Strings: Variational Deformable Models of Multivariate Continuous Boundary Features

Sennay Ghebreab and Arnold W.M. Smeulders, *Senior Member*, *IEEE*

Abstract—We propose a new image segmentation technique called strings. A string is a variational deformable model that is learned from a collection of example objects rather than built from a priori analytical or geometrical knowledge. As opposed to existing approaches, an object boundary is represented by a one-dimensional multivariate curve in functional space, a feature function, rather than by a point in vector space. In the learning phase, feature functions are defined by extraction of multiple shape and image features along continuous object boundaries in a given learning set. The feature functions are aligned, then subjected to functional principal components analysis and functional principal regression to summarize the feature space and to model its content, respectively. Also, a Mahalanobis distance model is constructed for evaluation of boundaries in terms of their feature functions, taking into account the natural variations seen in the learning set. In the segmentation phase, an object boundary in a new image is searched for with help of a curve. The curve gives rise to a feature function, a string, that is weighted by the regression model and evaluated by the Mahalanobis model. The curve is deformed in an iterative procedure to produce feature functions with minimal Mahalanobis distance. Strings have been compared with active shape models on 145 vertebra images, showing that strings produce better results when initialized close to the target boundary, and comparable results otherwise.

Index Terms—Machine learning, deformable models, energy minimization, multivariate statistics, shape analysis, functional data analysis, chemometrics, active shape models.

1 Introduction

In advanced image segmentation problems, object boundaries frequently have inhomogeneous characteristics. The shape of a target boundary may be blunt at some parts and strongly convoluted at other parts. Or, the image gradient along a target boundary may be clearly visible and pointing outward at some places, while it is hardly defined due to neighboring objects at other places. We note that, in many advanced segmentation problems, boundaries are fractured, occluded, convoluted, or inhomogeneous otherwise, requiring the definition of multiple features for accurate boundary description. For this reason, it is imperative to construct inhomogeneous boundary models for application in image segmentation. We learn such boundary models by exploring the information contained in large collections of example images rather than constructing them from analytical or geometrical knowledge rules.

Learning in the context of boundary-based image segmentation has received considerable attention in literature, e.g., in [2], [13], [22], [26], in particular within the active shape model framework [6]. In active shape models, the appearance of a boundary is learned by statistical analysis of feature values of a set of example boundaries. During segmentation, the learned boundary model is used as a reference for evaluation of potential boundaries recorded by an active shape. The generic approach of active shape models is very appealing. However, they often fail to fully exploit the multivariate continuous nature of boundaries. The question raised here is how multiple continuous boundary features, such as edge gradient and contour curvature, are exploited for learning a continuous variational image segmentation model.

We propose a unified approach to learning structurally different boundary features for multifeature image segmentation. The problem of learning is transposed into one of analyzing the closed functional curves in feature space that best describe average feature values and the most important variations therein. Image segmentation is conceived of as an iterative procedure of recording multiple continuous feature values, weighting these feature values to amplify the statistically most descriptive features, and qualifying the weighted feature values with respect to the values seen in the learning set. The proposed string segmentation model combines theory from functional data analysis [24] with theory from chemometrics [21].

2 Related Work

A number of active shape models have been presented in literature. The method introduced by Cootes et al. in [6] makes use of established statistical techniques to construct a shape model from examples. In the reference, a shape is represented as an N-vector of vertices

$$\mathbf{x} = [(x_1, y_1), \ldots, (x_N, y_N)]^T. \quad (1)$$

- *S. Ghebreab was with the Intelligent Sensory Information Systems Group, Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands. He is with the Biomedical Imaging Group Rotterdam, Room Ee2167, Departments of Medical Informatics and Radiology, Erasmus MC, University Medical Center Rotterdam, The Netherlands. E-mail: s.ghebreab@erasmusmc.nl.*
- *A.W.M. Smeulders is with the Intelligent Sensory Information Systems Group, Informatics Institute, Faculty of Sciences, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands. E-mail: smeulders@science.uva.nl.*

Manuscript received 1 Nov. 2001; revised 27 May 2002; accepted 18 Mar. 2003.
Recommended for acceptance by M.A.T. Figueiredo, E.R. Hancock, M. Pelillo, and J. Zerubia.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 115300.

Assuming two-dimensional vertices, each N vertex is a single point in a $2N$-dimensional vector space. The set of M shape examples subsequently forms a learning set of size M

$$\mathcal{L} = \{\mathbf{x}_1, \ldots, \mathbf{x}_M\}. \tag{2}$$

To remove variation attributed to stretching, shearing, and rotation of the shapes from the learning set, the example shapes are aligned by Procrustes analysis [15], aiming at minimizing the sum of distances of each example shape to the average

$$\epsilon_{\mathcal{L}} = \sum_{m=1}^{M} ||\mathbf{x}_m - \bar{\mathbf{x}}||^2, \tag{3}$$

where $\bar{\mathbf{x}}$ is an initial estimate of the average shape, with $||\bar{\mathbf{x}}|| = 1$. The M shape examples are aligned one-by-one, with each iteration refining the estimation of the average shape.

Assuming the cluster of aligned shapes forms an ellipsoid, in [6] principal components analysis is performed to reduce the dimensionality of the data using the covariance matrix

$$\mathbf{C}_{\mathcal{L}} = \frac{1}{M-1} \sum_{m=1}^{M} (\mathbf{x}_m - \bar{\mathbf{x}})(\mathbf{x}_m - \bar{\mathbf{x}})^T, \tag{4}$$

where $\bar{\mathbf{x}}$ now denotes the average shape of the aligned example shapes. The eigenvalues λ_n of $\mathbf{C}_{\mathcal{L}}$, with $\lambda_n \geq \lambda_{n+1}$ for $n = 1, \ldots, 2N$, tell the amount of variance captured by each principal component. The largest fraction of the total variance is given by the first $Q \leq 2N$ eigenvalues:

$$\lambda = \sum_{q=1}^{Q} \lambda_q. \tag{5}$$

The eigenvalues λ_q describe the most significant modes of variation and the corresponding eigenvectors describe the dimensions in which they occur. A shape instance $\mathbf{x}$ is explained as the mean shape plus some linear combination of these eigenvectors:

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}^T, \tag{6}$$

where the $Q \times 2N$ matrix $\mathbf{P}$ contains the first Q eigenvectors and $\mathbf{b}$ is a Q-vector of weighting coefficients. New plausible shapes are generated by varying $\mathbf{b}$ within suitable limits, derived by statistically examining the distribution of weighting coefficients required to generate the learning data. Based on feasible shapes, high gradient boundaries are sought in an image.

A number of shortcomings of active shape models have been recognized in literature. One problem arises when the learning set $\mathcal{L}$ is contaminated with outliers, influencing the statistics adversely. For this reason, Duta et al. [11] remove outliers in an iterative approach using an inter-shape distance matrix that defines the mean alignment error between a polygonal approximation of a shape instance and an original shape from the learning set. The reduced learning set is used to construct a less distorted statistical shape model. Application of the corrected shape model for image segmentation consequently leads to better performance. We adopt this idea and we remove outliers in one step using another distance measure.

Another problem is that minimization of (3) only takes into account pose and scale differences between instances and does not account for nonlinear shape differences. This is solved by Duta et al. [11] using a flexible point matching technique that performs global similarity registration of two arbitrary sets of points and nonlinear registration based on local similarity of two curves. Instead of minimizing (3), the trade off between a compensated mean alignment error and the number of correspondences is minimized. This way the effect that unconstrained linear registration of two sets of points tends to shrink [11] one set with respect to the other is avoided. In addition, in [11] no correspondence between points is required when performing nonlinear registration. Other work on the shape alignment problem includes [20], where model compactness is optimized rather than variance. We also aim at linear and nonlinear registration, while avoiding problems with the cardinality of point sets.

A restriction of point distribution models is that deformation of the shape model by adjusting weighting vector $\mathbf{b}$ only allows limited deformation reflecting the variation in the learning set. For this reason, Wang and Staib [28] use prior models based on principal component analysis of additional covariance matrices. By replacing (4) with a weighted and mixed covariance matrix that deals with independence and smoothness, they are capable of building a wide range of shape models even when there are few examples shapes or if the learning set exhibits small variation. An earlier attempt to add artificial variation to the statistical model is found in [5]. We take a different road aiming at building models also capable of explaining objects dissimilar to the ones in the learning set.

Another limitation of active shape models concerns their inability to statistically capture image features around shape models. Cootes et al. [7] solve this problem by taking samples of the image intensity perpendicularly to a shape. By recording intensity profiles for all labeled points of the shape model, they arrive at an augmented learning set. As with the shape data, they compute the statistics of the intensity data using (1), (2), (3), (4), (5), and (6). The image features improve model specificity and, hence, segmentation accuracy. Other work on modeling image features includes [27], where optimal features are selected by nonlinear classifiers. We adopt the idea of modeling both shape and image features.

In [7], it is assumed that image feature values, sampled at different points along the boundary, are independent of one another while commonly feature values are spatially highly correlated. For this reason, Haslam et al. [17] propose a probabilistic fitness measure using concatenated intensity profiles, bringing in some continuity in feature values. This way, in addition to resolving the issue of dependence, they reduce the number of point distributions to two. Other work acknowledging the often continuous nature of boundary features, particularly shape features, is e.g., [1], [8], [20], [22], [25]. We embrace the thought of capturing spatial correlations among image features and we extend this idea to arbitrary features.

To overcome the problem of high correlation between structurally different features [17], Edwards et al. [12], and more recently Cootes et al. [4], propose an active appearance model that couples shape and image models more explicitly by using a single vector containing both shape and image feature values. By concatenating the feature values into a single vector and representing this vector as a point in a very high-dimensional space, a learning cluster is formed of combined shape and image features. This way, a more specific boundary model is obtained and more accurate

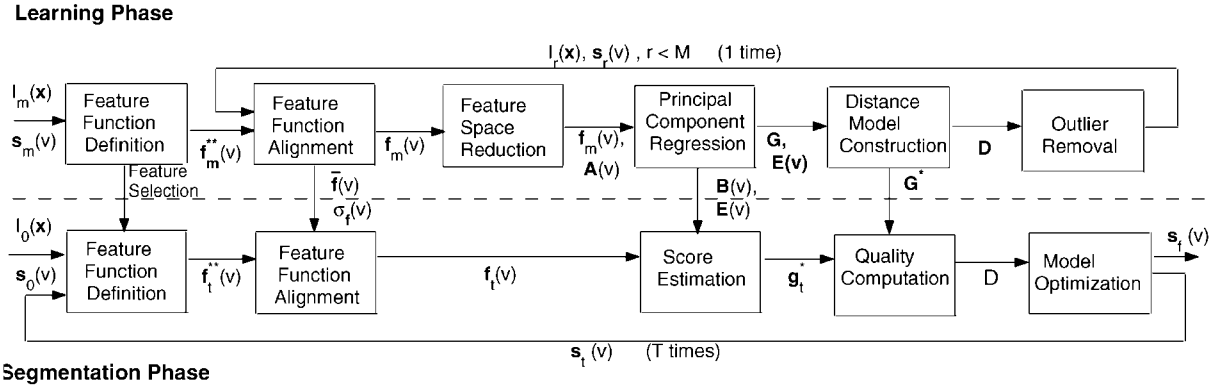

Fig. 1. Overview of the string segmentation technique. Note that at the start of the learning phase, we need a set of images $I_m(\mathbf{x}), m = 1 \ldots M$ with corresponding known segmentations $\mathbf{s}_m(v)$. Invariably, in this paper, boldface upper case indicates a matrix of functions, e.g., $\mathbf{A}(v), \mathbf{B}(v), \mathbf{E}(v)$, or in the case of $\mathbf{G}$ of scalars, boldface lower case indicates a vector of functions, e.g., $\mathbf{f}_t$, or scalars, e.g., $\mathbf{g}_t^*$, and regular lower case indicates a function or a scalar.

image segmentations are achieved. We adopt and generalize the idea of capturing correlations between structurally different boundary features.

In conclusion, we mark the following shortcomings addressed by the aforementioned methods:

1. the discrete point representation of boundaries, e.g., discrete points lead to discretization problems when point sets have different numbers of elements, requiring pseudocontinuous solutions,
2. insufficient correlation among structurally different features, e.g., disregarding interdependency between shape and image features may lead to unrealistic boundary models,
3. insubstantial exploitation of saliency, e.g., a priori specification of saliency rather than computation by optimality may result in indiscriminate features, and
4. redundant loss of feature information, e.g., currently applied procedures for dimensionality reduction do not attempt to minimize information loss.

It is the purpose of this study to integrally address these shortcomings. We do this with a technique we call *strings*.

3 Strings by Functional Data Analysis

Object boundaries are continuous. Hence, they should be represented by curves rather than by a collection of discrete points. Apart from this, boundaries are often described by multiple features. Hence, they should be represented by multivariate models rather than by a combination of univariate models. These boundary characteristics motivate the representation of boundary features by multivariate curves in functional space, rather than by points in vector space. When boundaries are represented this way, the learning and segmentation problems can be solved by functional data analysis [24] (see Fig. 1).

Each of the components in Fig. 1 will be described in more detail in the following sections. In Section 4, we will again elaborate on these components one by one to show the (intermediate) results for a vertebra application.

3.1 Feature Function Definition

In order to construct a statistical boundary model, example features need to be specified and their values computed. Commonly, the feature values are a set of coordinate values of boundary points and a set of image gradient values recorded orthogonally to those boundary points, see, for example, [4], [6], [11], [12].

We explore boundary features in example images using their known segmentation, represented by smooth curves $\mathbf{s} : \Re \rightarrow \Re^2$ parameterized by $v \in \Re$. Given the set of M input images $I_1(\mathbf{x}), \ldots, I_M(\mathbf{x})$ with corresponding segmentations $\mathbf{s}_1(v), \ldots, \mathbf{s}_M(v)$, the learning set consists of pairs of image and shape data (compare to (2))

$$\mathcal{L} = \{(I_1(\mathbf{x}), \mathbf{s}_1(v)), \ldots, (I_M(\mathbf{x}), \mathbf{s}_M(v))\}. \tag{7}$$

For the mth learning example, the shape $\mathbf{s}_m(v)$ relates to the image at points $I_m(\mathbf{s}_m(v))$. The relation is expressed in terms of N features derived from the shape (e.g., curvature) as well as from the image (e.g., isophote curvature). The mapping $\mathbf{f} : \Re \rightarrow \Re^N$ takes care of this, yielding feature functions $\mathbf{f}_m^{**}(v)$ in the N-dimensional functional space, where each dimension corresponds to one feature, i.e.,

$$\mathbf{f}_m^{**}(v) = [f_{m1}^{**}(v), \ldots, f_{mN}^{**}(v)]. \tag{8}$$

To capture features on and off a boundary, we exploit the local Taylor expansion [19] up to the second order, sampled at discrete but dense points along the boundary. This way, we have an approximately complete two-dimensional description of local boundary properties. Later, the principal components analysis, yet to be described, will cancel out linear dependencies in the Taylor set.

3.2 Feature Function Alignment

For statistical analysis of example boundaries, feature values at one point on an example boundary need to be compared with values at an equivalent point on an other example boundary. Commonly, this is achieved by scaling, rotating, and translating the examples so that they correspond as closely as possible. This reduces to aligning a set of discrete points when boundaries are represented by point distribution models as in e.g., [4], [6], [12].

In the context of our functional data, the alignment problem is a curve registration problem. Feature functions $\mathbf{f}_m^{**}(v)$ may differ due to the fact that they are not measured at the same path position v or due to small nonlinear differences. A shift of feature values along v and a nonlinear warping

account for this. Alignment reduces to finding the warping function $\omega_m(v)$ that produces the warped feature function

$$\mathbf{f}_m^*(v) = \mathbf{f}_m^{**}(\omega_m(v)). \tag{9}$$

The strictly monotonic warping function $\omega_m(v), v = 0, \ldots, \mathcal{V}$, is differentiable up to a certain order and has properties $\omega_m(0) = 0$ and $\omega_m(\mathcal{V}) = \mathcal{V}$. Consequently, corresponding starting points for and direction of alignment (clockwise or counterclockwise) need to be known. Here, $\omega_m(v)$ takes care of a shift and a nonlinear transformation by the roughness penalty approach described in [24]. In this case, we penalize by the size of the third derivative of $\omega_m(v)$.

Alignment of $\mathbf{f}_m^{**}(v)$ is done by the Procrustes method [15] using a global alignment criteria that computes the least squares distance to $\hat{\mathbf{f}}(v)$, the overall average feature function. This reduces to finding $\omega_m(v)$ such that

$$\omega_m(v) = \underset{\omega_m^*(v)}{\operatorname{argmin}} \sum_{m=1}^{M} \int_v ||\mathbf{f}_m^{**}(\omega_m^*(v)) - \hat{\mathbf{f}}(v)||^2 dv. \tag{10}$$

The warping functions are estimated in an iterative process where argument values for a feature function are shifted and transformed so as to minimize the least squares error. The estimated average $\hat{\mathbf{f}}(v)$ is updated by reestimating it from the partially aligned functions.

The final average feature function $\bar{\mathbf{f}}(v)$ is computed from the aligned set. It is subtracted from each feature function to normalize the range of feature values. This yields

$$\mathbf{f}_m(v) = \frac{\mathbf{f}_m^*(v) - \bar{\mathbf{f}}(v)}{\boldsymbol{\sigma}_{\mathbf{f}}(v)} \tag{11}$$

with units of variance due to normalization by the variance vector of functions

$$\boldsymbol{\sigma}_{\mathbf{f}}(v) = \left(\frac{1}{M-1} \sum_{m=1}^{M} ||\mathbf{f}_m^*(v) - \bar{\mathbf{f}}(v)||^2 \right)^{1/2}. \tag{12}$$

Normalization is required to reduce the influence of variational differences due to measurements in different units. The aligned normalized feature functions $\mathbf{f}_m(v)$ contain all information needed to statistically summarize features into a boundary model.

3.3 Feature Space Reduction

We perform principal components analysis to project the high-dimensional functional data to a smaller feature space expecting that the essential structure in the original data is preserved. This is admissible as long as the features exhibit a small number of modes of variation, covering a large part of the variability in the data.

Functional principal component analysis [24] computes the main modes of variation in the collection of N-dimensional feature functions $\mathbf{f}_m(v)$. The number of modes is derived from a given proportion of the variance as explained in the learning set. When the modes are numbered by $q = 1, \ldots, Q$, the central concept is that of taking the linear combination

$$g_{mq} = \sum_{n=1}^{N} \int_v f_{mn}(v) \alpha_{qn}(v) dv, \tag{13}$$

where $\alpha_{qn}(v)$ denotes a weighting function chosen so as to highlight variation in the data in dimension n. As before, $f_{mn}(v)$ is the nth dimension of the mth observed feature function. The principal component scores g_{mq} are used to produce more robust descriptions [10].

To obtain the value of g_{mq} for all q, the corresponding vectors of weighting functions $\boldsymbol{\alpha}_q(v) = [\alpha_{q1}(v), \ldots, \alpha_{qN}(v)]$ need to be computed. They are sought for one-by-one in such a way that they explain most of the variation in the learning data

$$\begin{aligned} \boldsymbol{\alpha}_q(v) &= \underset{\boldsymbol{\alpha}_q^*(v)}{\operatorname{argmin}} \frac{1}{M} \sum_{m=1}^{M} g_{mq}^2 \\ &= \underset{\boldsymbol{\alpha}_q^*(v)}{\operatorname{argmin}} \frac{1}{M} \sum_{m=1}^{M} \left(\sum_{n=1}^{N} \int_v f_{mn}(v) \alpha_{qn}(v) dv \right)^2, \end{aligned} \tag{14}$$

where $\boldsymbol{\alpha}_k(v)$, for each iteration k, is subject to the following orthonormal constraints

$$\sum_{n=1}^{N} \int_v \alpha_{qn}(v)^2 dv = 1 \tag{15}$$

$$\sum_{n=1}^{N} \int_v \alpha_{kn}(v) \alpha_{qn}(v) dv = 0, k \leq q. \tag{16}$$

Equations (15) and (16) ensure that the vector $\boldsymbol{\alpha}_1(v)$ contains most of the independent variation. After $\boldsymbol{\alpha}_1(v)$ has been established, the above process is continued until all Q significant modes of variation, each described by $\boldsymbol{\alpha}_q(v)$, are obtained: $\boldsymbol{\alpha}_1(v)$ captures the location with largest correlated variation, $\boldsymbol{\alpha}_2(v)$ the second largest of the remaining variance, etc., until most of the variation is explained.

The matrix of functions $\mathbf{A}(v) = [\boldsymbol{\alpha}_1(v), \ldots, \boldsymbol{\alpha}_Q(v)]^T$ indicates where along the boundary there is independent variation in the learning ensemble. Hence, the matrix $\mathbf{A}(v)$ of functional principal components captures the most important feature subspace.

3.4 Principal Components Regression

To construct an underlying model of feature values, the distribution of feature values seen in the learning set needs to be captured in statistical terms. In the active shape model literature [4], [6], [11], [12], feature reconstruction is done after projecting the features to the space spanned by the most important principal components. Hence, only part of the data in the learning set is subjected to modeling. We perform principal component regression [3] to obtain a predictive model from the feature functions in the learning set.

We define the matrix of functions $\mathbf{F}(v) = [\mathbf{f}_1(v), \ldots, \mathbf{f}_M(v)]^T$ and the matrix $\mathbf{G}$ by

$$\mathbf{G} = \mathbf{F}(v)\mathbf{A}(v) \tag{17}$$

with scalar elements according to the dot product defined in (13). We define a matrix of regression functions $\mathbf{B}(v) = [\boldsymbol{\beta}_1(v), \ldots, \boldsymbol{\beta}_Q(v)]^T$, with elements of the same N-dimensional functional form as the elements of $\mathbf{F}(v)$. To find the values of $\mathbf{B}(v)$, the matrix of feature functions $\mathbf{F}(v)$ is expressed as

$$\mathbf{F}(v) = \mathbf{G}\mathbf{B}(v) + \mathbf{E}(v) \tag{18}$$

with $\mathbf{E}(v) = [\boldsymbol{\epsilon}_1(v), \ldots, \boldsymbol{\epsilon}_M(v)]^T$ being the matrix of residual functions yet to be defined. Instead of regression on the original data, regression is performed on the principal

component scores containing information on how the feature samples correlate with one another. The matrix of regression functions $\mathbf{B}(v)$ gives an estimate of how the principal scores relate to the feature functions and what the contribution of each is toward defining an unknown feature function. The regression functions are computed by least squares minimization such that

$$\mathbf{B}(v) = \underset{\mathbf{B}^*(v)}{\operatorname{argmin}} \sum_{m=1}^{M} \int_v ||\mathbf{f}_m(v) - \mathbf{g}_m \mathbf{B}^*(v)||^2 dv. \tag{19}$$

Since there are no particular restrictions on the way in which the matrix of functions $\mathbf{B}(v)$ varies as a function of v, the solution can be obtained by minimizing the least squares difference for each v separately. After least squares minimization, we have

$$\mathbf{E}(v) = ||\mathbf{F}(v) - \mathbf{G}\mathbf{B}(v)||^2. \tag{20}$$

With the help of the estimated regression functions, we predict scores for an unknown feature function in the segmentation phase, reconstruct it in reduced space according the regression model, and evaluate it by examining the distance of its score to the cluster of scores corresponding to the feature functions in the learning set.

3.5 Mahalanobis Distance

In order to determine how a new boundary relates to the boundaries in the learning set, a distance measure needs to be defined. Commonly, the evaluation of boundary feature values is performed in terms of the Mahalanobis distance, i.e., the distance to the average normalized by the variation in each dimension. Following Cootes et al. [6], we use a Mahalanobis distance model [10] to compute the distance of a feature function to the average of the learning set.

The Mahalanobis distance model is obtained by augmenting $\mathbf{G}$ with the vector of residuals $\boldsymbol{\epsilon} = [\epsilon_1, \ldots, \epsilon_M]$ with elements defined as

$$\epsilon_m = \frac{1}{N} \sum_{n=1}^{N} \left(\int_v \epsilon_{mn}(v) dv - \frac{1}{M} \sum_{m=1}^{M} \int_v \epsilon_{mn}(v) dv \right). \tag{21}$$

This yields a $(Q+1) \times M$ augmented principal components scores matrix $\mathbf{G}^\star = [\mathbf{G}, \boldsymbol{\epsilon}^T]$. This improves the robustness of Mahalanobis distance calculation [10]. The Mahalanobis distance matrix is then defined as

$$\mathbf{D} = \frac{\mathbf{G}^{\star T} \mathbf{G}^\star}{(M+1)}. \tag{22}$$

The Mahalanobis distance of $\mathbf{f}_m(v)$ to the average $\bar{\mathbf{f}}(v)$ is computed using $\mathbf{g}_m^\star = [\mathbf{g}_m, \epsilon_m]$:

$$D^2(\mathbf{f}_m(v), \bar{\mathbf{f}}(v)) = \mathbf{g}_m^\star \mathbf{D}^{-1} \mathbf{g}_m^{\star T}. \tag{23}$$

Note that D depends on $\mathbf{g}_m^\star$. Hence, in addition to variation, the distance model also considers the valuable additional discriminating factor of residual information [21]. In the segmentation phase, the Mahalanobis distance model will be used as an objective function to find a boundary in a new image that resembles the ones seen in the learning set.

3.6 Stochastic Outlier Removal

As outlier feature functions in the learning set can have a severe influence on the discrimination ability of the Mahalanobis distance model, they are removed following [11]. Those with a Mahalanobis distance exceeding a threshold τ_M are considered outliers. For a Gaussian distribution, $\tau_M = 3$ corresponds to the removal of all instances that have a 1 percent probability of belonging to the class. Outlier removal results in a new learning set in which each element has a Mahalanobis distance of τ_M or less. Learning is done once again to obtain more appropriate regression and distance models. For simplicity of notation, we assume, in the following, that computations are performed on the reduced set rather than on the original learning set.

3.7 Deformable Strings

Having constructed a statistical boundary model in the learning phase, in the segmentation phase, we use this model as a reference for finding a boundary in a new image. We use the following information from the learning phase: average feature function $\bar{\mathbf{f}}(v)$, normalization function $\boldsymbol{\sigma}_\mathbf{f}(v)$, Mahalanobis distance model $\mathbf{G}^\star$, and regression functions matrix $\mathbf{B}(v)$.

Now, consider the active feature function $\mathbf{f}_t^{**}(v)$, which will be deformed in time t. We call it a string. The string lives in the N-dimensional feature space and, hence, we can compute its Mahalanobis distance to $\bar{\mathbf{f}}(v)$. As in the learning phase (see (9)), we first obtain

$$\mathbf{f}_t^*(v) = \mathbf{f}_t^{**}(\omega_t(v)), \tag{24}$$

where, similar to (11), the warping function $\omega_t(v)$ is computed such that

$$\omega_t(v) = \underset{\omega_t^*(v)}{\operatorname{argmin}} \int_v ||\mathbf{f}_t^{**}(\omega_t^*(v)) - \bar{\mathbf{f}}(v)||^2 dv. \tag{25}$$

We normalize the feature function in analogy to (11) and (12). Normalization yields

$$\mathbf{f}_t(v) = \frac{\mathbf{f}_t^*(v) - \bar{\mathbf{f}}(v)}{\boldsymbol{\sigma}_\mathbf{f}(v)}. \tag{26}$$

The quality of $\mathbf{f}_t(v)$ with respect to the matrix $\mathbf{F}(v)$ of feature functions is determined from the relation of its corresponding score vector $\mathbf{g}_t$ to the cluster of example scores contained in $\mathbf{G}$. The vector $\mathbf{g}_t$ is estimated by solving

$$\mathbf{f}_t(v) = \mathbf{g}_t \mathbf{B}(v)^T + \boldsymbol{\epsilon}_t(v). \tag{27}$$

That is, the equation estimates the score of a new feature function on the basis of the principal component regression model obtained in Section 3.4. The principal component scores are estimated by least squares minimization such that

$$\mathbf{g}_t = \underset{\mathbf{g}_t^*}{\operatorname{argmin}} \int_v ||\mathbf{g}_t^* \mathbf{B}(v)^T - \mathbf{f}_t(v)||^2 dv. \tag{28}$$

In analogy to (20), we augment $\mathbf{g}_t$ with the residual of the least squares minimization as an additional discriminating factor for the string. The $(Q+1)$-vector $\mathbf{g}_t^\star = [\mathbf{g}_t, \epsilon_t]$ is obtained on the basis of

$$\boldsymbol{\epsilon}_t(v) = ||\mathbf{g}_t \mathbf{B}(v)^T - \mathbf{f}_t(v)||^2 \tag{29}$$

TABLE 1
Features in Our Implementation Defining the Dimensions of $\mathbf{f}^{**}(v)$

Dimension	Feature	Definition
1	contour curvature	$\frac{\mathbf{s}_x(v)\mathbf{s}_{yy}(v)-\mathbf{s}_{xx}(v)\mathbf{s}_y(v)}{(\mathbf{s}_x^2(v)+\mathbf{s}_y^2(v))^{3/2}}$
2	isophote curvature	$\frac{I_{xx}(\mathbf{x})I_y^2(\mathbf{x})-2I_x(\mathbf{x})I_y(\mathbf{x})I_{xy}(\mathbf{x})+I_{yy}(\mathbf{x})I_x^2(\mathbf{x})}{(I_x^2(\mathbf{x})+I_y^2(\mathbf{x}))^{3/2}}$
3	directional correspondence	$\nabla I(\mathbf{s}(v))\cdot\mathbf{n}(v)$

The first dimension is the contour curvature, the second the isophote curvature, and the third the directional correspondence between the normal $\mathbf{n}(v)$ to the shape $\mathbf{s}$ at v and the image gradient at $\nabla I(\mathbf{x})$ with $\mathbf{x}=\mathbf{s}(v)$.

by adding to $\mathbf{g}_t$ the residual sum

$$\epsilon_t = \frac{1}{N}\sum_{n=1}^{N}\left(\int_v \epsilon_{tn}(v)dv - \frac{1}{M}\sum_{m=1}^{M}\int_v \epsilon_{mn}(v)dv\right). \tag{30}$$

The Mahalanobis distance of $\mathbf{f}_t^{**}(v)$ is defined as the distance of the newly augmented score $\mathbf{g}_t^{\star}$ to the average of the cluster of example scores, i.e.,

$$D^2(\mathbf{f}_t(v),\bar{\mathbf{f}}(v)) = \mathbf{g}_t^{\star}\mathbf{D}^{-1}\mathbf{g}_t^{\star T}. \tag{31}$$

We use this quality measure for image segmentation by strings. The string $\mathbf{f}_t^{**}(v)$ is defined by features extracted from an active shape model $\mathbf{s}_t(v)$ and from the image in which that shape model lives. The string vibrates in feature space due to deformation of the shape model in image space. The shape model is freely deformed rather than constructed in reduced space as in [6] and [12]. Shapes that are less plausible are punished rather than prohibited.

3.8 Optimization

Finally, we formulate the segmentation problem as an optimization problem. The objective is to find a boundary that gives rise to a feature function with minimal Mahalanobis distance $D(.)$ to the ones seen in the learning set. To this end, the deformable shape model $\mathbf{s}_t(v)$ is deformed in the image $I_0(\mathbf{x})$ to suggest a statistically optimal boundary described by the deformable string $\mathbf{f}_t^{**}(v)$. Starting from an initial shape configuration $\mathbf{s}_{t=0}(v)$ the shape model is deformed by tuning its shape parameters in such a way that the state of minimal energy provides the optimal feature function. This reduces to optimizing $\mathbf{s}_t(v)$ such that

$$\mathbf{f}(v) = \underset{\mathbf{f}_t(v)}{\operatorname{argmin}}\, D(\mathbf{f}_t(v),\bar{\mathbf{f}}(v)). \tag{32}$$

We use simulated annealing [18] for optimization as it distinguishes between different local minima in the energy landscape. Starting off at the initial configuration, a sequence of iterations is generated, where each iteration consists of the random selection of a configuration from the neighborhood of the current configuration and the calculation of the corresponding change in the energy value. A transition is achieved from one configuration into another one by a small perturbation of the x and y-coordinates of $\mathbf{s}_t(v)$. If the change from time t to $t+1$ yields negative $\Delta D = D(\mathbf{f}_{t+1}(v),\bar{\mathbf{f}}(v)) - D(\mathbf{f}_t(v),\bar{\mathbf{f}}(v))$, the transition is accepted unconditionally; if the cost function increases the transition is accepted with a probability based upon the Boltzmann distribution $p = e^{|-\frac{\Delta D}{kT}|}$, where k is a constant and the temperature T is a control parameter. This temperature is gradually lowered throughout the segmentation from a sufficiently high starting value, i.e., a temperature where almost every proposed transition, both positive and negative, is accepted to a freezing temperature, where no further changes occur. The temperature is decreased in stages and, at each stage, the temperature is kept constant until thermal quasi-equilibrium is reached.

4 Experiments and Results

We illustrate and discuss model construction and image segmentation, step by step according to Fig. 1. We use 145 annotated and digitized NHANES X-ray images [14] of normal cervical vertebrae, acquired from the National Library of Medicine. As can be seen from Fig. 1, the vertebra boundary in these images is ill-defined. It is characterized by the presence of interfering boundaries (e.g., the adjacent vertebra), convoluted boundary parts (e.g., tips of the vertebral body), missing image evidence (e.g., at the pedicles) and, in this case, very poor image quality. These characteristics complicate model construction and image segmentation.

4.1 Feature Instantiation

For boundary description, we use a repertoire of features. We confine ourselves to invariant features as they generalize applicability, but more importantly, as they minimize the need for feature alignment. We use the features listed in Table 1 and their derivatives up to second order. As a consequence, $N=9$ features are measured along each example vertebra boundary.

In our application, the vertebra boundary is represented by seven discrete points, manually marked in the image by a medical expert. For this reason, we compute a continuous approximation of the boundary by interpolation of a curve $\mathbf{s}(v)$ through the seven points. This curve and its corresponding nine-dimensional feature function $\mathbf{f}^{**}(v)$ are tensor product B-Splines [23]:

$$\mathbf{s}(v;\mathbf{p}) = \sum_{k=1}^{K} B_k(v)\mathbf{p}_k \tag{33}$$

$$\mathbf{f}^{**}(v;\mathbf{q}) = \sum_{j=1}^{J} B_j(v)\mathbf{q}_j. \tag{34}$$

Basis functions $B_k(v)$ correspond to the $K=7$ manually marked points $\mathbf{p}_k$. The $N=9$ features, and also point coordinates values, are computed along $\mathbf{s}(v;\mathbf{p}_k)$ at 100 samples. The 100 coordinate values are used for redefinition of the 7-vector of control points $\mathbf{p}_k$ to a 100-vector of control points, to be used in active shape model segmentation (yet to be

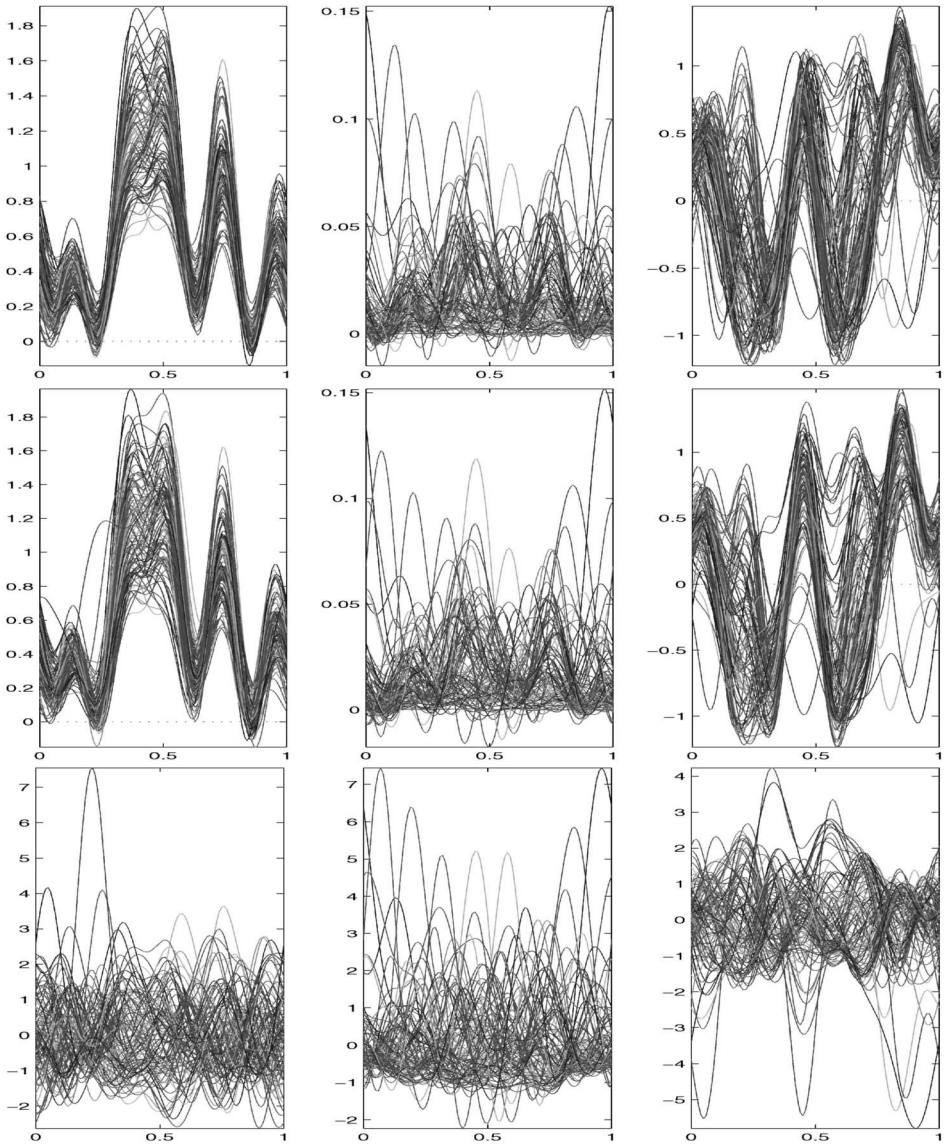

Fig. 2. **Upper row**: Feature values along path parameter v for all $M = 145$ examples. Left: Contour curvature values $f^{**}_{m1}(v)$. Middle: Isophote curvature values $f^{**}_{m2}(v)$. Right: Directional correspondence values $f^{**}_{m3}(v)$. **Middle row**: Results of registering feature functions vector $\mathbf{f}^{**}_{m}(v)$. Left: $f^{*}_{m1}(v)$. Middle: $f^{*}_{m2}(v)$. Right: $f^{*}_{m3}(v)$. Note that only three dimensions are shown out of the nine. **Lower row**: Results of normalizing feature functions $\mathbf{f}^{*}_{m}(v)$. Left: $f_{m1}(v)$ corresponding to contour curvature. Middle: $f_{m2}(v)$ corresponding to isophote curvature. Right: $f_{m3}(v)$ corresponding to directional correspondence. Note the increased variation in contour curvature values and directional correspondence values in comparison with the upper row due to centering and scaling.

described). Feature functions $\mathbf{f}^{**}(v;\mathbf{q}_j)$ are defined by basis functions $B_j(v)$, corresponding to the $J = 100$ N-dimensional control points $\mathbf{q}_j$, i.e., the sampled feature values. As these feature values are not always smooth, we impose regularity by using basis expansions with a relatively small number of basis functions [24].

4.2 Learning Phase

We discuss the results of the learning phase for the vertebra application. We only discuss the 0th-order derivative values of the features, i.e., only the first three dimensions of $\mathbf{f}_m(v)$.

The upper row of Fig. 2 illustrates contour curvature values $f^{**}_{m1}(v)$, isophote curvature values $f^{**}_{m2}(v)$, and directional correspondence values $f^{**}_{m1}(v)$ computed from $\mathbf{s}_m(v)$ and the image data arrays (see (8)). The presence of correlated structure in the contour curvature values is apparent. The peaks in curvature values correspond with the tips of the vertebral body and correlate with the peaks in $f^{**}_{m2}(v)$. This is more clearly seen from the average functions in the upper row of Fig. 3. Note that the isophote curvature is badly defined along most parts of the vertebra boundary. More correlated structure is seen in the feature values of directional correspondence.

The results of aligning feature functions $\mathbf{f}^{**}_{m}(v)$ by the iterative Procrustes procedure in (9) are illustrated in the middle row of Fig. 2. The majority of the feature functions already has a good alignment thanks to the a priori manual registration of the discrete points by medical experts (from which the continuous shapes and feature functions are derived). Consequently, alignment brings no significant changes to feature functions $\mathbf{f}^{**}_{m}(v)$. Alignment is more important during segmentation where the starting points for sampling are unknown.

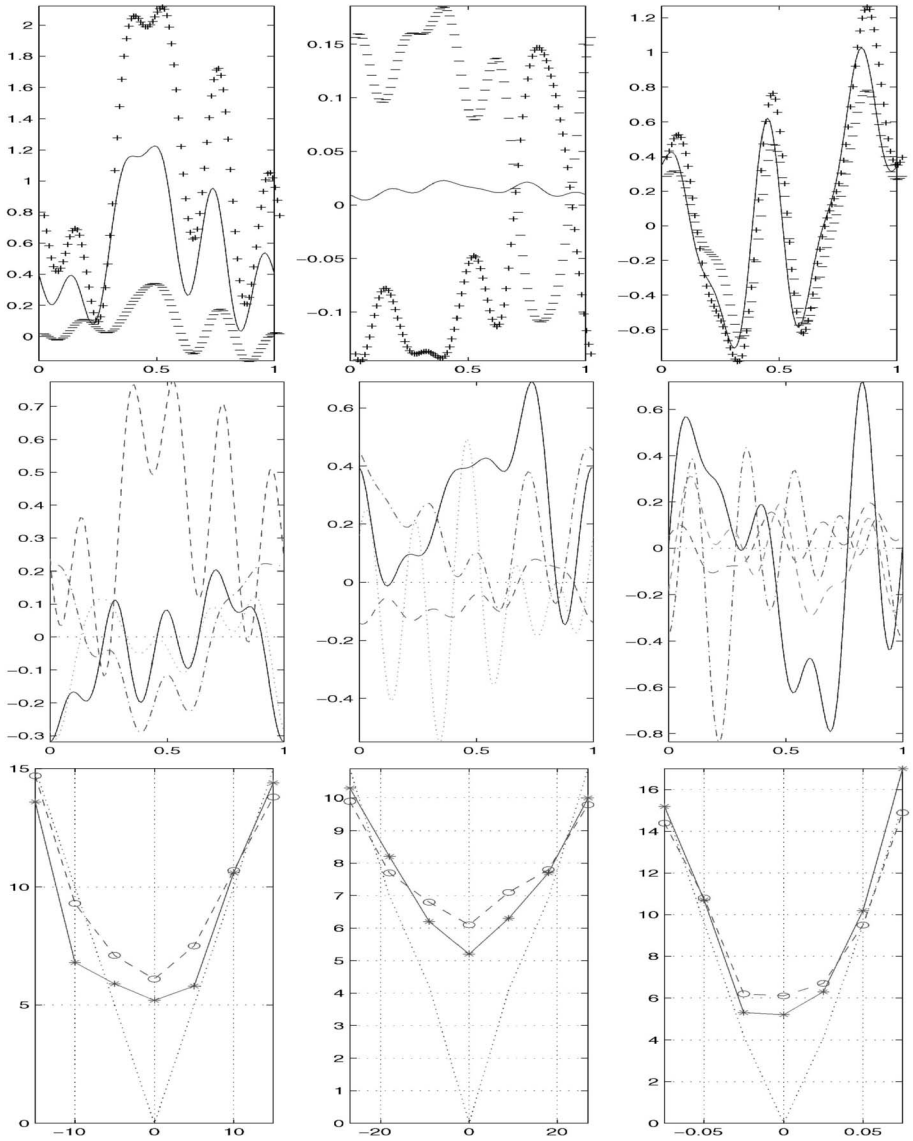

Fig. 3. **Upper row**: Average feature functions and the effects of adding (+) and subtracting (-) two standard deviations of the first principal component: $\bar{\mathbf{f}}(v) + / - 2$ standard deviations of $\boldsymbol{\alpha}_1(v)$. From left to right the effects for: contour curvature, isophote curvature, and directional correspondence. **Middle row**: The regression functions indicate for all $N = 3$ features the relative importance with respect to the $Q = 4$ principal components. Left: Functions $\beta_{11}(v), \ldots, \beta_{41}(v)$ for contour curvature. Middle: $\beta_{12}(v), \ldots, \beta_{42}(v)$ for isophote curvature. Right: $\beta_{13}(v), \ldots, \beta_{43}(v)$ for directional correspondence. **Lower row**: The average root mean square distance as a function of translation (left), rotation (middle), and scaling (right) for the initial shape (dotted line), and the final shape for active shape model segmentation (dashed line) and string segmentation (solid line).

The effects of normalizing $\mathbf{f}_m^*(v)$ are shown in the lower row of Fig. 2. Note that the feature functions exhibit considerable variation even though they have been aligned in the previous step. This is because centering and scaling according to (11) suppresses little variation and amplifies large variation. The feature functions with extreme variation are candidate outliers, to be removed later from the learning set.

The composite effect of adding and subtracting two standard deviations of the first principal component $\boldsymbol{\alpha}_1(v)$ to the average feature function $\bar{\boldsymbol{f}}(v)$ is shown in the upper row of Fig. 3. We reduce the feature space using $Q = 4$ principal components (see (14)), capturing 83.9 percent of the total variability in the data. Q has been set to four because, in this application, we expect four corners, hence four places, in the model where the learning data shows independent variation in feature values. Note, the displays remind of the diagrams in physics of modes of vibration of a string fixed at both ends, hence the name *strings*.

Finally, regression functions $\mathbf{B}(v)$, obtained using (19), are shown in the middle row of Fig. 3. The function $\beta_{qn}(v)$ indicates how the nth feature along the curve contributes to the qth principal component. Hence, the regression functions indicate which features are locally most important to define the boundary characteristics, implying that weighting is done in a way that exploits the most correlated and descriptive features along the boundary.

Not shown is the effect of the optional outlier removal. A total of 11 feature functions have been considered outliers and removed. After this, the whole learning procedure is repeated one more time. The average vector $\bar{\mathbf{f}}(v)$, the normalization vector $\boldsymbol{\sigma}_{\mathbf{f}}(v)$, the matrix of regression functions $\mathbf{B}(v)$, and of scalars $\mathbf{G}^\star$ are transferred to the segmentation phase.

4.3 Segmentation Phase

We perform string segmentation of the vertebra images using the above learning results as a reference. The B-spline $\mathbf{s}_t(v;\mathbf{p}_k), k = 1,\ldots,100$ is deformed in the image by repositioning the control points in an iterative procedure. Feature function $\mathbf{f}_t^{**}(v;\mathbf{q}_k)$ emanating from $\mathbf{s}_t(v;\mathbf{p}_k)$ is constructed from 100 samples along $\mathbf{s}_t(v;\mathbf{p}_k)$. The samples are taken at control points $\mathbf{p}_k$ to facilitate comparison with active shape model segmentation (described below), i.e., $\mathbf{p}_k$ constitute the point set of an active shape model. For computation of image features, derivatives are computed by convolution of the image with Gaussian derivatives at scale 4.0. Optimization of the active shape model is done by the simulated annealing optimization procedure, with the Boltzmann factor set to 0.99, thermal equilibrium defined as a 10 percent or smaller change in 10 random trials and a maximum of 50 iterations.

We also perform multiresolution active shape model segmentation [16]. The active shape model learns the distribution of points $\mathbf{p}_k$. The normalized gradient is captured perpendicularly to the boundary in profiles of length 3. The number of levels of resolution is set to 6, with level 0 the original image, level 2 the image with half the number of pixels, etc. After alignment of the shapes by rotation, translation, and scaling, a shape model is constructed in which 84.2 percent of the variance is explained by the first six principal components. We compare the performance of the string with that of the active shape model starting from the same initial shape. That shape is a transformation of the correct shape, projected onto the space spanned by the 4 and 6 principal components [16], respectively. This way, both models try to find an equilibrium starting from a configuration away from the average. The pixel neighborhood examined during optimization is nine and the maximum number of iterations 50.

We perform three experiments with strings and active shape models (software courtesy [16]). In each experiment, a shape is placed in the image on the correct position and transformed a known amount to verify robustness against initialization. The transformed shape forms the initial configuration. The first experiment concerns translations up to 15 pixels, the second rotations up to 30 degrees, and the third scalings with respect to a center point $\mathbf{c}$ with a factor up to 0.1. To measure the accuracy of the segmentation, the distance from the deformed shape, represented by the optimized control points, to the correct shape, represented by the original control points is computed using the root squared metric error. The experiments are performed systematically, excluding each learning instance from the learning set and using the excluded one to test the performance of the model built without it.

The lower row of Fig. 3 shows the average root squared metric of 145 segmentations for varying amounts of translation, rotation, and scaling. As can be seen from the reduction in the root mean squared error, the initial shape almost always moves to the correct boundary for both strings and active shape models. The string segmentation method outperforms active shape model segmentation when the initial shape is close to the correct boundary, in spite of the fact that the ground-truth is poorly sampled. If not initialized close to the target, string segmentation produces results similar to those produced by active shape model segmentation, occasionally worse. This sensitivity to initialization is attributed to the fact that no pose parameters are optimized to explicitly account for pose corrections, in contrast to active shape model segmentation. The active shape model finds the correct boundary even from a large distance if that boundary is well-defined. However, when the image evidence is vague, the active shape model tends to get trapped in a local minima far away from the correct boundary due to wrong pose optimization. For strings, we expect performance improvement by optimization of pose parameters in addition to shape parameters under the condition that pose parameters are restricted to admissible ranges. Equivalent improvements may be expected for active shape models by restricting pose parameters.

From the lower row in Fig. 3, we observe that even when the initial shape is the correct shape, optimization brings it to rest at an average of almost five pixel distance from the correct shape. We note that the points marking the vertebra boundaries in our learning set have been placed by a single medical experts and that variation of five or more pixels in manual point placement can be expected. Hence, the structural error of approximately five pixels is largely ascribed to intra-observer variability. We expect an improvement of performance proportional to the accuracy of manual demarcation, either by more precise individual assignment or by using larger amounts of salient points per vertebra. As a general model also contributes to the structural error, we expect that a dedicated boundary model, e.g., a C1 vertebra model, will bring in more accuracy and specificity to the boundary model.

The upper row in Fig. 4 shows an example result for the active shape model segmentation and the string segmentation of a NHANES II vertebra image. Segmentation typically takes about 15 seconds and 95 seconds, respectively, on a standard machine. In spite of the elaborate matrix function manipulations, performance is still very good. Much of the processing time is attributed to feature function alignment. Also, the exhaustive search of simulated annealing contributes to the high computational cost. We expect a significant reduction in computational cost with landmark-based alignment, a more efficient optimization technique and optimization of our code for speed. Note that much of the erroneous solutions of the active shape model [16] in Fig. 4 are global of nature. Apparently, the image evidence around the target boundary is too vague to be conclusive for such a model. Hence, the active shape model converges with a feasible shape in an unacceptable pose. For the string segmentation, the erroneous solutions are confined to segments, which are not on the sought boundary but rather on other visually-detectable edges, such as tissue/background edges or even edges produced by gray-scale intensity variation within a vertebra.

To demonstrate our method on a different set of images, we constructed a synthetic image containing an object with an inhomogeneous boundary (e.g., varying edge strength gradient, strongly curved parts). The object boundary was warped in four different ways with varying directions and magnitudes for the warping vectors. Each warp produced a set of five images. In all images, the object boundary was manually delineated to arrive at a set of 21 image and shape pairs with known number and modes of independent variation. We have used 16 pairs, including the original, for model construction and five for case segmentations. The five were chosen such that they have a different direction and magnitude of warping. An example is given in the middle row of Fig. 4, showing a warped image and object outline with

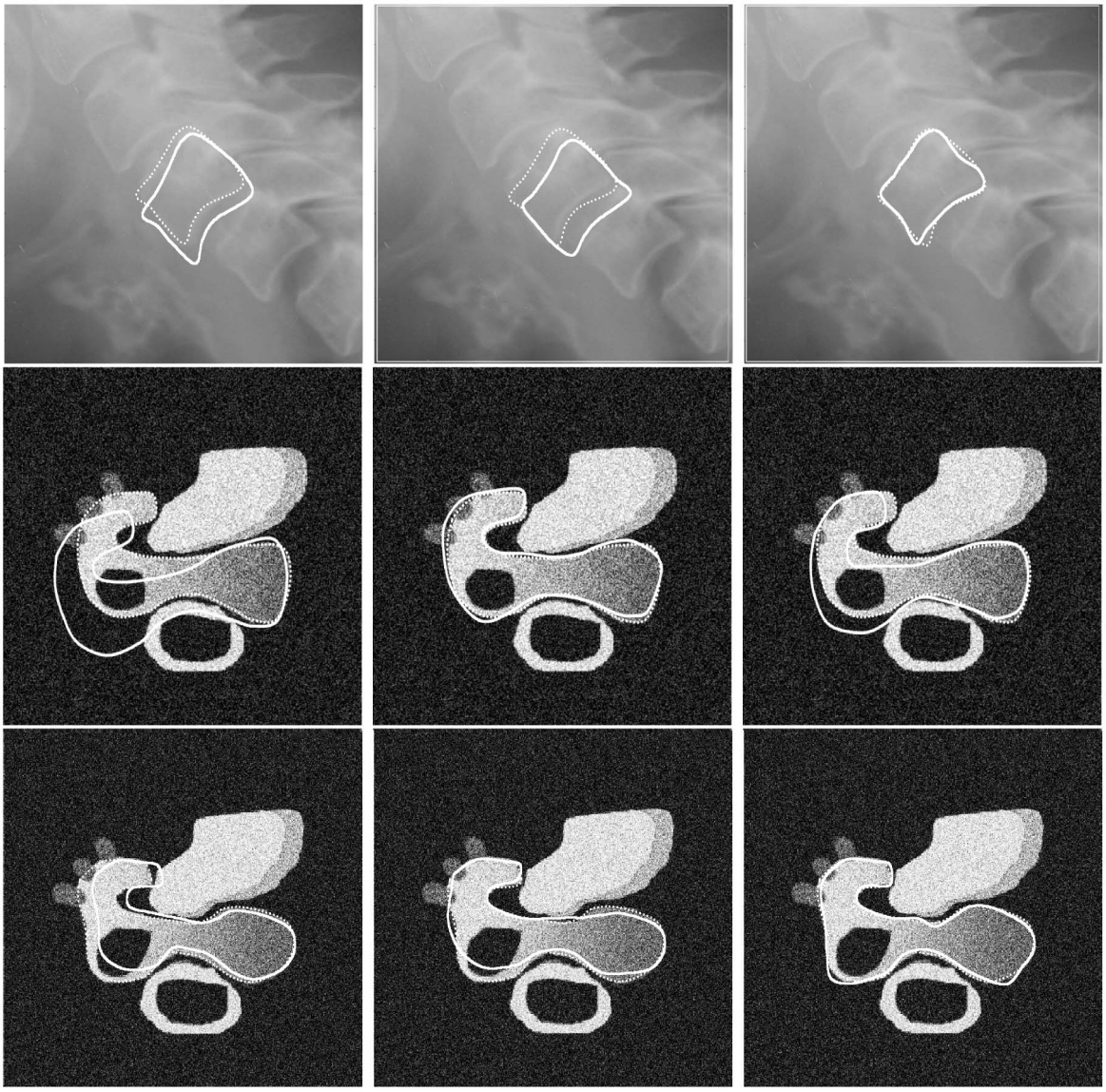

Fig. 4. **Upper row**: Segmentation of a NHANES cervical vertebra image (note the vague image evidence along the vertebra boundary due to the very low image quality). Left: An NHANES image with correct shape placed on it (dotted line) and transformation of this shape as the initial configuration for deformation (solid line). Middle: The correct shape (dotted line) and the active shape model segmentation (solid line). Right: The correct shape (dotted line) and string segmentation (solid line). **Middle row**: Segmentation of a warped synthetic image. Left: Synthetic image with correct shape placed on it (dotted line) and transformation of this shape as the initial configuration for deformation (solid line). Middle: The correct shape (dotted line) and the active shape model segmentation (solid line). Right: The correct shape (dotted line) and string segmentation (solid line). The bad initial configuration adversely influences the string segmentation. **Lower row**: Segmentation of another warped synthetic image with a different initialization. Left: image with correct shape (dotted line) and transformation of this shape as the initial configuration for deformation (solid line). Middle: The correct shape (dotted line) and the active shape model segmentation (solid line). Right: The correct shape (dotted line) and string segmentation (solid line). Note that the string correctly captures the bump in the upper boundary, where the active shape model fails to do so.

an appearance close to the average (left), an active shape model segmentation (middle) and a string segmentation (right). Here, the active shape outperforms the string, which fails to find the object boundary from a distance. The lower row of Fig. 4 shows the object most deviating from the average. The shape is initialized close to its boundary. The string finds that boundary even at places where it deviates much from average. On the basis of these cases, we note that active shape models outperform strings in case of distant initializations and objects with an appearance close to the average, whereas strings produce better results if the initialization is good and if the object of interest deviates much from the average.

5 Discussion and Conclusion

In conclusion, we have addressed a number of problems with statistical image segmentation models. We have represented boundaries by curves, solving some problems arising from discrete representations. By placing the curves in a multivariate functional feature space, we have properly dealt with the problem of spatial and feature interdependencies. We have performed curve registration, releasing us from the problem of missing points when aligning. We have substituted feature data by principal component scores, thereby not only reducing the feature space but also endorsing more robust computations. We have constructed a regression model for predicting unknown boundaries,

even ones dissimilar to the examples. We have built a distance model that also accounts for residual information. Finally, we have defined image segmentation as string optimization in multidimensional feature space.

In this paper, we have applied the string method for segmentation of rigidly shaped objects with ill-defined image evidence along their boundary, as often found in medical images. The method is also expected to work well for objects with more articulated shapes, provided there is correlated variation in the shape properties. When applied to objects of the real world, where objects often have an unknown and varying background, the definition of features should be such that values are computed from the inside of the object only. This provides a multivariate statistical description of the object area, rather than a description of the object boundary with its surrounding. It is important to note that continuous ground-truth segmentations are required to fully exploit the capabilities of strings.

With [4], [6], [11], [12], we share the observation that features should be learned rather than constructed from a priori geometrical or analytical knowledge. Geometrical and analytical features such as smoothness act as constraints on the solution since the resulting shape then will often be smooth at most places, regardless whether that is the appropriate solution or not. As in [6], we conceive of shape and image evidence as features and learn where they are effective in describing the statistics of the model. In fact, we have adopted the idea of mapping features to a space where the most important modes of variation are determined by principal components analysis and used for steering a shape for segmentation of an image.

The difference is that we conceive of a boundary as a multivariate continuous curve, requiring continuous functions to be learned. Cootes et al. [6] reduce the boundary to point sets. This introduces the following problems. First, points may be confused with other points if they are not labeled, leading to erroneous classification. Second, point location correspondence may be doted with error. Third, as recognized by Duta et al. [11], points may be missing, requiring a pseudocontinuous reconstruction of the boundary. Finally, measurements at discrete sample points are less reliable and lead to lose of spatial coherence.

We are less critically effected by these difficulties as we use closed continuous curves. We profit from functional data analysis in exploiting the spatial and feature correlations to explain the observed variation in the learning data rather than removing or down weighting such correlations. In explaining these variations, we also consider residual information, rather than omitting it. Contrary to Cootes et al. [6] we built a regression model from which we try to explain unknown boundaries, including the errors made in doing so. Another difference is that, instead of producing new models in reduced space and, hence, restricting the allowed shapes, we freely produce new segmentation shapes and punish implausible ones.

In comparing the performance of our string implementation with an active shape model implementation [16], we note the following. The active shape model performs better in the case the target object is visually well-defined and the initial shape is placed at a large distance from the target. This is thanks to the explicit optimization of pose parameters, e.g., rotation and scaling. When the initial shape is initialized close to the target object, the active shape model performs better in the case where the object has boundary properties like most objects in the learning set. If the target object has boundary properties that are dissimilar to the ones in the learning set, the string model produces more accurate results. We note that the prediction of feature values in reduced space, aids in explaining objects not seen in the learning set.

If the target boundary is visually ill-defined and the shape initialized distantly from the target, both string segmentation and active shape model segmentation perform poorly. For that condition, we found active shape models to produce much better or much worse results than string segmentation. On average, their performance is then similar. Model initialization remains a problem in view of the presence of disturbing attractors in the image. Strings outperform active shape models if the target object is visually ill-defined and the initialization is close to the target boundary because strings locally exploit the discriminative power of a repertoire of features. Also, the fact that strings cope better with spatial dependencies between feature values and the fact that they explicitly consider correlations among features improves segmentation under such highly demanding conditions.

Hence, we arrive at the conclusion that strings are particularly suited for learning variational models of objects with inherently multivariate continuous boundary properties. Strings are also very well suited for segmentation of complex scenes, where the visual evidence is vague, or where a multidimensional feature set is needed to capture an object boundary.

ACKNOWLEDGMENTS

The authors would like to thank Lister Hill National Center for Biomedical Communications, NLM/NIH for providing NHANES II images from the National Center for Health Statistics. They would also like to express gratitude to Dr. J.-M. Geusebroek, Dr. N. Vlassis, and Ir. J. van de Weijer for their assistance and the reviewers for their helpful comments and suggestions.

REFERENCES

[1] A. Baumberg and D. Hogg, "Learning Flexible Models from Image Sequences," *Proc. European Conf. Computer Vision,* pp. 299-308, 1994.

[2] M. Brejl and M. Sonka, "Object Localization and Border Detection Criteria Design in Edge-Based Image Segmentation: Automated Learning from Examples," *IEEE Trans. Medical Imaging,* vol. 19, pp. 973-985, 2000.

[3] S. Chatterjee, A.S. Hadi, and B. Price, *Regression Analysis by Example.* New York: John Wiley and Sons, 2000.

[4] T.F. Cootes, G.J. Edwards, and C. J Taylor, "Active Appearance Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 681-685, June 2001.

[5] T.F. Cootes and C.J. Taylor, "Combining Point Distribution Models with Shape Models Based on Finite Element Analysis," *Image Vision and Computing,* vol. 13, pp. 403-409, 1995.

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

[7] T.F. Cootes, C.J. Taylor, A. Lanitis, D.H. Cooper, and J. Graham, "Building and Using Flexible Models Incorporating Grey-Level Information," *Proc. IEEE Int'l Conf. Computer Vision,* pp. 242-246, 1993.

[8] R.H. Davies, T.F. Cootes, C. Twining, and C.J. Taylor, "An Information Theoretic Approach to Statistical Shape Modelling," *Proc. British Machine Vision Conf.,* pp. 3-11, 2001.

[9] I.L. Dryden and K.V. Mardia, *Statistical Shape Analysis.* New York: John Wiley and Sons, 1998.
[10] J. Duckworth, "What Is Chemometrics," *Proc. Int'l Conf. Near Infrared Spectroscopy,* 1995.
[11] N. Duta, A.K. Jain, and M.P. Dubuisson-Jolly, "Automatic Construction of 2D Shape Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 5, pp. 433-446, May 2001.
[12] G.J. Edwards, A. Lanitis, C.J. Taylor, and T.F. Cootes, "Statistical Models of Face Images: Improving Specificity," *Image and Vision Computing,* vol. 16, no. 3, pp. 203-211, 1998.
[13] S.D. Fenster and J.R. Kender, "Sectored Snakes: Evaluating Learned-Energy Segmentations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 9, pp. 1028-1034, Sept. 2001.
[14] "Plan and Operation of the Second National Health and Nutrition Examination Survey 1976-80," Programs and Collection Procedures, no. 15, DHHS publication no. (PHS) 81-1317, Nat'l Center for Health Statistics, July 1981.
[15] C. Goodall, "Procrustes Methods in the Statistical Analysis of Shape," *Royal Statistical Soc. Series B. Methodological,* vol. 53, no. 2, pp. 285-33, 1991.
[16] G. Hamarneh, "Active Shape Model: Modeling Shape Variations and Gray Level Information and an Application to Image Search and Classification," Technical Report R005/1998, Dept. of Signals and Systems, Chalmers Univ. of Technology, Sweden, 1998.
[17] J. Haslam, C.J. Taylor, and T.F. Cootes, "A Probabalistic Fitness Measure for Deformable Template Models," *Proc. British Machine Vision Conf.,* pp. 33-42, 1994.
[18] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by Simulated Annealing," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 5, no. 3, pp. 267-286, 1983.
[19] J.J. Koenderink, "The Structure of Images," *Biological Cybernetics,* vol. 50, pp. 363-370, 1984.
[20] A.C.W. Kotcheff and C.J. Taylor, "Automatic Construction of Eigenshape Models," *Medical Image Analysis,* vol. 2, pp. 303-314, 1998.
[21] R. Kramer, *Chemometric Techniques for Quantitative Analysis.* Marcel-Dekker, 1998.
[22] M. Leventon, E. Grimson, and O. Faugeras, "Statistical Shape Influence in Geodesic Active Contours," *Proc. Computer Vision and Pattern Recognition,* 2000.
[23] L. Piegl and W. Tiller, *The NURBS Book.* Berlin: Springer-Verlag, 1997.
[24] J. Ramsay and B.W. Silverman, *Functional Data Analysis.* Berlin: Springer-Verlag, 1997.
[25] S. Rowe and A. Blake, "Statistical Feature Modelling for Active Contours" *Proc. European Conf. Computer Vision,* vol. B, pp. 560-569, 1996.
[26] L.H. Staib and J.S. Duncan, "Boundary Finding with Parametrically Deformable Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 11, pp. 1061-1075, Nov. 1992.
[27] B. van Ginneken, A.F. Frangi, J.J. Staal, B. ter Haar Romeny, and M. Viergever, "Active Shape Model Segmentation with Optimal Features," *IEEE Trans. Medical Imaging,* vol. 21, no. 8, pp. 924-933, 2002.
[28] Y.M. Wang and L.H. Staib, "Boundary Finding with Prior Shape and Smoothness Models," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 7, pp. 738-743, July 2000.

Sennay Ghebreab received the MSc and PhD degrees from the University of Amsterdam, The Netherlands, in 1996 and 2002, respectively. From 1996 to 2002, he was a PhD student at the ISIS research group, Faculty of Sciences, University of Amsterdam, The Netherlands, where he did research in the fields of visual learning from examples, model-based analysis of images, and image retrieval by content. He is co-organizer of the first international VISIM Workshop on Information Retrieval and Exploration in Large Medical Image Collections. Since 2002, he has been with the BIGR research group, Departments of Medical Informatics and Radiology, Erasmus MC, University Medical Center Rotterdam, The Netherlands. His current research focuses on the fundamentals of information retrieval and knowledge discovery in biomedical multimedia databases.

Arnold W.M. Smeulders graduated from the Technical University of Delft in physics in 1977 (MSc) and in 1982 from Leyden University in medicine (PhD) on the topic of visual pattern analysis. He heads the ISIS research group with research on the theory, practice, and implementation of multimedia information analysis including image databases and computer vision and with an extensive record in co-operations with Dutch industry in the area of multimedia and video analysis. His current personal interest is characterized by semantic image retrieval, invariant representation of color and texture, the relation between pictures and language, learning vision, and the design of vision algorithms from first principles. He received a Fulbright grant to spend at Yale University in 1987, and a visiting professorship at the City University Hong Kong in 1996, and at ETL Tsukuba Japan in 1998. In 2000, he was elected fellow of the International Association of Pattern Recognition. He was an associated editor of *IEEE Transactions of Pattern Analysis and Machine Intelligence* from 1996-2001. He is a senior member of the IEEE and an honorary member of the Dutch Society for Pattern Recognition. He is director of the MultimediaN national initiative.