentation tasks in medical images [21]–[27]. Our contribution in this paper consists of a new type of appearance model for the gray-level variations around the border of the object. Instead of using the normalized first derivative profile, we consider a general set of local image structure descriptors, *viz.* the moments of local histograms extracted from filtered versions of the images using a filter bank of Gaussian derivatives. Subsequently a statistical analysis is performed to learn which descriptors are the most informative at each resolution, and at each landmark. This analysis amounts to feature selection with a k-nearest neighbors (kNN) classifier and sequential feature forward and backward selection. The kNN classifier with the selected set of features is used to compute the displacements of landmarks during optimization, instead of the Mahalanobis distance for the normalized first derivative profile. In this paper, we refer to this new segmentation method as "ASMs with optimal features," where the term optimal must be understood as described above.

TABLE I
PARAMETERS FOR ACTIVE SHAPE MODELS (ORIGINAL SCHEME AND NEW METHOD WITH OPTIMAL FEATURES). VALUES USED IN THE EXPERIMENTS ARE GIVEN BETWEEN PARENTHESES

General

s	number of training images

Shape model

n	number of landmark points
alignment	true or false, whether or not to align the shapes before computing the shape model
t	number of modes in the shape model (controlled by f_v)
f_v	part of variance to be explained by the shape model, determines number of modes (0.98 with alignment, 0.995 without)
m	bounds on eigenvalues λ_i (2.0)

Appearance model

k	number of points in profile on either side of the landmark point, giving profiles of length $2k + 1$ (2)
L_{max}	number of resolution levels for appearance model and search algorithm (4)
N_{grid}	*(optimal features method)* size of $N_{grid} \times N_{grid}$ grid of points from which features are sampled in training images (5)
f_{max}	*(optimal features method)* maximum number of features to select (10)
k_{NN}	*(optimal features method)* number of neighbors used in kNN classifier during searching and feature selection (5)

Search algorithm

n_s	number of new positions to evaluate during each iteration on either side of current landmark position; total number of positions evaluated is $2n_s + 1$ (4)
N_{max}	iterations per resolution level (10)

This paper is organized as follows. In Section II, there is a step-by-step description of the original ASM scheme. In Section III, some observations regarding ASMs are made and possible modifications to the method are discussed. In Section IV, the new method is explained. In Section V, the experiments on synthetic images, chest radiographs, and brain MR data with both the original ASM scheme and the new method with optimal features are described and the results are presented in Section VI. Discussion and conclusions are given in the Section VII.

II. ACTIVE SHAPE MODELS

This section briefly reviews the ASM segmentation scheme. We follow the description and notation of [2]. The parameters of the scheme are listed in Table I. In principle, the scheme can be used in nD, but in this paper we give a two-dimensional (2-D) formulation.

A. Shape Model

An object is described by n points, referred to as landmark points. The landmark points are (manually) determined in a set of s training images. From these collections of landmark points, a point distribution model [28] is constructed as follows. The landmark points $(x_1, y_1), \ldots (x_n, y_n)$ are stacked in shape vectors

$$\mathbf{x} = (x_1, y_1, \ldots, x_n, y_n)^T. \tag{1}$$

Principal component analysis (PCA) is applied to the shape vectors $\mathbf{x}$ by computing the mean shape

$$\overline{\mathbf{x}} = \frac{1}{s} \sum_{i=1}^{s} \mathbf{x}_i \tag{2}$$

the covariance

$$\mathbf{S} = \frac{1}{s-1} \sum_{i=1}^{s} (\mathbf{x}_i - \overline{\mathbf{x}}) (\mathbf{x}_i - \overline{\mathbf{x}})^T \tag{3}$$

and the eigensystem of the covariance matrix. The eigenvectors corresponding to the t largest eigenvalues λ_i are retained in a matrix $\mathbf{\Phi} = (\phi_1 | \phi_2 | \cdots | \phi_t)$. A shape can now be approximated by

$$\mathbf{x} \approx \overline{\mathbf{x}} + \mathbf{\Phi} \mathbf{b} \tag{4}$$

where $\mathbf{b}$ is a vector of t elements containing the model parameters, computed by

$$\mathbf{b} = \mathbf{\Phi}^T (\mathbf{x} - \overline{\mathbf{x}}). \tag{5}$$

When fitting the model to a set of points, the values of $\mathbf{b}$ are constrained to lie within the range $\pm m \sqrt{\lambda_i}$, where m usually has a value between two and three.

The number t of eigenvalues to retain is chosen so as to explain a certain proportion f_v of the variance in the training shapes, usually ranging from 90% to 99.5%. The desired number of modes is given by the smallest t for which

$$\sum_{i=1}^{t} \lambda_i \geq f_v \sum_{i=1}^{2n} \lambda_i. \tag{6}$$

Before PCA is applied to the shapes, the shapes can be aligned by translating, rotating and scaling them so as to minimize the sum of squared distances between the landmark points. An iterative scheme known as Procrustes analysis [29] is used to align the shapes. This transformation and its inverse are also applied before and after the projection of the shape model in (5). This alignment procedure makes the shape model independent of the size, position, and orientation of the objects. Alignment can also help to better fulfill the requirement that the family of point distributions is Gaussian, which is an underlying assumption of the PCA model. To this end, a projection into tangent space of each shape may be useful; see [2] for details.

However, the alignment can also be omitted. In that case, the result is a shape model that can generate only shapes with a size, position, and orientation that is consistent with the supplied examples. If, for a certain application, the objects occur only within a specific range of sizes, positions and orientations, such a model might lead to higher segmentation performance. In an unaligned shape model, the first few modes of variation are usually associated with variations in size and position and the variation seen in the first few modes had the shape model been constructed from aligned shapes, is usually shifted toward modes with lower eigenvalues. Therefore, the parameter f_v should be larger than in the case of aligned shapes in which no variation is present with respect to size and position. In our experience, building an unaligned shape model can improve segmentation performance provided that enough training data is available.

B. Gray-Level Appearance Model

The gray-level appearance model that describes the typical image structure around each landmark is obtained from pixel profiles, sampled (using linear interpolation) around each landmark, perpendicular to the contour.

Note that this requires a notion of connectivity between the landmark points from which the perpendicular direction can be computed. The direction perpendicular to a landmark (x_n, y_n) is computed by rotating the vector that runs from (x_{n-1}, y_{n-1}) to (x_{n+1}, y_{n+1}) over 90°. In the applications presented in this paper, all objects are closed contours, so for the first landmark, the last landmark and the second landmark are the points from which a perpendicular direction is computed; for the last landmark, the second to last landmark and the first landmark are used.

On either side k pixels are sampled using a fixed step size, which gives profiles of length $2k + 1$. Cootes and Taylor [2] propose to use the normalized first derivatives of these profiles to build the gray-level appearance model. The derivatives are computed using finite differences between the $(j-1)$th and the $(j+1)$th point. The normalization is such that the sum of absolute values of the elements in the derivative profile is 1.

Denoting these normalized derivative profiles as $\mathbf{g}_1, \ldots, \mathbf{g}_s$, the mean profile $\overline{\mathbf{g}}$ and the covariance matrix $\mathbf{S}_g$ are computed for each landmark. This allows for the computation of the Mahalanobis distance [30] between a new profile $\mathbf{g}_i$ and the profile model

$$f(\mathbf{g}_i) = (\mathbf{g}_i - \overline{\mathbf{g}}) \mathbf{S}_g^{-1} (\mathbf{g}_i - \overline{\mathbf{g}}). \tag{7}$$

Minimizing $f(\mathbf{g}_i)$ is equivalent to maximizing the probability that $\mathbf{g}_i$ originates from a multidimensional Gaussian distribution.

C. Multiresolution Framework

These profile models, given by $\overline{\mathbf{g}}$ and $\mathbf{S}_g$, are constructed for multiple resolutions. The number of resolutions is denoted by $L_{\max}$. The finest resolution uses the original image and a step size of one pixel when sampling the profiles. The next resolution is the image observed at scale $\sigma = 1$ and a step size of two pixels. Subsequent levels are constructed by doubling the image scale and the step size.[1]

The doubling of the step size means that landmarks are displaced over larger distances at coarser resolutions. The blurring causes small structures to disappear. The result is that the fitting

---

[1]Note that we do not subsample the images, as proposed by Cootes and Taylor.

at coarse resolution allows the model to find a good approximate location based on global images structures, while the later stages at fine resolutions allow for refinement of the segmentation result.

D. Optimization Algorithm

Shapes are fitted in an iterative manner, starting from the mean shape. Each landmark is moved along the direction perpendicular to the contour to n_s positions on either side, evaluating a total of $2n_s + 1$ positions. The step size is, again, $2^{(i-1)}$ pixels for the ith resolution level. The landmark is put at the position with the lowest Mahalanobis distance. After moving all landmarks, the shape model is fitted to the displaced points, yielding an updated segmentation. This is repeated $N_{\max}$ times at each resolution, in a coarse-to-fine fashion.

There is no guarantee that the procedure will converge. It is our experience, however, that in practice the scheme almost always converges. The gray-level model fit improves steadily and reaches a constant level within a few iterations at each resolution level. Therefore, we, conservatively, take a large value (10) for $N_{\max}$.[2]

III. IMPROVING ASMs

There are several ways to modify, refine and improve ASMs. In this section, we mention some possibilities.

- *Bounds for the Shape Model.* The shape model is fitted by projecting a shape in the $2n$-dimensional space (n the number of landmarks, the factor two is because we consider 2-D images) upon the subspace spanned by the t largest eigenvectors and by truncating the model parameters $\mathbf{b}$ so that the point is inside the box bounded by $\pm m\sqrt{\lambda_i}$. Thus there is no smooth transition; all shapes in the box are allowed, outside the box no shape is allowed. Clearly this can be refined in many ways, using a penalty term or an ellipsoid instead of a box, and so on [2], [16].

- *Nonlinear Shape Models.* The shape model uses PCA and, therefore, assumes that the distribution of shapes $\mathbf{x}$ in the Nn-dimensional space is normal. If this is not true, nonlinear models, such as mixture models, could be more suitable (see for example [31]).

- *Projecting the Shape Model.* By projecting a shape according to (5), i.e., fitting the shape model, the resulting model parameters $\mathbf{b}$ minimize the sum of squared distances between true positions and modeled positions. In practice, it can be desirable to minimize only the distance between true and model positions in the direction perpendicular to the object contour because deviation along the contour does not change whether pixels are inside or outside the object. In [32], it is demonstrated how to perform this projection on the contour.

- *Landmark Displacements.* After the Mahalanobis distance at each new possible position has been computed, Behiels *et al.* [25] propose to use dynamic programming to find new positions for the landmarks, instead of moving each point to the position with the lowest distance. This avoids the possibility that neighboring landmarks jump to new positions in different directions

and thus leads to a "smoother" set of displacements. This can lead to quicker convergence [25].

- *Confidence in Landmark Displacements.* If information is available about the confidence of the proposed landmark displacement, weighted fitting of the shape model can be used, as explained in [21].

- *Initialization.* Because of the multiresolution implementation, the initial position of the object (the mean shape, i.e., the mean location of each landmark) does not have to be very precise, as long as the distance between true and initial landmark positions is well within $kn_s 2^{(N_{\max}-1)}$ pixels. But if the object can be located anywhere within the input image, an (exhaustive) search to find a suitable initialization, e.g., as described in [15], can be necessary.

- *Optimization Algorithm.* Standard nonlinear optimization algorithms, such as hill climbing, Levenberg–Marquardt, or genetic algorithms can be used to find the optimal model parameters $\mathbf{b}$ instead of using the algorithm of alternating displacement of landmarks and model fitting. A minimization criterion could be the sum of the Mahalanobis distances, possibly complemented by a regularization term constructed from the shape model parameters. Note that a multiresolution approach can still be used with standard nonlinear optimization methods. Alternatively, a snake algorithm can be used in which the shape model provides an internal energy term and the gray-level appearance model fit is used as external energy term.

This list is not complete, but it is beyond the scope of this article to present a complete discussion of the strengths and weaknesses of the ASM segmentation method. The issues described above are not considered in this work. Instead, we focus on the following points:

- *Normalized First Derivative Profiles.* The original version of the gray-level appearance model is always based on normalized first derivative profiles. There is no *a priori* reason why this should be an optimal choice. In this paper, we propose an alternative.

- *Mahalanobis Distance.* The Mahalanobis distance in (7) assumes a normal distribution of profiles. In practice, the distributions of profiles will often be nonnormal, for example in cases where the background of the object may be one of several possible choices. The ASM scheme proposed here uses a nonlinear classifier in the gray-level appearance model and can, therefore, deal with nonnormal distributions.

IV. ASMs WITH OPTIMAL FEATURES

In this section, a new gray-level appearance model is described that is an alternative to the construction of normalized first derivative profiles and the Mahalanobis distance cost function of the original ASMs.

The aim is to be able to move the landmark points to better locations during optimization, along a profile perpendicular to the object contour. The best location is the one for which everything on one side of the profile is outside the object, and everything on the other side is inside of it.[3] Therefore, the probability that

---

[2]We always perform $N_{\max}$ iterations, contrary to Cootes and Taylor who move to a finer resolution if a convergence criterion is reached before the $N_{\max}$th iteration.

[3]This assumes that the thickness of the object, in the direction perpendicular to a landmark, is larger than half the length of the profile. We will return to this point later.

a location is inside/outside the object is estimated, for the area around each landmark separately. We base this classification on *optimal* local image features obtained by feature selection and a nonlinear kNN-classifier, instead of using the fixed choice of the normalized first derivative profiles and the Mahalanobis distance.

A. Image Features

We are looking for general image structure descriptors. A Taylor expansion approximates a function f around a point of interest x_0 by a polynomial of (some) order N. The coefficients in front of each term are given by the derivatives $f^{(n)}$ at x_0

$$f(x) \approx \sum_{n=0}^{N} \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n. \tag{8}$$

Derivatives of images are computed by convolution with derivatives of Gaussians at a particular scale σ. This motivates the use of a filter bank of multiscale Gaussian derivatives to describe local image structure. The Taylor expansion of images is known as the local jet, or multiscale local jet in the case of the Taylor expansion of the scale-space of the image [33], [34].

Given a set of filtered images, we will extract features for each location by taking the first few moments of the local distribution of image intensities (the histogram) around each location. The most suitable choice for a window function to compute this histogram, is a Gaussian, since every other choice induces spurious resolution [35]. The size of this window function is characterized by a second scale parameter α. The construction of local histograms, extracted from a Gaussian aperture function, is called a *locally orderless image* and discussed in [36]. The idea of using moments of histograms of responses of an image to a bank of filters is a standard technique in texture analysis; see, e.g., [37].

Notice that there are quite some parameters to vary: the order of the Taylor expansion (i.e., the number of filters in the filter bank), the number of scales σ to consider, the number of scales α to use for the local window, and the number of moments m to extract from the local histograms. It remains an open question which combinations are optimal for a given application and even a given location in the images. Our strategy is to compute an extensive set of features and use feature selection techniques in the subsequent classification stage to determine the optimal features. However, we must have $\alpha > \sigma$, otherwise the histogram will be computed over a homogeneous region and will, therefore, be uninteresting.

In this paper, we use only first and second moments ($m = 1, 2$), all derivatives up to second-order ($L, L_x, L_y, L_{xx}, L_{yy}, L_{xy}$), five inner scales ($\sigma = 0.5, 1, 2, 4, 8$ pixels), and a fixed relation between the inner scale σ and the histogram extent α of $\alpha = 2\sigma$. For the first moments this yields an effective scale of 1.12, 2.23, 4.47, 8.94, and 17.89 pixels, respectively (because the image is first blurred with a kernel of scale σ and subsequently with a kernel $\alpha = 2\sigma$). The total number of feature images is $2 \times 6 \times 5 = 60$.

Obviously the method can be extended by using more scales and higher-order derivatives, higher-order moments, or by releasing the fixed relation between σ and α.

B. Training and Classification

The next step is to specify how to construct a training set from the training images, which classifier to use, and how to perform feature selection.

Consider again the optimization procedure. At each iteration, each landmark is positioned at $2k + 1$ locations along a profile perpendicular to the current object location. Obviously the image structure is different for each landmark, but the positions that are evaluated are also different for each resolution. Therefore, we will select a distinct optimal set of features for each landmark *and* for each resolution, amounting to $nL_{\max}$ feature sets. Note that in the original ASMs the same strategy is followed: $nL_{\max}$ mean profiles and the covariance matrices $\mathbf{S}_g$ as they appear in (7) are computed: for each landmark, at each resolution.

From each training image and for each landmark a square grid of $N_{\text{grid}} \times N_{\text{grid}}$ points is defined with N_{grid} an odd integer and the landmark point at the center of the grid. The spacing is $2^{(i-1)}$ pixels for the ith resolution level.

N_{grid} is fixed to 5, which means that for each landmark and for each resolution level, a feature vector with 60 elements is sampled at 25 points. The output of each feature vector is either inside (1) or outside (0) the object. The landmark points themselves are considered to be inside the objects (this is an arbitrary choice). The set of training images is divided in two equal parts. This leads to two sets of samples, a training and a validation set. A kNN classifier [38] with weighted voting is used. $k_{\text{NN}} = 5$ was used and the weight of each vote is $\exp(-d^2)$, where d is the Euclidean distance to each neighbor in the feature space.

Sequential feature forward selection (also known as Whitney's method [39]–[41]) is used to find a feature set of at most $f_{\max}$ features. This set is subsequently trimmed by sequential feature backward selection, that is, features are removed if that improves performance. This procedure of forward selection followed by backward selection is as or almost as effective as optimal "floating" feature selection schemes [40], [41]. The resulting set is the "optimal" set of features that will be used during segmentation. After feature selection, the samples from the training and the validation set are merged and a list of the selected features for each landmark and each resolution is stored.

When the model is fitted to an input image, the scheme starts by computing the 60 feature images. Instead of sampling the normalized derivative profiles, the optimal feature set at each position along the profile is fed into a kNN classifier to determine the probability that this pixel is inside the object. The objective function $f(\mathbf{g})$ to be minimized is the sum of absolute differences between the expected probability (0 or one for points outside or inside the object, respectively) and the predicted probability, for each point g_i along the profile $\mathbf{g}$

$$f(\mathbf{g}) = \sum_{i=-k}^{-1} g_i + \sum_{i=0}^{+k} (1 - g_i) \tag{9}$$

where the index along the profile $\mathbf{g}$, that is oriented from the outside to the inside of the object, runs from $-k$, to $+k$. This metric replaces the Mahalanobis distance from (7).

C. Summary of Training and Segmentation Algorithm

Training the shape model.

1) Construct shape model $\overline{\mathbf{x}}$, $\mathbf{\Phi}$ [(2) and (3)].

Training the gray-level appearance model.

1) Compute the 60 feature images for each training image.
2) For each landmark, at each resolution, construct a set of training samples with as input the 60 features and as output zero or one depending on whether the sample is in or outside the object. Samples are taken from an $N \times N$ grid around the landmark in each training image (so each training set contains sN_{grid}^2 samples).
3) For each training set, construct a kNN classifier with selected optimal features. So the final result of the training phase is a set of $nL_{\max}$ classifiers.

Segmentation.

1) Initialize with the mean shape $\overline{\mathbf{x}}$.
2) Start the coarsest resolution level.
3) For each landmark, put it at $2n_s + 1$ new locations, evaluate (9) with the kNN classifier, move landmark to best new position.
4) Fit the shape model to displaced landmarks; cf. (5).
5) Iterate steps 3 and 4 $N_{\max}$ times.
6) If the current resolution is not yet the finest resolution, move to a finer resolution and go to Step 3.

D. Computational Considerations

One of the advantages of the original ASM scheme compared to other segmentation methods is its speed. The new method is considerably more computationally expensive. However, all the feature selection is to be done off-line (during training). An optimized kNN classifier [42] was used, available on the web at http://www.cs.umd.edu/~mount/ANN. We provide some benchmark figures, obtained with a 600-MHz Pentium III PC, a 256×256 image, and the same parameter settings that are used in all experiments in this study. The feature images have to be computed on-line (during segmentation), which required 8.0 s. For the original ASM method, a number of blurred images have to be computed, which required 0.35 s.

During optimization feature vectors must be classified by kNN classifiers and this requires more time than computing Mahalanobis distances. The total time for segmentation was 0.17 s for the original ASM scheme and 4.1 s for the method with optimal features.

Using a smaller feature set would reduce the computational cost of the method (almost linearly). An alternative would be to select a subset of the 60 features for all operations (each landmark, each resolution) so that it is no longer necessary to compute all 60 images for each input image. However, this speed improvement would probably come at the price of a decrease in performance.

V. EXPERIMENTS

A. Materials

Five different segmentation experiments have been performed with three types of data. The images and objects used in

TABLE II
DESCRIPTION OF THE OBJECTS AND IMAGES USED IN THE FIVE
SEGMENTATION EXPERIMENTS

#	object	resolution	no. of images	landmarks fixed	total
I	house (simulated data)	200×200	80	5	25
II	right lung (chest X-ray)	256×256	230	3	40
III	left lung (chest X-ray)	256×256	230	3	40
IV	corpus callosum (MRI)	320×256	90	3	50
V	cerebellum (MRI)	320×256	90	7	50
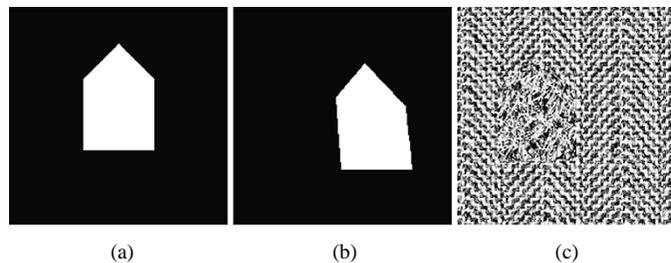
Fig. 1. (a) A generic house image. The image has a resolution of 200×200 pixels. New house images are randomly generated by adding a displacement (dx, dy) to each corner point where dx and dy are randomly selected from the interval $(-20, +20)$ pixels, and subsequently translating the distorted house by (dx, dy) where dx and dy are randomly selected from the interval $(-35, +35)$ pixels. (b) An example of a generated house image. (c) An example of a generated textured house image. Random parts from two different textures, both taken from the Brodatz set and histogram equalized, are used to fill the inside and outside of the house.

the experiments, that are labeled I to V, are briefly described in Table II. For all experiments, the images are randomly divided into training and test sets of equal size.

The data for Experiment I consisted of simulated images with an object that has the shape of a house from which the five corner points were perturbed. Details about the construction of the images are given in Fig. 1.

The image data for Experiments II and III are 230 standard PA chest radiographs selected from a tuberculosis screening program. The data contained both normal and abnormal cases of patients of 16 years and older. The images were taken with a mobile Electrodelca (Oldelft BV, Delft, The Netherlands). The tube voltage was 117 kV and the images were printed on 10×10 cm film and digitized with a Lumisys 100 scanner (Lumisys, Inc., Sunnyvale, CA) and subsampled to 256×256 pixels. Two observers independently segmented the right and left lung field.

For Experiments IV and V, a collection of 90 MRI slices of the brain was used, in which the corpus callosum and the cerebellum were segmented. Two to three slices were taken, on average, from the same patient. On average The images and segmentations were made available by the University of Iowa Hospitals and Clinics and were also used by Brejl and Sonka in [15]. The resolution is 320×256 pixels, 0.7 mm/pixel, obtained by interpolating the original volumetric data acquired with 1.5-mm-thick coronal slices.

The objects in the images were annotated by a number of *fixed* landmarks and a closed contour between those fixed points from which a number of equidistant landmark points were sampled. The number of fixed and total landmarks are given in Table II.

B. Methods

For each parameter of ASMs, a fixed setting was selected that yielded good performance, after initial pilot experiments. For the house and lung shapes, no shape alignment was performed and a shape model was constructed with $f_v = 0.995$. For these experiments, omitting alignment led to better segmentation performance. The number of modes in the shape models was six for the house images and ten and 11 for the right and left lung, respectively. For the brain structures, shape alignment was used (in this case better segmentation performance was obtained with the use of alignment) and a shape model explaining 98% of the variance ($f_v = 0.98$) was constructed. The number of modes in the shape model was six for the cerebellum and 20 for the corpus callosum.

The other settings were four levels of resolution ($L_{\max} = 4$), ten iterations/level ($N_{\max} = 10$), profiles of length five ($k = 2$) and evaluation of nine positions/iteration ($n_s = 4$). When fitting the shape model to the displaced landmarks, each mode was constrained within two times the standard deviation ($m = 2.0$). For the extended ASMs, at most ten features were selected for each landmark and each resolution ($f_{\max} = 10$). Training data were selected from 5×5 neighborhoods around each landmark ($N_{\mathrm{grid}} = 5$). In the kNN classifier, five neighbors were used ($k_{\mathrm{NN}} = 5$). All parameter settings are listed in Table I.

To compare different segmentations, the following "overlap" measure Ω was used

$$\Omega = \frac{TP}{TP + FP + FN} \qquad (10)$$

where TP stands for true positive (the area correctly classified as object), FP for false positive (area incorrectly classified as object), and FN for false negative (area incorrectly classified as background). $\Omega = 1$ for a perfect result and $\Omega = 0$ if there is no overlap at all between the detected and true object. This measure more closely reflects the idea of a good segmentation than the average distance between the true and detected landmark location, because the latter is not sensitive to shifts of the landmarks along the contour.

In all experiments, the performance when fitting the shape directly to the true landmarks [cf. (5)] was also computed. For Experiments II and III manual segmentations by a second observer were available. Therefore, Ω for the second observer can be compared with Ω for the automatic methods.

VI. RESULTS

The results of all experiments are given in Table III. The result of directly fitting the shape model to the landmark points cf. (5) is included because it indicates an upper bound for both the original method and the method with optimal features. Note that fitting the shape model minimizes the distance between the predicted landmark position and the true landmark position; it does not necessarily optimize Ω. Therefore, it is possible that an ASM scheme produces a set of model parameters **b** for which Ω is higher than Ω for fitting the shape model directly. This occurred in a few cases. Another practical measure of the optimal performance any automatic segmentation method that is trained with examples can achieve, is the variation between observers. This measure is given for Experiments II and III, where the me-

TABLE III
EXPERIMENTAL RESULTS OF ORIGINAL ASM, ASM WITH OPTIMAL FEATURES, DIRECTLY FITTING THE SHAPE MODEL AND A COMPARISON WITH A SECOND OBSERVER (EXP. II AND III)

Exp. I: Houses	$\mu \pm \sigma$	median
ASMs	0.536 ± 0.203	0.550
ASMs optimal features	0.895 ± 0.034	0.894
Fit of shape model	0.933 ± 0.029	0.937
Exp. II: Right lung field	$\mu \pm \sigma$	median
ASMs	0.882 ± 0.074	0.902
ASMs optimal features	0.929 ± 0.026	0.933
Fit of shape model	0.948 ± 0.030	0.955
Second observer	0.945 ± 0.017	0.948
Exp. III: Left lung field	$\mu \pm \sigma$	median
ASMs	0.861 ± 0.109	0.891
ASMs optimal features	0.887 ± 0.114	0.924
Fit of shape model	0.942 ± 0.090	0.955
Second observer	0.934 ± 0.021	0.938
Exp. IV: Corpus callosum	$\mu \pm \sigma$	median
ASMs	0.617 ± 0.206	0.535
ASMs optimal features	0.805 ± 0.093	0.837
Fit of shape model	0.887 ± 0.052	0.906
Exp. V: Cerebellum	$\mu \pm \sigma$	median
ASMs	0.870 ± 0.078	0.904
ASMs optimal features	0.910 ± 0.058	0.927
Fit of shape model	0.950 ± 0.014	0.950

dian Ω of the ASM method with optimal features is close to median Ω of a second human observer.

Experiment I was included to demonstrate the limitations of the original ASM method. In the case of texture boundaries, a pixel profile or a normalized first derivative of such a profile, will not produce a clear distinction between the inside and outside of the object. The optimal features are derived from local image structure measures and usually some of these will be different for two different textures. Consequently the method based on optimal features can deal with many different types of texture boundaries. For the other experiments the differences between the methods are smaller, but in all cases ASMs with optimal features produced significantly higher Ω values than the original scheme ($p < 0.001$ in a paired t-Test for all experiments). This is also clear from Fig. 2 which shows scatter plots for each segmentation task. In these plots, points which are above the diagonal line indicate images for which the segmentation with optimal features is better than the result of the original scheme. It is apparent that a substantial improvement is achieved for Experiment I and evident improvements for Experiments II, IV, and V. Only for Experiment III, the left lung fields, there is a considerable number of cases where the original method has better performance.

Example results are shown in Figs. 3–7. There were no significant differences between the results for normal and abnormal chest radiographs.

(a) (b) (c)

Fig. 3. Example result for the segmentation of the generated textured images with a house shape. Segmentations are given by the thick white line. (a) True shape. (b) ASMs ($\Omega = 0.219$). (c) Optimal ASMs ($\Omega = 0.861$).

Fig. 2. Scatter plots of each segmentation experiment. The overlap measure Ω for the original ASM scheme is plotted against Ω for the ASM method with optimal features for each segmented image. (a) Houses, $p = 1.64 10^{-13}$; (b) right lung fields, $p = 7.92 10^{-12}$; (c) left lung fields, $p = 8.63 10^{-4}$; (d) corpus callosum, $6.48 10^{-8}$; and (e) cerebellum, $p = 1.79 10^{-4}$.

VII. DISCUSSION

In this section, we discuss some properties and possible refinements of the proposed ASM method.

The largest improvement in performance is obtained for simulated data in which the textural appearance of the image inside and outside the object is different. This indicates that the proposed method may be especially useful to segment textured objects from textured backgrounds. An example could be segmentation in ultrasound images.

An important aspect is that an optimal set of features is selected for each landmark and each resolution separately. Alternatively, a single set of optimal features could be used, which would be equivalent to designing a pixel classifier that assigns each image location to one of two classes: outside or inside the object of interest. The segmentation method could be run on these "potential images" instead of on the real data. We have

conducted pilot experiments which indicated that the performance of such an approach is worse than that of the method presented here, ASMs with optimal features. The set of selected features varies considerably/landmark and resolution and is different for different applications. Had we used a standard feed-forward neural network to for the classification of locations in the image as inside or outside the object, instead of a kNN-classifier and feature selection, the particular combination of input features constructed by the network could be considered as the optimal filter for that particular landmark and resolution. In our case, the selected features cannot directly be interpreted as a single optimal filter, but the idea is similar. Note that the method does not require the use of a kNN classifier; any classifier could be used instead. Similarly, the method does not rely on the specific set of features used here. More feature images can be used, by using higher moments, more (higher-order) derivatives and by relaxing the fixed relation between σ and α.

The results of the improved ASM method approaches the median result of a second observer, which was available for Experiments II and III. However, the second observer performed still significantly better.

Both the original and improved ASM method contain a range of free parameters (Table I). Although we have found that segmentation results are not very dependent on the choice for these parameters, as long as they are within a sensible range, it could be desirable to use a straightforward iterative procedure to select optimal settings using the training set.

A more elaborate criterion for evaluating new landmark positions could be as follows. Currently landmarks are moved to those locations where the profile values are closest to zero for points outside the object and closest to one for points inside the object. In practice, the optimal profiles may be different. Especially if the object is very thin and the fitting occurs at a coarse resolution level, the innermost points of the profile may cross the border on the other side of the object! The actual profiles can be extracted from the training set and used to construct a model based on their mean and covariance matrices, that can steer the landmark displacement, in the same way as the original ASM scheme. Another enhancement would be to take into account the structure along the profile, instead of using local pixel classification for each position along the profile independently. This may improve performance. Consider a set of images, half show a black object on a white background, and the other half a white object on a black background. With local image features, it is impossible to classify locations as inside or outside the object. But a set of features measured along the profile can easily

(a)   (b)   (c)   (d)
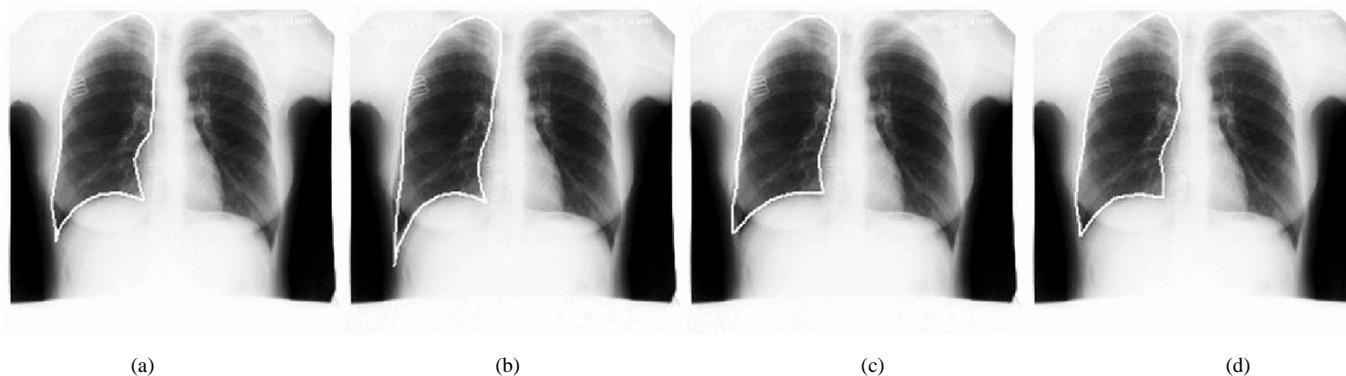
Fig. 4. Example result for the right lung field segmentation. Segmentations are given by the thick white line. (a) True shape. (b) ASMs, $\Omega = 0.884$. (c) ASMs, $\Omega = 0.945$. (d) ASMs, $\Omega = 0.952$.

(a)   (b)   (c)   (d)
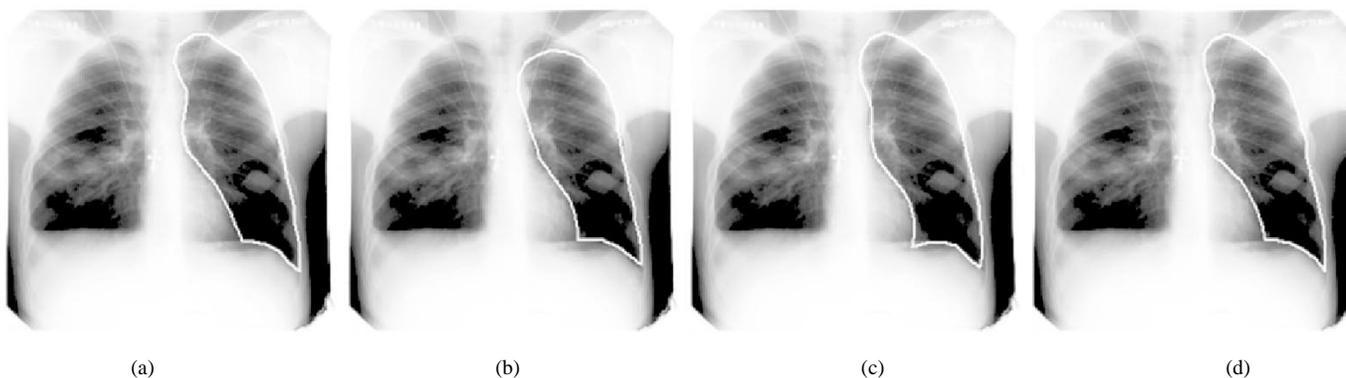
Fig. 5. Example result for the left lung field segmentation. Segmentations are given by the thick white line. (a) True shape. (b) ASMs, $\Omega = 0.873$. (c) ASMs, $\Omega = 0.935$. (d) ASMs, $\Omega = 0.961$.

(a)   (b)   (c)
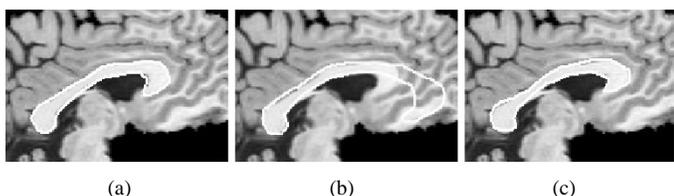
Fig. 6. Example result of segmenting the corpus callosum. Segmentations are given by the thick white line. (a) True shape. (b) ASMs, $\Omega = 0.474$. (c) ASMs, $\Omega = 0.828$.

(a)   (b)   (c)
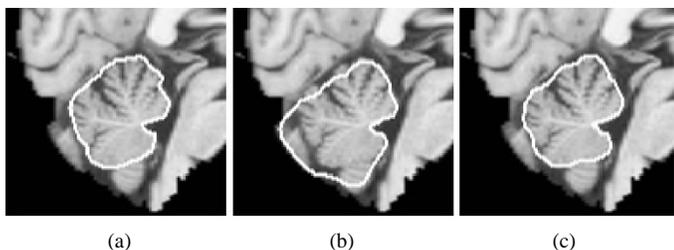
Fig. 7. Example result of segmenting the cerebellum. Segmentations are given by the thick white line. (a) True shape. (b) ASMs, $\Omega = 0.679$. (c) ASMs, $\Omega = 0.909$.

distinguish correct profiles, with an intensity jump at exactly the landmark location, from incorrect profiles, with no intensity jump or an intensity jump at a different location.

The computational complexity of the improved ASM method is roughly 20-fold that of the original scheme. The computational burden can be reduced if the feature images are computed only at those points where their value is required during seg-

mentation. Using a faster classifier will also reduce computation time. Nevertheless, the algorithm still requires only a few seconds on standard PC hardware.

We conclude by stating that active shape models provide a fast, effective, automatic, model-based method for segmentation problems in medical imaging. The new ASM method introduced in this paper significantly improves the original method through the use of an adaptive gray-level appearance model based on local image features.

REFERENCES

[1] T. F. Cootes, C. J. Taylor, D. Cooper, and J. Graham, "Active shape models—Their training and application," *Comput. Vis. Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.
[2] T. F. Cootes and C. J. Taylor, "Statistical models of appearance for computer vision," Wolfson Image Anal. Unit, Univ. Manchester, Manchester, U.K., Tech. Rep., 1999.
[3] ——, "Statistical models of appearance for medical image analysis and computer vision," *Proc. SPIE*, vol. 4322, pp. 236–248, 2001.
[4] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1988.
[5] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: A survey," *Med. Image Anal.*, vol. 1, no. 2, pp. 91–108, 1996.

[6] J. A. Sethian, *Level Set Methods and Fast Marching Methods*, 2 ed. Cambridge, U.K.: Cambridge Univ. Press, 1999.

[7] A. L. Yuille, P. W. Hallinan, and D. S. Cohen, "Feature extraction from faces using deformable templates," *Int. J. Comput. Vis.*, vol. 8, no. 2, pp. 99–112, 1992.

[8] A. Blake and M Isard, *Active Contours*. Berlin, Germany: Springer-Verlag, 1998.

[9] A. K. Jain, Y. Zhong, and M. P. Dubuisson-Jolly, "Deformable template models: A review," *Signal Processing*, vol. 71, no. 2, pp. 109–129, 1998.

[10] T. F. Cootes and C. J. Taylor, "A mixture model for representing shape variation," *Image Vis. Computing*, vol. 17, no. 8, pp. 567–574, 1999.

[11] A. Kelemen, G. Székely, and G. Gerig, "Elastic model-based segmentation of 3-D neuroradiological data sets," *IEEE Trans. Med. Imag.*, vol. 18, pp. 828–839, Oct. 1999.

[12] B. M. ter Haar Romeny, B. Titulaer, S. Kalitzin, G. Scheffer, F. Broekmans, E. te Velde, and J. J. Staal, "Computer assisted human follicle analysis for fertility prospects with 3D ultrasound," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 1999, vol. 1613, IPMI '99, pp. 56–69.

[13] A. K. Jain, Y. Zhong, and S. Lakshaman, "Object matching using deformable templates," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, pp. 267–278, Mar. 1996.

[14] R. Ronfard, "Region-based strategies for active contour models," *Int. J. Comput. Vis.*, vol. 13, no. 2, pp. 229–251, 1994.

[15] M. Brejl and M. Sonka, "Object localization and border detection criteria design in edge-based image segmentation: Automated learning from examples," *IEEE Trans. Med. Imag.*, vol. 19, pp. 973–985, Oct. 2000.

[16] S. M. Pizer, D. S. Fritsch, P. A. Yushkevich, V. E. Johnson, and E. L. Chaney, "Segmentation, registration, and measurement of shape variation via image object shape," *IEEE Trans. Med. Imag.*, vol. 18, pp. 851–865, Oct. 1999.

[17] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," in *Proc. Eur. Conf. Computer Vision*, vol. 2, H. Burkhardt and B. Neumann, Eds., 1998, pp. 484–498.

[18] ——, "Active appearance models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 681–685, June 2001.

[19] S. Sclaroff and A. P. Pentland, "Modal matching for correspondence and recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, pp. 545–561, July 1995.

[20] S. Sclaroff and J. Isodor, "Active blobs," in *Proc. IEEE ICCV*, 1998, pp. 1146–1153.

[21] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam, "The use of active shape models for locating structures in medical images," *Image Vis. Computing*, vol. 12, no. 6, pp. 355–366, 1994.

[22] P. P. Smyth, C. J. Taylor, and J. E. Adams, "Automatic measurement of vertebral shape using active shape models," in *Proc. Br. Machine Vision Conf.*, 1996, pp. 705–714.

[23] S. Solloway, C. J. Taylor, C. E. Hutchinson, and J. C. Waterton, "The use of active shape models for making thickness measurements from MR images," in *Proc. 4th Eur. Conf. Computer Vision*, 1996, pp. 400–412.

[24] N. Duta and M. Sonka, "Segmentation and interpretation of MR brain images: An improved active shape model," *IEEE Trans. Med. Imag.*, vol. 17, pp. 1049–1067, June 1998.

[25] G. Behiels, D. Vandermeulen, F. Maes, P. Suetens, and P. Dewaele, "Active shape model-based segmentation of digital X-ray images," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 1999, MICCAI '99, pp. 128–137.

[26] F. Vogelsang, M. Kohnen, J. Mahlke, F. Weiler, M. W. Kilbinger, B. Wein, and R. W. Günther, "Model based analysis of chest radiographs," *Proc. SPIE*, vol. 3979, pp. 1040–1052, 2000.

[27] G. Hamarneh and T. Gustavsson, "Combining snakes and active shape models for segmenting the human left ventricle in echocardiographic images," in *Computers in Cardiology*. Piscataway, NJ: IEEE, 2000, vol. 27, pp. 115–118.

[28] I. Dryden and K. V. Mardia, *The Statistical Analysis of Shape*. London, U.K.: Wiley, 1998.

[29] C. Goodall, "Procrustes methods in the statistical analysis of shapes," *J. Roy. Statist. Soc. B*, vol. 53, no. 2, pp. 285–339, 1991.

[30] C. Rao, *Linear Statistical Inference and Its Applications*. New York: Wiley, 1973.

[31] P. Sozou, T. F. Cootes, C. J. Taylor, and E. DiMauro, "A nonlinear generalization of point distribution models using polynomial regression," *Image Vis. Computing*, vol. 13, no. 5, pp. 451–457, 1995.

[32] A. Hill, T. F. Cootes, and C. J. Taylor, "Active shape models and the shape approximation problem," *Image Vis. Computing*, vol. 14, no. 8, pp. 601–607, 1996.

[33] J. J. Koenderink and A. J. van Doorn, "Representation of local geometry in the visual system," *Biological Cybern.*, vol. 55, pp. 367–375, 1987.

[34] L. M. J. Florack, B. M. ter Haar Romeny, J. J. Koenderink, and M. A. Viergever, "The Gaussian scale-space paradigm and the multiscale local jet," *Int. J. Comput. Vis.*, vol. 18, no. 1, pp. 61–75, 1996.

[35] J. J. Koenderink, "The structure of images," *Biological Cybern.*, vol. 50, pp. 363–370, 1984.

[36] J. J. Koenderink and A. J. van Doorn, "The structure of locally orderless images," *Int. J. Comput. Vis.*, vol. 31, no. 2/3, pp. 159–168, 1999.

[37] M. Unser, "Local linear transforms for texture measurements," *Signal Processing*, vol. 11, pp. 61–79, 1986.

[38] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[39] A. Whitney, "A direct method of non parametric measurement selection," *IEEE Trans. Computing*, vol. C-20, pp. 1100–1103, 1971.

[40] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recogn. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994.

[41] P. Somol, P. Pudil, J. Novovicova, and P. Paclik, "Adaptive floating search methods in feature selection," *Pattern Recogn. Lett.*, vol. 20, no. 11–13, pp. 1157–1163, 1999.

[42] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching in fixed dimensions," *J. ACM*, vol. 45, no. 6, pp. 891–923, 1998.