

Level Set Segmentation with shape and appearance models using affine moment descriptors

C. Platero, M.C. Tobar, J. Sanguino, J.M. Poncela, O. Velasco

Applied Bioengineering Group, Technical University of Madrid

Abstract. We propose a level set based variational approach that incorporates shape priors into edge-based and region-based models. The evolution of the active contour depends on local and global information. It has been implemented using an efficient narrow band technique. For each boundary pixel we calculate its dynamic according to its gray level, the neighborhood and geometric properties established by training shapes. We also propose a criterion for shape aligning based on affine transformation using an image normalization procedure. Finally, we illustrate the benefits of the our approach on the liver segmentation from CT images.

1 Introduction

Level set techniques have been adapted to segment images based on numerous low-level criteria. Several edge-based and region-based models have been proposed without information priors. More recently, shape prior has been integrated into the level set framework [1–3]. There are two topics in this area: a) shape alignment and b) shape variability. The first issue is to calculate the set of pose parameters (rotation, translation and scaling) used to align the template set, and hence remove any variations in shape due to pose differences. And the second question is the deformation of the shape, which is typical derived from a training set using Principal Component Analysis (PCA) on the distance surface to the object [4, 1].

In this paper, we introduce a new approach for shape variability, which combines a parametric registration of shapes by affine moment descriptors with shapes encoded by their signed distance functions. We solve the shape alignment using the image normalization procedure [5], which avoids increasing the number of coupled partial differential equations of the problem. Finally, a new active contour evolves using a combination the appearance terms and shape terms. The paper is organized as follows: in Section 2, we show the problem of the shape alignment and our approach based on image normalization. Section 3 describes the problem of building a shape term and present one based on the two shape distance measure. Section 4 presents our framework for image segmentation. Finally, in Section 5, we apply our procedure for liver segmentation from CT images.

2 Shape Alignment

Consider a training set consisting of N binary images $\{S_i\}_{i=1,\dots,N} : \Omega \subset \mathbb{R}^n \rightarrow \{0, 1\}$, $n = 2$ or 3 . Our first aim is to align it, in order to avoid artifacts due to different pose. The new aligned images are defined as $\tilde{S}_i = S_i \circ T_i^{-1}$, where T_i is an affine transformation, given by the composition of a rotation, a scaling transformation and a translation. Equivalently, S_i can be written in terms of \tilde{S}_i as: $S_i = \tilde{S}_i \circ T_i$. Traditionally, the pose parameters have been estimated minimizing the following energy functional, via gradient descent [2]:

$$E_{align} = \sum_{\substack{i,j=1 \\ i \neq j}}^N \frac{\int_{\Omega} (\tilde{S}_i - \tilde{S}_j)^2 dx}{\int_{\Omega} (\tilde{S}_i + \tilde{S}_j)^2 dx}. \quad (1)$$

Minimizing (1) is equivalent to simultaneously minimizing the difference between any pair of binary images in the training database. We propose to improve this approach by using a normalization procedure over the shape priors. An advantage is that the affine transformation is defined by closed-form expressions involving only geometric moments. No additional optimization over pose parameters is necessary. This procedure will be applied both to the N aligned training shapes and to the problem of aligning the active contour. Specifically, given a reference image S_{ref} , we propose a criterion for alignment based on a shape normalization algorithm. It is only necessary to compute the first and the second order moments. The first-order moments locate the centroid of the shape and the second-order moments characterize the size and orientation of the image. Given a binary image S_i , we compute the second-order moment matrix, and the image is rotated using the eigenvectors and it is scaling along the eigenvectors according to the eigenvalues of the second-order moment matrix of S_i and S_{ref} . Then, it is translated to the centroid. We do not consider the problem of reflection (for this see [6]).

If we only consider moments up to order two, S_i is approximated to an ellipse/ ellipsoids centered at the image centroid. The ellipse/ellipsoids rotate angles and the semi-axes are determined by the eigenvalues and the eigenvectors of the second-order moment matrix [5]. We denote R as the rotation matrix.

Let S_{ref} be a normalized binary image as reference and $\{\lambda_j^{ref}\}_{j=1,\dots,n}$ be the eigenvalues for the reference image. We consider one of the following scale matrices: a) $W = \sqrt{\frac{\lambda^{ref}}{\lambda^i}} \cdot I$ where $\lambda = (\prod_{j=1}^n \lambda_j)^{1/n}$ and I is the identity matrix or b) W is diagonal matrix where $w_{j,j} = \sqrt{\frac{\lambda_j^{ref}}{\lambda_j^i}}$. In the first case is a scaling identical in all directions, while in the second case the size fits in each principal axis as the reference. The first option is used for shape priors without privileged directions otherwise the second option should be used. Finally, if the reference centroid is \bar{x}_{ref} , the affine transformation translates the origin of the coordinate system to the reference centroid. Then, the affine transformation is defined as follows:

$$T_i^{-1}(x) = R \cdot W \cdot (x - \bar{x}_i) + \bar{x}_{ref} \quad (2)$$

This affine transformation aligns from S_i to S_{ref} . If we use a scaling identical in all directions, S_{ref} will be only a numeric artefact for the pose algorithm. The alignment error does not depend on the reference, S_{ref} . But when each principal axis is adjusted to the reference, the alignment error depends on the choice of the reference. We can not guarantee the optimal pose for any shape. But neither the gradient descent method guaranteed to find the optimum because there is no evidence that the functional (1) is convex. Our procedure is fast and optimum if the shapes are closed to ellipses or ellipsoids. Section 5 will compare our approach with the variational method of (1).

3 Handling shape variability

Early work on this problem involves the construction of shapes and variability based on a set of training shapes via principal component analysis (PCA). In recent years, researchers have successfully introduced prior information about expected shapes into level set segmentation. Leventon et al. [1] modeled the embedding function by principal component analysis of a set of training shapes and added appropriate driving terms to the level set evolution equation. Tsai et al. [2] suggested a more efficient formulation, where optimization is performed directly within the subspace of the first few eigenvectors. Following these works, suppose now that the N aligned shapes \tilde{S}_i define N objects, whose boundaries are embedded as the zero level set of N signed distance functions $\{\tilde{\phi}_1, \dots, \tilde{\phi}_N\}$ respectively, which assign positive distances to the inside of the object, negative to the outside. Now, we constrain the level set function $\tilde{\phi}$ to a parametric representation of the shape variability [2]:

$$\tilde{\phi}_{\alpha}(x) = \tilde{\phi}_0(x) + \sum_{i=1}^k \alpha_i \tilde{\psi}_i(x) \quad (3)$$

where $k \leq N$ is empirically chosen and $\alpha = \{\alpha_1, \dots, \alpha_k\}$ collects the weights for the first k eigenvectors $\tilde{\psi}_i(x)$, $i = 1, \dots, k$. We also have the mean level set function $\tilde{\phi}_0(x) = \frac{1}{N} \sum_{i=1}^N \tilde{\phi}_i(x)$ of the aligned shape database. This will drive the shape variability, where the parameter vector α models shape deformations. Experimentally, we have observed that given any parameter vector α , the shape generated by (3) is also a normalized shape. It preserves the centroid, the orientation and the product of the eigenvalues remains constant.

3.1 Shape model

Each aligned training shape $\tilde{\phi}_i$ can be represented by its corresponding shape parameter $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ik})$. Cremers et al. [3] have introduced nonlinear statistical shape priors based on kernel density estimation. The goal of statistical shape learning is to infer a statistical distribution $\mathcal{P}(\alpha)$ from these samples. Following [3], it considers a nonparametric density approximation:

$$\mathcal{P}(\boldsymbol{\alpha}) = \frac{1}{N\sigma} \sum_{i=1}^N K\left(\frac{\boldsymbol{\alpha} - \boldsymbol{\alpha}_i}{\sigma}\right) \quad (4)$$

where $K(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{|z|^2}{2})$, being $\sigma^2 = \frac{1}{N} \sum_{i=1}^N \min_{i \neq j} |\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_j|^2$ the average nearest neighbor distance. It combines the nonparametric shape prior and a data term within a Bayesian framework to form the energy functional for segmentation. The data term use to minimize the probability of misclassified pixels for two regions [7]. Let $u : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}$ be the image to be segmented and let $\tilde{\phi}$ be the level set function with $\tilde{\phi}(x) > 0$ if $x \in \Omega_{in}$ and $\tilde{\phi}(x) < 0$ if $x \in \Omega_{out}$. We also consider the regularized Heaviside function $H(s)$. The data energy functional is written as

$$\begin{aligned} E_{data}(\boldsymbol{\alpha}) = & - \int_{\Omega} H(\tilde{\phi}_{\boldsymbol{\alpha}}(x)) \log(p_{in}(u(T^{-1}(x)))) dx \\ & - \int_{\Omega} (1 - H(\tilde{\phi}_{\boldsymbol{\alpha}}(x))) \log(p_{out}(u(T^{-1}(x)))) dx \end{aligned} \quad (5)$$

being T^{-1} the affine transformation which accommodates shape variability due to differences in pose, it is also calculated using geometric moments. With the nonparametric models for shape and intensity introduced above, this leads to an energy of the form

$$E(\boldsymbol{\alpha}) = E_{data}(\boldsymbol{\alpha}) - \log(\mathcal{P}(\boldsymbol{\alpha})). \quad (6)$$

This approach is quite robust with respect to initialization and noise. However, it has also been observed that the evolution becomes inefficient when the object shape to be segmented varies significantly with respect to the training base. Indeed, when the dimension of the shape parameter vector is much lower than the number of elements of the active contour, then the decision of each element of the boundary can not be taken with a vector so generic. Therefore, we propose a signed distance-based measure. Let Φ be the level set function for segmentation, and $\tilde{\phi}_{\boldsymbol{\alpha}}$ be the one embedding the shape model as (3). Both are signed distance functions. Hence, we propose the following shape term:

$$E_{shape}(\Phi, \tilde{\phi}_{\boldsymbol{\alpha}}) = \frac{1}{2} \int_{\Omega} (\Phi(x) - \tilde{\phi}_{\boldsymbol{\alpha}}(T(x)))^2 dx. \quad (7)$$

It minimizes the dissimilarity measure between the target $\Phi(x)$ and the shape model $\tilde{\phi}_{\boldsymbol{\alpha}}(T(x))$. The pose parameter is calculated with the moments of $\Phi(x) \geq 0$ and $\tilde{\phi}_{\boldsymbol{\alpha}}(x) \geq 0$. The optimization criterion produces local pixel-wise deformation.

4 Our model

Several edge-based and region-based models have been proposed. These two types of models are usually based on the fact that the regions are piecewise constant and the borders have high slopes. However, the regions of interest are not

usually statistically homogeneous; noise, weak edges and small objects are also presented in most of the real images. We propose a new procedure that takes all these features into account. The basic idea is to pre-process the image with a filter based on the nonlinear diffusion techniques. The segmentation problem was formalized by Mumford and Shah as the minimization of an energy functional that penalizes deviations from smoothness within regions and the length of their boundaries. Under certain assumptions on the discretization, the nonlinear diffusion filters are connected with minimizing the Mumford Shah functional [8]. Following this idea, we have used a nonlinear diffusion filter and we have also applied a stopping time criteria for obtaining piecewise smooth images. We can certainly assume that the nonlinear diffusion produces a filtered image where the intensity distributions are closed to Gaussian functions, modeled with the mean and variance for each region. Let us consider a image made up of two regions and integrating the maximization of the a posteriori probability and the regularity constraint, we obtain the following energy [7]:

$$E_{region}(\Phi) = - \int_{\Omega} H(\Phi(x)) \log p_{in}(u(x)|\mu_{in}, \sigma_{in}^2) dx - \int_{\Omega} (1 - H(\Phi(x))) \log(p_{out}(u(x)|\mu_{out}, \sigma_{out}^2)) dx + \nu \int_{\Omega} |\nabla\Phi(x)| dx \quad (8)$$

where $p(u|\mu_i, \sigma_i)$ denotes the probability of observing the filtered image u when Ω_i is a region of interest and ν is a positive constant chosen empirically. On the other hand, Kimmel and Bruckstein [9] have developed a novel and robust edge integration scheme. The edge-based stopping term serves to stop the contour on the desired object boundary. The active contour evolves along the second-order derivative in the direction of the image gradient. The functional is given by

$$E_{edge}(\Phi) = - \int_{\Omega} \delta(\Phi(x)) \nabla u \cdot \mathbf{n} dx + \int_{\Omega} H(\Phi(x)) \operatorname{div} \left(\frac{\nabla u}{\|\nabla u\|} \right) \|\nabla u\| dx \quad (9)$$

being $\delta(\cdot)$ a regularized version of the Dirac function such as $\delta(s) = H'(s)$ and \mathbf{n} is the normal unit vector to zero level set. Finally, the global functional is a weighted sum of the above ones where is combined data terms and shape priors:

$$E(\Phi, \tilde{\phi}_{\alpha}) = \varrho_1 E_{region}(\Phi) + \varrho_2 E_{edge}(\Phi) + \varrho_3 E_{shape}(\Phi, \tilde{\phi}_{\alpha}) \quad (10)$$

where ϱ_i are positive constants chosen empirically. Obviously, there are two evolutions, the first one is the active contour for the object segmentation and the second one represents the deformation model evolution. The two evolutions are related through the proposed affine transformation between $\Phi(x) \geq 0$ and $\tilde{\phi}_{\alpha}(x) \geq 0$. Moreover, both of them use the same statistical data improving the algorithm.

4.1 Numerical algorithms

In this subsection, we show the numerical algorithms for minimizing the functionals presented. We can not guarantee that our functionals are convex. Therefore,

gradient descent process stops at a local minima. One challenge is tracking down the significant minimum. This is done by initializing the active contour near of the object of interest. Following these observations, we propose a method of two main stages: (i) an initial segmentation is generated using a combination of traditional techniques, and (ii) we deform locally this initial solution using a level set approach which combines edge alignment, homogeneity terms and shape dissimilarity measures.

The level set evolution algorithm uses an efficient distance preserving the narrow band technique[10]. The reinitialization of the level set is not longer necessary. The algorithm is implemented using a simple finite difference scheme. It is based on the following internal energy term: $\gamma \int_{\Omega} \frac{1}{2} (\|\nabla\Phi(x)\| - 1)^2 dx$. Here $\gamma > 0$ is a parameter that controls the effect of penalizing the deviation of Φ from a signed distance function. Finally, the global functional is a weighted sum of the above functionals. The resulting evolution of the level set function is the gradient flow that minimizes the overall energy functional:

$$\begin{aligned} \partial_t \Phi(x) = \delta(\Phi(x)) & \left[\varrho_1 \cdot \log \frac{p_{in}(u(x))}{p_{out}(u(x))} + \varrho_1 \cdot \nu \cdot \operatorname{div} \left(\frac{\nabla\Phi(x)}{\|\nabla\Phi(x)\|} \right) - \varrho_2 \cdot u_{\eta\eta}(x) \right] \\ & + \varrho_3 \cdot (\Phi(x) - \tilde{\phi}_{\alpha}(T(x))) + \gamma \cdot \left(\Delta\Phi(x) - \operatorname{div} \left(\frac{\nabla\Phi(x)}{\|\nabla\Phi(x)\|} \right) \right) \end{aligned} \quad (11)$$

where $u_{\eta\eta}$ is the second derivative of u in the gradient direction. Recall that the affine transformation $T(x)$ is defined by the pose parameter. It has connected the pose from the normalized shape model to the target $\Phi(x) \geq 0$. This property makes more efficient the algorithm since it allows to pre-load $\tilde{\phi}_0(x)$ and $\tilde{\psi}(x)_{i=1,\dots,k}$. Moreover, $\tilde{\phi}_{\alpha}(x)$ is calculated using the above data about $p_{in}(u(x))$ and $p_{out}(u(x))$. Gradient descent method is now used to find the shape parameter α that minimizes $E(\alpha)$ in equation (6):

$$\partial_t \alpha = \int_{\Omega} \delta(\tilde{\phi}(x)) \tilde{\psi}(x) \left[\log \frac{p_{in}(u(T^{-1}(x)))}{p_{out}(u(T^{-1}(x)))} \right] dx + \frac{\sum_{i=1}^N (\alpha_i - \alpha) K_i}{\sigma^2 \sum_{i=1}^N K_i} \quad (12)$$

with $K_i = K(\frac{\alpha - \alpha_i}{\sigma})$ and $\tilde{\psi} = \{\tilde{\psi}_1, \tilde{\psi}_2, \dots, \tilde{\psi}_k\}$.

5 Liver segmentation from 3D CT Images

Liver segmentation from 3D CT images is usually the first step in the computer-assisted diagnosis and surgery systems for liver diseases. Algorithms relying solely on image intensities or derived features usually fail. To deal with missing or ambiguous low-level information, shape prior information has been successfully employed. However, the lack of knowledge about the location, orientation and deformation of the liver, due to diseases or different acquisition procedures, adds another major challenge to any segmentation algorithm. In order to address these problems, we have applied our framework to liver segmentation. The proposed method has been trained on 20 patient CT slice set, and tested on another 10 specified CT datasets. The shape model is composed of 20 segmented

livers. The segmented livers are aligned by the proposed procedure. In this case, each principal axis is adjusted to the reference. Experimentally, S_{ref} has been tuned by minimizing (1). Using *Similarity Index*, $\left(SI = \frac{1}{N} \sum_i \frac{(S_i \circ T_i^{-1}) \cap S_{ref}}{(S_i \circ T_i^{-1}) \cup S_{ref}}\right)$, our approach gets a SI of 0.66 while variational method of (1) gives 0.54 over the training set. From aligned training set, we calculate and save the mean level set function $\tilde{\phi}_0(x)$, the first k eigenvectors $\tilde{\psi}(x)$ and the shape parameter α_i , for each sample. In this application, we use $k = 10$. Our approach starts filtering the CT image by a nonlinear diffusion filter with selection of the optimal stopping time. Once the image has been processed, an intensity model is built, $p_{in}(u)$ and $p_{out}(u)$. We also calculate $u_{\eta\eta}$. We load the mean level set function $\tilde{\phi}_0(x)$, the first eigenvectors $\tilde{\psi}(x)$ and the shape parameter α_i , $i = 1, \dots, 20$. Then, region growing and 3D edge detector are applied to the filtered image. Morphological post-processing merges the previous steps, giving the initial solution and initializing the contour. The zero level of Φ is employed as the starting surface. The evolution of $\Phi(x)$ follows (11) and α is calculated as (12). The constant parameters of the active contour were tuned to segmentation scores using the leave-one-out technique. The pose parameter is calculated with the moments of $\Phi(x) \geq 0$ and $\tilde{\phi}_\alpha(x) \geq 0$. Fig. 1 shows slices from two cases, drawing the result of the method (in blue) and the reference (in red). The quality of the segmentation and its scores are based on the five metrics[11]. Each metric was converted to a score where 0 is the minimum and 100 is the maximum. Using this scoring system one can loosely say that 75 points for a liver is comparable to human performance. Table 1 lists the average values of the metrics and their scores over the test data set.

Table 1. Average values of the metrics and scores for all ten test case: volumetric overlap error (m_1), relative absolute volume difference (m_2), average symmetric surface distance (m_3), root mean square symmetric surface distance (m_4) and maximum symmetric surface distance (m_5).

Type	m_1	m_2	m_3	m_4	m_5
metrics	12.6%	4.7%	1.84 mm	3.86 mm	21.9 mm
scores	51	75	54	46	71

6 Conclusion

We have presented two main contributions. Firstly, the shape alignment has been solved using an image normalization procedure. An advantage is that the proposed affine transformation is defined by closed-form expressions involving only geometric moments. No additional optimization over pose parameters is necessary. This procedure has been applied both to the training shapes and to the problem of aligning the active contour. Secondly, we have proposed a level set based variational approach that incorporates shape priors into edge-based

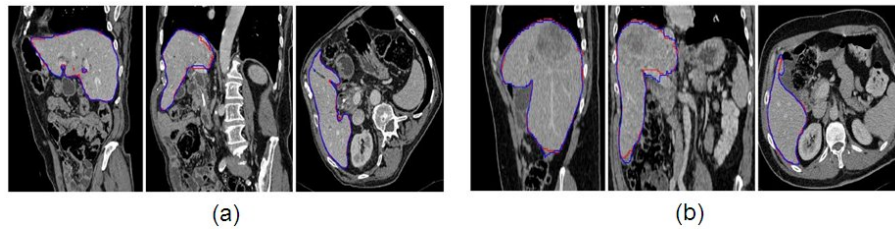


Fig. 1. From left to right, a sagittal, coronal and transversal slice for an easy case (a) and a difficult one (b). The outline of the reference standard segmentation is in red, the outline of the segmentation of the method described in this paper is in blue.

and region-based models. Using the Cremers' shape model, we have integrated a shape dissimilarity measure, a piecewise smooth region-based model and an edge alignment model. For each boundary pixel, our approach calculates its dynamic according to its gray level, the neighborhood and geometric properties established by training shapes.

References

1. Leventon, M., Grimson, W., Faugeras, O.: Statistical shape influence in geodesic active contours. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 1., IEEE Computer Society; 1999 (2000)
2. Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W., Willsky, A.: A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging* **22**(2) (2003) 137–154
3. Cremers, D., Rousson, M.: Efficient kernel density estimation of shape and intensity priors for level set segmentation. In Suri, J.S., Farag, A., eds.: *Parametric and Geometric Deformable Models: An application in Biomaterials and Medical Imagery*. Springer (2007)
4. Cootes, T., Taylor, C., Cooper, D., Graham, J., et al.: Active shape models-their training and application. *Computer vision and image understanding* **61**(1) (1995) 38–59
5. Pei, S., Lin, C.: Image normalization for pattern recognition. *Image and Vision computing* **13**(10) (1995) 711–723
6. Heikkila, J.: Pattern matching with affine moment descriptors. *Pattern recognition* **37**(9) (2004) 1825–1834
7. Rousson, M., Brox, T., Deriche, R.: Active unsupervised texture segmentation on a diffusion based feature space. In: *CVPR, IEEE Computer Society* (2003) 699–706
8. Kawohl, B.: From Mumford-Shah to Perona-Malik in image processing. *Mathematical Methods in the Applied Sciences* **27**(15) (2004) 1803–1814
9. Kimmel, R., Bruckstein, A.: Regularized laplacian zero crossings as optimal edge integrators. *International Journal of Computer Vision* **53**(3) (2003) 225–243
10. Li, C., Xu, C., Gui, C., Fox, M.: Level set evolution without re-initialization: A new variational formulation. In: *IEEE computer society conference on computer vision and pattern recognition*. Volume 1. (2005) 430–436
11. van Ginneken, B., Heimann, T., Styner, M.: 3D segmentation in the clinic: A grand challenge. In: *3D Segmentation in the Clinic: A Grand Challenge, MICCAI Workshop, Proceedings*. (2007) 7–15