

# Automatic Cerebrovascular Segmentation by Accurate Probabilistic Modeling of TOF-MRA Images

Ayman El-Baz<sup>1</sup>, Aly A. Farag<sup>1</sup>, Georgy Gimel'farb<sup>2</sup>, and Stephen G. Hushek<sup>3</sup>

<sup>1</sup> CVIP Laboratory, University of Louisville, Louisville, KY 40292, USA  
{elbaz, farag}@cvip.louisville.edu  
<http://www.cvip.louisville.edu>

<sup>2</sup> Department of Computer Science, Tamaki Campus,  
University of Auckland, Auckland, New Zealand

<sup>3</sup> Technical Director, iMRI Department, Norton Hospital, Louisville, KY, USA

**Abstract.** Accurate automatic extraction of a 3D cerebrovascular system from images obtained by time-of-flight (TOF) or phase contrast (PC) magnetic resonance angiography (MRA) is a challenging segmentation problem due to small size objects of interest (blood vessels) in each 2D MRA slice and complex surrounding anatomical structures, e.g. fat, bones, or grey and white brain matter. We show that due to a multi-modal nature of MRA data blood vessels can be accurately separated from background in each slice by a voxel-wise classification based on precisely identified probability models of voxel intensities. To identify the models, an empirical marginal probability distribution of intensities is closely approximated with a linear combination of discrete Gaussians (LCDG) with alternate signs, and we modify the conventional Expectation-Maximization (EM) algorithm to deal with the LCDG. To validate the accuracy of our algorithm, a special 3D geometrical phantom motivated by statistical analysis of the MRA-TOF data is designed. Experiments with both the phantom and 50 real data sets confirm high accuracy of the proposed approach.

## 1 Introduction

Accurate cerebrovascular segmentation is of prime importance for early diagnostics and timely endovascular treatment. Unless detected at early stage, serious vascular diseases like carotid stenosis, aneurysm, and vascular malformation may cause not only severe headaches but also a brain stroke or a life-threatening coma [1]. Non-invasive MRA is a valuable tool in preoperative evaluation of suspected intracranial vascular diseases. Three commonly used MRA techniques are TOF-MRA, phase contrast angiography (PCA), and contrast enhanced MRA (CE-MRA). Both TOF-MRA and PCA use flowing blood as an inherent contrast medium, while for CE-MRA a contrasting substance is injected into the circulatory system.

A variety of today's most popular techniques for segmenting blood vessels from TOF-MRA data can be roughly classified into deformable and statistical models.

The former methods iteratively deform an initial boundary surface of blood vessels in order to optimize an energy function which depends on image gradient information and surface smoothness [2]. Topologically adaptable surfaces make classical deformable models more efficient for segmenting intracranial vasculature [3]. Geodesic active contours implemented with level set techniques offer flexible topological adaptability to segment MRA images [4] including more efficient adaptation to local geometric structures represented e.g. by tensor eigenvalues [5]. Fast segmentation of blood vessel surfaces is obtained by inflating a 3D balloon with fast marching methods [6]. Two-step segmentation of a 3D vascular tree from CTA data sets in [7] is first carried out locally in a small volume of interest. Then a global topology is estimated to initialize a new volume of interest. A multi-scale geometrical flow is proposed in [8] to segment vascular tree from MRI images.

The statistical approach extracts the vascular tree automatically, but its accuracy depends on underlying probability data models. The TOF-MRA image is multi-modal in that signals in each region-of-interest (e.g. blood vessels, brain tissues, etc) are associated with a particular mode of the total marginal probability distribution of signals. To the best of our knowledge, up-to-now there exists only one adaptive statistical approach for extracting blood vessels from the TOF-MRA data proposed by Wilson and Noble [9]. They model the marginal data distribution with a mixture of two Gaussian and one uniform components for the stationary CSF, brain tissues, and arteries, respectively. To identify the mixture (i.e. estimate all its parameters), they use a conventional EM algorithm. Furthermore, a region-based deformable contour for segmenting tubular structures was derived in [10] by combining signal statistics and shape information.

## 2 LCDG-Model of a Multi-modal TOF-MRA Image

Let  $q; q \in \mathbf{Q} = \{0, 1, \dots, Q-1\}$ , denote the  $Q$ -ary gray level. The discrete Gaussian (DG) is defined as the probability distribution  $\Psi_\theta = (\psi(q|\theta) : q \in \mathbf{Q})$  on  $\mathbf{Q}$  such that  $\psi(q|\theta) = \Phi_\theta(q+0.5) - \Phi_\theta(q-0.5)$  for  $q = 1, \dots, Q-2$ ,  $\psi(0|\theta) = \Phi_\theta(0.5)$ ,  $\psi(Q-1|\theta) = 1 - \Phi_\theta(Q-1.5)$  where  $\Phi_\theta(q)$  is the cumulative Gaussian probability function with a shorthand notation  $\theta = (\mu, \sigma^2)$  for its mean,  $\mu$ , and variance,  $\sigma^2$ .

We assume the number  $K$  of dominant modes, i.e. regions, objects, or classes of interest in a given TOF-MRA image, is already known. In contrast to a conventional mixture of Gaussians and/or other simple distributions, one per region, we closely approximate the empirical gray level distribution for a TOF-MRA image with an LCDG having  $C_p$  positive and  $C_n$  negative components such that  $C_p \geq K$ :

$$p_{\mathbf{w}, \Theta}(q) = \sum_{r=1}^{C_p} w_{p,r} \psi(q|\theta_{p,r}) - \sum_{l=1}^{C_n} w_{n,l} \psi(q|\theta_{n,l}) \quad (1)$$

under the obvious restrictions on the weights  $\mathbf{w} = [w_{p,\cdot}, w_{n,\cdot}]$ : all the weights are non-negative and

$$\sum_{r=1}^{C_p} w_{p,r} - \sum_{l=1}^{C_n} w_{n,l} = 1 \quad (2)$$

To identify the LCDG-model including the numbers of its positive and negative components, we modify the conventional Expectation-Maximization (EM) algorithm to deal with the LCDG.

First the numbers  $C_p - K$ ,  $C_n$  and parameters  $\mathbf{w}$ ,  $\Theta$  (weights, means, and variances) of the positive and negative DG components are estimated with a sequential EM-based initializing algorithm. The goal is to produce a close initial LCDG-approximation of the empirical distribution. Then under the fixed  $C_p$  and  $C_n$ , all other model parameters are refined with an EM algorithm that modifies the conventional one in [11] to account for the components with alternating signs.

## 2.1 Sequential EM-Based Initialization

Sequential EM-based initialization forms an LCDG-approximation of a given empirical marginal gray level distribution using the conventional EM-algorithm [11] adapted to the DGs. At the first stage, the empirical distribution is represented with a mixture of  $K$  positive DGs, each dominant mode being roughly approximated with a single DG. At the second stage, deviations of the empirical distribution from the dominant  $K$ -component mixture are modeled with other, “subordinate” components of the LCDG. The resulting initial LCDG has  $K$  dominant weights, say,  $w_{p,1}, \dots, w_{p,K}$  such that  $\sum_{r=1}^K w_{p,r} = 1$ , and a number of subordinate weights of smaller values such that  $\sum_{r=K+1}^{C_p} w_{p,r} - \sum_{l=1}^{C_n} w_{n,l} = 0$ .

The subordinate components are determined as follows. The positive and negative deviations of the empirical distribution from the dominant mixture are separated and scaled up to form two new “empirical distributions”. The same conventional EM algorithm is iteratively exploited to find the subordinate mixtures of positive or negative DGs that approximate best the scaled-up positive or negative deviations, respectively. The sizes  $C_p - K$  and  $C_n$  of these mixtures are found by sequential minimization of the total absolute error between each scaled-up deviation and its mixture model by the number of the components. Then the obtained positive and negative subordinate models are scaled down and then added to the dominant mixture yielding the initial LCDG model of the size  $C = C_p + C_n$ .

## 2.2 Modified EM Algorithm for LCDG

Modified EM algorithm for LCDG maximizes the log-likelihood of the empirical data by the model parameters assuming statistically independent signals:

$$L(\mathbf{w}, \Theta) = \sum_{q \in \mathbf{Q}} f(q) \log p_{\mathbf{w}, \Theta}(q) \quad (3)$$

A local maximum of the log-likelihood in Eq. (3) is given with the EM process extending the one in [11] onto alternating signs of the components. Let  $p_{\mathbf{w}, \Theta}^{[m]}(q) = \sum_{r=1}^{C_p} w_{p,r}^{[m]} \psi(q|\theta_{p,r}^{[m]}) - \sum_{l=1}^{C_n} w_{n,l}^{[m]} \psi(q|\theta_{n,l}^{[m]})$  denote the current LCDG at iteration  $m$ . Relative contributions of each signal  $q \in \mathbf{Q}$  to each positive and negative DG at iteration  $m$  are specified by the respective conditional weights

$$\pi_{\mathbf{p}}^{[m]}(r|q) = \frac{w_{\mathbf{p},r}^{[m]} \psi(q|\theta_{\mathbf{p},r}^{[m]})}{p_{\mathbf{w},\Theta}^{[m]}(q)}; \quad \pi_{\mathbf{n}}^{[m]}(l|q) = \frac{w_{\mathbf{n},l}^{[m]} \psi(q|\theta_{\mathbf{n},l}^{[m]})}{p_{\mathbf{w},\Theta}^{[m]}(q)} \quad (4)$$

such that the following constraints hold:

$$\sum_{r=1}^{C_{\mathbf{p}}} \pi_{\mathbf{p}}^{[m]}(r|q) - \sum_{l=1}^{C_{\mathbf{n}}} \pi_{\mathbf{n}}^{[m]}(l|q) = 1; \quad q = 0, \dots, Q-1 \quad (5)$$

The following two steps iterate until the log-likelihood changes become small:

- E-step**<sup>[m+1]</sup>: Find the weights of Eq. (4) under the fixed parameters  $\mathbf{w}^{[m]}$ ,  $\Theta^{[m]}$  from the previous iteration  $m$ , and
- M-step**<sup>[m+1]</sup>: Find conditional MLEs  $\mathbf{w}^{[m+1]}$ ,  $\Theta^{[m+1]}$  by maximizing  $L(\mathbf{w}, \Theta)$  under the fixed weights of Eq. (4).

Considerations closely similar to those in [11] show this process converges to a local log-likelihood maximum. Let the log-likelihood of Eq. (3) be rewritten in the equivalent form with the constraints of Eq. (5) as unit factors:

$$L(\mathbf{w}^{[m]}, \Theta^{[m]}) = \sum_{q=0}^Q f(q) \left[ \sum_{r=1}^{C_{\mathbf{p}}} \pi_{\mathbf{p}}^{[m]}(r|q) \log p^{[m]}(q) - \sum_{l=1}^{C_{\mathbf{n}}} \pi_{\mathbf{n}}^{[m]}(l|q) \log p^{[m]}(q) \right] \quad (6)$$

Let the terms  $\log p^{[m]}(q)$  in the first and second brackets be replaced with the equal terms  $\log w_{\mathbf{p},r}^{[m]} + \log \psi(q|\theta_{\mathbf{p},r}^{[m]}) - \log \pi_{\mathbf{p}}^{[m]}(r|q)$  and  $\log w_{\mathbf{n},l}^{[m]} + \log \psi(q|\theta_{\mathbf{n},l}^{[m]}) - \log \pi_{\mathbf{n}}^{[m]}(l|q)$ , respectively, which follow from Eq. (4). At the E-step, the conditional Lagrange maximization of the log-likelihood of Eq. (6) under the  $Q$  restrictions of Eq. (5) results just in the weights  $\pi_{\mathbf{p}}^{[m+1]}(r|q)$  and  $\pi_{\mathbf{n}}^{[m+1]}(l|q)$  of Eq. (4) for all  $r = 1, \dots, C_{\mathbf{p}}$ ;  $l = 1, \dots, C_{\mathbf{n}}$  and  $q \in \mathbf{Q}$ . At the M-step, the DG weights  $w_{\mathbf{p},r}^{[m+1]} = \sum_{q \in \mathbf{Q}} f(q) \pi_{\mathbf{p}}^{[m+1]}(r|q)$  and  $w_{\mathbf{n},l}^{[m+1]} = \sum_{q \in \mathbf{Q}} f(q) \pi_{\mathbf{n}}^{[m+1]}(l|q)$  follow from the conditional Lagrange maximization of the log-likelihood in Eq. (6) under the restriction of Eq. (2) and the fixed conditional weights of Eq. (4). Under these latter, the conventional MLEs of the parameters of each DG stem from maximizing the log-likelihood after each difference of the cumulative Gaussians is replaced with its close approximation with the Gaussian density (below “c” stands for “p” or “n”, respectively):

$$\mu_{\mathbf{c},r}^{[m+1]} = \frac{1}{w_{\mathbf{c},r}^{[m+1]}} \sum_{q \in \mathbf{Q}} q \cdot f(q) \pi_{\mathbf{c}}^{[m+1]}(r|q)$$

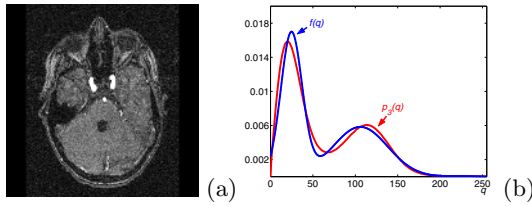
$$(\sigma_{\mathbf{c},r}^{[m+1]})^2 = \frac{1}{w_{\mathbf{c},r}^{[m+1]}} \sum_{q \in \mathbf{Q}} (q - \mu_{\mathbf{c},i}^{[m+1]})^2 \cdot f(q) \pi_{\mathbf{c}}^{[m+1]}(r|q)$$

This modified EM-algorithm is valid until the weights  $\mathbf{w}$  are strictly positive. The iterations should be terminated when the log-likelihood of Eq. (3) does not change or begins to decrease due to accumulation of rounding errors.

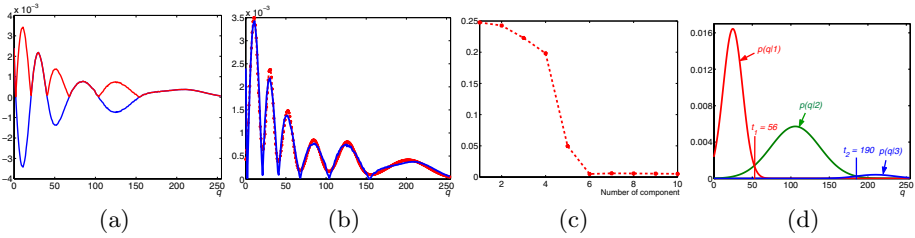
The final mixed LCDG-model  $p_{\mathbf{C}}(q)$  is partitioned into the  $K$  LCDG-submodels  $P_{[k]} = [p(q|k) : q \in \mathbf{Q}]$ , one per class  $k = 1, \dots, K$ , by associating the subordinate DGs with the dominant terms so that the misclassification rate is minimal.

### 3 Experimental Results

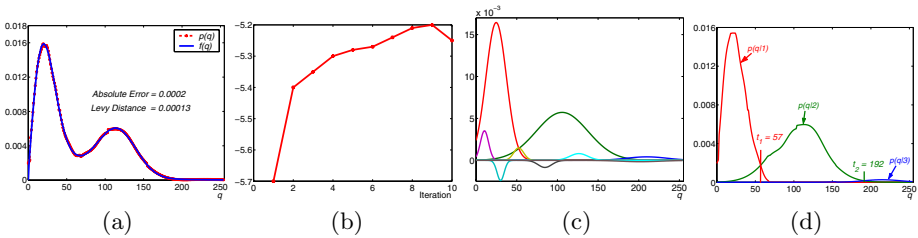
Experiments were conducted with the TOF-MRA images acquired with the Picker 1.5T Edge MRI scanner having spatial resolution of  $0.43 \times 0.43 \times 1.0$  mm. The size of each 3D data set is  $512 \times 512 \times 93$ . The TOF-MRA images contain three classes ( $K = 3$ ), namely, darker bones and fat, brain tissues, and brighter blood vessels. A typical TOF-MRA slice, its empirical marginal gray level distribution  $f(q)$ , and the initial 3-component Gaussian dominant mixture  $p_3(q)$  are shown in Fig. 1. Figure 2 illustrates basic stages of our sequential EM-



**Fig. 1.** Typical TOF-MRA scan slice (a) and deviations between the empirical distribution  $f(q)$  and the dominant 3-component mixture  $p_3(q)$  (b)

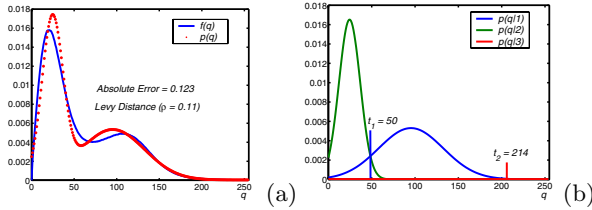


**Fig. 2.** Deviations and absolute deviations between  $f(q)$  and  $p_3(q)$  (a), the mixture model (b) of the absolute deviations in (a), the absolute error (c) as a function of the number of Gaussians approximating the scaled-up absolute deviations in (a), and the initial estimated LCDG-models for each class (d)

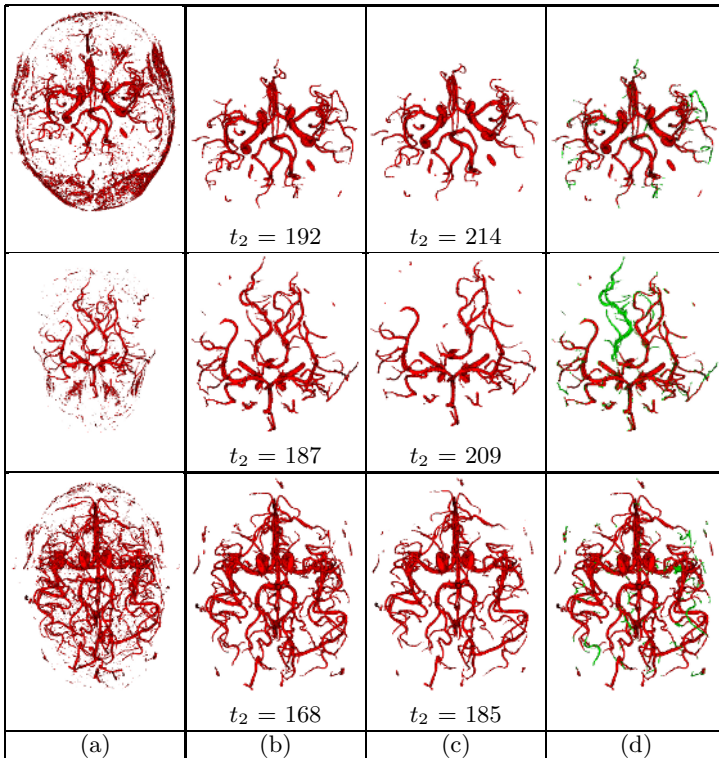


**Fig. 3.** Final 3-class LCDG-model overlaying the empirical density (a), the log-likelihood dynamics (b) for the refining EM-iterations, the refined model components (c), and the class LCDG-models (d)

based initialization by showing the scaled-up alternating and absolute deviations  $f(q) - p_3(q)$ , the best mixture model estimated for the absolute deviations (these six Gaussian components give the minimum approximation error), and the initial LCDG-models for each class. The scaling makes the sums of the positive or absolute negative deviations for  $q = 0, \dots, Q - 1$  equal to one. Figure 3



**Fig. 4.** Wilson-Noble's model [9]: the estimated distribution (a) and the class models (b)



**Fig. 5.** Each row relates to one patient: our segmentation before (a) and after (b) noise and small fat voxels are eliminated with the connectivity filter, the Wilson-Noble's segmentation (c) after the connectivity filter, and the differences between (b) and (c); the green voxels are missed by the Wilson-Noble's approach and the red ones are detected by the both approaches

presents the final LCDG-model after refining the initial one with the modified EM-algorithm and shows successive changes of the log-likelihood at the refinement iterations. The final LCDG-models of each class are obtained with the best separation thresholds  $t_1 = 57$  and  $t_2 = 192$ . First nine refining iterations increase the log-likelihood from  $-5.7$  to  $-5.2$ .

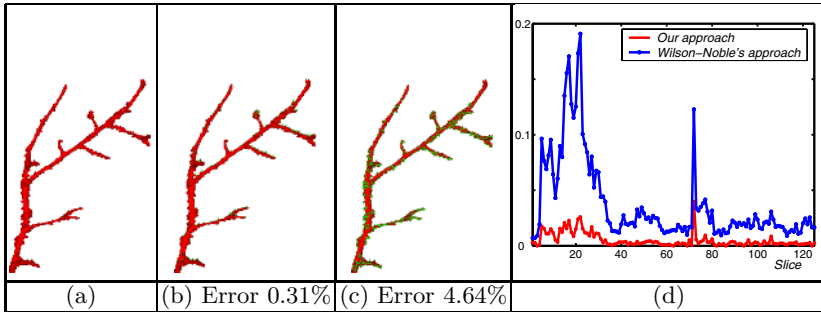
To highlight the advantages of our approach over the existing one, Fig. 4 shows results obtained with the model of Wilson and Noble [9]. To measure the estimation quality, we use the Levy distance between two distributions [12] and the absolute error. The Levy distance between the empirical distribution and its estimated model is 0.11 and 0.00013 and the absolute errors are 0.123 and 0.0002 for the Wilson-Noble's and our approach, respectively. The larger Levy distance and absolute error indicate the notably worse approximation which strongly affects the accuracy of separating the blood vessels from the background. Because of a typically higher separation threshold, e.g.  $t_2 = 214$  versus our  $t_2 = 192$  in this particular example, the Wilson-Noble's approach misses some blood vessels, as shown in Fig. 5.

Both the approaches have been compared on 50 data sets. Results of the three tests are depicted in Fig. 5. As the first column, (a), suggests, TOF-MRA is sensitive to tissues like subcutaneous fat with a short T1 response that may obscure the blood vessels in the segmented volume. To eliminate them, the volume is processed with an automatic connectivity filter which selects the largest connected tree structures using a 3D volume growing algorithm [13]. The results after applying such a filter to our and Wilson-Noble's segmentation in Fig. 5 show that the latter approach fails to detect sizeable fractions of the vascular trees which are validated by the expert (radiologist) that the green parts which are detected by our approaches follow the topology of the brain vascular tree.

## 4 Validation

It is very difficult to get accurate manually segmented complete vasculatures to validate our algorithm. Thus to evaluate its performance, we have created a wooden phantom shown in Fig. 6(a) with topology similar to the blood vessels. Furthermore, the phantom mimics bifurcations, zero and high curvature that exist in any vasculature system, and it has a varying radius to simulate both large and small blood vessels. The phantom was scanned by CT and then manually segmented to obtain the ground truth. The blood vessel and non-vessel signals for the phantom are generated according to the distribution  $p(q|3)$  and  $p(q|1)$ ,  $p(q|2)$ , respectively, in Fig. 3(d) using the inverse mapping methods. The resulting phantom's histogram was similar to that in Fig. 3(a).

Let the total segmentation error be a percentage of erroneous voxels with respect to the overall number of voxels in the manually segmented 3D phantom. Figure 6 shows our approach is 15 times more accurate than the Wilson-Noble's one (the total errors 0.31% and 4.64%, respectively). The error constituents per each 2D slice for both the approaches are also plotted in Fig. 6.



**Fig. 6.** The 3D geometrical phantom (a), our (b) and Wilson-Noble's (c) segmentation, and total errors per each phantom's slice for both the approaches (d)

## 5 Conclusions

We presented a new statistical approach to find blood vessels in multi-modal TOF-MRA images. The LCDG-model accurately approximates the empirical marginal gray level distribution yielding the high quality segmentation. The accuracy of our approach is validated using a specially designed 3D geometrical phantom.

Our present implementation on C++ programming language using a single 2.4 GHZ Pentium 4 CPU with 512 MB RAM takes about 49 *sec* for 93 TOF-MRA slices of size 512x512 pixels each.

The LCDG-model is also suitable for segmenting PC-MRA and CTA images which are not presented due to space limitations.

## References

1. Health Resources. Patient Resources, *Disorder of the Month. Cerebrovascular Disease. Neurosurgery: On-Call* [serial online], July 1999.
2. V. Caselles, R. Kimmel, and G. Sapiro., "Geodesic active contours," *Int. J. Computer Vision*, vol. 22, pp. 61–79, 1997.
3. T. McInerney and D. Terzopoulos, "Medical image segmentation using topologically adaptable surface", *Proc. CVRMed-MRCAS'97*, pp. 23–32, 1997
4. L. M. Lorigo, O. D. Faugeras, W. E. L. Grimson, and R. Keriven, "Curves: Curve evolution for vessel segmentation", *Medical Image Analysis*, vol. 5, pp. 195–206, 2001.
5. O. Wink, W. J. Niessen, and M. A. Viergever, "Fast delineation and visualization of vessels in 3-D angiographic images," *IEEE Trans. Med. Imaging*, vol. 19, pp. 337–346, 2000.
6. T. Deschamps and L. D. Cohen, "Fast extraction of tubular and tree 3D surfaces with front propagation methods", *Proc. 16<sup>th</sup> ICPR*, pp. 731–734, 2002.
7. R. Manniesing and W. Niessen, "Local speed functions in level set based vessel segmentation", *Proc. MICCAI'04*, pp. 475–482, 2004.



8. M. Descotesux, L. Collins, and K. Siddiqi, "Geometric flows for segmenting vasculature in MRI: Theory and validation", *Proc. MICCAI'04*, pp. 500–507, 2004.
9. D. L. Wilson and J. A. Noble, "An adaptive segmentation algorithm for time-of-flight MRA data", *IEEE Trans. Med. Imaging*, vol. 18, pp. 938–945, 1999.
10. D. Nain, A. Yezzi, and G. Turk, "Vessels segmentation using a shape driven flow", *Proc. MICCAI'04*, pp. 51–59, 2004.
11. M. I. Schlesinger and V. Hlavac, *Ten Lectures on Statistical and Structural Pattern Recognition*, Kluwer Academic, Dordrecht, 2002.
12. J. W. Lamperti, *Probability*, J. Wiley & Sons, New York, 1996.
13. A. A. Farag, A. El-Baz, and G. Gimel'farb, "Density estimation using modified Expectation Maximization for a linear combination of Gaussians", *Proc. IEEE Int. Conf. Image Processing, Oct. 24–27, 2004, Singapore*, IEEE CS Press: Los Alamitos, vol. 1, pp. 194–197, 2004.