Global Active Contour-based Image Segmentation via Probability Alignment

Andriy Myronenko and Xubo Song
Dept. of Science and Engineering, School of Medicine
Oregon Health and Science University
20000 NW Walker Rd, Beaverton , OR, 97006, USA
myron@csee.ogi.edu, xubosong@csee.ogi.edu

Abstract

Active contours is a popular technique for image segmentation. However, active contour tend to converge to the closest local minimum of its energy function and often requires a close boundary initialization. We introduce a new approach that overcomes the close boundary initialization problem by reformulating the external energy term. We treat the active contour as a mean curve of the probability density function $p(\mathbf{x})$. It moves to minimize the Kullback-Leibler (KL) divergence between $p(\mathbf{x})$ and the probability density function derived from the image. KL divergence forces $p(\mathbf{x})$ to "cover all image areas" and the uncovered areas are heavily penalized, which allows the active contour to go over the edges. Also we use deterministic annealing on the width of $p(\mathbf{x})$ to implement a coarse-to-fine search strategy. In the limit, when the width of $p(\mathbf{x})$ goes to zero, the KL divergence function converges to the conventional external energy term (which can be seen a special case) of active contours. Our method produces robust segmentation results from arbitrary initialization positions.

1. Introduction

Deformable models for image segmentation became popular since 1987, when Kass, Witkin and Terzopoulos [11] introduced their seminal work on *snakes*. Applications of snakes include medical image analysis [18] and moving object tracking [19, 2]. Snake, or active contour, is a curve in the image plane, which moves to minimize its internal and external energy functions. The internal energy function regularizes the shape of the contour, which is defined as first and second order derivative regularization. The external energy function is a modified image itself, such that the smaller intensity values correspond to the desirable contour location.

Close boundary initialization of snakes is crucial. First, snakes rely on the gradient of the external energy to move; thus, a contour initialized far from the boundary in a homogeneous area will make no move. Second, noise and false boundaries create numerous local minima, where snakes can be easily trapped. Third, snakes can not move over the boundaries and thus requires initialization around or within the required segmentation boundaries.

To overcome the close boundary initialization problem, many region-based methods, including Chan and Vese [5], Li and Yezzi [13], Bresson et al. [3], An et al. [1], have been proposed based on Mumford-Shah formulation [16]. Region-based information helps to overcome local minima and in some case achieve a globally optimal solution regardless of the initialization. Bresson et al. [3] introduced a fast method for global minimization of active contours by combining variational image segmentation and denoising methods. An et al. [1] proposed to use a Γ-convergence approximation for a multi-scale piecewise smooth model to overcomes the limitations of global region models while avoiding the high sensitivity of local approaches. Other researchers, including Cremers et al. [6] and Leventon et al. [12], proposed to utilize the shape prior to find a globally optimal solution, given the statistical shape prior is available. For a thorough overview of the active contour methods see [13, 6, 3] and references therein.

We believe the difficulty of snakes lies in the definition of its external energy function, which is a preprocessed image itself, and has numerous local minima and flat regions. It is hard to find a global optimum of such function from the optimization perspective. The energy is minimized only locally, which requires appropriate initialization and has no guarantee of reaching the correct solution.

In this paper, we take a different approach to the formulation of the active contour and its external energy. We treat the active contour as a continuous mean curve $\mathbf{y}(s) \in \mathcal{R}^2$ ($s \in [0,1]$) of the probability density function $p(\mathbf{x})$ ($\mathbf{x} \in \mathcal{R}^2$). We derive another pdf $F(\mathbf{x})$ from the input image, so that $F(\mathbf{x})$ has larger values at the features of interest. We align $p(\mathbf{x})$ to $F(\mathbf{x})$ by minimize the Kullback-Leibler (KL) divergence between the distributions subject to the contour

position $\mathbf{y}(s)$. We gradually anneal the width of $p(\mathbf{x})$ to implement a coarse-to-fine search strategy. In the limit, when the width of $p(\mathbf{x})$ goes to zero, the KL divergence converges to the conventional external energy term of active contours. Thus the conventional external energy can be seen as a special case of our formulation. Our method overcomes the close boundary initialization problem of conventional active contours and allows robust segmentation from an arbitrary initialization position.

2. Fundamentals

Conventional snake is a curve $\mathbf{y}(s) \in \mathcal{R}^2$, $s \in [0,1]$, which moves over the image domain to minimize its energy function [11]:

$$E(\mathbf{y}) = \int_0^1 [E_{int}(\mathbf{y}(s)) + E_{ext}(\mathbf{y}(s)]ds \qquad (1)$$

The internal energy, E_{int}, keeps the snake short and smooth:

$$E_{int}(\mathbf{y}(s)) = \frac{1}{2}\left(\alpha \left\|\frac{\partial \mathbf{y}(s)}{\partial s}\right\|^2 + \beta \left\|\frac{\partial^2 \mathbf{y}(s)}{\partial s^2}\right\|^2\right) \qquad (2)$$

The first and second order derivative regularization force the snake to have the smallest length (tension) and the smoothest shape (rigidity). The α and β weights control the snake's tension and rigidity respectively.

The external energy, E_{ext}, is a modified image itself, such that the lower intensity values correspond to the desirable snake location. The negative norm of image gradient is usually used as the external energy:

$$E_{ext}(\mathbf{x}) = -|\nabla[G(\mathbf{x}) * I(\mathbf{x})]|^2 \qquad (3)$$

where a Gaussian filter G is used to presmooth the image I.

A snake that minimizes $E(\mathbf{y})$ must satisfy the Euler-Lagrange differential equation:

$$\alpha\frac{\partial^2 \mathbf{y}(s)}{\partial s^2} - \beta\frac{\partial^4 \mathbf{y}(s)}{\partial s^4} - \nabla E_{ext}(\mathbf{y}(s)) = 0 \qquad (4)$$

Approximating the derivatives with finite differences, the non-linear system of equations can be iteratively solved by by treating $\mathbf{y}$ as a function of time (see [11] for details).

3. Method

Key to our method is the notation of the *probability-map* $F(\mathbf{x})$ derived from the original image $I(\mathbf{x})$, so that $F(\mathbf{x})$ has larger values at the features of interest. For instance, if we are interested in the image edges, we can use a norm of the image gradient as a probability-map:

$$F(\mathbf{x}) = \frac{1}{\int |\nabla I(\mathbf{x})|^2}|\nabla I(\mathbf{x})|^2 \qquad (5)$$

This notation is similar to the snakes' external energy in Eq. 3. The higher probability areas of $F(\mathbf{x})$ represent the desirable segmentation regions.

We represent our active contour $\mathbf{y}(s) \in \mathcal{R}^2$, $s \in [0,1]$ as a mean curve of the probability density function $p(\mathbf{x})$:

$$p(\mathbf{x}) = \int_0^1 \frac{1}{2\pi\sigma^2}\exp^{-\frac{\|\mathbf{x}-\mathbf{y}(s)\|^2}{2\sigma^2}} ds, \qquad (6)$$

$p(\mathbf{x})$ can be seen as a continuous analog of Gaussian mixture model with equal isotropic covariances. We also add one more term to the $p(\mathbf{x})$, a uniform distribution $\frac{1}{N}$ (where N is a size of the image domain), in order to account for noise and outliers. We rewrite $p(\mathbf{x})$ as:

$$p(\mathbf{x}) = (1-w)\int_0^1 \frac{1}{2\pi\sigma^2}\exp^{-\frac{\|\mathbf{x}-\mathbf{y}(s)\|^2}{2\sigma^2}} ds + w\frac{1}{N}, \qquad (7)$$

Parameter w represents the contribution of the uniform distribution ($0 < w < 1$).

We search for the active contour position, such that the distribution $p(\mathbf{x})$ maximally approximates $F(\mathbf{x})$, by minimizing Kullback-Leibler (KL) divergence between the distributions:

$$E_{fit}(y) = \int_{\mathcal{R}^2} F(\mathbf{x}) \log \frac{F(\mathbf{x})}{p(\mathbf{x})} d\mathbf{x} \qquad (8)$$

which we call the *fitness term*. To regularize the shape of the contour we keep the original active contour regularization:

$$E_{reg}(y) = \frac{1}{2}\int_0^1 (\alpha \left\|\frac{\partial \mathbf{y}(s)}{\partial s}\right\|^2 + \beta \left\|\frac{\partial^2 \mathbf{y}(s)}{\partial s^2}\right\|^2) ds \qquad (9)$$

which we call the *regularization term*. The goal is to find the function $\mathbf{y}(s)$ that minimizes the energy functional

$$E(\mathbf{y}) = E_{fit}(\mathbf{y}) + E_{reg}(\mathbf{y}). \qquad (10)$$

The influence of each term is controlled by the weights α and β.

We note that the probabilistic approach to active contours in Blake and Isard [2], where the mean curve is used to regularize the possible active contour shape, should not be confused with our work where the mean curve is a papameter itself that alters the form of the distribution $p(\mathbf{x})$ to align it with the probability-map $F(\mathbf{x})$.

3.1. Optimization

We rewrite the energy functional ignoring the terms independent of $\mathbf{y}$ as:

$$E(\mathbf{y}) = -\int_{\mathcal{R}^2} F(\mathbf{x}) \log(\int_0^1 e^{-\frac{\|\mathbf{x}-\mathbf{y}(s)\|^2}{2\sigma^2}} ds + \frac{2\pi\sigma^2 w}{(1-w)N}) d\mathbf{x} + \frac{1}{2}\int_0^1 \left(\alpha \left\|\frac{\partial \mathbf{y}(s)}{\partial s}\right\|^2 + \beta \left\|\frac{\partial^2 \mathbf{y}(s)}{\partial s^2}\right\|^2\right) ds \qquad (11)$$

The function $\mathbf{y}(s)$ that minimizes $E(\mathbf{y})$ must satisfy the Euler-Lagrange differential equation[1], which we obtain by taking the functional derivative of $E(\mathbf{y})$ with respect to $\mathbf{y}$:

$$-\frac{1}{\sigma^2}\int F(\mathbf{x})\frac{(\mathbf{x}-\mathbf{y}(s))\exp^{-\frac{\|\mathbf{x}-\mathbf{y}(s)\|^2}{2\sigma^2}}}{\int_0^1 \exp^{-\frac{\|\mathbf{x}-\mathbf{y}(k)\|^2}{2\sigma^2}}dk+\frac{2\pi\sigma^2 w}{(1-w)N}}d\mathbf{x} + \alpha\frac{\partial^2\mathbf{y}(s)}{\partial s^2}+\beta\frac{\partial^4\mathbf{y}(s)}{\partial s^4}=0 \quad (12)$$

Discretizing the functions $F(\mathbf{x})$ and $\mathbf{y}(s)$, we can rewrite Eq. 12 as a non-linear system of equations:

$$-\frac{1}{\sigma^2}\left(\mathbf{P}\,\mathrm{diag}\,(\mathbf{F})\,\mathbf{X}-\mathrm{diag}\,(\mathbf{PF})\,\mathbf{Y}\right)+\mathbf{AY}=0 \quad (13)$$

where $\mathbf{Y}_{M\times 2}=(\mathbf{y}_1,\ldots,\mathbf{y}_M)^T$ is the $\mathbf{y}(s)$ function discretized at M points. $\mathbf{X}_{N\times 2}=(\mathbf{x}_1,\ldots,\mathbf{x}_N)^T$ are the coordinates of image pixels. $\mathbf{F}_{N\times 1}$ denotes a column vector of the probability-map intensity values. $\mathbf{A}_{M\times M}$ is a pentadioganal matrix of discrete derivative approximation, the $\mathrm{diag}\,()$ notation defines a square diagonal matrix formed by the vector in parentheses. Matrix $\mathbf{P}_{M\times N}$ is defined as

$$p_{mn}=\frac{\exp^{-\frac{\|\mathbf{x}_n-\mathbf{y}_m\|^2}{2\sigma^2}}}{\sum_{k=1}^{M}\exp^{-\frac{\|\mathbf{x}_n-\mathbf{y}_k\|^2}{2\sigma^2}}+\frac{2\pi\sigma^2 w}{(1-w)N}} \quad (14)$$

To solve the non-linear system in Eq. 13, we use a fixed point optimization by iteratively solving the linear system:

$$\left(\mathrm{diag}\,(\mathbf{P}_t\mathbf{F})+\sigma^2\mathbf{A}\right)\mathbf{Y}_{t+1}=\mathbf{P}_t\,\mathrm{diag}\,(\mathbf{F})\,\mathbf{X} \quad (15)$$

3.2. Deterministic annealing

The value of σ serves as a capture range for the active contour. Smaller σ indicates more localized capture range. We use deterministic annealing for σ: We start from a large value of σ and gradually reduce it, tracking the local minimum of E. This way we hope to achieve the global minimum or to be very close to it at a small scale.

We use half of the image size as the initial σ, for which, the contour converges to the only optimum of its energy function: the mean of the probability-map. We use 0.95 annealing rate, so that the annealing process is slow enough for the algorithm to be robust. The final value of σ is a half a pixel (0.5).

3.3. Relation to conventional external energy

In the limit, when σ approaches zero, the KL divergence becomes equivalent to conventional snakes external energy term.

$$\lim_{\sigma\to 0}E_{fit}(y)=\lim_{\sigma\to 0}\int_{\mathcal{R}^2}F(\mathbf{x})\log\frac{F(\mathbf{x})}{p(\mathbf{x})}d\mathbf{x}\sim$$
$$\lim_{\sigma\to 0}-\int F(\mathbf{x})\log(\frac{(1-w)}{2\pi\sigma^2}\int_0^1 e^{-\frac{\|\mathbf{x}-\mathbf{y}(s)\|^2}{2\sigma^2}}ds+\frac{w}{N})d\mathbf{x}=$$
$$\lim_{\sigma\to 0}-\int F(\mathbf{x})\log((1-w)\int_0^1\delta(\|\mathbf{x}-\mathbf{y}(s)\|)ds+\frac{w}{N})d\mathbf{x}=$$
$$-(\int_{\mathbf{x}=\mathbf{y}(s)}F(\mathbf{x})\log(1-w+\frac{w}{N})d\mathbf{x}+\int_{\mathbf{x}\neq\mathbf{y}(s)}F(\mathbf{x})\log\frac{w}{N}d\mathbf{x})=$$
$$-(\int_{\mathbf{x}=\mathbf{y}(s)}F(\mathbf{x})\log((1-w+\frac{w}{N})\frac{N}{w})d\mathbf{x}+\int F(\mathbf{x})\log\frac{w}{N}d\mathbf{x})\sim$$
$$-\int_0^1 F(\mathbf{y}(s))\log\frac{N(1-w)+w}{w}ds\propto-\int_0^1 F(\mathbf{y}(s))ds \quad (16)$$

Here we used the fact that Gaussian approaches delta function when its width approaches zero. Thus, the conventional external energy term $-\int F(\mathbf{y}(s))ds$ can be seen a special case of the distance between two distributions $p(\mathbf{x})$ and $F(\mathbf{x})$.

3.4. Computational complexity

The computational bottleneck of our method is in solving the linear system of equations (Eq. 15), which is in general $\mathcal{O}(M^3)$, and in updating the matrix $\mathbf{P}$, which is $\mathcal{O}(MN)$.

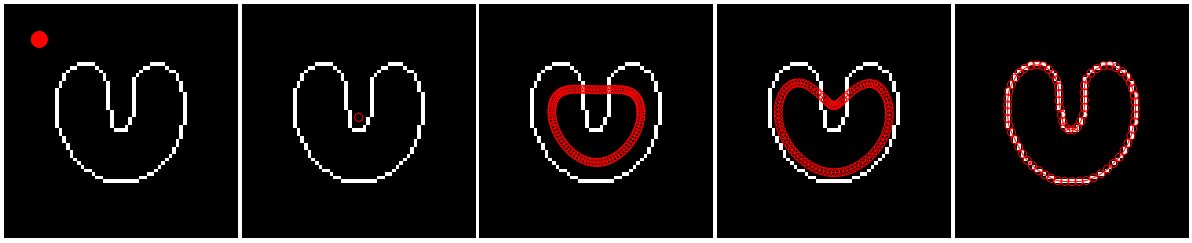
To accelerate the linear system solution, we note that the left-hand side matrix in Eq. 15 is sparse, banded and symmetric. We use a linear conjugate gradient (CG) method [17], which we can initialize from the $\mathbf{Y}$ value of the previous iteration (which will be close to the solution) rather than solving each system anew. CG requires $\mathcal{O}(M^2)$ to solve a sparse banded system.

To accelerate the $\mathbf{P}$ matrix computation, we can truncate the Gaussian kernel so that p_{mn} are nonzero only for the k nearest neighbors (obtained for free from the pixel grid), which is $\mathcal{O}(kM)$; or we can use the fast Gauss transform [9], which is $\mathcal{O}(2(N+M))$.

4. Results

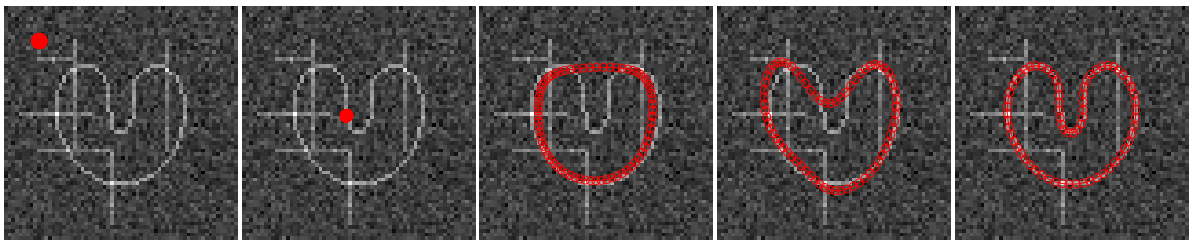
We implemented the algorithm in Matlab, and tested it on a Pentium4 CPU 3GHz with 4GB RAM. We show the method's performance on several toy and real images with average size of 150×150 pixels. For all experiments, we used α and β parameters equal one, unless explicitly stated otherwise. We set the weight w of the uniform distribution term equal 0.5. The algorithm requires about 50 iterations to converge to the tolerance 10^{-5}, for a single value of σ. Overall, the computational time of the method is about 1 minute.

[1] Strictly speaking, the Euler-Lagrange differential equation requires a standard form of the functional, which we do not have. However, we still can take the functional derivative and equate it to zero.



Initialization Iteration 1 Iteration 10 Iteration 20 Iteration 40

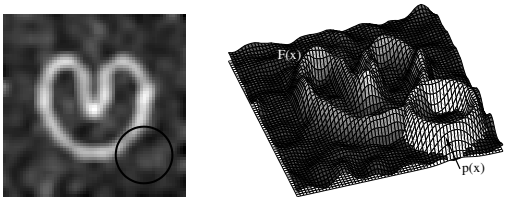
Figure 1. U-shaped object segmentation. The first figure shows the image and the active contour (red circles) initialization at the upper left corner. We use the image itself as a probability-map $F(\mathbf{x})$. At the first iteration, the active contour goes to the mean of the probability-map. As σ decreases, the active contour enlarges and segments the probability-map. This example shows the accurate method performance in presence of cavities and outside initialization without any image smoothing or preprocessing.



Initialization Iteration 1 Iteration 10 Iteration 20 Iteration 40

Figure 2. Corrupted U-shaped object segmentation. We corrupt the image by Gaussian noise and add several random lines. We use the image itself as a probability-map $F(\mathbf{x})$. Consecutive figures show the convergence progress of the method (please view this figure in color).

Synthetic examples. We start from the famous U-shape example [20]. The U-shaped image is a line-drawing, thus we use the image itself as a probability-map $F(\mathbf{x})$. Figure 3 shows an example of the contour initialization, and the corresponding 3D view of $p(\mathbf{x})$ and $F(\mathbf{x})$. The conventional and Gradient Vector Flow (GVF) snakes [20] have to be initialized around the object (or inside) in order to succeed. In our experiment, we initialize the active contour far from the object (a small red circle in the left upper corner in Fig. 1). Figure 1 demonstrates the accurate method performance without any extra image smoothing or preprocessing,



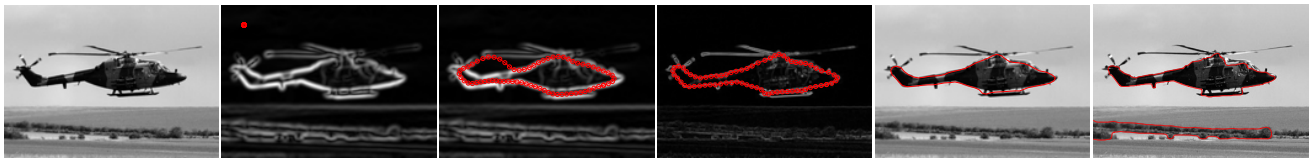
(a) (b)

Figure 3. (a) Probability-map $F(\mathbf{x})$ and the location of the mean curve $y(\mathbf{s})$ (black circle). (b) Both $F(\mathbf{x})$ and $p(\mathbf{x})$ in the the 3D view. We align $p(\mathbf{x})$ to $F(\mathbf{x})$ by minimizing the KL divergence.

even though it is initialized far on a side of the object.

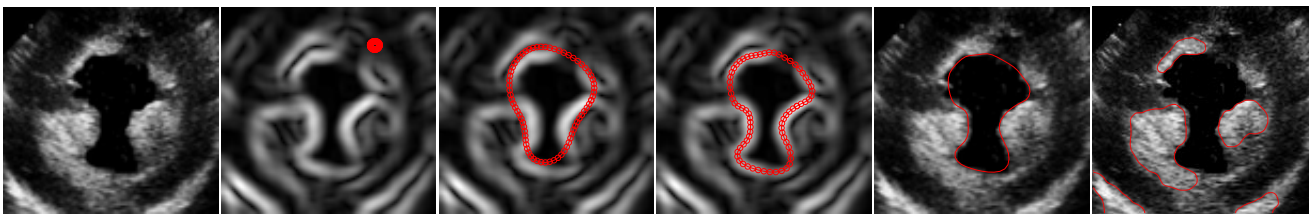
We corrupted the U-shaped image with noise and random lines to make the segmentation task more challenging (Figure 2). We initialized the active contour far from the object. First, the active contour converges to the mean of the probability-map. As we anneal on σ, the active contour enlarges to approximate the probability-map for a given scale σ. The uniform distribution (second term in Equation 7) accounts for all regions not described by the active contour distribution. As σ decreases the active contour achieves the optimum of its energy functional when it lies at the highest non-overlapping areas of the probability-map and the curve produced by the active contour is short and smooth. The active contour does get attracted to the strait lines, but the shape regularization constraints forces to choose the smoothest and the shortest match.

Real-world examples. We tested the method on several real images. For all of them we were interested in the object edges, thus we computed the probability-map, $F(\mathbf{x})$, as the norm of the image gradient preprocessed by the Gaussian filter. We again initialized the active contour at an arbitrary location far from the edges. We compare our method to that of fast global minimization of the active contour model (FGMAM) [3]. Figure 4 shows the helicopter segmenta-



Input image Initialization Iteration 20 Iteration 40 Our result FGMAM result

Figure 4. The helicopter segmentation. The first figure shows the input image. The second figure shows the probability-map $F(\mathbf{x})$ and the active contour initialization (far from the object). As we anneal on σ, the active contour adapts to the high intensity region of the probability-map (please view this figure in color).



Input image Initialization Iteration 20 Iteration 40 Our result FGMAM result

Figure 5. Ultrasound image segmentation. The first figure shows the input image with short-axis view of the left ventricle. The second figure shows the probability-map $F(\mathbf{x})$ and the initial location of the active contour. Despite many high intensity regions of the probability-map, the method produces an accurate segmentation result (please view this figure in color).

tion result. As σ decreases, the active contour adapts to the probability-map and shows the accurate segmentation. We note, that other image detail, including the edges from the ground level, also attract the active contour, but the desired helicopter shape wins, because of the higher probability and smoother form. The FGMAM method produced satisfactory helicopter segmentation, which, however, includes part of the ground level objects.

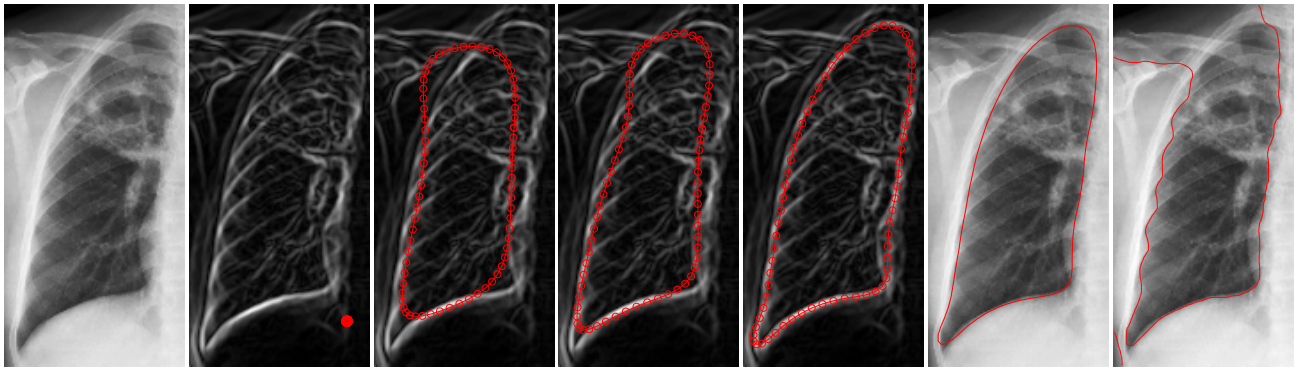
Figure 5 shows the ultrasound image of the left ventricle short-axis view. Here, the probability-map has many high probability regions, which attract the active contour. However the shortest and the smoothest path corresponds to the wall boundary, which makes it possible for the active contour to produce the accurate segmentation. The segmentation result is the same for various different initialization positions we have tried. The FGMAM method has not produced a reasonable segmentation result, even though different combinations of parameters have been tried. We note that FGMAM has not been designed to segment only a single closed contour, which partially explain its poor performance. Some other local-based segmentation methods succeed to segment this example only when given the reasonably close-boundary initialization [14].

Finally, Figure 6 shows the X-ray image segmentation of the lung. This is a challenging example, because many high probability regions attract the active contour. Here, we set β equal 5 to add some rigidity to the active contour. The active contour segments the lung boundary accurately, starting from an arbitrary location. Once again, the algorithm picks the smoothest, shortest path with higher probability. However, if one is interested in weaker edges (e.g. ribs), then different construction of the probability map is required, as we shall discuss in Section 6. The FGMAM method has not produced accurate segmentation result, even after heavy parameter tuning.

We note, that the conventional snake algorithms can also accurately segment the images in our experiments, only if the snake is initialized reasonably close to the boundary and on the correct side of the object. Our method does not have such limitation. The snakes' algorithms that incorporate a shape prior can overcome the initialization problem, given that the shape prior is available. In this paper, we consider general segmentation task, where no prior information is available. However, for the future work we can incorporate the shape prior in our approach as well.

5. Related work

Conventional Active Contours: Comparing to the conventional snakes, our method evaluate the fitness of the contour over the *whole image domain* at every iteration. This allows us to keep track of all possible segmentation regions simultaneously, whereas the conventional snake uses only local information, e.g. local intensity gradient, to direct its movement. From the computational viewpoint the conventional snakes requires numerical computation of the external energy gradient, ∇E_{ext} (which is errorprone), and image interpolation at the snake locations, whereas our method doesn't require any of such operations.



Input image Initialization Iteration 10 Iteration 20 Iteration 40 Our result FGMAM result

Figure 6. X-ray image segmentation. The first figure shows the input image. The second figure shows the probability-map $F(\mathbf{x})$ and the initial active contour location. The method produces an accurate segmentation result despite faraway initialization and numerous obstacles, such as the edges from the ribs and shoulder bones (please view this figure in color).

Region-based Methods: In region-based segmentation methods [5, 13, 3] the energy function depends on all image information, and thus operates on a "global"image scale. From this perspective, our method is global, because the data-driven (fitness) term depends on all image pixels. In formulation, our method resembles edge-based methods, such as conventional snake and GVF-snake. However, our fitness term depends on all image pixels, and thus the method is not simply edge-based.

Multi-scale Approach: Our deterministic annealing approach resambles the multiscale active contour algorithms, where the image is iteratively segmented at increasing scales [14]. The main difference here is two-fold: First, we maintain the overlap of $p(\mathbf{x})$ and $F(\mathbf{x})$ over the whole image domain at any scale. Second, at the fine scales of deterministic annealing, our active contour converges to the closed contour with 1's along the mean curve $\mathbf{y}(s)$ and 0's elsewhere (no noise or artifacts), whereas multiscale approach converges to the noisy image itself with irrelevant for segmentation objects.

Elastic Net: We would also like to mention the relation of our method to the elastic net approach [7, 4]. Elastic net is a Gaussian mixture model with a prior on its centroids, which regularizes the shape and defines the connectivity between the centroids, similar to the differential regularization of active contours. The elastic net fits the data points by minimizing its negative log-likelihood together with the regularization term. The log-likehood function is equivalent to the KL divergence, when the other distribution is a mixture of delta functions centered at the data points. In this sense, our active contour method can be seen as a continuous analog of the elastic net. Similarly, elastic net uses deterministic annealing approach on the width of its centroids. Applications of the elastic net mainly include multidimensional data reduction and manifold learning [4]. Elastic net was originally designed to find the shortest path through a set of points with no repetitions. From this point of view, we can formulate our active contour segmentation as a task to find a shortest and smoothest path through the high density regions of probability-map, which gives further insight into the method.

6. Discussion

Motivation to use the KL divergence. Our method can be seen as an alignment of two distributions $p(\mathbf{x})$ and $F(\mathbf{x})$ subject to the transformation parameters (mean curve position $\mathbf{y}(s)$) of $p(\mathbf{x})$). Alignment of two functions has been heavily studied in image registration community, where both functions represent images, which is similar two our formulation [10]. Many similarity measures, which are common for image registration, such as sum-of-squared differences, correlation coefficient, correlation ratio and mutual information, can be also applied to align $p(\mathbf{x})$ and $F(\mathbf{x})$ with relative level of success.

Our motivation to use KL divergence instead of other similarity measures is as follows. KL divergence, KL($F||p$), is a special case of $\alpha-$divergence for $\alpha = 1$. Following the terminology of Frey et al [8], this divergence is inclusive; it requires $p > 0$ whenever $F > 0$, thus avoiding "false negatives". This divergence tries to cover as much of F as possible, and the uncovered areas of F are heavily penalized, which forces the active contour to go over the edges if necessary [15]. It is important to emphasize that we use KL($F||p$), which is inclusive as oppose to KL($p||F$). Among other things, KL divergence can be derived from log-likelihood maximization and allows us apply similar to Expectation Maximization (EM) optimization technique.

Segmentation challenges. We also discuss the situations where the method does not work: if the disired edges appear weaker next to stronger edges or if there are several shape choices. In these cases we have to construct the probability-map $F(\mathbf{x})$ in a different way than a simple image gradient (for instance, by using more advanced preprocessing techniques including anisotropic diffusion, total-variation denoising, etc.). If $F(\mathbf{x})$ is properly constructed such that the desired segmentation region includes higher values of $F(\mathbf{x})$ and smother shape, then our method is not sensitive to initialization and will perform well. Additionally, one can add the shape prior information, however this approach is beyond the scope of this paper.

7. Conclusion

We introduce a novel method for image segmentation based on the active contour approach. We treat the active contour as a mean curve of the probability density function $p(\mathbf{x})$. It moves to minimize the Kullback-Leibler (KL) divergence between $p(\mathbf{x})$ and the probability density function derived from the image. The advantage of the KL divergence function is that it requires all regions of the probability-map (image regions) to be approximated by the model proportional to the probability-map values, which makes active contour to account for all image regions.

We use deterministic annealing on the width of $p(\mathbf{x})$ to implement a coarse-to-fine search strategy. We start from a large width σ and gradually reduce it, tracking down the optimum. Deterministic annealing does not guarantee to find the global optimum of the function, however empirically, the optimum found is within a few percent from the global optimum. In the limit, when $\sigma \rightarrow 0$, the KL divergence converges to the conventional external energy term of active contours. Thus, conventional active contours external energy term can be seen as a special case of distance function between two distribution: one derived from the image and the other represented by a delta functions along the contour.

Our method is less sensitive to noise and outliers presented in the image due its probabilistic nature and the ability of the contour to go over the boundaries. The segmentation performance will still depend on the accuracy of feature-map construction, but less heavily comparing to the conventional active contours. We test our method on several toy and real images. The method shows accurate and robust segmentation results even with initialization far away from the target boundary, which makes the segmentation procedure almost fully automated.

References

[1] J. An, M. Rousson, and C. Xu. Γ-convergence approximation to piecewise smooth medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention*, volume 4792, pages 495–502, 2007.

[2] A. Blake and M. Isard. *Active Contours*. Springer-Verlag New York, Inc., 1998.

[3] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2):151–167, June 2007.

[4] M. Carreira-Perpiñán, P. Dayan, and G. J. Goodhill. Differential priors for elastic nets. In *International Conference on Intelligent Data Engineering and Automated Learning (IDEAL'05)*, volume 3578, pages 335–342, 2005.

[5] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2), 2001.

[6] D. Cremers, F. R. Schmidt, and F. Barthel. Shape priors in variational image segmentation: Convexity, lipschitz continuity and globally optimal solutions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6, June 2008.

[7] R. Durbin, R. Szeliski, and A. Yuille. An analysis of the elastic net approach to the travelling salesman problem. *Neural Computation*, 1:348–358, 1989.

[8] B. J. Frey, R. Patrascu, T. S. Jaakkola, and J. Moran. Sequentially fitting inclusive trees for inference in noisy-or networks. In *Advances in Neural Information Processing Systems (NIPS)*, 2000.

[9] L. Greengard and J. Strain. The fast Gauss transform. *SIAM J. Sci. Statist. Comput.*, 12(1):79–94, Jan. 1991.

[10] D. L. G. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Physics in medicine and biology*, 46(3):R1–R45, Mar. 2001.

[11] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, Jan. 1988.

[12] M. E. Leventon, W. Eric, L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *International Conference on Computer Vision*, volume 1, pages 317–323.

[13] H. Li and A. Yezzi. Local or global minima: Flexible dual-front active contours. *PAMI*, 29(1):1–14, 2007.

[14] N. Lin, W. Yu, and J. S. Duncan. Combinative multi-scale level set framework for echocardiographic image segmentation. *Medical Image Analysis*, 7:529–537, 2003.

[15] T. Minka. Divergence measures and message passing. Technical Report MSR-TR-2005-173, Microsoft Rsrch, 2005.

[16] D. Mumford and J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math*, 42:577–685, 1989.

[17] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer-Verlag, New York, 1999.

[18] A. Singh, D. Terzopoulos, and D. B. Goldgof. *Deformable Models in Medical Image Analysis*. IEEE Press, 1998.

[19] D. Terzopoulos and R. Szeliski. Tracking with kalman snakes. In *Active Vision*, pages 3–20. MIT press, 1992.

[20] C. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Trans. Image Processing*, 7(3):359–369, 1998.