Local or Global Minima: Flexible Dual-Front Active Contours

Hua Li[1,2] and Anthony Yezzi[1]

[1] School of ECE, Georgia Institute of Technology, Atlanta, GA, USA
[2] Dept. of Elect. & Info. Eng., Huazhong Univ. of Sci. & Tech., Wuhan, China
hua.li@ece.gatech.edu

Abstract. Most variational active contour models are designed to find the "desirable" local minima of data-dependent energy functionals with the hope of avoiding undesirable configurations due to noise or complex image structure. As such, there has been much research into the design of complex region-based energy functionals that are less likely to yield undesirable local minima. Unfortunately, most of these more "robust" region-based energy functionals are applicable to a much narrower class of imagery due to stronger assumptions about the underlying image data. Devising new implementation algorithms for active contours that attempt to capture more global minimizers of already proposed image-based energies would allow us to choose an energy that makes sense for a particular class of energy without concern over its sensitivity to local minima. However, sometimes the completely-global minimum is just as undesirable as a minimum that is too local.

In this paper, we propose a novel, fast and flexible dual front implementation of active contours, motivated by minimal path techniques and utilizing fast sweeping algorithms, which is easily manipulated to yield minima with variable "degrees" of localness and globalness. The ability to gracefully move from capturing minima that are more local (according to the initial placement of the active contour/surface) to minima that are more global makes it much easier to obtain "desirable" minimizers (which often are neither the most local nor the most global). As the examples, we illustrate the 2D and 3D implementations of this dual-front active contour for image segmentation from MRI imagery.

1 Introduction

Since the introduction of snakes [1], active contours have become particularly popular for segmentation applications. Most variational active contour models [2,3,4,5] are designed to find local minima of data-dependent energy functionals with the hope that reasonable initial placement of the active contour will drive it towards a "desirable" local minimum rather than an undesirable configuration that can occur due to the noise or complex image structure.

As such, there has been much research [6,7,8,9,10,11,12] into the design of complex region-based energy functionals that are less likely to yield undesirable local minima when compared to simpler edge-based energy functionals whose sensitivity to noise and texture is significantly worse. Unfortunately,

Y. Liu, T. Jiang, and C. Zhang (Eds.): CVBIA 2005, LNCS 3765, pp. 356–366, 2005.

most of these more "robust" region-based energy functionals are applicable to a much narrower class of imagery compared to typical edge-based energies due to stronger assumptions about the underlying image data.

Devising new implementation algorithms for active contours that attempt to capture more global minimizers of already proposed image-based energies would allow us to choose an energy that makes sense for a particular class of energy without concern over its sensitivity to local minima. The minimal path technique proposed by Cohen et al. [13,14] is one such implementation. It attempts to capture the global minimum of an active contour model's energy between two points. However, for this minimal path technique, the initial points should be located exactly on the boundary to be extracted. Also, a topology-based saddle search routine is needed when they extended this technique to closed curve extraction. And it is not easy to expand to general 3D case [15].

Although many researchers keep their efforts on the design of robust active contour models to find the global minima and avoid the local minma, sometimes the completely global minimum is just as undesirable as a minimum that is too local. In this paper, we propose a novel, fast and flexible dual front implementation of active contours, motivated by minimal path technique and utilizing fast sweeping algorithms. In this model, the segmentation objective is achieved by iteratively dilating the initial curve to form a narrow region and then finding the new closest potential weighted minimal partition curve inside.

This dual-front active contour is easily manipulated to yield minima with variable "degrees" of localness and globalness. The degree of global or local minima can be controlled in a graceful manner by adjusting the width of the dilated narrow region. This ability to gracefully move from capturing minima that are more local (according to the initial placement of the active contour/surface) to minima that are more global makes it much easier to obtain "desirable" minimizers (which often are neither the most local nor the most global). This model guarantees the continuity and smoothness of the evolving curve with the capability to handle topology changes. In addition, it is easy to extend to the 3D case.

2 Dual-Front Active Contours

2.1 Background – The Minimal Path Technique

Given a potential $P > 0$ that takes lower values near desired boundaries, for example, $P = 1/(1 + ||\nabla I||^2)$, the objective of the minimal path method [13,14] is to look for a path (connect the pre-defined two points) along which the integral of $\widetilde{P} = P + w$ (w is the constant) is minimal. First, the minimal action map $U_0(p)$ is defined as the minimal energy integrated along a path between the starting point p_0 and any point p, which is

$$U_0(p) = \inf_{A_{p_0,p}} \{E(C)\} = \inf_{A_{p_0,p}} \{\int_\Omega \widetilde{P}(C(s))ds\}, \tag{1}$$

where $A_{p_0,p}$ is defined as the set of all paths between p_0 and p. Then, if given the minimal action maps U_0 to p_0 and U_1 to p_1, the minimal path between p_0 and p_1 is exactly the set of points p_g which satisfy

$$U_0(p_g) + U_1(p_g) = \inf_p \{U_0(p) + U_1(p)\}, \tag{2}$$

and this minimal path between p_0 and p_1 is determined by calculating U_0 and U_1 and then sliding back from the saddle point p', which is the first point that two action maps U_0 and U_1 meet each other, on the action map U_0 to p_0 and on the action map U_1 to p_1 according to the gradient descent.

Because the action map U_0 has only one minimum value at the starting point p_0 and increases from the starting point outwards, it can be easily determined by solving the Eikonal Eq. (3) using fast marching algorithm introduced by Sethian et al. [16]. The detailed explanation is shown in [13].

$$||\nabla U_0|| = \widetilde{P} \qquad \text{with} \qquad U_0(p_0) = 0 \tag{3}$$

2.2 Principle of Dual-Front Active Contours

Now we suppose that the image has two regions R_0 and R_1, and we choose one point p_0 inside the region R_0 and another point p_1 inside the region R_1. We still define two minimal action maps $U_0(p)$ and $U_1(p)$ according to the same definition as that in minimal path theory. The potential is also decided by the image features, for example, the potential takes lower values on the boundary of R_0 and R_1.

In the minimal path theory, the points satisfying the Eq. 2 are considered. Contrary to that, we consider the points satisfying the equation $U_0(p) = U_1(p)$. At these points, the level sets of the minimal action map U_0 meets the level sets of the minimal action map U_1. These meeting points form the Voronoi diagram of the image, decompose the whole image into two regions containing the point p_0 and the point p_1 respectively. One region containing the point p_0 is called as region R'_0, and the other region containing the point p_1 is called as region R'_1. All the points in the region R'_0 is closer to p_0 than p_1 in terms of the action map. All the points in the region R'_1 is closer to p_1 than p_0 in terms of the action map. Because the action maps are defined as the potential weighted distance maps, the boundary of regions R'_0 and R'_1 is called as the potential weighted minimal region partition related to the two points p_0 and p_1.

Actually the level sets of the action map U give the evolution of the front. The velocity of the evolving front is decided by the potential. When the evolving front arrives the boundary, the velocity is much lower, and the evolving front almost stops at the boundary. The same situation are for the action maps U_0 and U_1. When choosing the appropriate potentials for calculating the two minimal action maps, it is possible that these two action maps will meet each other at the actual boundary of two regions R_0 and R_1. In other words, we can search the two regions' boundary by calculating the minimal region partition related to the two points inside the two regions respectively.

Without loss of generality, we suppose X be a set of points (for example, a 2D curve or a 3D surface) in the image, U_X is the minimal action with potential $\widetilde{P}_X$ and starting points $\{p, p \in X\}$. Clearly, $U_X = min_{p \in X} U_p$. Then considering two sets $U_{X_i} = min_{p \in X_i} U_p$ and $U_{X_j} = min_{p \in X_j} U_p$, all points satisfying $U_{X_i}(p) = U_{X_j}(p)$ form the boundary of the two regions related to the two point sets X_i and X_j. Because the two action maps are the potential weighted distance maps, this formed boundary is also the potential weighted minimal partition of the regions related to the two point sets X_i and X_j.

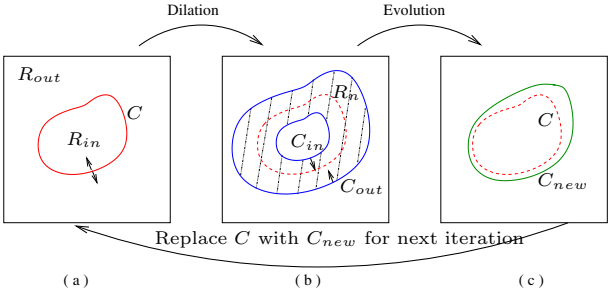


Fig. 1. Iteration process of dual front evolution and dilation. (a): an original contour C separating the region R to regions R_{in} and R_{out}; (b): the curve C is dilated to form the narrow active region R_n; (c): the inner and outer boundaries C_{in} and C_{out} propagate to form the new minimal partition curve C_{new} separating the region R to two regions, and then the curve C is replaced by the curve C_{new} for processing the next iteration.

Therefore, we propose the dual front evolution based on the above analysis to find the potential weighted minimum partition for a defined active region. The evolution principle is shown in Fig. 1. The narrow active region R_n, which is formed by expending the initial curve C, has the inner boundary C_{in} and the outer boundary C_{out}. Then the minimal action maps U_{in} and U_{out} are calculated with different potentials $\widetilde{P}_{in}$ and $\widetilde{P}_{out}$ respectively. When these two action maps meet each other, both evolutions of the level sets of the action maps stop automatically and a minimal partition boundary is formed in region R_n. All points p_g on this minimal partition boundary satisfy the following Eq. (4):

$$\begin{cases} |\nabla U_{in}| = \widetilde{P}_{in} & with\ U_{in}(C_{in}) = 0 \\ |\nabla U_{out}| = \widetilde{P}_{out} & with\ U_{out}(C_{out}) = 0 \\ U_{in}(p_g) = U_{out}(p_g) \end{cases} \tag{4}$$

The dual front evolution is implemented by labeling the initial curves with different labels, than evolving the labeled curve with different potentials to the unlabeled region until all the points are assigned an unique label. Dual front evolution provides us a method to find the minimal partition curve within a narrow active region. Here this minimal partition is the potential weighted global minima partition only inside the narrow active region, not in the whole image.

Clearly, the degree of this globalness may be changed flexibly according to the size of the narrow active region.

In the dual front evolution, the region-based information and the edge-based information may be unified in the potentials for guiding the curve evolution. The mean values u_{in}, u_{out}, the variance values σ_{in}^2, σ_{out}^2, of the inside region ($R_{in} - R_{in} \cap R_n$) and outside region ($R_{out} - R_{out} \cap R_n$) are calculated. The evolution speeds (or potentials) for the labeled points (x, y) are decided by Eq. (5).

$$\begin{aligned} \widetilde{P}_{in}(x,y) &= w_{in}^r \times f(|I(x,y) - \mu_{in}|, \sigma_{in}^2) + w_{in}^b \times g(\nabla I(x,y)) + w_{in} \\ \widetilde{P}_{out}(x,y) &= w_{out}^r \times f(|I(x,y) - \mu_{out}|, \sigma_{out}^2) + w_{out}^b \times g(\nabla I(x,y)) + w_{out} \end{aligned} \quad (5)$$

where $I(x, y)$ is the value of the image intensity at the examined point. $g(\nabla I(x, y))$ is a function of the image gradient. By choosing different functions f and g, and the different weight for each component of the potentials, this model can be used for different segmentation objectives.

Based on this dual front evolution, we propose the dual-front active contour model. It is an iterative process including dual front evolution and morphological dilation. First, we choose an initial curve, dilate it to form the narrow active region, and use dual-front evolution to find the minimal partition within this active region. Then, we expend the obtained minimal partition curve to form a new narrow active region, and the new potentials for the boundaries of the new active region are also calculated. We repeat this process until the difference between the consecutive obtained minimal partition curves less than the predefined threshold.

So, in dual-front active contours, the segmentation objective to find the minima with variable "degrees" globalness in defined region is transferred to find the global minimum partition curve within a narrow active region expanded from the initial contour, and then iteratively replace the current contour with the obtained global minimum partition curve until the final segmentation objective is achieved.

3 The Properties of Dual-Front Active Contours

3.1 Flexible Local or Global Minima

In dual-front active contours, the degree of global or local minima can be controlled in a graceful manner by adjusting the width of the narrow active region for the dual fronts' evolution. This ability to gracefully move from capturing minima that are more local (according to the initial placement of the active contour/surface) to minima that are more global makes this model much easier to obtain "desirable" minimizers (which often are neither the most local nor the most global).

The result of the dual front evolution is the global minimal partition curve inside the active region. So the size and the shape of the narrow active region will affect the final segmentation result. If the size of the active region is small, it possible leads to the problem of local minima because of the local noise. But if the

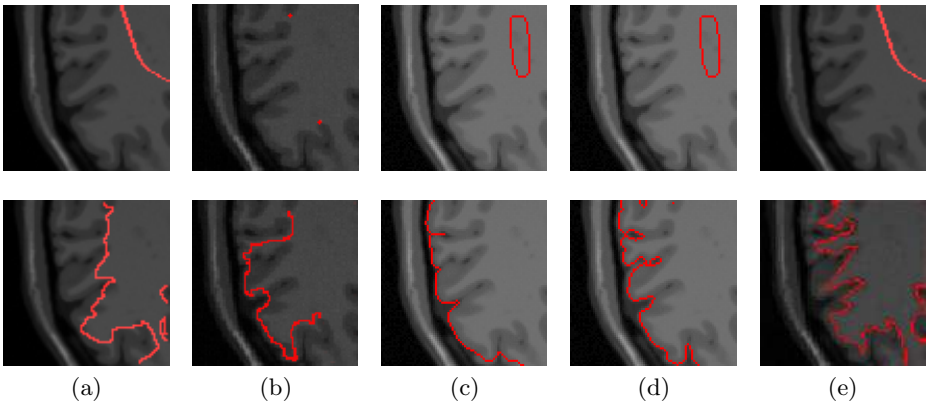


Fig. 2. Comparison of the different segmentation results of the interface of white matter/gray matter from different active contour models with different degrees of local minima and global minima. The gradient information used by panel (a),(b), and (e) is shown in Fig. 3. The top row shows the original image and the initialization for the curve evolution, and the bottom row shows the corresponding edge segmentation results from geodesic active contours (a), the minimal path technique (b), Chan-vese's method (c), Mumford-Shah's method (d) and dual-front active contours (e).

size of the active region is too large, the actual object boundary will be missed. The size of the active region should be selected based on the shape and the size of the detected object, the image quality and the background information, and so on.

In dual-front active contours, we provide very flexible method to define the active region. In fact, the active region is a kind of restricted searching space. The restricted space can be formed by choosing the automatic thresholding, calculating the distance map, using the length of the initial contour's normal, performing the morphological dilation and so on. All these methods can be used for defining the active region. Normally, we use the morphological dilation to obtain the narrow active region because the size of the active region can be controlled easily by adjusting the size of the structure element and the dilation times for the requirement from a given segmentation, or class of images.

The size of active region also can be changed during the evolution process. For example, when the initial curve is far from the object, we may first use the wider active region to expend the searching scale for one iteration, speed up the computation time and avoid the effect of the noise. When the curve, which is obtained after a number of iterations, near the object boundary, we may use the narrow active region for refining the accurate boundary. By the way, if the detected object is bigger, we may use wider active region, otherwise, we should use narrower active region.

In Fig. 2, we compare the different edge detection results by two of the edge-based methods, geodesic active contours [2] and the minimal path technique [13],

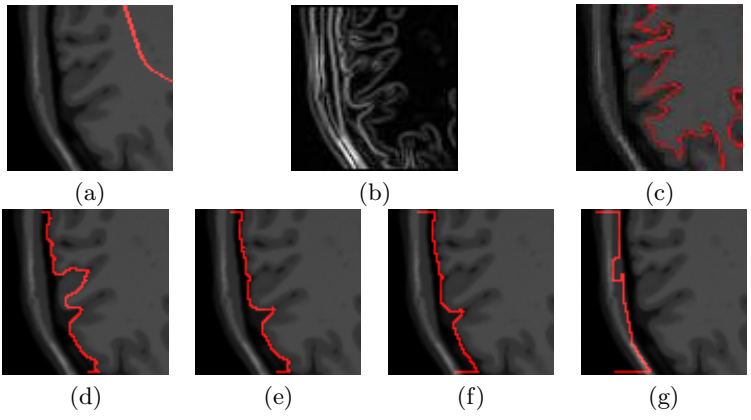


Fig. 3. By choosing the different size of the narrow active region, the dual-front active contour model achieves different minima.
(a): the original image with the initialization; (b): the corresponding gradient information; (c)-(g): the segmentation result using 5×5, 7×7, 11×11, 15×15, 23×23 pixels circle structuring elements for morphological dilation after 15 iterations.

two of the region-based methods, Chan-Vese's method [8] and Mumford-Shah methods [17], and the dual-front active contours proposed in this paper. This figure is the part of one 2D human brain MRI image, and the segmentation objective is to find the interface of the gray matter and the white matter. We use the above five methods to process this image and obtain the different segmentation result. We can see that, geodesic active contours suffer from undesirable local minima, and the "global minima" found by the minimal path technique is also not exactly what we want, which is effected by the location of the pre-defined two points. Chan-Vese's method and Mumford-Shah's method also found the incorrect global minima. As this figure indicates, our dual-front active contours can control the degree of global or local minima in active contour model, find the correct boundary, and perform better than other methods which find the local minima or global minima.

In Fig. 3, we give another example to demonstrate that, by choosing different narrow active regions with different sizes, the dual-front active contour model achieves different degree's global minima in the whole image. The potential for each point is $\widetilde{P}(x, y) = 1/(|I(x, y) - I_{mean}| + (1 + \nabla I)^2/10) + 0.1$.

3.2 Other Nice Properties

First, the dual front evolution provides the automatic stop criterion in each iteration. The dual front evolution also guarantees the continuity and smoothness of the curve with the capability to handle topology changes. Second, the dual-front active contour model provides automatic stop criterion by comparing the result from consecutive iterations. Third, the dual-front evolution combines the advan-

tages of level-set methods and fast marching methods, avoids the disadvantages of them, and transfers the point-to-point evolution to noncontinuous band-to-band evolution. In this manner, the computational cost is reduced significantly. The detailed information of these properties was shown in [18].

4 Experimental Results

In Fig. 4, we give two other examples to compare the different edge detection results by Chan and Vese's method [8], the Mumford shah algorithm [17], and dual-front active contours. Because Chan's method and Mumford's method are all designed for finding the global minima in the whole image, sometimes, they cannot receive the correct boundary. But for dual-front active contour, the degree of the global minima can be controlled by the size of narrow active region, the model can achieve flexible global degree's minima. In these two examples, the potential for each point is $\widetilde{P}(x,y) = 1/(|I(x,y) - I_{mean}|) + 0.1$.

In Fig. 5, we show that our model can be used for extract the object without clearly gradient information. The panel (a) is a noisy mammogram showing a cyst in the breast, the panel (b) is the corresponding gradient image, the panel (c) shows the middle step of the segmentation process, the panel (d) shows the final segmentation result. In this example, the potential for each point is $\widetilde{P}(x,y) = 1/(|I(x,y) - I_{mean}|) + 0.1$. The structuring element is 5×5 pixels circle for morphological dilation. It is clear from the results that the segmentation of the cyst is refined even with high noise level.

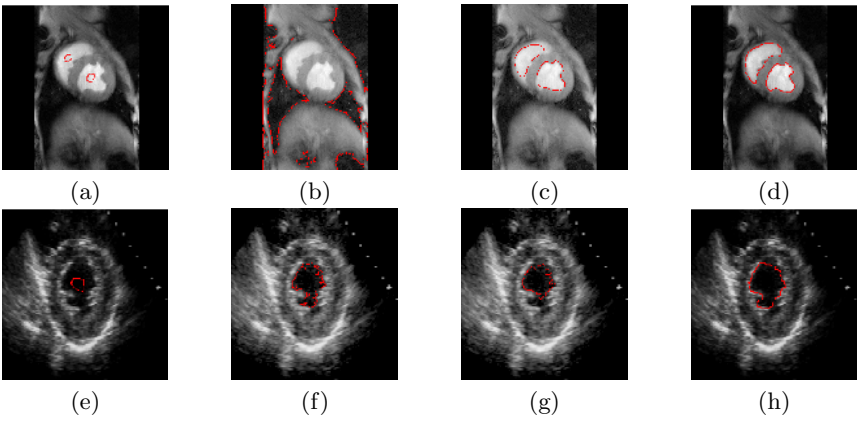
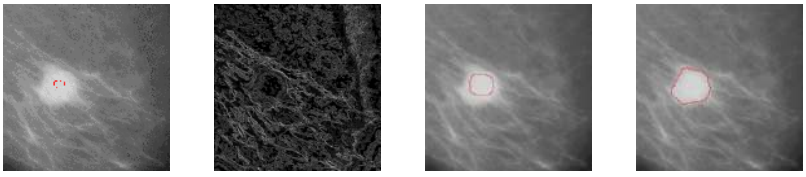


Fig. 4. Comparison of different region-based active contours related to the degree of global minima.
(a) and (e): two 2D medical images with the initializations; (b) and (f): the results from Chan-Vese's model suffer from undesirable global minima; (c) and (g): the results from Mumford-Shah model also suffer from the smoothing constraints; (d) and (h): the correct edge extractions from dual-front active contours using 7×7 pixels circle structuring element for morphological dilation.

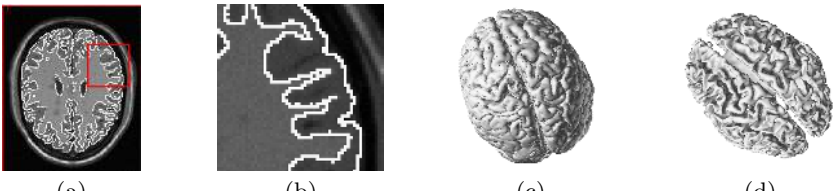


(a) (b) (c) (d)

Fig. 5. The segmentation result on 2D cyst image without gradient informations. The size of the dilation structure element is 5×5 pixels. Panel (a) shows the original image with the initialization. Panel (b) shows the gradient image. Panel (c) shows the middle step of the segmentation process after 5 iterations. Panel (d) shows the segmentation result after 15 iterations.

We also test the dual-front active contours on the simulated MRI 3D brain image data set and extract the interface of gray matter/white matter, as well as applications to specific cortical studies.

Because of the properties of dual-front active contours, the whole segmentation process is considered as a hierarchical decomposition. We assume that the normal brain includes three tissues: GM (gray matter), WM (white matter), and CSF (cerebral spinal fluid). After skull stripping and non-brain tissue removing, we separate the brain region and the background first. Then, we just consider the brain region and use dual-front active contours to segment the brain into CSF and WM+GM two regions. The third step is to restrict the WM+GM region and use dual-front active contours again to separate the WM region and GM region.



(a) (b) (c) (d)

Fig. 6. The segmented outer and inner cortical surfaces from the MRI brain image with our method

In Fig. 6, we present the segmented outer (CSF-GM interface) and inner (GM-WM interface) cortical surfaces in one slice of the 3D simulated brain image, and a zoom-in of extracted boundaries for this slice. We also show the 3D models of the cortical surfaces. The test image is available from the BrainWeb [19], which is generated from the MS Lesion brain database using the $T1$ modality, $1mm$ slice thickness, 3% noise level and 20% intensity nonuniformity settings. The image size is $181 \times 217 \times 217$. The initialization for the hierarchical

segmentation is a sphere centered at $(100, 100, 95)$, and the size is $75 \times 75 \times 150$. The potential for different point is $\widetilde{P}(x, y) = 1/(|I(x, y) - I_{mean}|) + 0.1$, and the size of the dilation structure element is $5 \times 5 \times 5$ pixels.

For a 3D brain image ($181 \times 217 \times 217$ voxels in total), Normally, our method requires only 10-20 iterations for segmenting one tissue type (CSF, GM or WM). For each iteration, the computation procedure includes one curve dilation and one dual-front evolution, one iteration last around 15 seconds.

5 Conclusions and Future Work

In this paper, we propose a novel, fast and flexible dual front implementation of active contours, which is easily manipulated to yield minima with variable "degrees" of localness and globalness. This ability to gracefully move from capturing minima that are more local (according to the initial placement of the active contour/surface) to minima that are more global makes it much easier to obtain "desirable" minimizers (which often are neither the most local nor the most global). As the examples, we illustrate the 2D and 3D implementations of this model for object extraction from MRI imagery.

While the underlying principle of the dual front evolution algorithm presented here is based on the authors' earlier work in [18], there are several novelties in the present work which do not appear in the earlier work. Other than the extension to the three dimensions, we have made the very important observation that the dual-front approach may be customed tailored to capture minimizers that are flexible in their degrees of localness and globalness.

As such, we have constructed around this basic building block and algorithm that may be controlled and adapted in ways that other active contour models cannot. This key point, which is not addressed at all in the earlier work [18] greatly extends the usefulness of their model to many important applications in computer vision, especially medical imaging, where user control and interaction is highly desirable.

Furthermore, the 3D algorithm presented in this paper is also quite novel in that is not a mere extension of the 2D algorithm presented in [18]. In fact, the hierarchical decomposition procedure used in our 3D algorithm could also be incorporated to improve even the original 2D algorithm. We believe that this analysis and interpretation of the original algorithm will be of great service to the computer vision community.

References

1. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision **1** (1988) 321–332
2. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. International Journal of Computer Vision **22** (1997) 61–79
3. Yezzi, A., Kichenassamy, S.: A geometric snake model for segmentation of medical imagery. IEEE Trans. on Medical Imaging **16** (1997) 199–209

4. Cohen, L., et al.: On active contour models and balloons. Computer Vision Graphics Image Processing: Image Understanding **53** (1991) 211–218
5. Tek, H., Kimia, B.: Image segmentation by reaction diffusion bubbles. In: Proceedings of International Conference of Computer Vision. (1995) 156–162
6. Zhu, S., Yuille, A.: Region competition: unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. IEEE Trans. on PAMI **18** (1996) 884–900
7. Chakraborty, A., Staib, L., Duncan, J.: Deformable boundary finding in medical images by integrating gradient and region information. IEEE Trans. on Medical Imaging **15** (1996) 859–870
8. Chan, T., Vese, L.: Active contours without edges. IEEE Trans. on Image Processing **10** (2001) 266–277
9. Paragios, N., Deriche, R.: Geodesic active regions: A new framework to deal with frame partition problems in computer vision. Journal of Visual Communication and Image Presentation **13** (2002) 249–268
10. Samson, C., Blanc-Feraud, L., et al.: A level set method for image classification. In: Int. Conf. Scale-Space Theories in Computer Vision. (1999) 306–317
11. Yezzi, A., Tsai, A., Willsky, A.: A statistical approach to image segmentation for bimodal and trimodal imagery. In: Proceedings of ICCV. Volume 2. (1999) 1–5
12. Yezzi, A., Tsai, A., Willsky, A.: A fully global approach to image segmentation via coupled curve evolution equations. Journal of Visual Communication and Image Representation **13** (2002) 195–216
13. Cohen, L., Kimmel, R.: Global minimum for active contour models: A minimal path approach. In: IEEE International Conference on CVPR (CVPR'96). (1996) 666–673
14. Cohen, L.: Multiple contour finding and perceptual grouping using minimal paths. Journal of Mathematical Imaging and Vision **14** (2001) 225–236
15. Ardon, R., Cohen, L.: Fast constrained surface extraction by minimal paths. In: International Conference on Computer Vision-Workshop on VLSM. (2003) 233–244
16. Sethian, J.: Fast marching methods. SIAM Review **41** (1999)
17. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problems. Commun. Pure Appl. Math. **42** (1989) 577–685
18. Li, H., Elmoataz, A., Fadili, J., Ruan, S.: Dual front evolution model and its application in medical imaging. In: MICCAI2004. Volume 3216., Rennes/Saint-Malo, France (2004) 103–110
19. Cocosco, C., Kollokian, V., Kwan, R., Evans, A.: Brainweb: Online interface to a 3D MRI simulated brain database. NeuroImage **5** (1997) S425