Leukocyte Segmentation in Blood Smear Images Using Region-Based Active Contours

Seongeun Eom, Seungjun Kim, Vladimir Shin, and Byungha Ahn

Department of Mechatronics,
Gwangju Institute of Science and Technology,
1 Oryong-dong, Buk-gu, Gwangju 500-712, South Korea
`{seueom, zizone, vishin, bayhay}@gist.ac.kr`

Abstract. In this paper, we propose a segmentation method for an automated differential counter using image analysis. The segmentation here is to extract leukocytes (white blood cells) and separate its constituents, nucleus and cytoplasm, in blood smear images. For this purpose, a region-based active contour model is used where region information is estimated using a statistical analysis. The role of the regional statistics is mainly to attract evolving contours toward the boundaries of leukocytes, avoiding problems with initialization. And contour deformation near to the boundaries is constrained by an additional regularizer. The active contour model is implemented using a level set method and validated with a leukocyte image database.

1 Introduction

A leukocyte (white blood cell) differential count as a percentage is one of the most frequently performed blood tests and plays an important role in the diagnosis of diseases such as anemia. In hospital, manual differential is usually performed by taking a drop of blood, spreading it on a slide, staining it, and evaluating around 100 cells for quantity and quality. However, it is tedious and time consuming to locate, classify and count leukocytes. An automated differential counter using image analysis makes it possible to replace the work, reducing reporting time and increasing precision with the larger of number of cell counted. The counter system is normally performed in this procedure: localization, segmentation, feature extraction and classification. We here deal with the leukocyte segmentation problem which is most difficult and error-prone.

In an attempt to solve the problem, several approaches have been presented in the literature. Wermser et al. [1] and Cseke [2] use hierarchical thresholding with two color features which discriminate between constituents of blood smear images. A more sophisticated algorithm is proposed by Sinha and Ramakrishnan [3]. It estimates a mixed Gaussian model using Expectation-Maximization (EM) algorithm in HSV color space, and thresholds and labels each Gaussian component with a prior knowledge. It is obvious that the algorithm is able to improve performance over linear thresholding, but it still does not use spatial information and so is limited to segment leukocytes.

J. Blanc-Talon et al. (Eds.): ACIVS 2006, LNCS 4179, pp. 867–876, 2006.

To overcome the limitation, another approach using region information is proposed. Haussman et al. [4] uses region labeling by relaxation operation. Initial regions are obtained from split-merge method and then relaxed using relationship between adjacent regions. [5] is also a similar method based on fuzzy rules. Although this approach uses both local and global information, it is not straightforward to design rules (or relationship) covering various situations.

More recent literature has applied active contour models since it is useful to detect objects in an image by evolving contours. One of the advantages of the model is that contours can be controlled by its geometric properties and external properties from an input image in a single framework. Furthermore higher level information such as shape priors also can be incorporated into the same framework. Hence we can design an active contour model using information suitable for a specific application.

The active contours proposed in [6,7,8] are based on the gradient of an input image. However a high gradient is often shown at the cytoplasm of polymorphonuclear cells (granulocytes). It means that evolving contours are likely to be stuck at local minima and so the performance is heavily influenced by initial contours. [9] instead uses the gradient vector flow (GVF) calculated from boundaries of cells including erythrocytes (red blood cells). The method relieves the initialization problem but contours still could be stuck in the area of small GVF force or attracted to one side of the boundary depending on initial contours

In this paper, we propose a more robust method using a region-based active contour model. Region information here is estimated through statistical analysis on intensity features of input images. Thanks to the regional statistics, initial contours are placed safely on each leukocyte and propagate toward its boundary, avoiding the local minima problem. At earlier stage, evolving contours are mainly under influence of the region statistics, but their deformation is controlled by an additional regularizer as they come close to the boundaries of leukocytes. The regularizer contributes to forming the shape of a leukocyte

2 Proposed Method

The proposed method is composed of five steps as shown in Fig. 1. We first select significant intensity features in blood smear images and estimate a finite mixed Gaussian model using the EM algorithm where the number of components and initial parameters are chosen with intensity priors. With the estimation, Bayes probabilities of each constituent of a leukocyte are computed on piecewise constant partition and then are embedded into an active contour model as region information. In the following subsections the procedure is explained in detail.

2.1 Regional Statistics

Each constitute of blood smear images shows the characteristics discriminating one another such as color, texture, and shape. But it seems that color information is the most significant. Depending on color model, various combination of

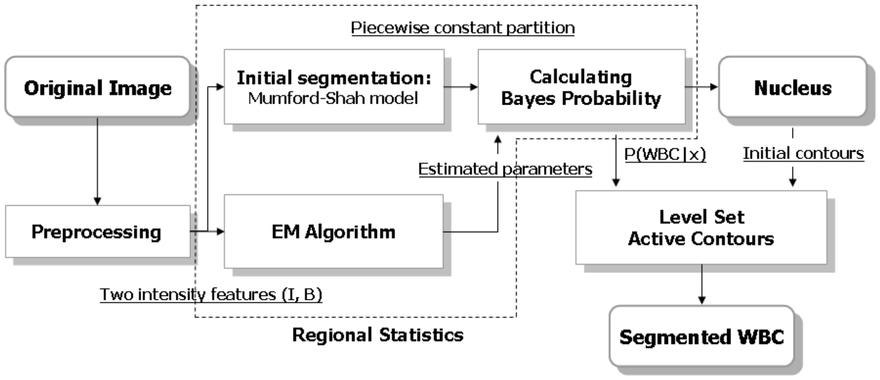


Fig. 1. Block diagram of the proposed segmentation method for leukocyte images. It is based on region-based active contours where the region information is estimated using a statistical analysis.

channels can be used. The experiments shows that among them grayscale and blue intensity space (I, B) is the most promising. Fig. 2 shows the typical pattern of the features extracted from a smoothed image with a Gaussian filter, and also exhibits the main location of each constituent.

As can be seen in the feature pattern, it is difficult to discriminate between different components through simple thresholding or clustering approach. For this reason, we estimate parameters of a finite mixed Gaussian model with the Expectation-Maximization (EM) algorithm, assuming all the components have a Gaussian distribution.

The EM algorithm [10] consists of two steps, Expectation and Maximization. The first step is to calculate the posterior probabilities with initial parameters for each data point, and then update the means, the covariance matrices and the mixing coefficients for each component in the second step. The procedure is repeated until a variation is less than some fixed threshold.

An initial guess, however, should be carefully determined particularly for blood smear images, because there often exit one or two more Gaussian components in the feature space: for example, at the granule in cytoplasm or outer boundary of an erythrocyte. At first we could assume five or six components, but it does not work well in two reasons. (1) Additional components are likely to have a negative effect on the estimation of the others, and (2) even not so it is not straightforward to label to four partitions.

Therefore the number of components is set to be four and now a way of preventing components from converging to unexpected local minima is required. The K-means algorithm [10] are commonly used to select initial parameters. However, it is not appropriate in our case for the same reason mentioned earlier: existence of undesirable Gaussian components. Instead those are determined based on the characteristics of the feature pattern and intensity priors of each constituent. Left-lower (A) and right-upper (B) feature point are selected as

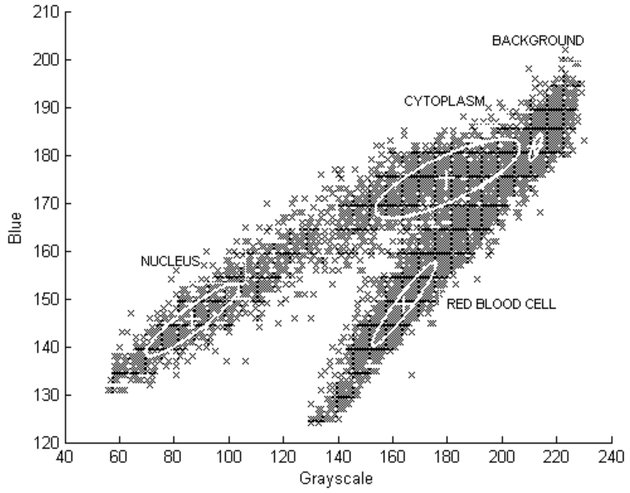


Fig. 2. Typical example of the feature patten and the main location of each constituent: nucleus, cytoplasm, red blood cell, and background

an initial guess of nucleus and background respectively. And an erythrocyte takes a point (C) located the farthest from the line going through two points, A and B. the last one (D) is decided middle point of two points, A and C. This simple method has shown more better results than the K-means algorithm. Furthermore, labeling is easily done by the same way with the final mean points

With the estimated and labeled Gaussian components, the posterior probabilities for each constituent can be calculated using Bayes' theorem [10]:

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)},$$

where

$$p(x) = \sum_{j=1}^{4} p(x|\omega_j)P(\omega_j)$$

and ω_j is a jth class, namely, nucleus, cytoplasm, erythrocyte or background, and x is a feature vector (I, B). It could be applied to each pixel itself, but we here compute the Bayes probabilities on piecewise constant partition which is obtained by minimizing the Mumford-Shah functional [11]:

$$F_1(u, B) = \int_{R-B} (u - u_0)^2 + \lambda \text{ length}(B),$$

where u_0 is an input image defined on a set R and u and B are a piecewise constant function and the boundaries between regions, respectively. This is useful to avoid many isolated points and small holes, and form more accurate boundaries.

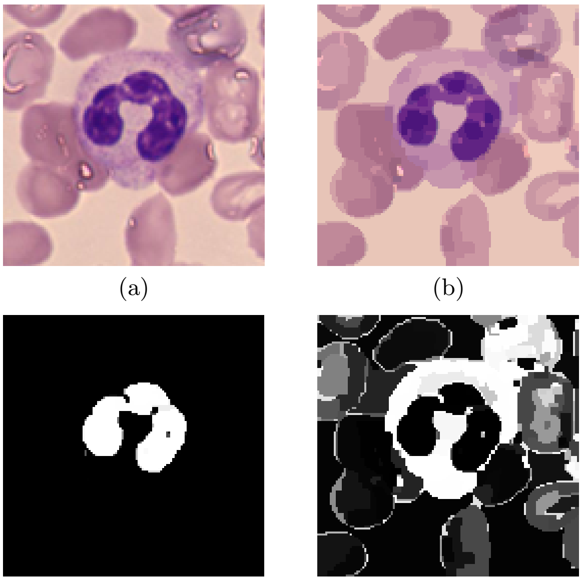


Fig. 3. Regional statistics. (a) input image, (b) piecewise constant partition, (c-d) posterior probability map of nucleus and cytoplasm. Courtesy of CellAtlas, CellaVision AB [12] (with the following figures).

Fig. 3 shows an example of initial partition and probability map of a leukocyte. The probability map is used in the region-based active contours framework as region information in the next stage.

2.2 Active Contours for Leukocyte Segmentation

Active contour models [13], also called snakes, are a powerful segmentation tool in a variety of image processing applications. Many models have been developed based on their representation (explicit or implicit) and information (edge and/or region) used. Hence, from the point of view of application the selection of an appropriate model is very important.

In our case the following implicit model is applied which is based on region information computed from the intensity features:

$$\frac{\partial \phi}{\partial t} = \delta(\phi)[\alpha_1(P - C_1)^2 + \alpha_2(P - C_2)^2] + \beta\kappa|\nabla\phi| + \gamma\boldsymbol{V} \cdot \nabla\phi,$$

where ϕ is a implicit level set function [14,15] where the boundary (interface) is defined by $\phi = 0$, and α_1, α_2, β, and γ are positive parameters. An input image P, here, is the posterior probability of a leukocyte $P(\texttt{WBC}|x)$ computed earlier and C_1, C_2 are the averages of P inside and outside of propagating contours. After all the first two terms [16] attract contours toward the boundaries of a leukocyte. As shown in Fig. 3, however, part of neighboring erythrocytes also could have a high probability and so regularization is needed.

A leukocyte is normally close to the shape of a circle but is likely to deform. In this case, however, the deformation is “smooth”, while undesirable regions having a high probability tend to be “rough” as shown in Fig. 3. Therefore a regularization method for discriminating the two cases is required. First a parameter of mean curvature term $\kappa|\nabla\phi|$ is set to be high so that the shape is close to a circle, and then a deformation beyond that is driven by the gradient vector flow (GVF) $\boldsymbol{V}(u,v)$ [17] obtained by minimizing

$$F_2(\boldsymbol{V}) = \int\int \mu({u_x}^2 + {u_y}^2 + {v_x}^2 + {v_y}^2) + |\nabla f|^2 |\boldsymbol{V} - \nabla f|^2 dxdy,$$

where an input image f is the boundaries of regions Bayes classified as a leukocyte. In spite of constraint of curvature, the contour now tries to propagate toward smoothly deformed boundaries thanks to this externally generated velocity field. On the other hand, it does not try in roughly deformed regions because the normal force of GVF is very weak at the boundary between a leukocyte and an erythrocyte [9].

Besides this role, GVF contributes to extracting the contours with the first two terms in the initial stages and so seems to be able to take the place of the first two ones. However, in this case part of the contours could be stuck in the area of small GVF force. Gradient information also causes the similar stuck problem at the local minima and so is not used in the proposed model.

Another important factor in the use of active contours is initial contours. In our case it is relatively insensitive to the initial contours which just need to be located inside leukocytes. Fortunately Bayes classified nuclei are satisfactory in the regional statistics stage. Therefore nuclei regions are extracted at the stage and are used for selection of initial contours. We put circles at the centroid of each nucleus, whose diameters are 0.3 times ones of circles with the same area as each nucleus. If there are multiple nuclei in a leukocyte they are merged automatically thanks to level set implementation. Fig. 4 shows the propagating contours on the posterior probability map.

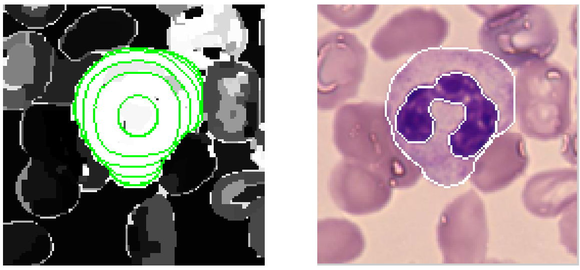

(a) (b)

Fig. 4. Segmentation using region-based active contours. (a) propagating contour, (b) segmentation result.

3 Experimental Results

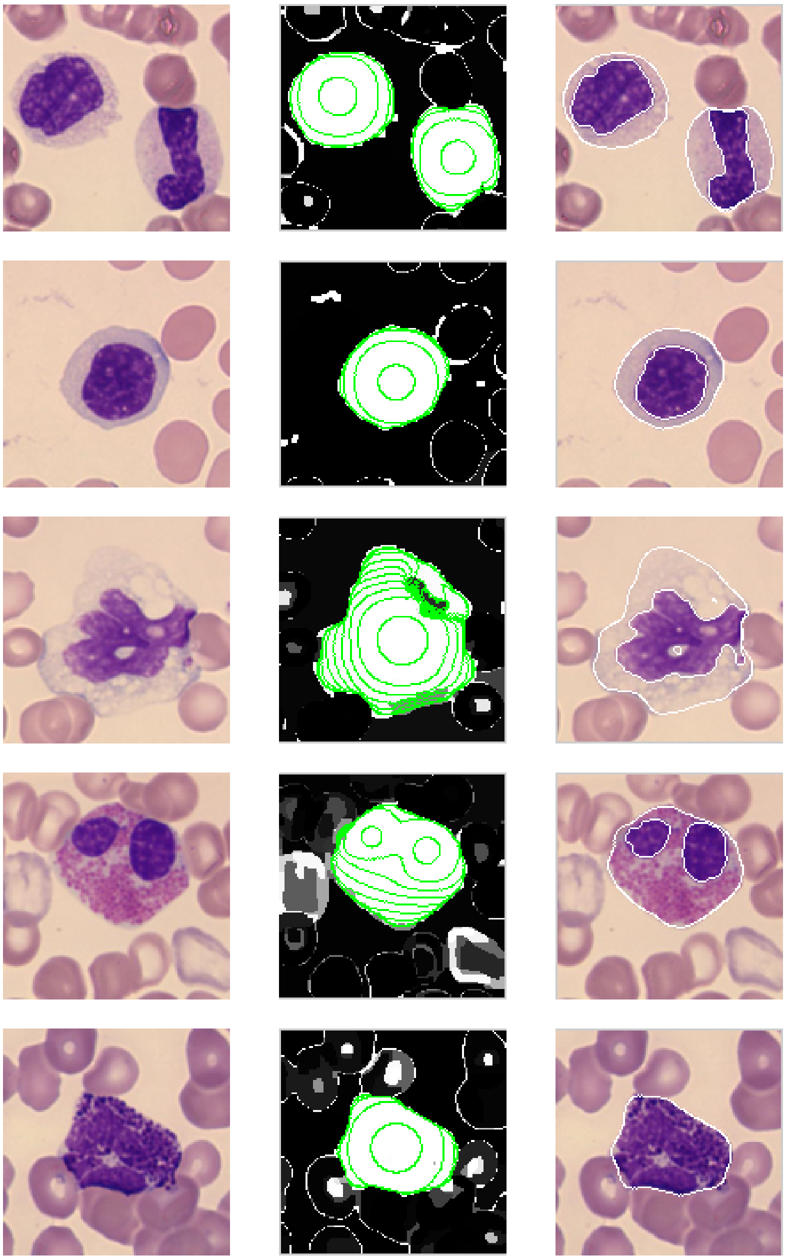
The experiments were carried out on the cell image database which was provided by CellAtlas [12]. All the images in the database were stained by the May-Grünwald-Giemsa (MGG) method and already classified by type by experts within the field of hematology. The types of leukocytes used in our experiments are five: neutrophil, lymphocyte, monocyte, eosinophil, and basophil. Assuming that leukocytes were localized, a region of interest (ROI) containing leukocytes was extracted with 128×128 size.

These images first were smoothed with a 3×3 Gaussian filter with a standard deviation of 0.8 and in the Mumford-Shah functional for a piecewise constant partition, a region-growing method [18] was applied, where the number of final regions (1000) was used as a stopping criterion instead of a scale λ. In the last stage, there are three important parameters which were set ($\alpha_1 = \alpha_2 = 0.005, \beta = 4.5, \gamma = 0.6$) as follows: (1) with $\gamma = 0$, adjust the first two parameters such that a contour becomes similar to a circle in shape as it approaches to the boundary of a leukocyte and then (2) γ value is chosen for the GVF force to push the contour toward the smoothly deformed regions.

Fig. 5 shows the process of segmentation applying the proposed method under the conditions above. The first column is the original image for each type of leukocytes, and the second shows evolving contours on the Bayes probability map of a leukocyte. The final segmentation results are shown in the last column. As already expected, errors happen in parts of images in Fig. 5(b), but in the most of cases the errors happen between cytoplasm and red blood cells because of similarity of their intensity features (I, B). However, it is noteworthy that the probability of errors in cytoplasm regions is very low and so true leukocytes are mostly included in the regions having high probability $P(\mathtt{WBC}|x)$ like in Fig. 5(b). The used features (I, B) were selected among two channels or full channels of RGB, HSV, L*a*b*, and L*u*v* space, not scaled, using the probability of error in the Bayes classification. More sophisticated methods could be used for optimal feature extraction, but it is now acceptable.

As explained earlier Bayes classified nuclei regions are satisfactory but sometimes there are undesirable small segments and holes and so we applied the morphological operators, area opening and closing, in order to remove the undesirable regions. The postprocessed nuclei were then used for initial contours. If there are multiple nuclei the same number of initial contours are placed as shown in Fig. 5(b). At an earlier stage these contours propagate mainly under influence of the regional statistics, and deform under influence of the curvature and GVF force as they come close to the boundary of a leukocyte. Fig. 5(b) shows the process of propagation. It exhibits that a contour is able to cross small holes inside a leukocyte and multiple ones are finally merged.

To validate the segmentation results shown in Fig. 5(c), we adopted the most common approach. The boundaries of leukocytes were first extracted manually and compared to the obtained results. The following metric [19] was used as a measure of segmentation accuracy: $S = 2 \cdot n\{R \cap T\}/(n\{R\} + n\{T\})$. This measure quantifies similarity $S \in [0, 1]$ between two segmentations R and T.



(a) (b) (c)

Fig. 5. Process of segmentation using the proposed method. (a) original images: neutrophil, lymphocyte, monocyte, eosinophil, and basophil, (b) intermediate results on the Bayes probability of a leukocyte, (c) final segmentation results.

Table 1. Average segmentation accuracy rates and their standard deviations (%)

Cell type	Neutrophil	Lymphocyte	Monocyte	Eosinophil	Basophil
# of images	80	125	120	124	154
Nucleus	91.40 (13.94)	95.45 (2.94)	90.83 (11.03)	88.43 (15.26)	- (-)
Cytoplasm	93.60 (6.85)	95.10 (4.45)	91.82 (10.24)	91.41 (6.91)	93.02 (9.92)

$n\{R\}$ indicates the area of R. Table 1 provides the average segmentation accuracy rates and their standard deviations of nucleus and cytoplasm for all the images; the nucleus of basophil was not included in the evaluation because it is often obscured by cytoplasmic granules and so is difficult to discern. We have obtained the reasonable results with an overall average error less than 8%.

The major errors were caused by the following reasons. First, when touching regions of erythrocytes have a high probability $P(\texttt{WBC}|x)$ and the regions look like part of a true leukocyte, no distinction is made between them and so a contour propagates into the regions. Second, when cytoplasmic granules are as dark as a nucleus or a cytoplasm as bright as a background, they are classified as a nucleus and a background, respectively. Third, when the shape of a leukocyte is deformed excessively, a contour fails to deform that much. The parameters involving deformation was selected as tradeoff between this case and the first one. In future research we should focus on these problems in order to improve performance more accurately.

A natural extension of the proposed method is to deal with touching leukocytes using a single level set representation. It is obvious that the separate leukocytes one another can be segmented well and also even touching ones can be segmented individually, but not at a time in a single level set framework. So we are currently under investigation on this point and expect to be useful for other applications, too.

4 Conclusion

We have proposed a segmentation method of leukocytes in blood smear images. It is based on a region-based active contour model driving initial contours toward the boundary of a leukocyte, avoiding problem with initialization and local minima. Region information here is estimated from intensity features which discriminate effectively between the constituents of blood smear images. And a regularizer has also applied in the same model to constrain excessive deformation of a evolving contour. In the experiments with a public image database, we have obtained the segmentation results with an overall average error less than 8%. Although it can be more improved by decreasing the errors mentioned earlier, the results are reasonable as an input for leukocytes classification.

References

1. Wermser, D., Haussmann, G., Liedtke, C.E.: Segmentation of blood smears by hierarchical thresholding. Computer Vision, Graphics, and Image Processing **25** (1984) 151–168
2. Cseke, I.: A fast segmentation scheme for white blood cell images. In: Proc. 11th IAPR Int. Conf. Pattern Recognition, Conf. C: Image, Speech and Signal Analysis. Volume 3. (1992) 530–533
3. Sinha, N., Ramakrishnan, A.G.: Automation of differential blood count. In: Proc. Conf. Convergent Technologies for Asia-Pacific Region TENCON 2003. Volume 2. (2003) 547–551
4. Haussmann, G., Liedtke, C.E.: A region extraction approach to blood smear segmentation. Computer Vision, Graphics, and Image Processing **25** (1984) 133–150
5. Park, J., Keller, J.M.: Fuzzy patch label relaxation in bone marrow cell segmentation. In: Proc. IEEE Int. Conf. Systems, Man, and Cybernetics. (1997) 1133–1138
6. Ongun, G., Halici, U., Leblebicioglu, K., Atalay, V., Beksac, M., Beksac, S.: An automated differential blood count system. In: Proc. 23rd EMBS Int. Conf. (2001) 2583–2586
7. Nilsson, B., Heyden, A.: Segmentation of dense leukocyte clusters. In: Proc. IEEE Workshop on Mathematical Methods in Biomedical Image Analysis. (2001) 221–227
8. Nilsson, B., Heyden, A.: Model-based segmentation of leukocytes clusters. In: Proc. 16th International Conf. Pattern Recognition. Volume 1. (2002) 727–730
9. Theerapattanakul, J., Plodpai, J., Pintavirooj, C.: An efficient method for segmentation step of automated white blood cell classifications. In: Proc. IEEE Region 10 Conf. TENCON 2004. Volume 1. (2004) 191–194
10. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. 2nd edn. Wiley-Interscience, New York (2000)
11. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. Communications on Pure and Applied Mathematics **42** (1989) 577–684
12. (http://www.cellatlas.com)
13. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. Int. J. Computer Vision **1** (1988) 321–331
14. Osher, S., Fedkiw, R.: Level Set Methods and Dynamic Implicit Surfaces. 1st edn. Springer, New York (2002)
15. Sethian, J.A.: Level Set Methods and Fast Marching Methods. 2nd edn. Cambridge University Press, Cambridge (1999)
16. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Process. **10(2)** (2001) 266–277
17. Xu, C., Prince, J.L.: Snakes, shapes, and gradient vector flow. IEEE Trans. Image Process. **7(3)** (1998) 359–369
18. Aubert, G., Kornprobst, P.: Mathematical Problems in Image Processing. Volume 147 of Applied Mathematical Sciences. Springer, New York (2001)
19. Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., Palmer, A.C.: Morphometric analysis of white matter lesions in mr images: method and validation. IEEE Trans. Medical Imaging **13(4)** (1994) 716–724