

Interactive Image Segmentation Using Level Sets and Dempster-Shafer Theory of Evidence*

Björn Scheuermann and Bodo Rosenhahn

Leibniz Universität Hannover, Germany

{scheuermann,rosenhahn}@tnt.uni-hannover.de

Abstract. Variational frameworks based on level set methods are popular for the general problem of image segmentation. They combine different feature channels in an energy minimization approach. In contrast to other popular segmentation frameworks, e.g. the graph cut framework, current level set formulations do not allow much user interaction. Except for selecting the initial boundary, the user is barely able to guide or correct the boundary propagation. Based on Dempster-Shafer theory of evidence we propose a segmentation framework which integrates user interaction in a novel way. Given the input image, the proposed algorithm determines the best segmentation allowing the user to take global influence on the boundary propagation.

1 Introduction

Mumford and Shah [13] formalized the problem of image segmentation as the minimization of an error functional. Independent of each other, Chan and Vese [3] and Tsai et al. [21] proposed level set implementations of the Mumford-Shah functional. The boundary between object and background is represented by the zero-level set of a signed distance function $\varphi : \Omega \rightarrow \mathbb{R}$ [15]. The boundary evolution is modeled by a partial differential equation coming from the corresponding Euler-Lagrange equation. In contrast to the frameworks propagating explicit boundary points [9], the implicit level set representation has several well known properties, e.g. it can handle topological changes elegantly and it can easily be extended to higher dimensions.

Due to their popularity, region-based segmentation frameworks have been refined continuously [11, 2, 18, 14, 5, 22, 17, 6] to increase the number of scenes, where this framework can succeed. E.g. authors proposed statistical modeling of regions [22], additional feature channels like texture [17] or shape priors [6].

Yet, most existing level set methods are not qualified as an interactive segmentation tool. The corresponding initial value problem propagates the region boundary to a local minimum of the energy function without allowing the user to correct the final segmentation result. In contrast to the variational approaches, discrete energy minimization segmentation frameworks such as graph cut approaches [20, 1, 16] provide an elegant way to treat user interaction to guide or correct the segmentation process.

* This work is partially funded by the German Research Foundation (RO 2497/6-1).



Fig. 1. Segmentation results using the proposed user-interactive segmentation framework. The left example is a James Bond photo from the internet and the right one is from the Berkeley Segmentation Database.

A simple rule-based reasoning is usually used to integrate user interaction into the segmentation process: if the user marks a pixel as an object, then it is forced to be object, on the contrary a pixel is background if the user marks it as background. These so-called hard constraints can also be found in [1,16,4].

Level set methods for interactive segmentation have been proposed earlier by Cremers et al. [4]. They developed a statistical framework integrating user interaction. In addition to an initial boundary they provide a framework, where the user is able to mark object and background regions in terms similar to a shape prior. Thus the user can indicate which areas are likely to be part of the object or the background. In contrast, the key contribution of our paper is to develop a framework based on Dempster-Shafer theory of evidence which actually also uses the intensity information contained in the user defined regions. Thus, the evolving boundary is directly driven by the following three terms:

- the intensity information contained in the image [22],
- the user labeling in terms similar to a shape prior [4] and
- the intensity information of the user labeling (**our contribution**).

The proposed contribution results in global influence of the user defined regions, while other frameworks only allow local refinement of the segmentation (e.g. [4,1]). The different features contained in the image and the user defined regions are combined according to Dempster's rule of combination.

We continue with a review of the variational approach for image segmentation, which is the basis of our segmentation framework. Section 3 introduces the proposed segmentation method which integrates the user interactivity and the final workflow of the segmentation is recapitulated. Experimental results in Section 4 demonstrate the advantages of the proposed method in comparison to other state-of-the-art segmentation methods. Section 5 concludes the paper.

2 Level Set Segmentation with Evidence Theory

The variational approach for image segmentation used in our framework is based on the works of [3,5,22]. An extension of these works introducing the Dempster-Shafer theory of evidence is presented in [19]. In this section we will shortly review this framework and the key advantages of using evidence theory for level set based segmentation methods instead of the traditional Bayesian framework.

The key idea, which makes it different from other Bayesian segmentation frameworks, is the use of Dempster's rule of combination to fuse informations arising from different feature channels [7]. This framework is often described as a generalization of the Bayesian theory to represent inaccuracy and uncertainty at the same time. With the Bayesian theory, feature channels with low support have a high influence on the decision. In contrast, with Dempster-Shafer theory of evidence feature channels with high support for a specific region have a higher influence on the evolving boundary [19]. We make use of these two properties to introduce the proposed user interactivity.

The basis for our segmentation method is the following energy-functional:

$$\begin{aligned} E(\varphi) = & \underbrace{- \int_{\Omega} H(\varphi) \log m_{im}(\Omega_1) d\Omega - \int_{\Omega} (1 - H(\varphi)) \log m_{im}(\Omega_2) d\Omega}_{\text{data term}} \\ & + \lambda_1 \underbrace{\int_{\Omega} |\nabla H(\varphi)| d\Omega}_{\text{curve constraint}}, \end{aligned} \quad (1)$$

where λ_1 is a weighting parameter between the given constraints, $H(s)$ is a regularized Heaviside function and $m_{im} : \varphi(\Omega) \rightarrow [0, 1]$ is the mass function defined over the hypothesis set $\Omega = \{\Omega_1, \Omega_2\}$. In this context the region Ω_1 denotes the object region and Ω_2 the background. The mass function m_{im} , fuses k feature channels with Dempster's rule of combination:

$$m_{im} = m_1 \otimes m_2 \otimes \dots \otimes m_k, \quad (2)$$

where the single mass functions m_j are defined by probability densities for background and object. Dempster's rule of combination is given by

$$m(A) = m_1(A) \otimes m_2(A) = \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B)m_2(C)}. \quad (3)$$

The probability densities are computed by a histogram analysis for each region. They are defined by

$$\begin{aligned} m_j(\Omega_1) &= p_{1,j}(I(x)), \quad m_j(\Omega_2) = p_{2,j}(I(x)), \\ m_j(\emptyset) &= 0, \quad m_j(\Omega) = 1 - (p_{1,j}(I(x)) + p_{2,j}(I(x))), \end{aligned} \quad (4)$$

for $j \in \{1, \dots, k\}$. For the quality of the segmentation process, it is very important how the probability densities for each region are modeled. In this paper we restrict to the nonparametric Parzen density estimates [10]. Using discrete histograms this approach comes down smoothing the histograms computed for each region i and channel j by a Gaussian kernel. Other cues and models that can be integrated in such a framework are texture information [17] and shape priors [11].

Minimizing the energy (1) with respect to φ using variational methods and gradient descent leads to the following partial differential equation:

$$\frac{\partial \varphi}{\partial t} = \delta(\varphi) \left(\log \frac{m_{im}(\Omega_1)}{m_{im}(\Omega_2)} + \lambda_1 \operatorname{div} \left(\frac{\nabla \varphi}{|\nabla \varphi|} \right) \right). \quad (5)$$

Thus the segmentation process works according to the well known expectation-maximization principle [8] with an initial partitioning (Ω_1, Ω_2) .

3 Interactive Variational Image Segmentation

Analogue to Cremers et al. [4] we assume a given image $I : \Omega \rightarrow \mathbb{R}$ and a user input L marking certain image locations as object or background regions.

$$L : \Omega \rightarrow \{-1, 0, 1\}, \quad (6)$$

where the label values reflect the user input:

$$L(x) = \begin{cases} 1, & x \text{ marked as object,} \\ -1, & x \text{ marked as background,} \\ 0, & x \text{ not marked.} \end{cases} \quad (7)$$

Using the user-defined labeling $L(x)$ and defining a new segmentation constraint on this leads to the proposed energy function:

$$E(\varphi) = E_{im}(\varphi) + \lambda_1 E_{curve}(\varphi) + \underbrace{E_{user}(\varphi)}_{new} \quad (8)$$

where $E_{user} = \nu \cdot E_{user-shape} + E_{user-image}$. The first term of the *user-defined* energy function is defined according to [4] by

$$E_{user-shape} = -\frac{1}{2} \int_{\Omega} L_{\sigma}(x) \operatorname{sign}(\varphi(x)) d\Omega, \quad (9)$$

with a Gaussian-smoothed label function

$$L_{\sigma}(x) = \int_{\Omega} L(x) k_{\sigma}(x) d\Omega \quad (10)$$

and the Gaussian kernel function $k_{\sigma}(x)$.

This model has two free parameters ν and σ which can be interpreted as follows. The parameter ν provides the overall weight of the user interaction and determine how strongly the user input will affect the segmentation. The parameter σ defines the spatial range within which a point labeled as object or background will affect the segmentation. It can therefore be interpreted as a *brush size*.

The novel second term $E_{user-image}$ of E_{user} is inspired by the image energy E_{image} and is defined as follows:

$$\begin{aligned} E_{user-image} = & \int_{\Omega} H(\varphi) \log m_{user}(\Omega_1) d\Omega \\ & - \int_{\Omega} (1 - H(\varphi)) \log m_{user}(\Omega_2) d\Omega. \end{aligned} \quad (11)$$

The mass function m_{user} is, in contrast to the mass function m_{image} , defined by the marked regions while the function m_{image} is defined by the image regions divided by the curve.

$E_{user-shape}$ can be interpreted as a user-defined shape prior, while $E_{user-image}$ takes the image information on the marked regions into account and can therefore be interpreted as an indicator for the appearance of a region.

Fusing the mass functions m_{im} and m_{user} contained in E_{image} and $E_{user-image}$ respectively, with Dempster's rule of combination we obtain an energy-functional of the form:

$$\begin{aligned} E(\varphi) = & \underbrace{- \int_{\Omega} H(\varphi) \log m(\Omega_1) d\Omega - \int_{\Omega} (1 - H(\varphi)) \log m(\Omega_2) d\Omega}_{\text{data term + user defined term}} \\ & + \lambda_1 \underbrace{\int_{\Omega} |\nabla H(\varphi)| d\Omega}_{\text{curve constraint}} - \lambda_2 \nu \underbrace{\int_{\Omega} L_{\sigma} H(\varphi) d\Omega}_{\text{user-shape}}, \end{aligned} \quad (12)$$

where the mass function $m = m_{im} \otimes m_{user}$ fuses the image data given by m_{im} and the user data given by m_{user} according to Dempster's rule of combination. Minimizing (12) using variational methods and gradient descent leads to the following partial differential equation:

$$\frac{\partial \varphi}{\partial t} = \delta(\varphi) \left[\log \frac{m(\Omega_1)}{m(\Omega_2)} + \lambda_1 \operatorname{div} \left(\frac{\nabla \varphi}{|\nabla \varphi|} \right) + \lambda_2 (\nu L_{\sigma}) \right]. \quad (13)$$

The Dempster-Shafer theory of evidence is used to fuse the information because feature channels with low support have a lower influence on the evolving boundary as shown in [19]. This is helpful because the user-defined regions can be very sparse, which means that the resulting channel-histograms can have regions where neither the object nor the background region is supported. Using the Bayesian framework for fusing this information would lead to small probabilities for both regions, ignoring all other feature channels, especially the image feature channels. With the proposed framework based on Dempster's rule of combination, this would be interpreted as uncertainty meaning that the other feature channels are not affected by this feature.

In contrast to the work of Cremers et al. [4] the proposed framework not merely provides an indication in terms of a shape prior for the segmentation,

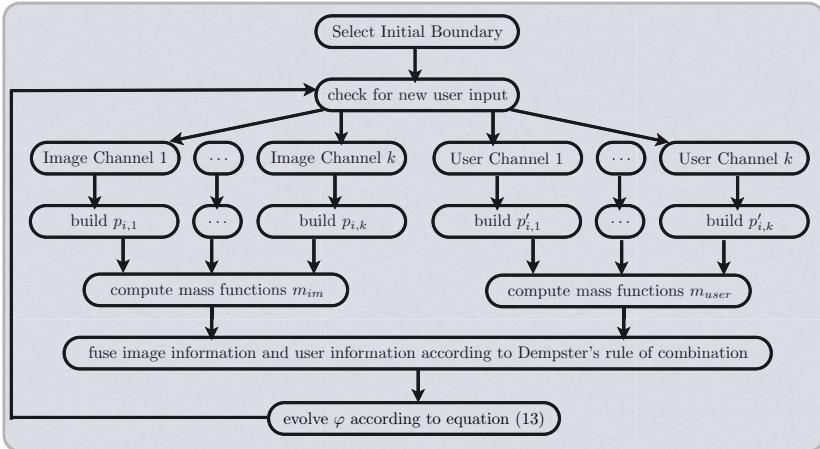


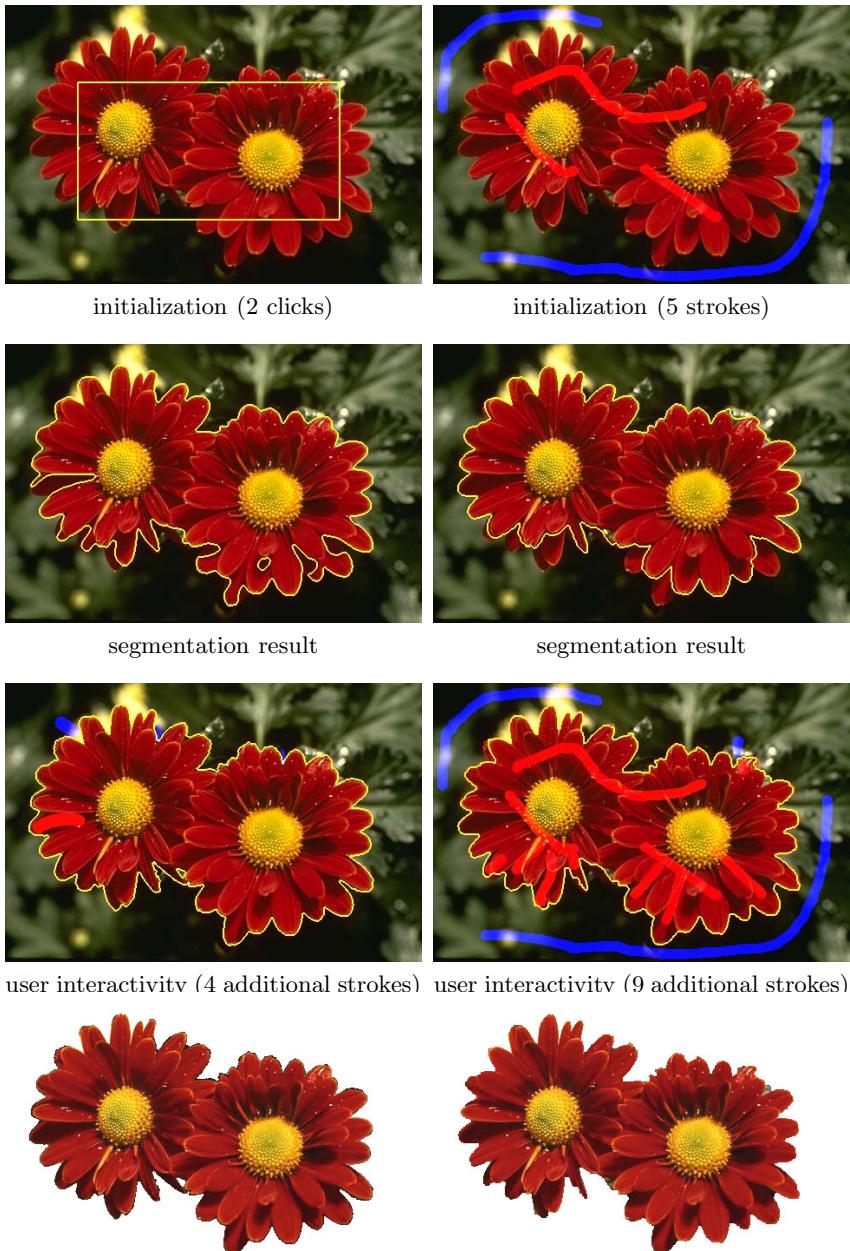
Fig. 2. General workflow of the proposed variational, user interactive segmentation framework

but actually uses the intensity information given by the user labeling. This information is further combined with Dempster's rule of combination, instead of multiplying the different probabilities, to represent inaccuracy and uncertainty. While the user labels in [4] have only local support to the evolving boundary and thus to the final segmentation, our framework allows global support for user defined regions. Figure 2 shows the proposed general interactive segmentation workflow.

4 Experiments

In this section we test the proposed user-interactive segmentation framework experimentally. Several results are shown in Figures 1 and 3 and compared to the graph cut framework by Boykov et al. [1]. The images used for the experiments are natural images taken from the berkeley segmentation dataset [12]. Furthermore we compared the proposed framework with the user-interactive framework in [4]. While user-interaction only has local influence on the final segmentation in the framework proposed by Cremers et al., our framework allows global influence on the segmentation result with a small stroke, e.g. the global influence of the small foreground stroke in Figure 3.

We also performed a user study, where six persons segmented five real images with the proposed framework and the graph cut segmentation tool. In these moderately difficult examples (e.g. the soldier in Figure 4) the proposed framework needed significantly fewer user-interactions while the mean F_1 measure over all segmented images is comparable. The result of our user study is shown in Table 1. Figure 4 shows some of the segmented images. We have to distinguish that users tend to make longer strokes using graph cuts. Especially the two initial strokes are very large (see Figure 3 and 4) compared to the small strokes in our



Final segmentation result after user refinement (Initialization (2 clicks, 1 foreground and 3 background strokes)

Segmentation result after user refinement (11 foreground- and 3 background strokes)

Fig. 3. Segmentation results using the proposed interactive segmentation framework (left) and graph cut (right). The proposed segmentation algorithm needs significant fewer user interactions (red and blue strokes) to produce a slightly better result.

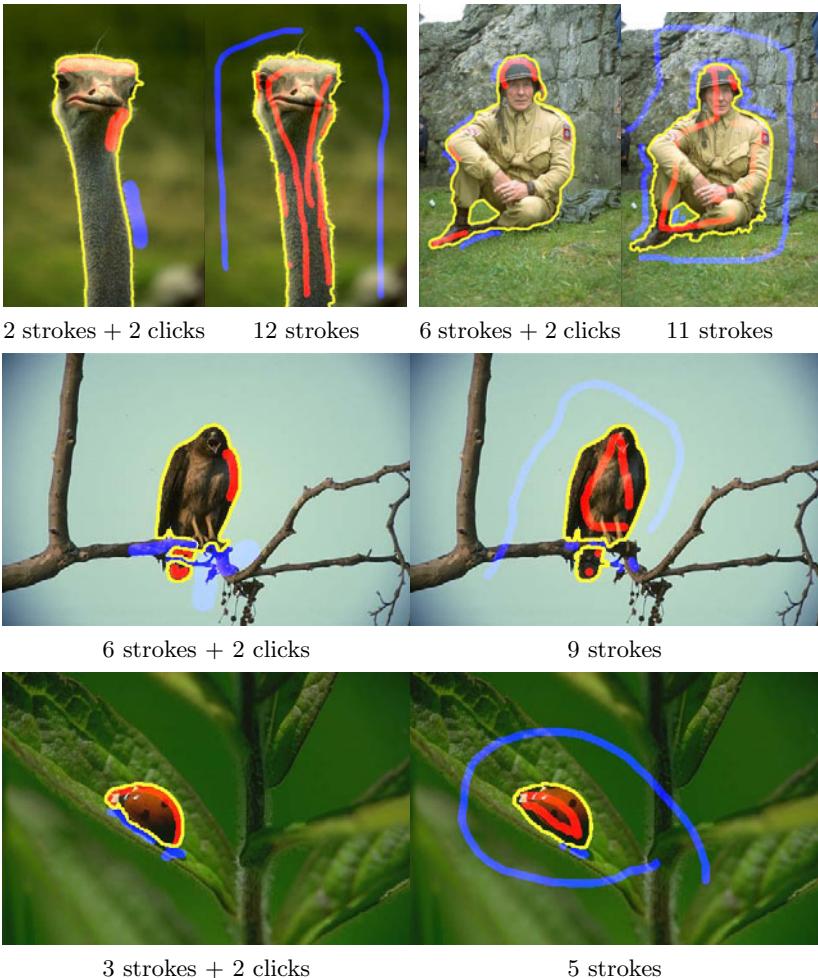


Fig. 4. Segmentation results using the proposed variational framework (left) and GraphCut (right). The yellow curves describe the segmentation boundaries, while the blue and red strokes mark the user defined regions.

framework. The average stroke size with the proposed method is approximately half of the stroke size with graph cuts. Although the proposed method is not implemented on the GPU, the total time for segmenting the images was almost the same for both methods. In addition the users are able to guide the evolving boundary instead of changing the final segmentation.

We have to mention that we did not test other segmentation frameworks like e.g. GrabCut [16], but the results presented in [16] are very close to what we and the users have achieved in our study.

Table 1. Results of our user study. While the mean F_1 -measure is comparable for all methods, the proposed method needed significantly fewer user interactions, comparing the average number of strokes.

Image	Graph Cuts	F_1	[4]	F_1	proposed Method	F_1
Lady Bug	4.33 str.	0.9366	1.3 str. + 2 klicks	0.8943	1.3 str. + 2 klicks	0.9011
Eagle	9.33 str.	0.9676	7 str. + 2 klicks	0.9265	5.5 str. + 2 klicks	0.9472
Bird	7.83 str.	0.9610	2.3 str. + 2 klicks	0.9541	1.83str. + 2 klicks	0.9624
Flowers	7.17 str.	0.9808	6.6 str. + 2 klicks	0.9891	4 str. + 2 klicks	0.9892
Soldier	9.33 str.	0.9654	10.3 str. + 2 klicks	0.9814	7.33 str. + 2 klicks	0.9736

5 Conclusion

We presented a new framework for foreground extraction based on level set methods and Dempster-Shafer theory of evidence. The framework extends the traditional framework by means of user-interactivity to allow more precise segmentations. The user-interactivity (strokes) is integrated into the traditional framework by a user-defined shape prior (local influence) and by user-defined image features (global influence). The impact of the new framework is demonstrated by several experiments on natural images and a user study in comparison to the well known graph cut framework. With the new extensions the level set based segmentation framework allows small user interactions having global influence on the evolving boundary. In comparison to graph cut, the presented framework needs significantly fewer user interactions to produce high-quality segmentations.

References

1. Boykov, Y., Jolly, M.: Interactive graph cuts for optimal boundary and region segmentation of objects in nd images. In: International Conference on Computer Vision, vol. 1, pp. 105–112 (2001)
2. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. International Journal of Computer Vision 22(1), 61–79 (1997)
3. Chan, T., Vese, L.: Active contours without edges. IEEE Transactions on Image Processing 10(2), 266–277 (2001)
4. Cremers, D., Fluck, O., Rousson, M., Aharon, S.: A probabilistic level set formulation for interactive organ segmentation. In: Proc. of the SPIE Medical Imaging, San Diego, USA (2007)
5. Cremers, D., Rousson, M., Deriche, R.: A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. International Journal of Computer Vision 72(2), 195–215 (2007)
6. Cremers, D., Schnörr, C., Weickert, J.: Diffusion Snakes: Combining statistical shape knowledge and image information in a variational framework. In: Paragios, N. (ed.) IEEE First Int. Workshop on Variational and Level Set Methods, Vancouver, pp. 137–144 (2001)

7. Dempster, A.P.: A generalization of bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)* 30(2), 205–247 (1968)
8. Dempster, A., Laird, N., Rubin, D., et al.: Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 39(1), 1–38 (1977)
9. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active Contour models. *International Journal of Computer Vision* 1(4), 321–331 (1988)
10. Kim, J., Fisher III, J., Yezzi Jr., A., Cetin, M., Willsky, A.: Nonparametric methods for image segmentation using information theory and curve evolution. In: IEEE International Conference on Image Processing (ICIP), pp. 797–800 (2002)
11. Malladi, R., Sethian, J.A., Vemuri, B.C.: Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 158–175 (1995)
12. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc. 8th Int'l Conf. Computer Vision, vol. 2, pp. 416–423 (2001)
13. Mumford, D., Shah, J.: Boundary detection by minimizing functionals. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 22–26. IEEE Computer Society Press, Springer, San Francisco, CA (1985)
14. Osher, S., Paragios, N.: Geometric level set method in imaging, vision, and graphics. Springer, Heidelberg (2003)
15. Osher, S., Sethian, J.: Fronts propagating with curvature dependent speed: Algorithm based on hamilton-jacobi formulation. *Journal of Computational Physics* 79, 12–49 (1988)
16. Rother, C., Kolmogorov, V., Blake, A.: Grabcut: Interactive foreground extraction using iterated graph cuts. In: ACM SIGGRAPH 2004 Papers, p. 314. ACM, New York (2004)
17. Rousson, M., Brox, T., Deriche, R.: Active unsupervised texture segmentation on a diffusion based feature space. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Madison, WI, pp. 699–704 (2003)
18. Rousson, M., Paragios, N.: Shape priors for level set representations. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2351, pp. 78–92. Springer, Heidelberg (2002)
19. Scheuermann, B., Rosenhahn, B.: Feature quarrels: The dempster-shafer evidence theory for image segmentation using a variational framework. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part II. LNCS, vol. 6493, pp. 426–439. Springer, Heidelberg (2011)
20. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22(8), 888–905 (2000)
21. Tsai, A., Yezzi Jr., A., Wells III, W., Tempany, C., Tucker, D., Fan, A., Grimson, W., Willsky, A.: Model-based curve evolution technique for image segmentation. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 463–468 (2001)
22. Zhu, S.C., Yuille, A.: Region competition: unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 18(9), 884–900 (1996)