Published in final edited form as:
Proc IAPR Int Conf Pattern Recogn. 2010 August 23; : 3955–3958. doi:10.1109/ICPR.2010.962.

Locally Deformable Shape Model to Improve 3D Level Set based Esophagus Segmentation

Sila Kurugol[1], **Necmiye Ozay**[1], **Jennifer G. Dy**[1], **Gregory C. Sharp**[2], and **Dana H. Brooks**[1]

Sila Kurugol: kurugol.s@neu.edu; Necmiye Ozay: ozay.n@neu.edu; Jennifer G. Dy: jdy@ece.neu.edu; Gregory C. Sharp: gcsharp@partners.org; Dana H. Brooks: brooks@ece.neu.edu

[1] Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA

[2] Dept. of Radiation Oncology, Mass. General Hospital and Harvard Medical School, Boston, MA, USA

Abstract

In this paper we propose a supervised 3D segmentation algorithm to locate the esophagus in thoracic CT scans using a variational framework. To address challenges due to low contrast, several priors are learned from a training set of segmented images. Our algorithm first estimates the centerline based on a spatial model learned at a few manually marked anatomical reference points. Then an implicit shape model is learned by subtracting the centerline and applying PCA to these shapes. To allow local variations in the shapes, we propose to use nonlinear smooth local deformations. Finally, the esophageal wall is located within a 3D level set framework by optimizing a cost function including terms for appearance, the shape model, smoothness constraints and an air/contrast model.

Keywords

3D medical image segmentation; shape model; level sets

I. Introduction

Contouring anatomic structures at risk in thoracic CT images is critical for radiotherapy treatment planing. Since manual contouring is labor intensive, considerable attention has been devoted to automating the process. Locating the esophageal wall is particularly challenging. Inherent difficulties in segmentation of the esophagus include the lack of consistent intensity contrast and variable appearance in regions which contain air bubbles or remains of oral contrast agent. Thus, a simple appearance model of esophagus tissue density which ignores this variability in appearance will not suffice. In addition, the segmentation algorithm needs a good model for the 3D esophagus shape, which is complex and variable.

Only a few studies have been reported on esophagus segmentation. Rousson et al. [1] located the esophageal wall in a limited range of the cranio-caudal axis (where the Left Atrium (LA) was present) by fitting a 2D ellipse shape model to each slice using an appearance based cost function with a regularization term for slice-to-slice smoothness. More recent work by Feulner et al. [2] used a fast algorithm based on probabilistic classification to find candidate 2D ellipses in each slice and combined them using a Hidden Markov Model. This method requires considerable labeled training data for correct classification of the esophagus, in part due its variable appearance in the presence of air bubbles, contrast agent, or both.

In previous work [3], we proposed a 3D esophagus segmentation technique for the entire thoracic range using a 3D level set framework. Unlike the previous approaches that used a collection of 2D ellipse models, no slice-to-slice smoothing was needed. We tried to incorporate as much prior information as possible from a small set of training data. In this paper, we follow a similar learning based approach. We improve the accuracy of our earlier work by proposing a new shape model, which includes a nonlinear local transformation component. Furthermore, we include a simple shape prior learned from the eigenmode weight distributions of training shapes to further constrain the shape space.

Level set methods have been used widely in medical image segmentation. Various energy functions, based on a variety of features, have been proposed to drive the deformable curve/surface to the desired location in the image. However, existing energy terms are insufficient for segmentation of certain structures in medical images where there is a lack of both consistent intensity contrast between the object and the background and of clear discriminative features. To overcome these difficulties, shape models capturing the geometry of the structure of interest have been incorporated into the level set framework as a shape energy term [4]. This term drives the evolving contour to be similar to a shape model learned from the training samples, using parametric/non-parametric density estimation or PCA [5]. When a limited amount of training data is available, PCA is preferred over density estimation. However, as noted in [6], PCA models only global variation of the samples within the training set.

In this paper, we propose a new shape model and a shape matching term. In order to circumvent the limitations of PCA based shape models, we separate the shape model into global and local components. We use PCA modes to represent the global component and we incorporate a nonlinear local transformation component to our shape model to account for local deformations. In order to limit the variability in the training data set, we center the training data at each slice around the estimated center location through the new spatial model that we introduce. We then incorporate the smooth local deformations and the shape model into the energy functional that the level set evolves accordingly. The main contributions of this work are two-fold: The introduction of nonlinear local transformations to the PCA-based shape model to account for local deformations and the efficient utilization of prior information within a 3D variational framework to address the very challenging task of esophagus segmentation.

II. Proposed Method

A. Shape Model

Global shape registration—For our shape model, we used a set of manually annotated training data sets. We first apply a global registration step to register each training shape with respect to a reference shape. The standard approach used for shape registration before constructing a shape model is matching the shapes using an affine transformation. However, for deformable structures such as esophagus, a more complex model might be necessary. Thus, we use a global registration method consisting of two steps: 1) First we apply a normalization in the cranio-caudal (z) axis on all the training data sets by matching 7 anatomical landmark locations and interpolating slices in between the landmarks. 2) Then, we subtract the centerline from the data sets, i.e. shifting each slice such that the shape center lies at the origin (See Fig 1(a)).

The global registration method applied to the training data sets will also be applied to a test set given the locations of anatomical landmark points in z axis and the estimated centerline location. For centerline estimation we proposed an algorithm in our previous work [3]. We built a spatial model of esophagus center location with respect to neighboring anatomical

structures and an appearance model of the esophagus from annotated training sets. Based on these models, we estimated the center location for anatomical landmark slices and to estimate the centerline for the rest of the slices we used our centerline model. Once we had the estimated centerline for a test set, we applied our global shape registration method to bring the test data into a common reference framework with the framework of our shape model.

Shape prior in the level set space—For our shape model, we first normalized training data sets by applying our global shape registration method. Then, we calculated the signed distance functions of the training data. We subtracted the mean shape from each data set and applied PCA to obtain the k most important modes of variation (U_i) of the shape. The esophagus shape without the centerline was represented by this shape model:

$$\psi=\psi_m+\sum_{i=1}^{k} w_i U_i \tag{1}$$

We also introduced a prior on weights of the modes (w_i). We calculated the histogram of each w_i over the training set of shapes. The ranges of w_is are calculated from these histograms and we assumed a uniform density within these ranges

B. Level set formulation of surface evolution

We use a 3D level set algorithm to find the esophagus surface by optimizing an energy function. However, due to variability in appearance and inconsistent contrast, classical appearance and edge based energy functions did not perform well. Thus, prior information from the training datasets was incorporated into the energy function. We propose a shape model from manually annotated esophagus data sets and constrain the segmented esophagus to stay close to this model by including a shape term in the energy function. This energy function (E) has an appearance term, air/contrast terms, a neighboring structures term, a shape fitting term, a level set regularization and a smoothing (curvature) term:

$$E=E_{app}+E_{air}+E_{nb}+E_{shape}+E_{reg}+E_{sm} \tag{2}$$

Each term in E is explained next.

We learned an appearance model from the training data as two intensity probability density functions (pdf), one inside (p_i) and one outside (p_o) the esophagus, calculated using kernel density estimation. The main challenge in estimating the intensity distribution is that the esophagus appearance varies in the presence of air bubbles (very dark compared to the esophagus) or oral contrast agent (very bright) or both. Hence, we first detect those regions using thresholding and a region growing algorithm inside a window of size 40×40 pixels around a preliminary esophagus center estimate. This preliminary estimation is based on only our spatial model. We discard detected air/contrast regions when calculating the esophagus pdf. At the time of center estimation, we substitute the intensities of these regions with the most likely esophagus intensity value. The energy term for appearance including these pdfs is:

$$E_{app}(p_i, p_o)= -\int_{\Omega}[H_{\varepsilon}(\varphi)\log(p_i(I))d\Omega - \int_{\Omega}(1 - H_{\varepsilon}(\varphi))\log(p_o(I))]d\Omega \tag{3}$$

where H_{ε} is the heaviside function of the level set φ.

We make use of the presence of air and contrast regions by incorporating them into our level set framework with an additional air/contrast energy term:

$$E_{air}(p_{air}) = -\int_{\Omega}[H_{\varepsilon}(\varphi)\log(p_{air}(I))d\Omega - \int_{\Omega}(1 - H_{\varepsilon}(\varphi))\log(p_{air}(I))]d\Omega \quad (4)$$

where p_{air} probability function that indicates the probability of a voxel being esophagus. This function is close to 1 if the pixel is air/contrast and 0.5 otherwise.

We use a similar energy term (E_{nb}, same form as Eqn 4) to exclude neighboring structures from the segmented esophagus. We create a probability function that takes low values for the neighboring structure voxels and 0.5 otherwise. The neighboring structure mark-ups are given for landmark locations. Contours for intermediate slices are calculated by shape interpolation.

We also include a shape term in our energy function. Our assumption is that the object to be segmented belongs to the family of shapes that are represented by our shape model up to a global affine transformation. We handle this global shape matching using our global registration method described earlier. However since we are looking for a nonrigid structure, in addition to global registration, we need local deformations. Due to high data variability in x-y direction and rough center estimation in our global registration step, we include into our shape model a local transformation function that acts on x and y directions. The local deformations allow the algorithm to correct for the inaccuracies in center estimation (See Fig 1(b)) and better capture local variabilities that cannot be represented by the global PCA-based shape model. We follow the locally affine transformation model in [7]. We choose N uniformly sampled *action points* z_k through the centerline. We allow a local transformation, in the form of a translation in x-y plane, to be applied to each z_k. This translation affects the neighboring slices and this effect smoothly dies off as one moves away from the action points in z-direction. Such a local deformation $\mathbf{A}$ can be formally defined as follows:

$$\mathbf{A}\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \sum_{k=1}^{N} \left\{ \begin{bmatrix} \lambda(z - z_k) \\ 0 \\ 0 \end{bmatrix} a_k + \begin{bmatrix} 0 \\ \lambda(z - z_k) \\ 0 \end{bmatrix} b_k \right\} \quad (5)$$

where $\lambda(r) = e^{-\frac{r^2}{2\sigma^2}}$, a_k and b_k ($k \in 1, \ldots, N$) are x and y translations at action point k and σ is a scale parameter. Next, we define the shape energy:

$$E_{shape}(\varphi, w) = \int_{\Omega}\delta_{\varepsilon}(\varphi)(\varphi - \psi(\mathbf{A}))^2 d\Omega \quad (6)$$

that drives the level set function φ to be similar to our shape model $\psi(\mathbf{A})$. δ_{ε} is the dirac function. After including a level set regularization term E_{reg} from Li et al. [8] and a smoothness term E_{sm} from [6], the overall energy functional in Eqn 2 is obtained.

Next, we apply the 3D level-set method described to locate the esophagus wall in the 3D CT stack. We initialize the shape prior as the mean shape. The initial level set function representing the esophagus boundary and the shape prior level set function are updated at each step t by minimizing E. The equation of evolution for φ is given by calculus of its variations; the optimization of E with respect to mode weights w_i can be obtained by solving a linear system [6]. Adding the weight priors results in a constrained least square minimization that is solved by convex optimization [9]. Finally, the minimization of E with

respect to local deformation parameters a_k, b_k is carried out using calculus of variations and following time evolution equation is obtained:

$$\begin{aligned}\frac{d}{dt}a_k &= 2\int_\Omega \delta_\varepsilon(\varphi)(\varphi - \psi(\mathbf{A}))(\nabla\psi(\mathbf{A})\frac{\partial}{\partial a_k}\mathbf{A})d\Omega \\ &= -2\int_\Omega \delta_\varepsilon(\varphi)(\varphi - \psi(\mathbf{A}))(\psi_x \lambda(z - z_k))d\Omega\end{aligned} \tag{7}$$

The equation for b_k is similar.

III. Experimental Results

We report experiments on 8 thoracic CT scans from different subjects using a leave-one-out scheme. The inputs to the algorithm are the anatomical landmark points and segmentation of the neighboring structures for only 7 slices at these landmarks. Everything else is obtained automatically. We use 6 action points separated by 10 slices. Energy function terms are weighted and the weights are chosen experimentally. We first show results of the global registration algorithm that centers the data to the estimated centerline. The complete center estimation algorithm applied to all slices achieved less than 1.9 mm average error in the x-direction but was less accurate in the y-direction, with a 4.1 mm average error. To improve the accuracy of our global registration algorithm we applied a local registration algorithm during the level set surface evolution and shape matching step. As a result, the centerline estimation error reduce to 1.1 mm average error in the x-direction and 2.7 in the y-direction.

We use a point-wise distance metric to evaluate the results of the outer boundary surface segmentation algorithm. For each slice we find the point correspondence between the expert labeled contour and the contour computed by the algorithm by comparing the points on both contours at the same angle from the x axis to obtain a point-wise distance metric. Fig. 2 shows point-wise mean and max distance errors of each data set. We obtained a point-wise mean error of 2.6 ± 2.1 mm and maximum error of 17.6 mm over all data sets, which indicates an improvement in accuracy in comparison to our previous work [3], where we applied a heuristic approach to correct for the errors in global registration and obtained a point-wise mean error of 3.3 ± 2.8 mm and maximum error of 19.2 mm over all data sets.

The results for various axial slices are shown and compared with ground truth in Fig 3. The resultant esophagus surface rendered in 3D in comparison to the ground-truth is shown in Fig. 4 for 2 different data sets.

IV. Conclusions and Future Work

We propose a supervised 3D segmentation algorithm to locate the esophagus in thoracic CT scans using a variational framework. We improved our results compared to previous work by introducing nonlinear smooth transformations to the shape model to correct for the inaccuracies of global transformation. In addition we added more constraints on the shape model through our shape prior. We plan to test our algorithm on a larger database. We eventually expect to be able to additionally decrease the required user input by automating the landmark selection and neighboring structure identification processes.

Acknowledgments

Support for the work of SK and DHB provided in part by the NIH/NCRR Center for Integrative Biomedical Computing (CIBC), P41-RR12553-09, JD was also partly supported by NSF IIS-0347532

References

1. Rousson M, Ying B, Chenyang X, Sauer F. Probabilistic minimal path for automated esophagus segmentation. Proc of SPIE. 2006:6144.
2. Feulner J, Feulner S, Zhou SK, Cavallaro A, Seifert S, Hornegger J, Comaniciu D. Fast automatic segmentation of the esophagus from 3d ct data using a probabilistic model. MICCAI. 2009
3. Kurugol, S.; Dy, JG.; Sharp, GC.; Brooks, DH. Level set esophagus segmentation in thoracic CT images using spatial, appearance and shape models. ISBI; 2010.
4. Rousson M, Paragios N. Shape priors for level set representations. ECCV. 2002:416–418.
5. Cremers D, Rousson M, Deriche R. A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape. Int J of Comp Vis. 2007; 72(2):195–215.
6. Rousson M, Paragios N, Deriche R. Implicit active shape models for 3D segmentation in MR imaging. MICCAI. 2004; 3216:209–216.
7. Narayanan R, Fessler JA, Park H, Meyer CR. Diffeomorphic nonlinear transformations: A local parametric approach for image registration. IPMI. 2005:174–185.
8. Li C, Xu C, Gui C, Fox MD. Level set evolution without re-initialization: A new variational formulation. CVPR. 2005; 1:430–436.
9. Grant, M.; Boyd, S. CVX: Matlab software for disciplined convex programming. June. 2009

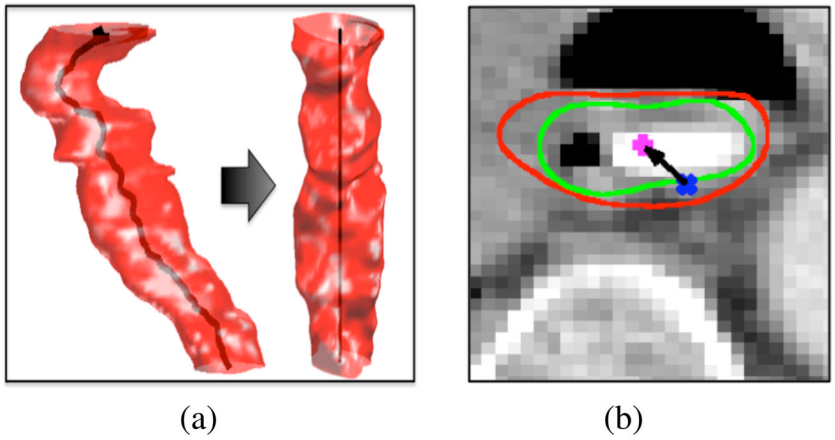

Figure 1.
(a) Global registration (centerline subtraction) (b) Initial center estimate (blue star) and final center after local registration (pink plus).

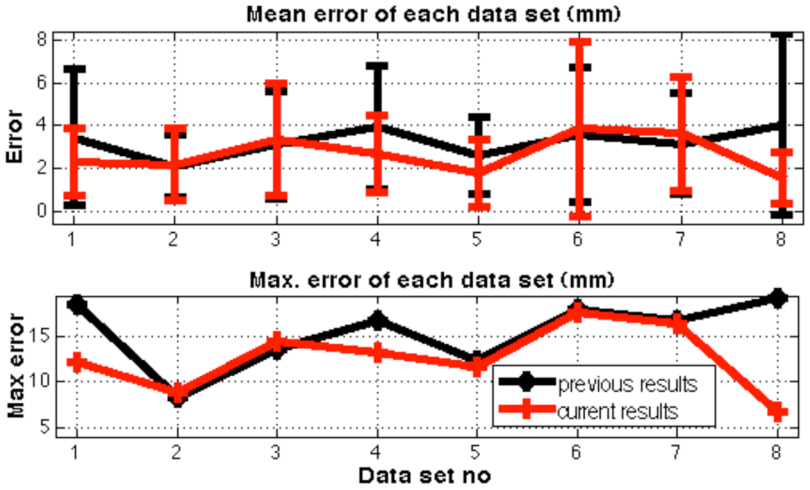

Figure 2.
Mean ± std of error (top figure) and maximum error of each data set (bottom figure)

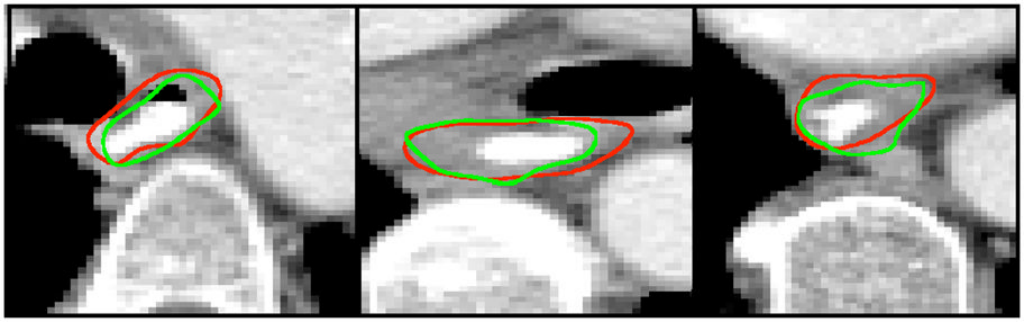

Figure 3.
Three axial slices showing the result of the algorithm (green) and expert (red)

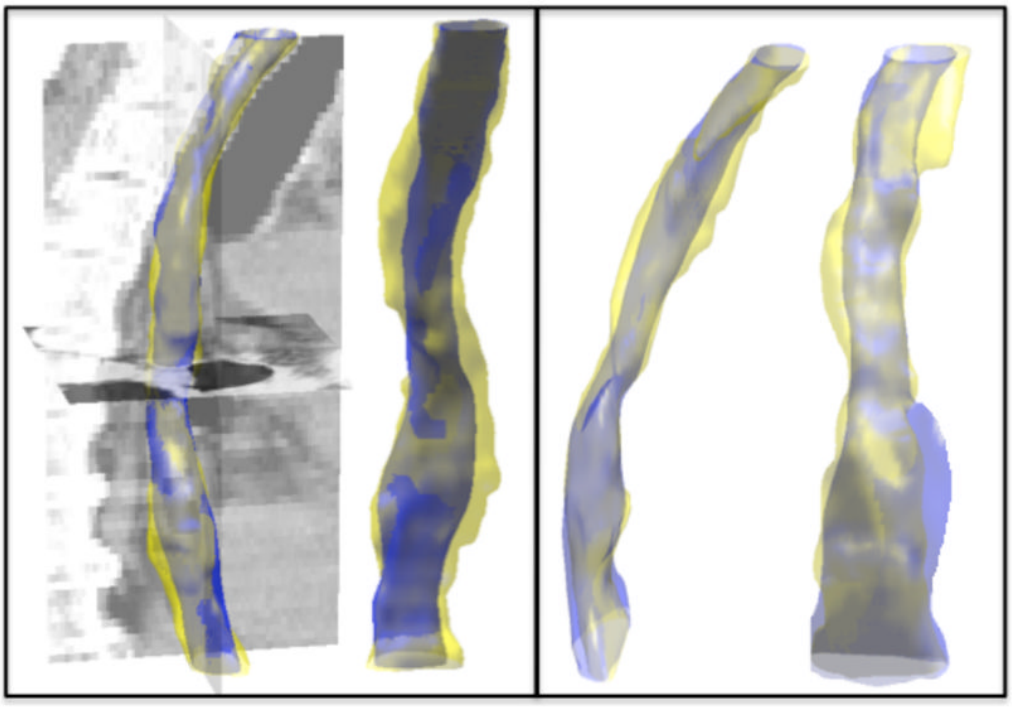

Figure 4.
Resultant esophagi in 3D (blue-ground truth, yellow-algorithm). Left and right boxes show two representative views for data set 8 and 2 respectively. z-direction is scaled by 0.5 for better visualization. Example orthogonal CT slices (from original input data) is superimposed in the leftmost result.