

Discriminant snakes for 3D reconstruction of anatomical organs

X.M. Pardo^{a,*}, P. Radeva^b, D. Cabello^a

^aDept. Electrónica e Computación, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain

^bComputer Vision Centre. Universitat Autònoma de Barcelona, 089193 Bellaterra, Barcelona, Spain

Received 26 October 2001; received in revised form 25 July 2002; accepted 19 December 2002

Abstract

In this work a new statistic deformable model for 3D segmentation of anatomical organs in medical images is proposed. A statistic discriminant snake performs a supervised learning of the object boundary in an image slice to segment the next slice of the image sequence. Each part of the object boundary is projected in a feature space generated by a bank of Gaussian filters. Then, clusters corresponding to different boundary pieces are constructed by means of linear discriminant analysis. Finally, a parametric classifier is generated from each contour in the image slice and embodied into the snake energy-minimization process to guide the snake deformation in the next image slice. The discriminant snake selects and classifies image features by the parametric classifier and deforms to minimize the dissimilarity between the learned and found image features. The new approach is of particular interest for segmenting 3D images with anisotropic spatial resolution, and for tracking temporal image sequences. In particular, several anatomical organs from different imaging modalities are segmented and the results compared to expert tracings.

© 2003 Elsevier B.V. All rights reserved.

Keywords: Snakes; Fisher linear discriminant analysis; Supervised learning; Segmentation; 3D medical images

1. Introduction

Three-dimensional segmentation of anatomical organs is widely used for diagnosis of diseases and planning of surgical interventions. However, the segmentation of medical images is still a difficult task due to the complexity and variability of the anatomic shapes, sampling artifacts, spatial aliasing and noise. Deformable models stand out from other segmentation techniques due to their ability to interpret sparse features and link them to obtain object contours.

Different examples of the application of deformable models can be found in the medical image literature (McInerney and Terzopoulos, 1996; Terzopoulos, 1987; Amini et al., 1990; Chen and Huang, 1992; Cohen and Cohen, 1993; Cootes et al., 1995; Blake and Isard, 1998;

Niessen et al., 1998, 1999; Zeng et al., 1999, Leventon et al., 2000). Obtaining a successful segmentation is intricately tied to the choice of features and the criteria used for discriminating the pattern classes (Jain et al., 1997).

Usually, image features are not selected in accordance with the specific organ and image modality; thus too many feature points that do not belong to the organ of interest are enhanced, while other important ones go unnoticed. Blake and Isard (1998) and Yuille et al. (1992) proposed the combination of different image features depending on the application domain. However, the best way of selecting and integrating different features remains an open problem. So, the feature space must be capable of representing all image features of interest. To achieve such a broad description, we propose employing a bank of Gaussian filters. Numerous physiological measurements support the theory that receptive field profiles can be modelled by Gaussian filters of various scales (Koenderink, 1984; Young, 1986; Florack et al., 1996). From a mathematical point of view, it has been proven that the complete family of scale-space Gaussian kernels defines a sufficient Car-

*Corresponding author. Tel.: +34-981-563-100; fax: +34-981-528-012.

E-mail address: pardo@dec.usc.es (X.M. Pardo).

tesian family for complete determination of local image structure (Florack et al., 1992).

A very important issue concerns the construction of an irreducible set of differential invariants in a fixed N -jet (the set of partial derivatives up to N th order). This irreducible set should provide optimal tools to measure the amount of *featureness* at a certain position for any generic image configuration (Niessen et al., 1997). We consider the problem of constructing an optimal subspace of the complete feature space which will better discriminate the different parts of the object boundary. For example, it is well-known that to discriminate texture patterns only second Gaussian derivatives are needed (Malik and Perona, 1990).

An important issue which must be addressed is the selection of filter scales. In general, an optimal single scale does not exist. One alternative lies in searching for the best local scale, but it depends on the imaging modality and the specific image feature (Subirana-Vilanova and Sung, 1992; Elder and Zucker, 1998). Multiscale analysis is usually used when there exists no clear way to choose the right scale. The decisions of how many scales should be used and how to combine them are the main issues of this approach (Pardo et al., 2001; and references therein).

The integration of the selection of features, and their scales, in a statistical framework is the most convenient approach to relate the segmentation technique to the specific task. Our approach combines the selection of the best scale, for the fine local fitting, with a multiscale approach to cope with cases where there is a big change in contour location from one slice to the next slice.

We can find in the literature statistical-based approaches, which preferably attract the curve towards image features consistent with their trained shapes (Yuille et al., 1992; Staib and Duncan, 1992; Chakraborty et al., 1996; Cootes et al., 1995; Nastar and Ayache, 1996; Blake and Isard, 1998) and/or greylevel appearance. Turk and Pentland (1991) used principal component analysis to segment face images in terms of a set of basis functions. Paragios and Deriche (1999) proposed a geodesic active contour based on a supervised statistical modeling, which combines a priori knowledge about the desired boundary, and the exterior and interior region properties. Yezzi et al. (1999) have presented active contours with a natural use of global and local information to segment regions distinguishable by a given set of statistics. Some authors consider the full appearance (shape and grey-level) of the object (Lanitis et al., 1995; Leventon et al., 2000).

Our discriminant approach describes and learns contour grey level appearances and integrates them into the snake formulation. The method extracts knowledge from samples on the adjacent slice in the same 3D image sequence. The contour configurations in a slice are good training samples for the algorithm that operates in the next slice, because there is a high correlation between adjacent slices. Unlike other probabilistic approaches, the statistic discriminant

snake (STD-snake) learns both the optimum features and discriminants. The external energy of the snake is defined as the distance from each curve part to its target contour in the feature space. Thus, each part of the snake curve makes its own interpretation of the image data and deforms thereby minimising this distance.

The paper is organized as follows. In Sections 2 and 3 we introduce the STD-snake based on Fisher linear discriminant analysis. In Section 4 we compare our proposal to classical snake, and give some results in the field of medical image. Finally, in Section 5 we outline the main conclusions of the work and future research.

2. Supervised feature learning

We represent the snake curve by means of a cubic B-spline, and therefore each snaxel (snake control point) influences four curve segments. For each snaxel, we define a *contour patch* as the central part of its four curve segments of influence, avoiding the overlap between adjacent contour patches. Each snaxel moves in accordance with the energy accumulated along its contour patch.

The STD-snake has a triple objective: to obtain a general representation of any contour class in terms of a set of features (Section 2.1), to determine the optimum discriminant for each desired contour part (Sections 2.2 and 2.3), and to deform towards the target object boundary (Section 3).

To make the snake discriminant, each snaxel must be able to distinguish between its corresponding contour target and other structures (other parts of the contour and contours of nearby objects). The discriminant snake should learn the more relevant features for each specific segmentation problem, and generalize the classical approach that only uses gradient information. For each desired contour part (class), a set of samples is obtained and then a supervised learning aimed at the maximization of the between-class scattering of image features, is accomplished. We define a parametric potential that depends on the features of each part of the target boundary. By parametric we mean that each snaxel sees the image features in a different way, because each one is aimed to fit different object boundary parts.

Fig. 1 illustrates the learning process carried out for each contour part. The features of the target contour part and the complementary class are extracted from the training contour and background. Then, Fisher linear discriminant analysis (FLDA) gives the optimal discriminant. During the optimization process each snaxel will evolve guided by its discriminant. The following sections describe in detail these steps.

2.1. Feature extraction

We propose to use a bank of Gaussian filters to extract

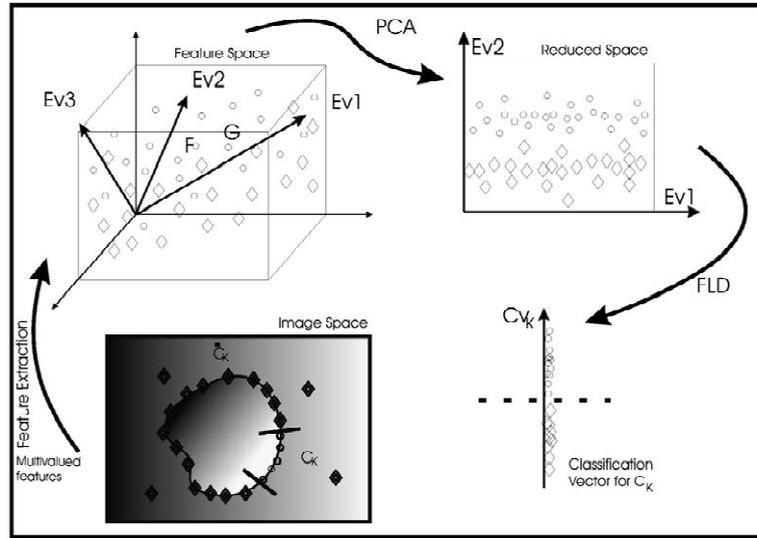


Fig. 1. From feature space to classification vector for each contour part.

the image features. Since the directional derivative operator is steerable, each filter $G^d(x, y, \sigma, \phi)$ of degree d and orientation ϕ can be synthesized by interpolating $d + 1$ basis functions $\{G^d(x, y, \sigma, \theta_k)\}_{k=1}^d$ (Freeman and Adelson, 1991). We define $\mathcal{G}_{\mathcal{D}\Sigma}$ as the \mathcal{D} -jet filter bank with N_Σ scales: $\sigma \in \{2^0, 2^1, \dots, 2^{N_\Sigma-1}\}$. The size of the bank of filters is a function of \mathcal{D} and N_Σ :

$$d_g = \dim(\mathcal{G}_{\mathcal{D}\Sigma}) = N_\Sigma \sum_{d=0}^{\mathcal{D}} (d + 1) = \frac{N_\Sigma}{2} (\mathcal{D}^2 + 3\mathcal{D} + 2). \quad (1)$$

Our bank of filters contains derivatives up to degree three, because variance of higher-order filters can be expected to approach that of image noise, and tend to be highly correlated to the outputs of lower order filters (Rao and Ballard, 1995). The impulse responses of the Gaussian filters are given by

$$\begin{aligned} G^0(x, y, \sigma) &= \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{1}{2} \left[\frac{x^2 + y^2}{\sigma^2}\right]\right\}, \\ G^1(x, y, \sigma, \theta_1) &= R\left(\left(-\frac{x}{\sigma^2}\right)G^0(x, y, \sigma)\right)^{\theta_1}, \\ G^2(x, y, \sigma, \theta_2) &= R\left(\left(-\frac{1}{\sigma^2} + \frac{x^2}{\sigma^4}\right)G^0(x, y, \sigma)\right)^{\theta_2}, \\ G^3(x, y, \sigma, \theta_3) &= R\left(\left(-\frac{3x}{\sigma^4} + \frac{x^3}{\sigma^6}\right)G^0(x, y, \sigma)\right)^{\theta_3}, \end{aligned}$$

where the superscript in $G^d(\cdot)$ represents the derivative degree, and $R(\cdot)^{\theta_d}$ represents a rotation of an angle $\theta_d \in \{k\pi/d + 1 | k = 0, 1, \dots, d\}$.

The convolution of $\mathcal{G}_{\mathcal{D}\Sigma}$ with the image \mathcal{I} allows the generalized description of edge, ridge, valley or whatever image intensity features. We define a multivalued feature \mathcal{P}_f as the result of this convolution:

$$\begin{aligned} \mathcal{P}_f: \mathbf{R} &\rightarrow \mathbf{R}^{d_g}, \\ \mathcal{I} &\rightarrow \mathcal{G}_{\mathcal{D}\Sigma} * \mathcal{I}. \end{aligned}$$

The image potential will refer to what each snaxel interprets on this multivalued feature.

2.2. External forces by statistic classifiers

Each snaxel will give different relevance to each component of the feature vector \mathcal{P}_f . In that sense, we shall talk about variable (parametric or locally defined) external energy. To determine the relative importance of each component of the feature vector, a supervised learning is performed. The algorithm starts with a set of N sample feature vectors on object contour and non-object contour $\{s_1, s_2, \dots, s_N\}$, $s_i \in \mathbf{R}^{d_g}$. For each contour patch k , each multivalued feature s_j is assigned to one of two classes $\{C_k, \bar{C}_k\}$, representing the pixels belonging to the contour k and the complementary class.

Then a FLDA provides the optimal linear discriminant for each patch k that maps the original d_g -dimensional space into a 1-dimensional space (see Appendix A). The discriminant for each patch k consists of the classifier vector \mathcal{V}_{C_k} and the class center \mathcal{O}_{C_k} of the samples of patch k in the feature space:

$$\begin{aligned} \mathcal{V}_{C_k} &= V_{k_opt}, \\ \mathcal{O}_{C_k} &= \mathcal{V}_{C_k}^T \mu_{C_k}, \end{aligned}$$

where V_{k_opt} maximizes the ratio between-class/within-class discrimination, and μ_{C_k} is the mean feature vector of the patch k .

We define the local external energy of the snake in terms of the distance D_{C_k} from the image features \mathcal{P}_f in the

current location of the snake to the desired contour features:

$$\begin{aligned} D_{C_k} &= (\mathcal{V}_{C_k}^T (\mathcal{P}_f - \mu_{C_k})) (\mathcal{V}_{C_k}^T (\mathcal{P}_f - \mu_{C_k})) \\ &= (\mathcal{V}_{C_k}^T \mathcal{P}_f - \mathcal{O}_{C_k})^2. \end{aligned} \quad (2)$$

We give an explicit expression for the external energy and forces in Section 3.

Each snaxel j learns different contour types $(\mathcal{V}_{C_j}, \mathcal{O}_{C_j})$ in slice $i - 1$, and therefore each snaxel interprets in different way the image features \mathcal{P}_f in slice- i . This behaviour of our method is illustrated in the synthetic example of Fig. 2. Fig. 2(a) contains a synthetic image of two circles on a

white background. Fig. 2(b,c) corresponds to the initial contour and final fitting of the snake guided by the classifiers (one per contour patch) in the multivalued feature space. Fig. 2(d,e) contains the components of the multivalued feature for two scales and derivatives up to degree three. Fig. 2(f) shows the image potentials as are seen by each snaxel.

A real example in the field of the medical images is shown in Fig. 3. Fig. 3(a) shows the training contour, where alternating colours represent different contour patches. Fig. 3(b–e) represents the distances (in feature space) to contour classes placed on the west, east, north and south sides of the training contour. As is desirable, the minimum of each distance map is around each learning contour class, and there is a maximum in the diametrically opposed part of the contour.

2.3. Continuity and regularization of classification vectors

In general, features smoothly vary along the boundary of an object in a medical image. If the contour is densely sampled, smooth variation of the classifiers can be expected to occur along the learning contour. Fig. 4 illustrates this hypothesis. It contains the representation of components of the classification vector along the internal contour of a coronary vessel in an intravascular ultrasound (IVUS) image. As can be seen, the features and the components of the classification vectors (V_{ij}) vary smoothly along the sequence of contour patches.

If the inter-slice resolution is close to the intra-slice resolution, the similarity between adjacent slices will be high, which will allow us to learn the contour features in a slice and use it to segment an adjacent slice. Fig. 5 shows the continuity of the classifiers between adjacent slices in a CT image sequence of femur.

We assume that features vary smoothly along the object surface (inter- and intra-slices) to improve the robustness of the learning process. Due to possible mislocation of snaxels at the beginning of the learning step, some spurious contour configurations could be learned. We introduce a regularizing process which increases the continuity of the classification vector, and makes the learning step more robust. In this process, the classifier of each contour patch p is regularized taking into account its direct neighbour classifiers:

$$\begin{aligned} \mathcal{V}'_{C_p} &= \zeta_p \mathcal{V}_{C_p} + \frac{1 - \zeta_p}{2} (\mathcal{V}_{C_{p-1}} + \mathcal{V}_{C_{p+1}}), \\ \mathcal{O}'_{C_p} &= \zeta_p \mathcal{O}_{C_p} + \frac{1 - \zeta_p}{2} (\mathcal{O}_{C_{p-1}} + \mathcal{O}_{C_{p+1}}), \end{aligned} \quad (3)$$

where the superscripts indicate new values, and $\zeta_p \in [0,1]$ is a regularizing parameter. From the regularized classifiers, we define a continuous classification function as follows:

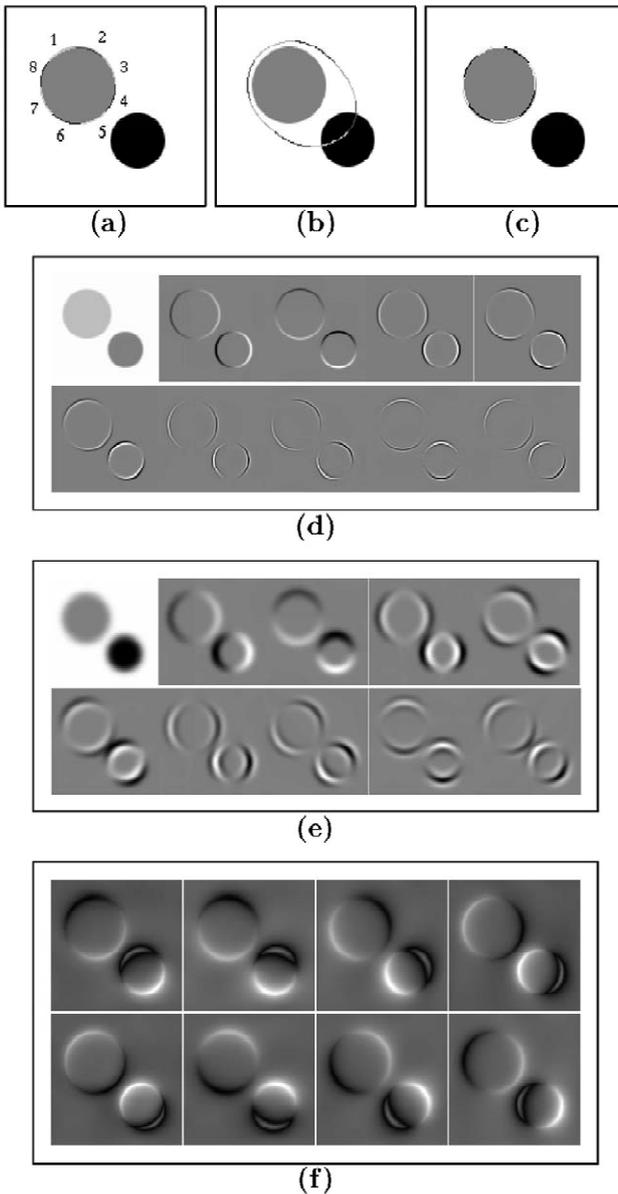


Fig. 2. Training contour and patch numbering (a); initial contour (b); final contour (c); filter responses for scales $\sigma = 1, 4$ (d,e); distance map to the different contour classes of the top left circle, top left map correspond to patch 1 and bottom right map correspond to patch 8 (f).

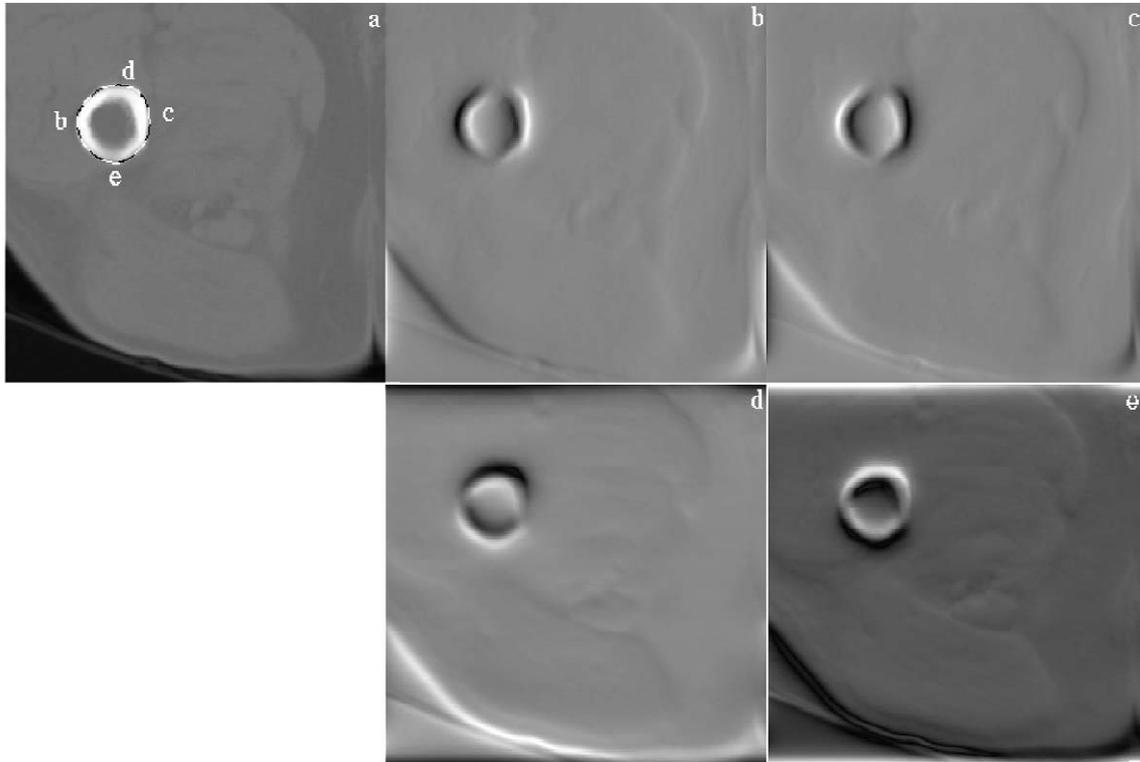


Fig. 3. Training contour and selected patches (b–e) (a); distance maps for contour patches on the west (b), east (c), north (d) and south (e) of the training contour (b)–(e).

$$\mathcal{V}(s) = \sum_{p=0}^{N-1} \mathcal{V}'_{C_p} \varphi_p(s),$$

where φ is considered in terms of finite-element nodal shape functions. For linear functions:

$$\varphi_p(s) = \begin{cases} |u(s) - u_p|, & \text{if } u \in [u_{p-1}, u_{p+1}], \\ 0, & \text{otherwise,} \end{cases}$$

where $u(s)$ represents the contour curve and u_p are the nodal points assuming uniform knots ($|u_j - u_{j+1}| = 1$) of the same curve.

To correct erroneous vectors without seriously affecting their neighbors, we compute parameter ζ_p as follows:

$$\zeta_p = \max \left\{ \mathcal{V}_{C_p}^T \mathcal{V}_{C_{p-1}}, \mathcal{V}_{C_p}^T \mathcal{V}_{C_{p+1}} \right\}. \tag{4}$$

Provided that classifiers are normalized, ζ_p is the cosine of the angle between both vectors:

$$\begin{aligned} \mathcal{V}_{C_p}^T \cdot \mathcal{V}_{C_{p\pm 1}} &= \|\mathcal{V}_{C_p}^T\| \cdot \|\mathcal{V}_{C_{p\pm 1}}^T\| \cos \angle_{\mathcal{V}_{C_p}, \mathcal{V}_{C_{p\pm 1}}}, \\ \|\mathcal{V}_{C_p}^T\| = \|\mathcal{V}_{C_{p\pm 1}}^T\| = 1 &\Rightarrow \mathcal{V}_{C_p}^T \cdot \mathcal{V}_{C_{p\pm 1}} = \cos \angle_{\mathcal{V}_{C_p}, \mathcal{V}_{C_{p\pm 1}}}, \end{aligned}$$

where $\angle_{\cdot, \cdot}$ denotes the angle between vectors. If ζ_p takes values close to 1 it means that classifier \mathcal{V}_{C_p} has at least one similar neighbour classifier, so it is not greatly

smoothed. This prevents that contours with real sharp changes smooth severely, and only erroneous vectors are significantly corrected.

In Fig. 6 one can see the 0, 1st and 3rd derivatives with respect to the coordinate x in the classifiers of snake patches ($p = 0, \dots, 45$). Note that different weights form a continuous function of the internal parameter of the snake curve. In patches 35 and 37 smoothing of the weights is observed after applying Eqs. (3) and (4).

3. Snake optimization

Let $C(s) = \sum_i V_i B_i(s)$ be a B-spline representation of the curve, where V_i are the control points and s is the curve parameter. The total energy can be written (Menet et al., 1990):

$$\begin{aligned} E_{\text{total}}(C) &= \int_0^1 (E_{\text{int}}(C(s)) + E_{\text{ext}}(C(s))) ds \\ &= \int_0^1 \left(\alpha \left| \frac{\partial C(s)}{\partial s} \right|^2 + \beta \left| \frac{\partial^2 C(s)}{\partial s^2} \right|^2 + E_{\text{ext}}(C(s)) \right) ds \\ &= \sum_j \left[\alpha \left(\sum_i V_i \frac{\partial B_i(s_j)}{\partial s} \right)^2 + \beta \left(\sum_i V_i \frac{\partial^2 B_i(s_j)}{\partial s^2} \right)^2 + E_{\text{ext}}(C(s_j)) \right]. \end{aligned}$$

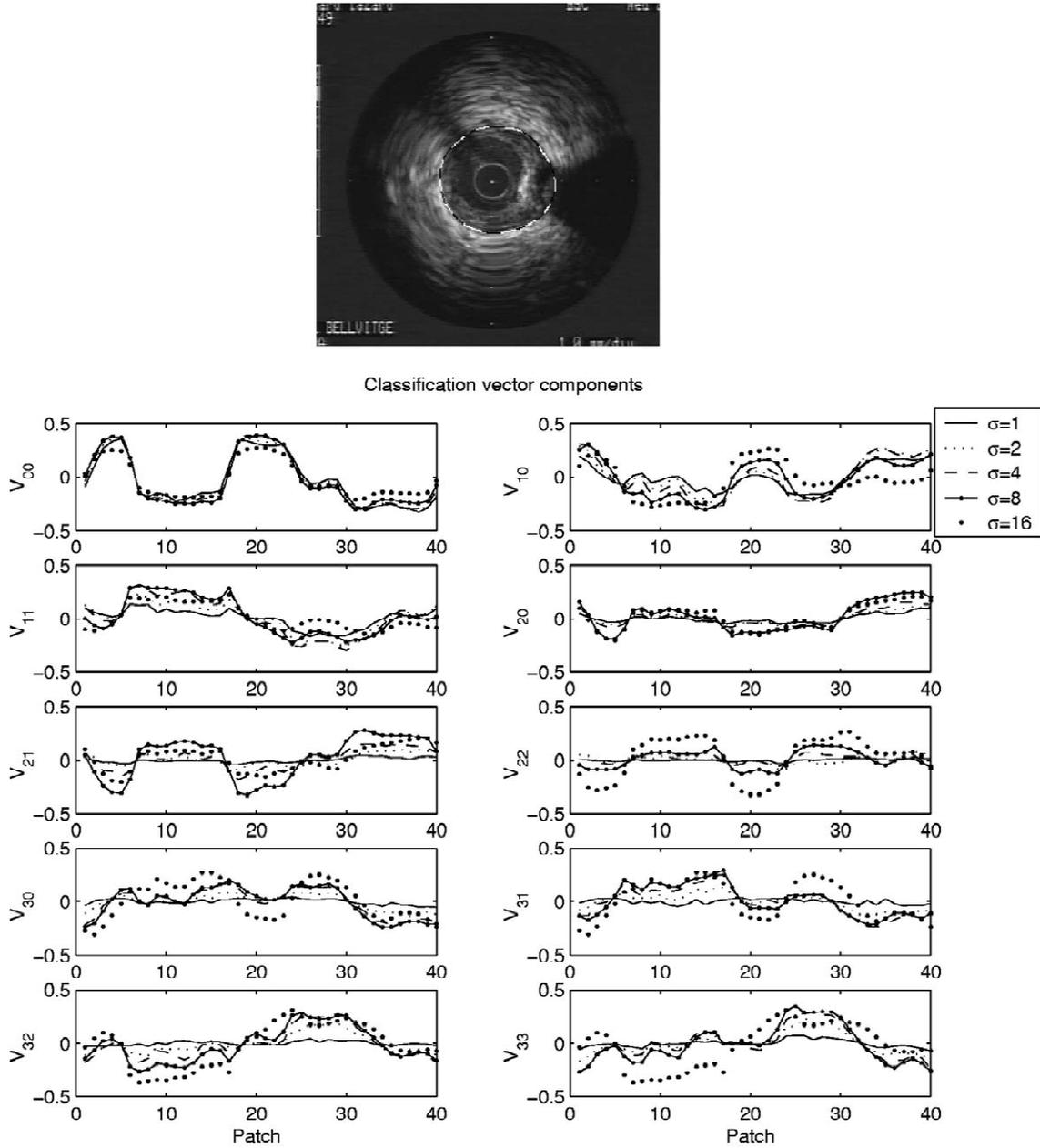


Fig. 4. Components of the classifier along the patches of the learning contour. In V_{ij} , $i \in [0, 1, 2, 3]$ indexes derivative degree and j indexes orientation ($\theta_i \in \{j\pi/(i+1) | j = 0, 1, \dots, i\}$), and each line corresponds to a different scale.

We define the external energy as a function of image features and the training contour classes (Section 2.2). For a point in the patch k , the external energy is

$$\begin{aligned} E_{\text{ext}}(C(s_j)) &= (\mathcal{V}_{C_k}^T \mathcal{P}_j(C(s_j)) - \mathcal{O}_{C_k})^2 \\ &= (\mathcal{V}_{C_k}^T (\mathcal{G}_{D\Sigma}^* \mathcal{I}(C(s_j))) - \mathcal{O}_{C_k})^2. \end{aligned} \quad (5)$$

We are looking for control points V_i , $i = 0, \dots, N$ that minimize the total energy, i.e., that satisfy

$$\frac{\partial E_{\text{total}}}{\partial V_i} = 0, \quad \forall i \in \{0, \dots, N\}.$$

These equations can be written in a matrix form similar to the classical snake:

$$A_b V + G = 0, \quad (6)$$

where the stiffness matrix A_b for B-spline snakes is still a banded matrix and G plays a role equivalent to forces in the classical snake. The i th element of G is

$$\begin{aligned} G_i &= \sum_j B_i(s_j) \nabla E_{\text{ext}}(C(s_j)) \\ &= -2 \sum_j B_i(s_j) (\mathcal{V}_{C_k}^T \cdot (\mathcal{G}_{D\Sigma}^* \mathcal{I}(C(s_j))) \\ &\quad - \mathcal{O}_{C_k}) (\mathcal{V}_{C_k}^T (\nabla \mathcal{G}_{D\Sigma}^* \mathcal{I}(C(s_j)))). \end{aligned} \quad (7)$$

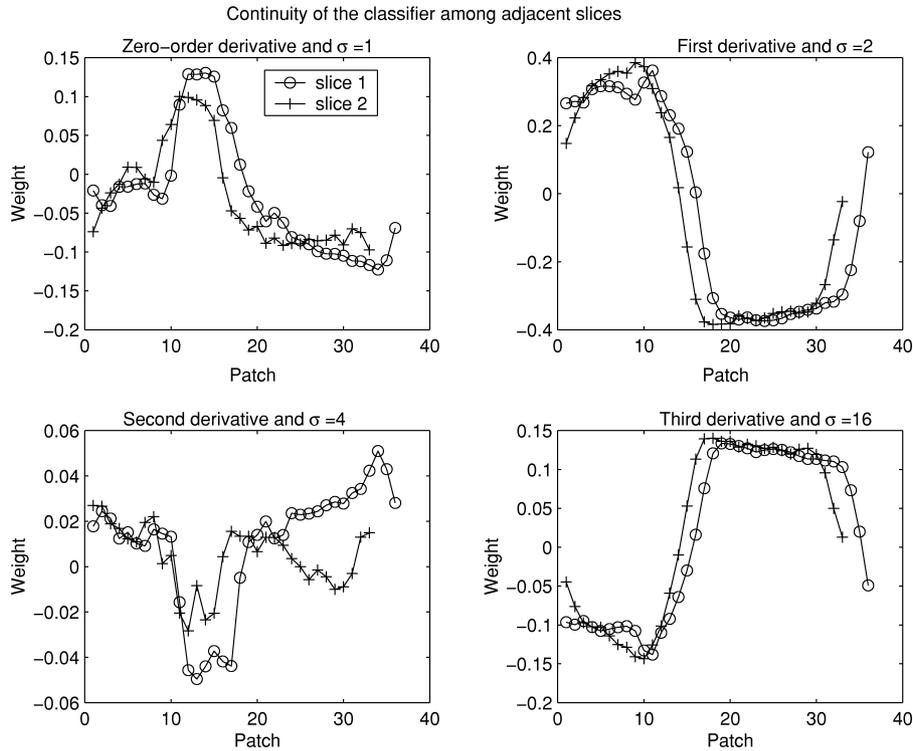


Fig. 5. Continuity of the classifiers among adjacent slices for different derivatives and scales.

The solution of the segmentation problem is found in an iterative and incremental way:

$$V_{i+1} = (A_b + \gamma I)^{-1} (\gamma V_i + G(C_i)),$$

where γ is the damping parameter that determines the convergence rate and is equal to 1 in our implementation.

The method is incremental because it begins with the highest scale components of the classification vector, incorporates a new scale after convergence, and finishes with all the scales. The weights corresponding to the scales that are not being considered are set to zero. In each incremental step the algorithm iterates until all the control points stop evolution or oscillate.

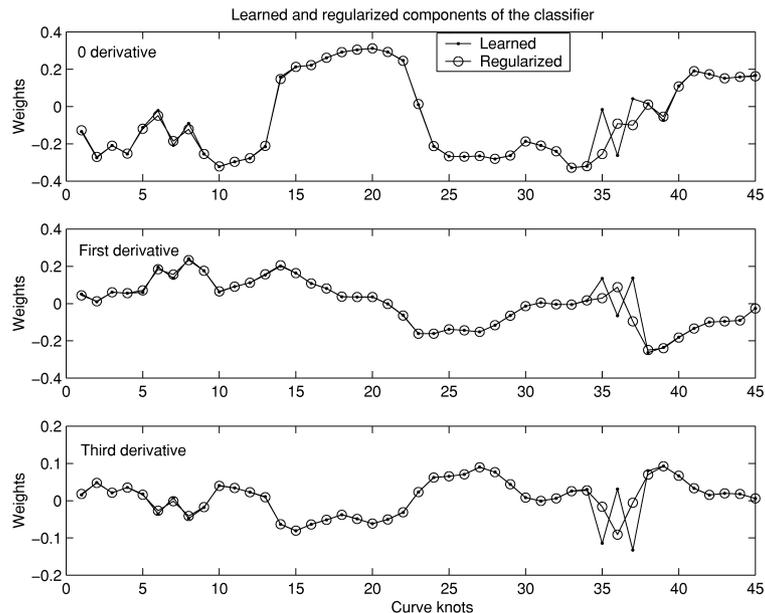


Fig. 6. Learned and regularized components of the parametric classifier.

As the number of control points increases, the delineations of convoluted shapes improve. However, when the number of control points is increased, the size of the patches decreases and discrimination power decreases. We found that a distance of about 5 pixels between adjacent control points worked well in all tested examples. To maintain these distances we augmented the algorithm with a reparameterization capability during the deformation process.

4. Results

In order to validate the method we have tested STD-snake on organs from different imaging modalities. First, we are going to show some properties of our approach with respect to classical implementations, and afterwards we will show a comparison to expert tracings.

4.1. Comparison to classical snakes

We are going to compare the method with classical

snakes to both prove the advantages of our method and to justify the choices made. We will begin by showing the advantages of using multiple scales and a continuous classifier, and will finish by demonstrating the ability of our method to describe and learning different image features.

4.1.1. Integration of multiple scales

The use of multiple scales has two objectives: first, the greater the scale, the wider the spatial effect of the potential. Hence high scales allow large displacement of the snake contour to reach the desired object. Second, multiple scales greatly reduce the number of local minima.

Fig. 7 shows an example that illustrates this property of the multiscale vector. We have computed the distance map generated by the classification vectors over all the image. Fig. 7(b) shows the computed distance map using only the lowest scale in the classification vector, and Fig. 7(c) contains the distance map computed by the multiscale classification vector. As can be seen in Fig. 7(d,e) the approximate Gaussian shape of the histogram has lesser variance and higher mean distance in Fig. 7(e) than in Fig.

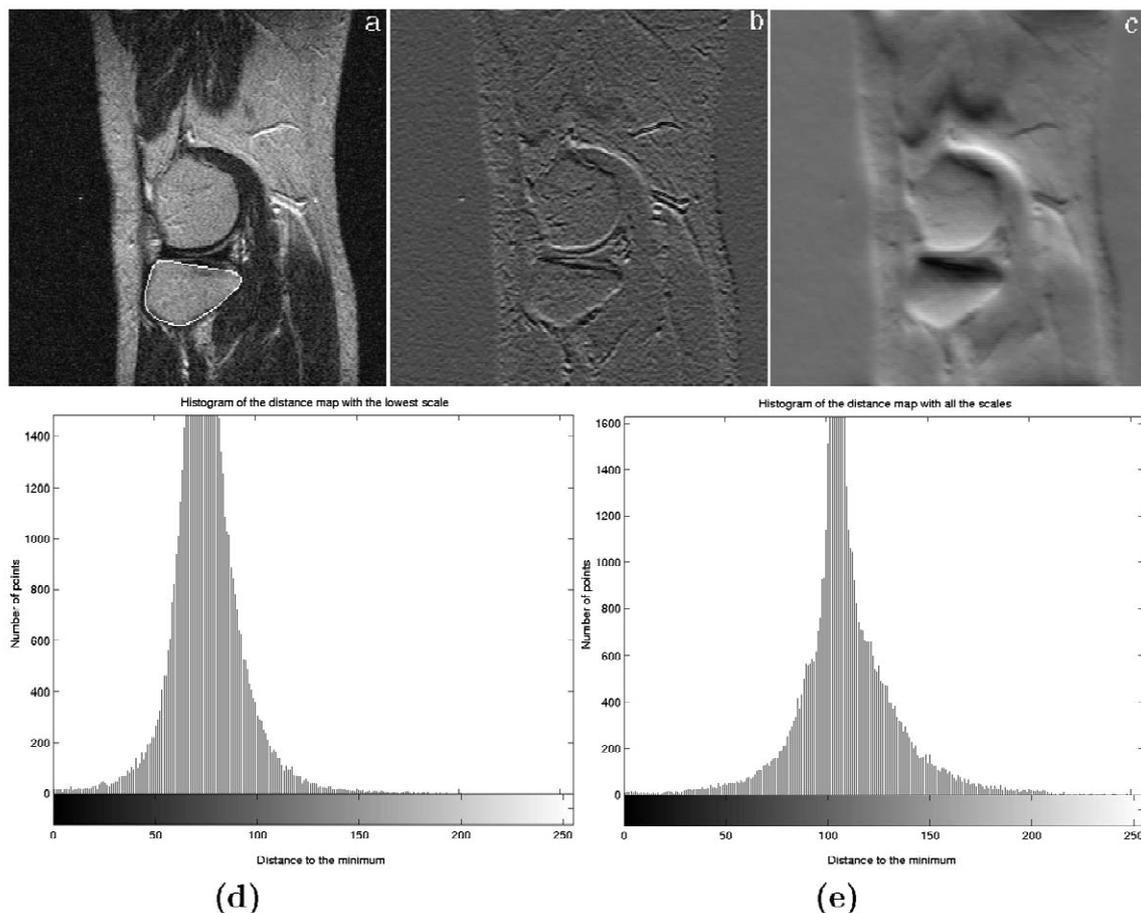


Fig. 7. Training contour: MR Image of the knee joint (a); distance map images generated by the classifier of the contour patch placed in the middle top of the bone contour in (a) considering: only the smallest scale (b), all the scales used in learning phase (c); histograms for the distance map: for single scale (d), and all scales (e).

7(d). This means that the number of convex components in the distance map diminishes and there are less potential minima in Fig. 7(c) (multiscale) than in Fig. 7(b) (single scale). In all the examples shown in this work, we used scale parameters with values $\sigma = 1, 2, 4, 8$.

4.1.2. Continuous classifier

The definition of a classifier that continuously varies along the contour has several advantages. One of them is the capability of filtering noise in the learned classifiers without excessive smoothing.

In Fig. 8 the regularized classifier is compared to simple average. The contour of a rectangular object against a background is learned in a synthetic binary image. The contour presents sharp changes in features (the filter weights change significantly in corners) which is in harmony with a square wave shape of the first and third derivatives. Simple constant average ($\zeta_p = 0.5, \forall p$) blurs the sharp changes, while the new regularizing process preserves them (the learned and regularized values agree). Fig. 8(c,d) shows the final fitting for both cases and Fig.

8(e) shows two samples of the smoothed and the regularized components of the classifier along the contour.

The continuous classifier also allows us to add new control points to fit to concave shapes. After convergence, the algorithm tests the distance between adjacent control points; when this distance is greater than two times the initial distance a new control point is introduced. If new control points were added the optimization would be done again. The smoothed continuous classifier (Eq. 4) allows the snake to track changes in shape and texture features along the image sequence.

Fig. 9 illustrates this ability: Fig. 9(a) contains a binary synthetic image where the object contour is learned, Fig. 9(b) shows the initial curve and Fig. 9(c) contains the final curve. If few contour patches are considered, erroneous fitting is achieved (Fig. 9(c)), but relying on the continuity of the classification vector, more control points can be dynamically introduced to allow a correct fitting (Fig. 9(d)). Fig. 9(e–g) shows the inverse case (deformation in the opposite direction).

The classical snake has the well known difficulty in

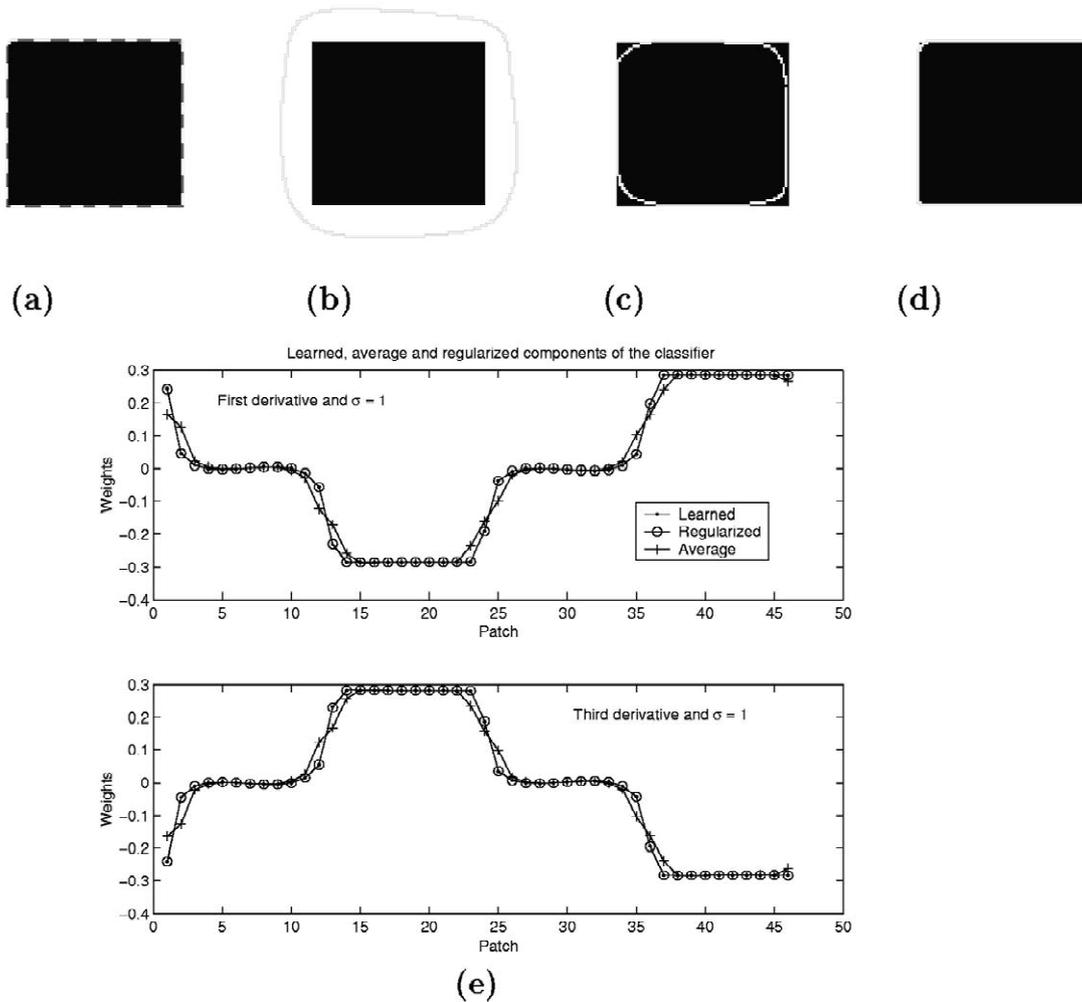


Fig. 8. Learned contour (a); initial contour (b); final contour using smoothed (average) classifiers (c); final contour using regularized classifiers (d); average ($\zeta_p = 0.5$) and regularized (preserving real sharp changes) components of the continuous classifier for the rectangle contour (e).

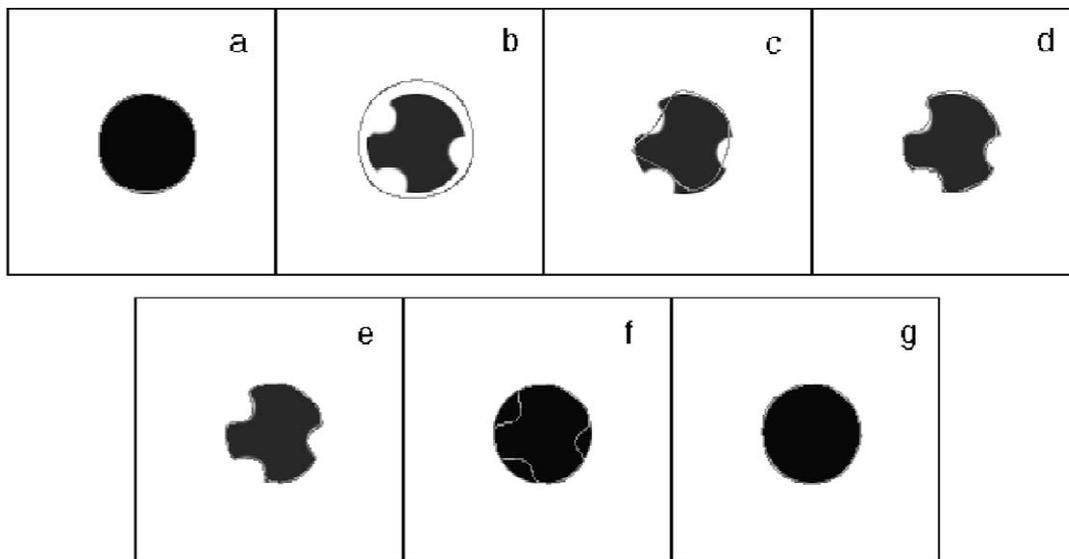


Fig. 9. Training contour (a); initial contour (b); intermediate contour (c); final contour after interpolation (d); training contour (e); initial contour (f); final contour (g).

segmenting concave objects due to the curve is equally attracted by different contour configurations (Fig. 10). As a consequence, the curve falls in a zone of equal distance to different contour configurations and stabilizes far from the desired contour. Fig. 10(d) shows the final fitting reached with classical snake where alternating colours of contour patches allow us to see that contour patches near to concavities increase in size because they are attracted by the ends of concavities.

Several methods have been proposed in the literature to cope with this problem. A well illustrated example about the behaviour of traditional snakes in the presence of concavities and revision of previous proposed solutions can be found in (Neuenschwander et al., 1997) and (Xu and Prince, 1998). We think that our new approach is more general in the sense that it handles non-convex shapes in a natural way, and has the ability to adapt to concave contours due to the selectivity of external energy and the smoothness of the discriminant function (classifier) along the contour.

The discriminant snake allows the tracking of contours with some abrupt changes in shape or texture through slices. Fig. 11 illustrates this property in a CT femur slice.

Fig. 11(a) shows the learning contour where alternating colours indicate different contour patches. Between this slice and the next one, a sharp change exists around the pointed patch; Fig. 11(e) shows the profiles normal to the pointed patch in the two adjacent images. However, the potential minimum for the corresponding patch in the next slice is in the correct place as Fig. 11(b) shows. Fig. 11(c) contains the translation of learning contour in the previous slice to the next slice, and Fig. 11(d) shows the correct final delineation.

4.1.3. Generalized description of features

Our snake does not have a predefined goal, but it learns the desired contour features for each specific segmentation problem. Moreover, each snake patch searches for its specific contour class, which avoids ambiguities in the event of there existing several neighbouring objects.

Fig. 12(a–d) illustrates the second capability. Fig. 12(a) contains the original image and Fig. 12(b) shows the contour used for a supervised learning of different parts of the external cortical bone of femur in an CT image. To show the robustness of our proposal we consider another slice (non-adjacent to that) which has different contour

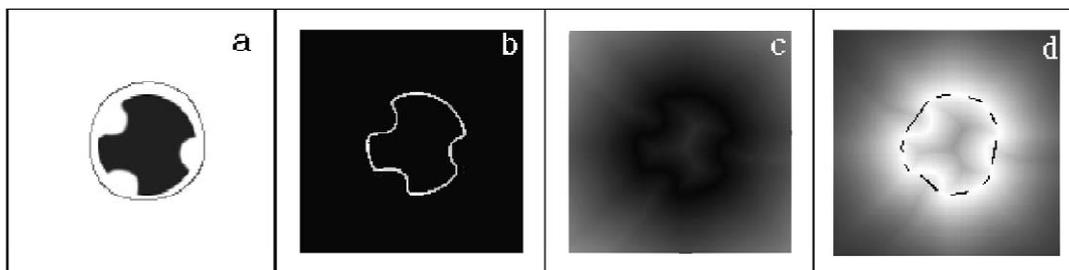


Fig. 10. Initial contour (a); edges (b); classical distance map (c); fitting of the classical snake in black over the inverted potential (d).

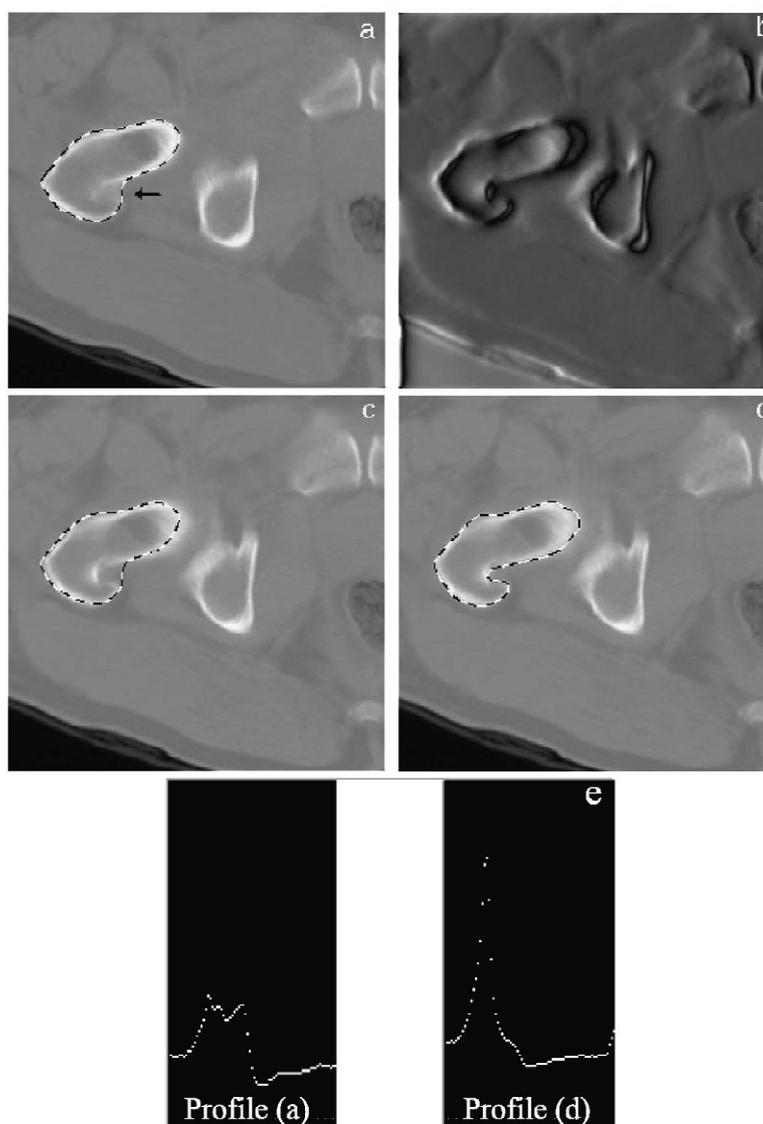


Fig. 11. Training contour (a); energy image generated by the classifier of the pointed patch in the following slice (b); initial contour (c) and final fitting (d). Corresponding profiles normal to the pointed patch (e).

configuration (texture and contrast). We use the classification vector in the previous slice, but with the initial contour deformed, and placed between two bones, see Fig. 12(c). Fig. 12(d) represents the final contour fitted to the bone of interest by our method. Although the initial contour is not close to the final contour and the characteristics of the contour are different, the snake found the contour more similar to the learned one, and this provided a good result. Different parts of the cortical bone have different filter responses, hence they belong to different classes and the snake moves towards the desired contour even in the case of nearby objects.

If the texture features and the geometry changes dramatically then the segmentation fails. Fig. 12(e) shows the learned contour in Fig. 12(b) over the slice to be segmented. The final result shown in Figs. 12(f) is incorrect because the learned contour and the actual

contour of interest are very different. In a normal medical image sequence the distances between adjacent slices are short so we can hope not to find such strong changes.

Now, we are going to show the ability to learn different features by segmenting medical images (from several modalities) where the desired object contours have different characteristics.

Fig. 13 shows the segmentations of a set of slices where a considerable change in contour shape takes place. The full sequence is made up of 60 slices with a distance between slices of 5 mm. This is a typical case where the 3D image has not the same resolution in all the spatial dimensions, and, therefore, a multislice segmentation approach is the best option. The supervised learning takes place in the first slice and converged active contours and updated classifiers are propagated from slice to slice. The snake is capable of following texture and shape changes.

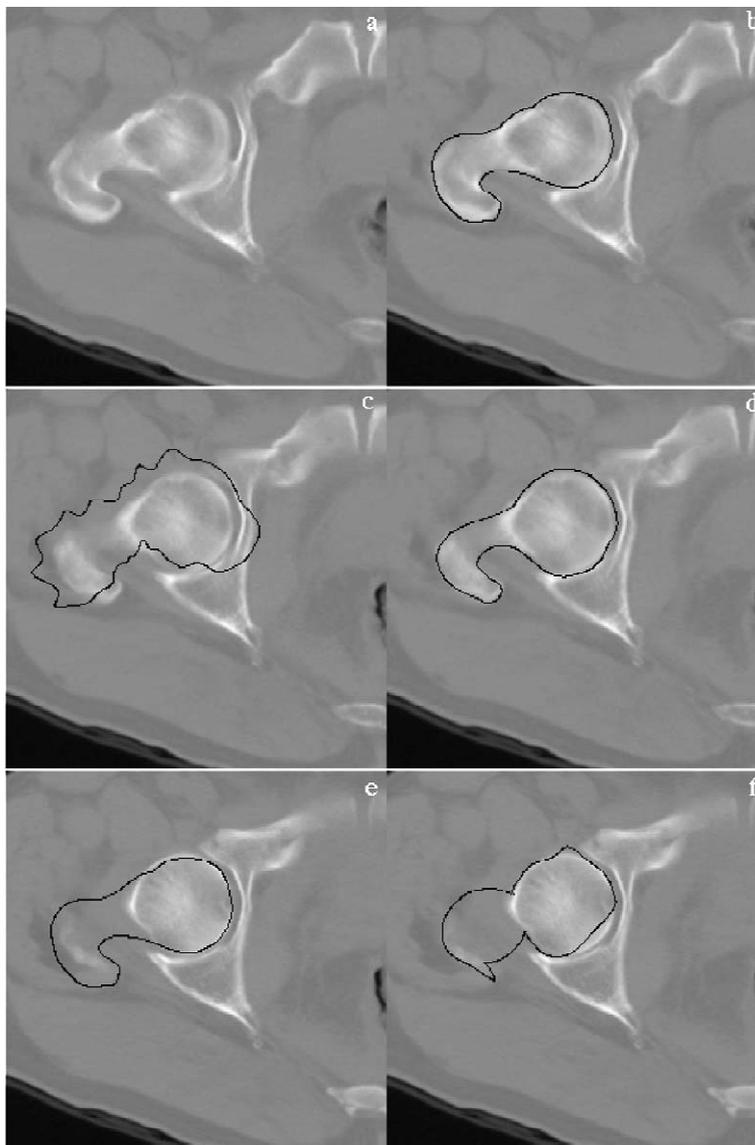


Fig. 12. Original image (a); learned contour (b); initial contour and final contour for a moderately distant slice (c,d); initial contour and final contour for a very distant slice (e,f).

Note that the new approach naturally fits convoluted shapes. Moreover, the new external energy eliminates ambiguities between internal–external contours and contours of nearby structures. Fig. 14 shows the 3D reconstruction of the femur.

To give an idea about the temporal cost of our algorithm, we measured the processing times for the complete sequence in a Pentium III at 400 Mhz. The algorithm took, on average, 25 s per scale to compute the multivalued features for the full 256×256 slice images, but this figure decreased (3 s/scale) when only a bounding box centered in the learning contour was considered. The bounding box was defined as a function of the maximum and the minimum coordinates of the control points, and of the maximum scale (+24 pixels, for maximum sigma equals to 8). The learning step took 8 s per slice for the four

scales, and the fitting process took 3 s per scale and slice. All these figures give a total of 36 s per slice and 36 min for the 3D femur sequence.

We also apply our method to MR images. Fig. 15 shows three samples of the segmentations of a sequence of 15 MR images of the distal femur. In these images contour has mostly the appearance of an intensity valley, unlike CT images where the contour features are mainly edges. Fig. 16 shows the 3D reconstruction obtained from the segmentation.

The next imaging modality that we consider is the IVUS in two different cases: abdominal aorta (AA) and coronary vessel. In these images the contours have less contrast and the signal to noise ratio is higher. Fig. 17 illustrates the segmentation of a sequence of IVUS images of the abdominal aorta of a pig. This figure contains the first

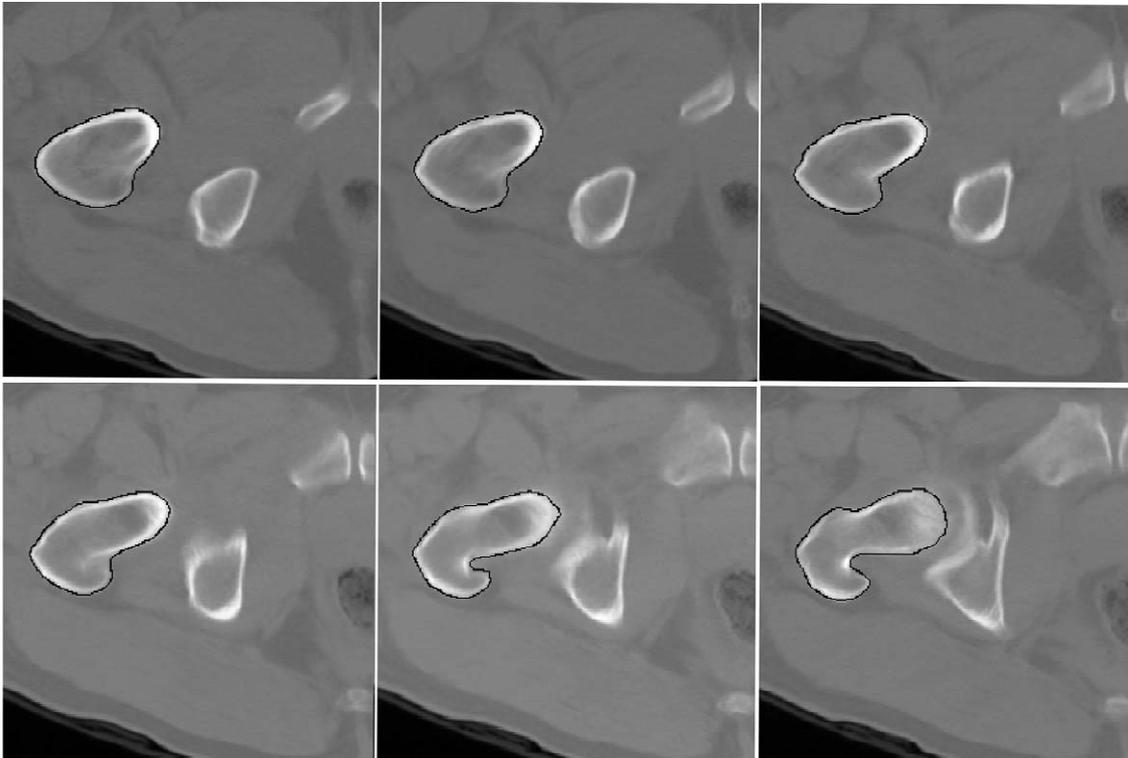


Fig. 13. CT images of slices 35–40.

image with the initial training contour, and other intermediate ones of a sequence made up of 100 slices. Classifiers are learned in the first slice and the deformed contour is carried to the next slice. Subsequently, the classifiers are again updated and transferred together with the final contour to the next slice. Although there are several different contours nearby, the snake fits the learned one. Fig. 18 shows the reconstruction of the 3D surface derived from the segmentation.

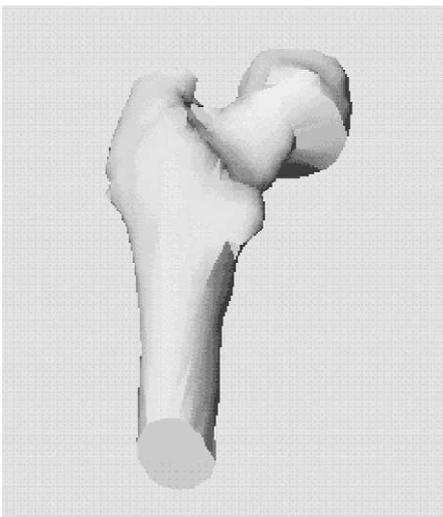


Fig. 14. 3D reconstruction of the proximal femur.

Finally, Fig. 19 shows the segmentation in intermediate slices of a sequence of 400 IVUS images of a coronary vessel. These images are very difficult to segment because some contour parts of the blood vessel are very difficult, if not impossible, to delineate even for a human operator. In that case we used a shape model that avoided severe changes between contours of adjacent slices where there existed no cues about the location of some contour parts. We have measured the ratio between horizontal and vertical diameters of the vessel in a set of sample images, and then computed their mean and variance values. A new internal energy term penalized the distance between the ratios of the contour and the model. The 3D reconstruction of the coronary arteries can be seen in Fig. 20.

4.2. Comparison to expert tracings

To quantitatively evaluate the new approach, five experts (e_1, e_2, e_3, e_4, e_5) manually segmented images from different sequences and imaging modalities (IVUS, CT, MRI). Model contours (M) were computed as the average operator contours and used as the ground truth segmentation contours. For each test image (j), we computed the distance between each pair of contours (e_i^j, M^j) and (STD^j, M^j), where STD^j represents the contour fitted by the STD-snake in image j .

We developed an application to allow an interactive manual segmentation: the expert selects a set of control points, then the application adds several control points so

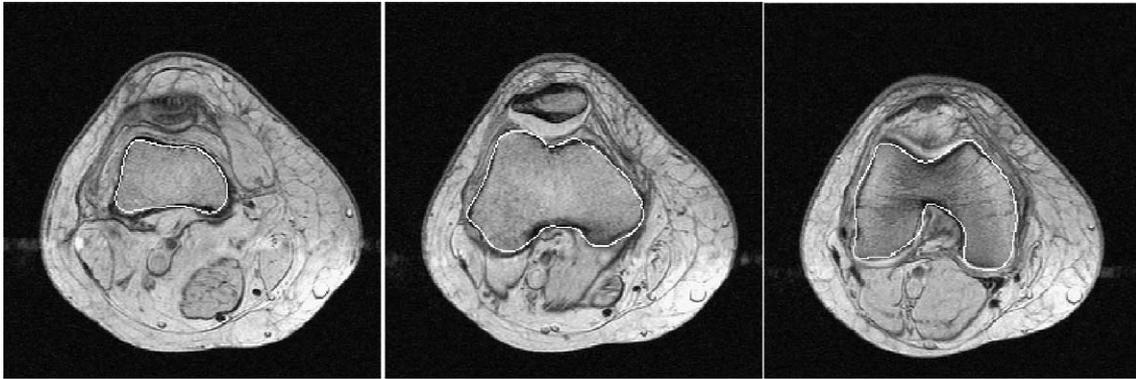


Fig. 15. MR images of the distal femur.

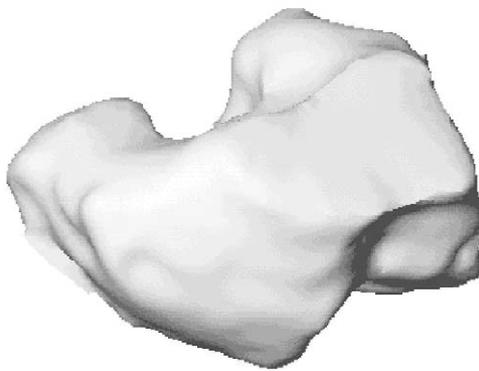


Fig. 16. 3D reconstruction of the distal femur.

that the distance between adjacent control points is less than 5 pixels. After that the application paints the B-spline and allows the expert to change the location of all the control points to get the best delineation of the object of interest. After each change the application repaints the curve.

We measure the distance between two contours in the following way: (1) we obtain an image where each pixel value is the distance to the model contour, the distance from the model contour to another contour is the sum of the values of distance in the locations of the contour; (2) then, the distance from each contour to the model is also computed; (3) finally, we compute the average of these

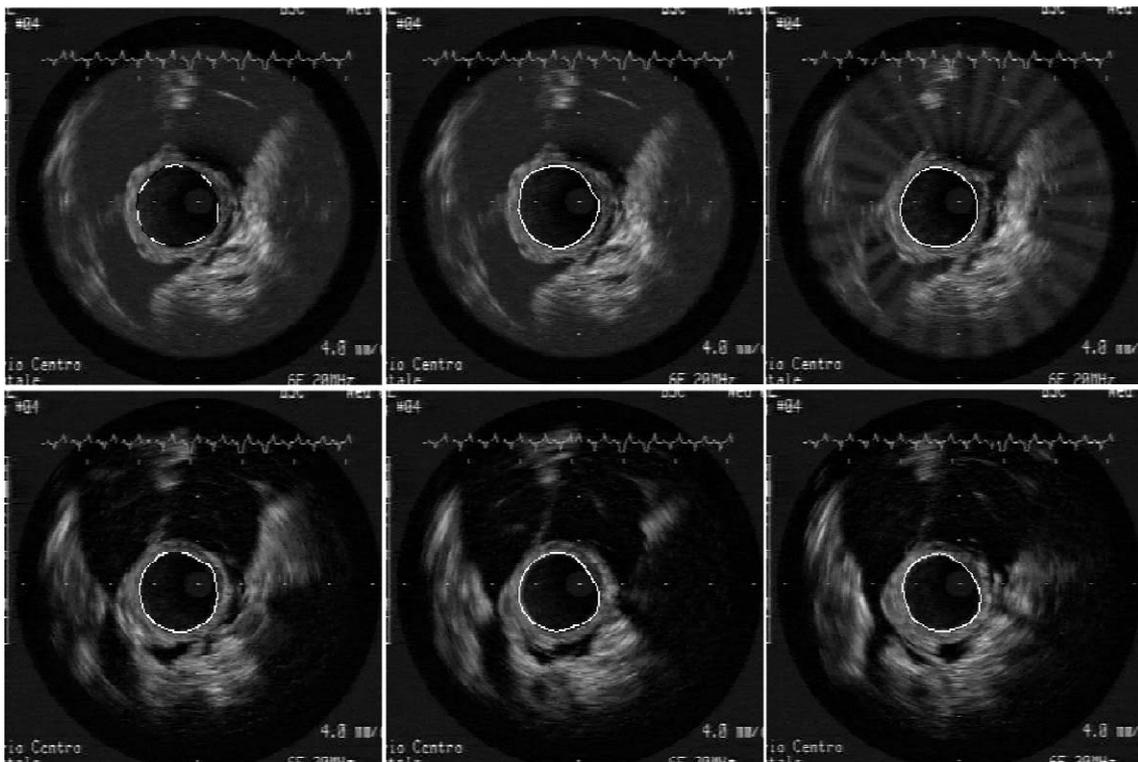


Fig. 17. Training contour and several segmentation examples in a sequence of 100 IVUS images of the abdominal aorta of a pig.

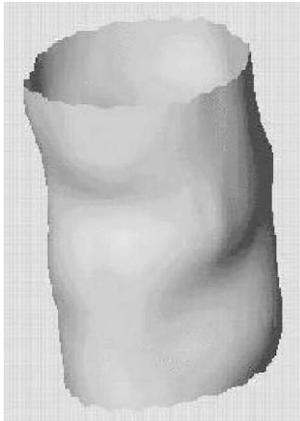


Fig. 18. Reconstructed internal surface of the abdominal aorta.

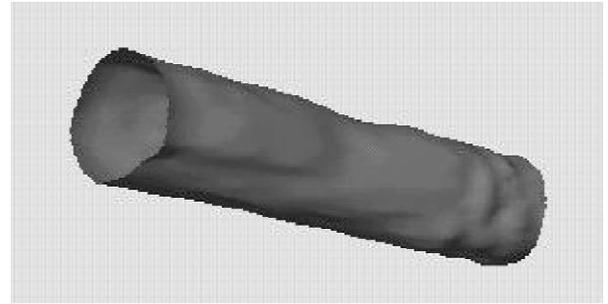


Fig. 20. Surface of the coronary artery.

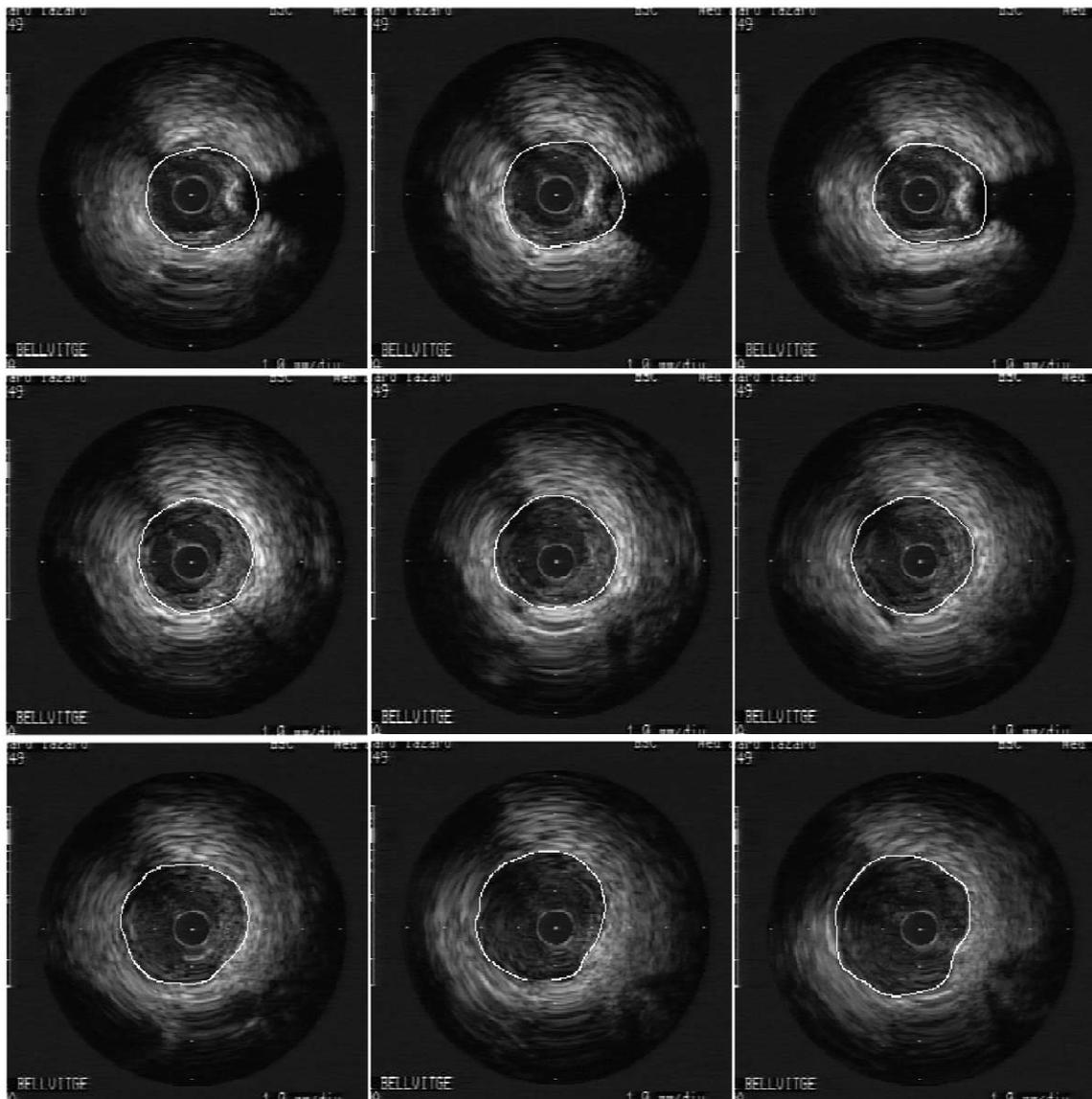


Fig. 19. Samples of the segmentation of a coronary vessel in an IVUS image sequence.

Table 1

Distance between the model segmentation contour and the contours provided by the different experts (inter-expert variability)

Sequence	e_1	e_2	e_3	e_4	e_5
	\bar{d}, σ^2				
CT femur	1.22, 0.86	1.02, 0.76	1.03, 0.74	0.99, 0.80	0.99, 0.67
IVUS vessel	1.77, 2.63	1.50, 1.68	1.51, 2.05	1.73, 2.06	1.43, 1.72
IVUS AA	1.28, 0.97	0.98, 0.62	1.17, 0.70	1.26, 0.80	1.15, 0.77

Table 2

Distance between the model segmentation contour and the contours provided by the human experts and the snake

Sequence	e	STD
	\bar{d}, σ^2	\bar{d}, σ^2
CT femur	1.04, 0.75	1.36, 0.94
IVUS vessel	1.59, 2.05	2.30, 3.46
IVUS AA	1.18, 0.78	1.62, 1.25

two distances. From this value and the contour size, we obtain the mean distance and the variance. The values of mean distance and variance for full sequences are obtained taking into account the size, the mean distance and the variance computed for each contour of the sequence.

The distance is characterized by its mean and variance (\bar{d}^j, σ^j) in pixels. Table 1 contains the measured values for each expert in three image sequences, and Table 2 shows the mean error of the method versus the mean error of the human segmentations. As can be seen, the average distance given by the snake is similar to the distances given by manual segmentations. It is important to remark that the ground truth segmentation was built from all manual segmentations, therefore the ground truth segmentation depends on human delineations but it does not depend on snake segmentations. We consider that this is the reason for slightly higher values on snake evaluation parameters, since there is no relevant difference in the visual inspection of results. Note that in the case of IVUS images the resolution is 1 mm for 45 pixels, hence 2 pixels of difference represents 0.044 mm, which is considered as insignificant by the medical experts.

5. Conclusions

The goal of this work is the generalization of the classical feature-based snake technique in the field of segmentation of spatial or temporal image sequences. We have designed a parametric expression for external energy so that different patches of the snake look for different parts of the contour. Thus, the STD-snake is more selective and robust than the classical snake.

We have applied the method to the reconstruction of anatomical organs in images from different modalities. Frequently, 3D image data sets have no isotropic spatial resolution; voxel size is bigger along the longitudinal axis in order to maintain as low a radiation dose as possible, to

increase the ratio number-of-scanned-patients/time, etc. When this happens, features cannot be extracted in 3D, and therefore a slice-to-slice framework is more adequate.

In most general case, organs are complex structures and, as a consequence, their boundaries are rarely well defined by single features. Because of this, we consider a multi-valued feature capable of representing any contour configuration. From it, our method finds the most relevant features for each case. As a result, our model is a natural generalization of contour-, valley- and crest-based snakes.

The extension of the method to deal with 3D operators and learning surfaces (instead of learning contours) will imply learning the features from 3D volume data samples of the same organ and imaging modality. Moreover, a new problem arises with the location of the initial surface, because it is difficult to presuppose a high correlation in location between different image sequences. This is one of the main issues of our current research.

Although the approach is very promising, it is not free of limitations. One of the most important challenges in our immediate research is to cope with the problem shown in Fig. 2(f). A comparison between the large and small circle shows that contour parts with similar orientation on the small circle get a white center and two dark sidelobes. This effect is generated by the difference in contrast. Slopes have the same sign in the border of both circles but different magnitudes, so circle features are different in the positions of the maximum slopes. However, features near the black circle are close to features at the stronger transition in the white circle. Nevertheless, this behaviour is less important than it seems at first glance because the undesired behaviour takes place far from the desired contour part and our method is aimed at locating, in the next slice, the contour part most similar to a contour part in the current slice, and we assume that they are close in location and features.

The current method only considers the features along the object contour (one dimension). We think that a method which considers the features in a certain region (two dimensions) would be more reliable. We are exploring one possible solution: to deal with features in the frequency transformed domain (inherently 2D).

User interaction is still required when shape changes considerably between slices. Fig. 12 illustrates this problem. To completely automate the segmentation process a high degree of specificity is required, which is usually accomplished by means of task oriented procedures. The

robustness of the method could be improved with the incorporation of shape knowledge, mainly in slices where strong changes in shape occur or there are not image clues about boundary location. The combined learning and use of statistical shape and image features seem very promising. In fact, we have used shape constraints to segment the example in Fig. 19, as aforementioned.

The application of all the filters has large requirements of storage and time. To reduce them, given the iterative scene of snake deformation, we perform these computations in the neighborhood of the moving contour, as mentioned in the previous section. Further savings could be achieved by implementing ideas from the front propagation problem to define tight areas of interest.

Acknowledgements

This work was supported by CICYT TIC2000-0399-C02-02 and Xunta de Galicia grant PGIDT99PXI20606B.

Appendix A. Fisher linear discriminant analysis

The Fisher linear discriminant analysis is the optimal solution to the problem of assigning patterns to two classes $\{C_k, \bar{C}_k\}$, if the underlying distributions of the observations are multivariate normal (Duda and Hurt, 1973). The Fisher discriminant function is a scalar function:

$$\mathcal{F}_k(\mathcal{P}_f) = V_k^T \cdot \mathcal{P}_f,$$

where $\mathcal{P}_f \in \mathbf{R}^{d_g}$ is the multivalued feature, and $V_k \in \mathbf{R}^{d_g}$ is a vector that projects the features of pixel j into the reduced feature space \mathbf{R} , to determine the similarity to the class C_k .

The optimal projection V_{k_opt} is defined as the one which maximizes the ratio of the between-class to the within-class scatter of the projected samples:

$$V_{k_opt} = \operatorname{argmax}_{V_k} \frac{|V_k^T S_{bc} V_k|}{|V_k^T S_{wc} V_k|},$$

where

$$S_{bc} = N_{C_k}(\mu_{C_k} - \mu)(\mu_{C_k} - \mu)^T + N_{\bar{C}_k}(\mu_{\bar{C}_k} - \mu)(\mu_{\bar{C}_k} - \mu)^T,$$

$$S_{wc} = \sum_{s_j \in C_k} (\mu_{C_k} - s_j)(\mu_{C_k} - s_j)^T + \sum_{s_j \in \bar{C}_k} (\mu_{\bar{C}_k} - s_j)(\mu_{\bar{C}_k} - s_j)^T$$

are the between-class S_{bc} and within-class S_{wc} scatter matrices; being N_{C_k} and $N_{\bar{C}_k}$ the number of samples in class C_k and \bar{C}_k , respectively; μ_{C_k} and $\mu_{\bar{C}_k}$ are their mean feature

vectors, and μ is the mean vector over all samples. V_{k_opt} is the S_{wc} -generalized eigenvector of S_{bc} with the largest corresponding eigenvalue (Belhumeur et al., 1997).

If S_{wc} is nonsingular we can obtain a conventional eigenvalue problem by writing

$$V_{k_opt} = S_{wc}^{-1}(\mu_{C_k} - \mu_{\bar{C}_k}).$$

Before computing V_{k_opt} , it is necessary to prevent the existence of a singular S_{wc} . To this purpose, principal component analysis (PCA) is firstly performed on all training samples. As a result, we obtain V_{pca} as a matrix of dimension $d_g \times m$ formed by the m most important eigenvectors of the scatter matrix S_T of the full set of samples. PCA is then followed by FLDA, to obtain the best discriminant, V_{k_opt} , for each contour part C_k :

$$V_{k_opt}^T = V_{k_flda}^T V_{pca}^T,$$

$$V_{k_flda} = \operatorname{argmax}_{V_k} \frac{|V_k^T V_{pca}^T S_{bc} V_{pca} V_k|}{|V_k^T V_{pca}^T S_{wc} V_{pca} V_k|}.$$

Finally, we can write the optimal projection as

$$V_{k_opt} = V_{pca}(V_{pca}^T S_{wc} V_{pca})^{-1} V_{pca}^T (\mu_{C_k} - \mu_{\bar{C}_k}). \quad (A.1)$$

References

- Amini, A.A., Weymouth, T.E., Jain, R.C., 1990. Using dynamic programming for solving variational problems in vision. *IEEE Trans. Patt. Anal. Machine. Intell.* 12 (9), 855–867.
- Blake, A., Isard, M., 1998. In: *Active Contours*. Springer, Berlin.
- Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J., 1997. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. Patt. Anal. Machine. Intell.* 7, 711–720.
- Chakraborty, A., Staib, L.H., Duncan, J.S., 1996. Deformable boundary finding in medical images by integrating gradient and region information. *IEEE Trans. Med. Imag.* 15 (6), 859–870.
- Chen, Ch.-W., Huang, Ch.-L., 1992. Human face recognition from a single front view. *Int. J. Patt. Recognition Artif. Intell.* 6 (4), 571–594.
- Cohen, L.D., Cohen, I., 1993. Finite-element methods for active contour models and balloons for 2-D and 3-D images. *IEEE Trans. Patt. Anal. Machine Intell.* 15 (11), 1131–1147.
- Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active shape models-their training and application. *Comput. Vis. Image Underst.* 61 (1), 38–59.
- Duda, R.O., Hurt, P.E., 1973. In: *Pattern Classification and Scene Analysis*. Wiley-Interscience, New York.
- Elder, J.H., Zucker, S.W., 1998. Local scale control for edge detection and blur estimation. *IEEE Trans. Patt. Anal. Machine Intell.* 20 (7), 699–716.
- Florack, L.M.J., ter Haar Romeny, B.M., Koenderink, J.J., Viergever, M.A., 1992. Scale and the differential structure of images. *Image Vision Comput.* 10 (6), 376–388.
- Florack, L.M.J., ter Haar Romeny, B.M., Viergever, M.A., Koenderink, J., 1996. The Gaussian scale-space paradigm and the multiscale local jet. *Int. J. Comput. Vision* 18, 61–75.
- Freeman, W.T., Adelson, E.H., 1991. The design and use of steerable filters. *IEEE Trans. Patt. Anal. Machine Intell.* 13 (9), 891–906.
- Jain, A.K., Ratha, N.K., Lakshmanan, S., 1997. Object detection using Gabor filters. *Pattern Recognition* 30 (2), 295–309.

- Koenderink, J.J., 1984. The structure of image. *Biol. Cybern.* 50, 363–370.
- Lanitis, A., Taylor, C.J., Cootes, T.F., 1995. Automatic face identification system using flexible appearance models. *Image Vision Comput.* 13 (5), 393–401.
- Leventon, M.E., Faugeras, O., Grimson, W.E.L., Wells, W.M., 2000. Level set based segmentation with intensity and curvature priors. In: *Proc. of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MBBIA)*, pp. 4–11.
- McInerney, T., Terzopoulos, D., 1996. Deformable models in medical image analysis: a survey. *Medical Image Analysis* 1 (2), 91–108.
- Malik, J., Perona, P., 1990. Preattentive texture discrimination with early vision mechanisms. *J. Optical Soc. Am.* 7, 923–932.
- Menet, S., Saint-Maric, P., Medioni, G., 1990. B-snakes: implementation and application to stereo. In: *Proc. DARPA Image Understanding Workshop*, pp. 720–726.
- Nastar, C., Ayache, N., 1996. Frequency-based nonrigid motion analysis: application to four dimensional medical images. *IEEE Trans. Patt. Anal. Machine Intell.* 18 (11), 1067–1079.
- Neuenschwander, W., Fua, P., Székely, G., Kübler, O., 1997. Velcro surfaces: fast initialization of deformable models. *Comput. Vis. Image Underst.* 65 (2), 237–245.
- Niessen, W.J., Duncan, J.S., Nielsen, M., Florack, L.M.J., ter Haar Romeny, B.M., Viergever, M.A., 1997. A multiscale approach to image sequence analysis. *Comput. Vis. Image Underst.* 65 (2), 259–268.
- Niessen, W.J., ter Haar Romeny, B.M., Viergever, M.A., 1998. Geodesic deformable models for medical image analysis. *IEEE Trans. Med. Imag.* 17 (4), 634–641.
- Niessen, W.J., Romeny, B.M.H., Viergever, M.A., 1999. Geodesic deformable models for medical image analysis. *IEEE Trans. Med. Imag.* 17 (4), 634–641.
- Paragios, N., Deriche, R., 1999. Unifying boundary and region-based information for geodesic active tracking. In: *Proc. of Computer Vision and Pattern Recognition*, Vol. 2, Colorado, pp. 300–305.
- Pardo, X.M., Carreira, M.J., Mosquera, A., Cabello, D., 2001. A snake for CT image segmentation integrating region and edge information. *Image Vision Comput.* 19, 461–475.
- Rao, R.P.N., Ballard, D.H., 1995. Natural basis functions and topographic memory for face recognition. In: *Proc. International Joint Conf. on Artificial Intelligence*, pp. 10–17.
- Staib, L.H., Duncan, J.S., 1992. Boundary finding with parametrically deformable models. *IEEE Trans. Patt. Anal. Machine Intell.* 14 (11), 1061–1075.
- Subirana-Vilanova, J.B., Sung, K.K., 1992. Multi-Scale Vector-Ridge-Detection for Perceptual Organization without Edges. Technical Report AIM-1318, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.
- Terzopoulos, D., 1987. On matching deformable models to images: direct and iterative solutions. *Topical Meeting on Machine Vision*, Vol. 12, Technical Digest, Optical Society of America, pp. 160–116.
- Turk, M., Pentland, A., 1991. Eigenfaces for recognition. *J. Cogn. Neurosci.* 3 (1), 71–86.
- Xu, C., Prince, J.L., 1998. Snakes, shapes, and gradient vector flow. *IEEE Trans. Image Process.* 7 (3), 359–369.
- Yezzi, A., Tsai, A., Willsky, A.S., 1999. A statistical approach to snakes for bimodal and trimodal imagery. *Proc. of 7th Int. Conf. on Computer Vision* 1, 898–903.
- Young, R.A., 1986. Simulation of human retinal function with the gaussian derivative model. In: *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Miami, FL, pp. 564–569.
- Yuille, A.L., Hallinan, P.W., Cohen, D.S., 1992. Feature extraction from faces using deformable templates. *Int. J. Comput. Vision* 8 (2), 133–144.
- Zeng, X., Staib, L.H., Schultz, R.T., Duncan, J.S., 1999. Segmentation and measurements of the cortex from 3D MR images using coupled-surfaces propagation. *IEEE Trans. Med. Imag.* 18 (10), 927–937.