Cerebrovascular segmentation from TOF using stochastic models

M. Sabry Hassouna a,*, A.A. Farag a, Stephen Hushek b, Thomas Moriarty c

a Computer Vision and Image Processing Laboratory, University of Louisville, Louisville, KY 40292, USA
b iMRI Department, Norton Hospital, Louisville, KY 40202, USA
c Department of Neurological Surgery, University of Louisville, Lutz Hall, Room 415, 2301 South 3rd Street, Louisville, KY 40292, USA

Received 15 September 2003; received in revised form 28 July 2004; accepted 16 November 2004
Available online 12 May 2005

Abstract

In this paper, we present an automatic statistical approach for extracting 3D blood vessels from time-of-flight (TOF) magnetic resonance angiography (MRA) data. The voxels of the dataset are classified as either blood vessels or background noise. The observed volume data is modeled by two stochastic processes. The low level process characterizes the intensity distribution of the data, while the high level process characterizes their statistical dependence among neighboring voxels. The low level process of the background signal is modeled by a finite mixture of one Rayleigh and two normal distributions, while the blood vessels are modeled by one normal distribution. The parameters of the low level process are estimated using the expectation maximization (EM) algorithm. Since the convergence of the EM is sensitive to the initial estimate of the model parameters, an automatic method for parameter initialization, based on histogram analysis, is provided. To improve the quality of segmentation achieved by the proposed low level model especially in the regions of significantly vascular signal loss, the high level process is modeled as a Markov random field (MRF). Since MRF is sensitive to edges and the intracranial vessels represent roughly 5% of the intracranial volume, 2D MRF will destroy most of the small and medium sized vessels. Therefore, to reduce this limitation, we employed 3D MRF, whose parameters are estimated using the maximum pseudo likelihood estimator (MPLE), which converges to the true likelihood under large lattice. Our proposed model exhibits a good fit to the clinical data and is extensively tested on different synthetic vessel phantoms and several 2D/3D TOF datasets acquired from two different MRI scanners. Experimental results showed that the proposed model provides good quality of segmentation and is capable of delineating vessels down to 3 voxel diameters.

Keywords: Magnetic resonance angiography; Time-of-flight; Expectation maximization algorithm; Markov random fields; Statistical segmentation

1. Introduction

Large numbers of people suffer a major cerebrovascular event, usually a stroke, each year. Serious types of vascular diseases such as carotid stenosis, aneurysms, and arterio-venous malformations (AVM) may lead to brain stroke unless they are detected at early stages. Stenosis is a narrowing of the artery which results from a partial or complete blockage of blood supply. Aneurysms are balloons of blood that occurs due to weakness in the arterial wall. The rupture of an aneurysm can cause severe headaches or even a life-threatening coma. AVM’s are abnormal connections of an artery, vein, or both, which deprive the tissue of its normal blood supply. Therefore, accurate cerebrovascular segmentation is the key to accurate diagnoses as well as endovascular treatment.

Magnetic resonance angiography (MRA) is a non-invasive MRI-based flow imaging technique. Its wide variety of acquisition sequences and techniques, beside
its ability to provide detailed images of blood vessels, enabled its use in the diagnosis and surgical planning of the aforementioned diseases. There are three techniques commonly used in performing MRA: time-of-flight (TOF) angiography, phase contrast angiography (PCA), and contrast enhanced MRA (CE-MRA). Both TOF and PCA utilize the flowing blood as an inherent contrast medium, and as such, can be considered non-invasive techniques, while CE-MRA requires the injection of a paramagnetic substance (commonly gadolinium), which provides contrast upon the introduction into the circulatory system. PCA exploits the change in phase of the transverse magnetization as flowing spins move through a magnetic field gradient. It also provides good background signal suppression and can quantify the flow velocity vectors for each voxel. On the other hand, TOF relies on the difference in the amplitude of longitudinal magnetization between flowing and static spins. The TOF technique is not as quantitative but it is widely used clinically because it is fast and provides high contrast images, which is the main motivation behind our work.

A variety of techniques have been proposed for segmenting blood vessels from MRA. Most of the 2D approaches are not applicable to 3D images. 3D techniques can be classified under the following categories: scale space analysis, deformable models, statistical models, and Hybrid methods.

### 1.1. Scale space analysis

In multiscale filtering, each image is convolved with a series of Gaussian filters at different scales. The eigenvalues of the Hessian matrix at each voxel in the image is analyzed to determine whether it belongs to blood vessels or a background noise. The output of this filter is used to define an enhanced set of images in which blood vessels are brightened, while background noise and planar structures such as skin are darkened. Enhanced images are either visualized directly (Frangi et al., 1998), thresholded (Sato et al., 1998), or segmented using an active contour method (Lorenz et al., 1997). The eigenvalues are used to define a candidate set of voxels which could belong to the centerlines of the vessels (Krissian et al., 1999, 1998). Multiscale response functions are evaluated at each of those voxels to determine the likelihood that the voxel is a vessel of various diameters. The maximal response over all choices of diameters is retained at each voxel. Finally, a surface model of the entire vascular structure is reconstructed from both the centerlines and diameters. As a different scale space approach, vessel centerlines are assumed to be very bright and are detected as intensity ridges (Aylward et al., 1996). The width of the vessel is then determined by a multiscale response function.

### 1.2. Deformable models

The main idea behind deformable model approaches is that an initial boundary estimate of the vessel is deformed iteratively to optimize an energy function which depends on the image gradient information and the smoothness of the surface (Caselles et al., 1997). Topologically adaptive surfaces (McInerney and Terzopoulos, 1997) are a variant of the classical deformable models but have an efficient topologically adaptable property for segmenting intracranial vasculature. Geodesic active contours were proposed to segment MRA speed images (Lorigo et al., 2001), where the contour is implemented using the level set methods to offer flexible topological adaptability, which has been extended to be locally more adaptable according to the properties of local geometrical structures such as the eigenvalues of the tensor (Wink et al., 2000). Deschamps and Cohen (2002) presented a fast approach for vessel surface segmentation by inflating a balloon from a user given single point utilizing fast marching methods.

### 1.3. Statistical models

Wilson and Noble (1999) developed a statistical model for extracting blood vessels from TOF data, based on the physical model of blood flow. Two different statistical models for segmenting PCA are suggested to provide a single global threshold (Chung and Noble, 1999) and an adaptive local threshold (Hassouna et al., 2002). Both speed and phase information provided by PCA are fused together to enhance the vessel segmentation (Chung et al., 2002) especially in the nearby of an aneurysm where the signal is very low.

### 1.4. Hybrid methods

Blood vessels are extracted iteratively from rotational angiography by combining a Gaussian statistical model with the maximum intensity projection (MIP) images acquired at three orthogonal directions (Gan et al., 2004). The MIP Z-buffer is segmented using a continuity criterion to generate candidate sets of seed voxels, which are then coupled with a global threshold to extract the whole tree using region growing (Parker et al., 2000). The accuracy of the MIP Z-buffer technique is later studied (Chapman et al., 2004). Vessels are detected by cylinder matching (Reuzé et al., 1993; Hernandez-Hoyos et al., 1999). The method is based on minimizing the inertia moments of a cylinder and a priori knowledge of the intensity profiles in and at the edge of a vessel. A more generalized technique approximating the vessel cross-section by a polygon has been suggested (Verdonck et al., 1995). Continuity and orientation between consecutive slices are used to calculate a locally optimal shape for the polygon with good accuracy. Summers
et al. (1997) proposed an octree decomposition of a velocity field image of PCA in order to find an optimal tessellation. Each block of the octree contains at most one feature defined by a gray level function and orientation. Masutani et al. (1998) proposed another method, where the vessels initial shape is extracted by thresholding followed by a region growing to extract locally smooth surface by using binary mathematical morphological operations. As another level set approach, the speed function of the level set surface evolution is controlled by the intensity distribution of the data (Farag et al., 2004). A recursive hybrid segmentation framework has been proposed by Chen and Metaxas (2000, 2003), that combines the Gibbs prior model, marching cubes, and deformable models. In the first step, the Gibbs model is used to estimate the object boundaries using region information from 2D image slices. The estimated boundaries are then used to construct a 3D mesh using marching cubes, which specifies the initial geometry of a deformable model. In the second step, the deformable model deforms and fits to the data under the influence of 3D image gradient forces. This recursive approach usually does not require more than two iterations to give a good estimate of the desired object’s surface.

Existing vascular segmentation methods have at least one of the following limitations:

1. Rely on the intensity gradient field to estimate the vessel boundary, however, in practice, gradient values are not sufficiently high in the low or complex flow regions.
2. The vessel cross-section is assumed to be circular, which is true for healthy arteries but not in the nearby of a stenosis or an aneurysm.
3. Assume that the intensity distribution of each structure in the image has a Gaussian distribution, which is not necessarily true, resulting in an error between the proposed model and the clinical data.
4. Suitable only for specific modality.
5. Require user interaction to either insert a seed inside a vessel of interest or select its ending points.
6. Have many tuning parameters whose estimation process is hard or not applicable.
7. Computationally expensive.

In this paper, we present a new statistical approach for segmenting 3D cerebrovascular system from TOF-MRA data, which extends our prior work (Hassouna et al., 2003). The voxels of the observed dataset are classified as either background or blood vessels classes. Each class is modeled by a low level stochastic process that describes its intensity distribution across the volume and a high level stochastic process that describes its statistical dependence among neighboring voxels. The low level process of the background signal is modeled by a finite mixture of one Rayleigh and two normal distributions, while the blood vessels are modeled by one normal distribution. The parameters of the proposed mixture density model for the low level processes are estimated using the expectation maximization (EM) algorithm. The convergence of the EM algorithm is sensitive to the initial estimate of the parameters. Therefore, we present an automatic method based on histogram analysis to find a good initial estimate of them. To improve the quality of the statistical segmentation, spatial contextual information has been incorporated through 3D Markov random field (MRF), whose parameters are estimated using maximum pseudo likelihood method. The experimental results on different synthetic vessel phantoms and several 2D/3D TOF datasets acquired from two different MRI scanners showed that the proposed model provides good quality of segmentation and is capable of delineating vessels down to 3 voxel diameters.

This paper is organized as follows. In Section 2, we give a quick overview on the different TOF acquisition techniques, a derivation of the low level model for both blood vessels and background signal, and finally show how to estimate the parameters of that model. In Section 3, we give a brief introduction to MRF models and its usage in image segmentation, and then show how to combine it with our low level model to improve segmentation results. We validate our method using different synthetic phantoms in Section 4. In Section 5, we present our segmentation results on clinical datasets. In Section 6, we conclude with a discussion of the current and future work.

2. TOF statistical segmentation

In this section, we derive the low level model of both blood vessels and background signal and then estimate their parameters using the EM algorithm. First, we will give a quick overview on the different TOF acquisition techniques.

2.1. TOF acquisition techniques

The basic idea behind TOF is that, the longitudinal magnetization of stationary tissues in a slice are saturated by multiple radio frequency (RF) pulses, whereas the moving spins are not, which results in a stronger signal from flow while entering the slice. TOF acquisitions can be classified into three major categories, a 2D technique, where thin slices are acquired sequentially, a 3D technique, where the whole volume is excited simultaneously, and will then be subdivided into thin slices by using an additional phase encoding scheme. 3D techniques provide higher SNR and resolution than 2D techniques, but the contrast of vessel/background is
getting smaller with the spins penetrating through the imaging volume, this is why they are used in combination with fast flow situations. On the other hand, 2D techniques provide high contrast because the moving spins are fully relaxed, so they are used to visualize slower flow. The last technique is MOTSA (multiple overlapped thin slices acquisition) technique, where several overlapped slabs are acquired to take the advantages of both 2D and 3D acquisitions.

2.2. Low level model

To the best of our knowledge, there is only one related work (Wilson and Noble, 1999) that segments blood vessels from TOF data using a statistical approach. In that study, the TOF data histogram was divided into three regions based on voxel intensity. The lowest intensity region corresponds to cerebrospinal fluid (CSF), bone, and the background air. The middle intensity region corresponds to brain tissues, including both the grey and white matter, and parts of the eyes. The third high intensity region corresponds to subcutaneous fat and arteries. A normal distribution is used to model each of the low and middle intensity regions, while a uniform distribution is used to model the vessels class as shown in Fig. 1(a). In their model, high accuracy was not required because they were looking for large vessels suitable for view selection cues. However, in this research we are interested in reconstructing the whole vasculature, so accuracy is required. We have tested various probability density models for the low intensity range of the TOF histogram of different clinical datasets that are acquired from two different MRI scanners, and found that the Rayleigh distribution provides an accurate fit when compared with the normal distribution proposed by the related work, as shown in Fig. 1(b). Theoretically, the vessel intensities are uniformly distributed over the high intensity range motivated by a physical model of the blood flow (Chung and Noble, 1999), but since the starting point of that range is unknown, it is extended over the whole intensity range.

When we first modeled the vessel class by a uniform distribution over the whole intensity range, we found that the decision level exceeds the expected value found from manual segmentation by 5–15 pixels, which leads to exclusion of part of the small and medium sized blood vessels. Therefore, the assumption that vessels exist in

Fig. 1. (a) Model by Wilson and Noble (1999). (b) Inaccuracy of one Rayleigh and two normal distributions. (c) The proposed model (accurate fitting). (d) Initial histogram of each distribution.
the low intensity range disturbs the fitting process, and results in inaccurate decision level. We found that modeling the vessels class by a normal distribution reduces the absolute error between the observed data histogram and the proposed model and improves the decision level as well. The reason behind these improvements is that the Gaussian peak of the vessels class exists in the upper part of the high intensity region, thus ensuring very low contribution by its long left tail to the low and middle intensity regions.

Modeling the middle intensity region by one normal distribution, leads to an accurate fitting at both ends of the histogram but not at the middle as marked by the circle in Fig. 1(b). To correct such a problem, we can add as many normal distributions to the low and middle intensity regions, which will reduce the absolute error between the model and the observed histogram. However, this will complicate the model as well as the parameter estimation process. Therefore, we will restrict ourselves to only one extra normal distribution, which is shared between both the low and middle intensity regions as shown in Fig. 1(c). The total probability density function of the mixture is given by

\[
\begin{align*}
  f(x) &= w_R f_R(x) + \sum_{l=1}^{3} w_G f_G(x),
\end{align*}
\]

The functions \( f_R(x) \), \( f_G(x) \), \( f_{G1}(x) \), and \( f_{G3}(x) \) are the Rayleigh and normal density functions, respectively. The quantities \( w_R \), \( w_{G1} \), \( w_{G2} \), and \( w_{G3} \) are the class proportions which sum is unity. The probability density function of the Rayleigh and normal distributions are given by Eqs. (2) and (3), respectively:

\[
\begin{align*}
  f_R(x) &= \frac{x}{\beta^2} \exp \left( -\frac{x^2}{2\beta^2} \right), \\
  f_G(x) &= \frac{1}{\sqrt{2\pi}\sigma_G} \exp \left( -\frac{(x - \mu_G)^2}{2\sigma_G^2} \right), \quad l \in [1, 3].
\end{align*}
\]

According to the maximum a posteriori (MAP) classification, a voxel \( x \) belongs to the vessels class if its probability is greater than the background probability

\[
\begin{align*}
  f(G_3|x) > f(R|x) + f(G_1|x) + f(G_2|x),
\end{align*}
\]

which can be rewritten as

\[
\begin{align*}
  w_{G3} f_{G3}(x) > w_R f_R(x) + w_{G1} f_{G1}(x) + w_{G2} f_{G2}(x).
\end{align*}
\]

The class labels of the background mixture components are denoted by \( R \), \( G_1 \), and \( G_2 \), respectively, while \( G_3 \) for the vessels class.

Before applying MAP segmentation, the parameters of each density function of Eq. (5) should be estimated. These parameters are the proportion \( w_R \) and the mode \( \beta \) of the Rayleigh distribution, and the proportion \( w_{G_l} \) mean \( \mu_{G_l} \) and variance \( \sigma_{G_l}^2 \) where \( l \in [1, 3] \) of each Gaussian distribution. We will estimate the eleven parameters using the expectation maximization algorithm (EM) (Dempster et al., 1977; MacLachlan and Krishnan, 1997).

### 2.3. Parameter estimation using the EM algorithm

The EM algorithm is a general method of finding the maximum-likelihood estimate of the parameters of an underlying distribution from a given data set when the data is incomplete or has missing values, which is the class label in our case. The mixture density parameter estimation is one of the most widely used applications of the EM algorithm. The update equations of the parameters of our mixture model are derived in Appendix A. The EM algorithm is an iterative technique that starts with an initial estimate of the model parameters. During its operation it searches for those parameters that maximize the conditional expectation of the log-likelihood function of the mixture distribution, thus, it may converge to local maxima if the initial set of the model parameters are not selected properly (MacLachlan and Krishnan, 1997).

The common approach is to run the EM algorithm more than once starting from different sets of initial parameter values and then select the estimated set that maximizes the conditional expectation, which is computationally expensive in our problem since we are dealing with large scale medical data volume. In addition, convergence is still not guaranteed. Therefore, we developed an automatic method for finding a good initial estimate to those parameters. The initial values of the parameters are set according to Table 1. Let \( h(x) \) be the normalized observed histogram and \( h_{R}^{\text{init}}(x) \) and \( h_{G}^{\text{init}}(x) \) be the initial histograms of the Rayleigh and normal distributions, respectively, as defined by the following equations:

\[
\begin{align*}
  h_{R}^{\text{init}}(x) &= C_R f_R(x) \beta_{R}^{\text{init}}, \\
  h_{G_l}^{\text{init}}(x) &= C_l f_{G_l}(x) \mu_{G_l}^{\text{init}}, \sigma_{G_l}^{2\text{init}}, \quad l \in [1, 3],
\end{align*}
\]

where

\[
\begin{align*}
  C_R &= \frac{h(I_{\text{peak}})}{f_R(I_{\text{peak}}) \beta_{\text{R}}^{\text{init}}}, \\
  C_l &= \frac{h_{\text{init}}(\mu_{G_l}^{\text{init}})}{f_{G_l}(\mu_{G_l}^{\text{init}}) \sigma_{G_l}^{2\text{init}}}. \quad (9)
\end{align*}
\]
The constants $C_r$ and $C_t$ ensure that the peaks of the initial histograms have the same height as $h(x)$. Let $I_{peak1}$ and $I_{peak2}$ be the intensities at which $h(x)$ achieves its two global peaks, and $I_{min}$ be the intensity at which $h(x)$ achieves its minimum value between those peaks as shown in Fig. 1(d). $I_{peak2}$ and $I_{min}$ can be achieved by smoothing $h(x)$ couple of times using an average filter. The residual histograms are calculated according to:

\[
\hat{h}_{res1}^{init}(x) = h(x) - h_{R}^{init}(x) \mid u(x - I_{peak1}),
\]

\[
\hat{h}_{res2}^{init}(x) = h_{res1}^{init}(x) - h_{G2}^{init}(x) \mid u(x - I_{min}).
\]

The unit step ensures that $\mu_{G1}^{init}$ and $\mu_{G2}^{init}$ have values greater than $I_{peak1}$ and $I_{min}$, respectively. Once the parameters are estimated, we carry out the maximum likelihood segmentation (ML) using Eq. (5).

The signature of both 2D and 3D TOF volume histogram is the same except for the middle intensity region. The reason is, in 2D acquisition, the moving spins experiences only a very few excitation pulses as it flows through the slice, so that most of the signal from stationary tissues are suppressed. In 3D, as we go deeper into the volume the blood becomes more saturated, so we use small tip angle to preserve the signal from blood, which will also preserve signal from tissues, so the middle intensity region in 3D has large peak than that of 2D. All the formulas presented in Table 1 are applicable for both acquisitions except for $\mu_{G1}^{init}$, which is set to the intensity at which $h_{res1}^{init}$ achieves maximum for 2D case, and $I_{min}$ for 3D one. It is worth noting that the proposed formulas of Table 1 are not dedicated to any specific dataset, and is tested heavily on many clinical datasets and synthetic phantoms as will be seen later.

### 3. Enhancing segmentation

Although the low level model provides a good fit to the observed data, we may still have some misclassified voxels because classification is based only on voxel intensity. For example, some vessel voxels may be classified as non-vessel class. This happens in regions with significant vascular signal loss due to complicated flow conditions including slowly and turbulent blood flow, which is a typical problem with TOF acquisitions.

Also, some background voxels may be classified as blood vessels class, when the noise has high intensity value. We can improve the segmentation process by taking into account the spatial information (statistical dependence) among neighboring voxels. The concept of contextual information enters the segmentation process though Markov Random Field (MRF) models (Besag, 1974; Dubes and Jain, 1989; Dubes et al., 1990; Geman and Geman, 1984), which serve as a prior distribution of the true label of the class of interest. MRF models are appropriate because they specify the local properties of image regions through Markovian property; the true label of a voxel is dependent on the labels of the spatially neighboring voxels. As noted by Dubes and Jain (1989), MRF model need not be an accurate model of the true labels to have good quality of segmentation, but it is a convenient model of introducing context, or dependence among neighboring voxels. An introduction to the theory of MRF is presented in Appendix B.

#### 3.1. MRF-based segmentation

The observed dataset is modeled as a composite of two random processes, low level process $Y$, which characterizes the statistical distribution of the data based on their intensity and a high level process $X$, which characterizes the statistical dependence among neighboring voxels. Both the two processes are random fields defined on the lattice $S$, which is the MRA volume. Let $Y = \{Y_1, Y_2, \ldots, Y_N\}$ be a set of observed random variables, where $Y_s$ is the random variable representing the intensity of voxel $s$. Assume that $X = \{X_1, X_2, \ldots, X_N\}$ is a MRF, where $X_s$ takes a value from the label set

### Table 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{G1}^{init}$</td>
<td>$I_{min}$</td>
</tr>
<tr>
<td>$\mu_{G2}^{init}$</td>
<td>$I_{peak2}$</td>
</tr>
<tr>
<td>$\mu_{G3}^{init}$</td>
<td>Calculated using the MLE from the last 3% of the high intensity data of the observed histogram</td>
</tr>
<tr>
<td>$\mu_{init}$</td>
<td>$I_{peak1}$, the value at which Rayleigh achieves maximum value</td>
</tr>
<tr>
<td>$\alpha^{2}_{G1}^{init}$</td>
<td>Calculated using MLE from the samples in the region $[\mu_{G1}^{init} - \Delta, \mu_{G1}^{init} + \Delta]$ of $h_{res1}^{init}(x)$, where $\Delta = (\mu_{G1}^{init} - I_{peak1}) / 2$</td>
</tr>
<tr>
<td>$\alpha^{2}_{G2}^{init}$</td>
<td>Calculated using MLE from the samples in the region $[\mu_{G2}^{init} - \Delta, \mu_{G2}^{init} + \Delta]$ of $h_{res2}^{init}(x)$, where $\Delta = (\mu_{G2}^{init} - \mu_{G1}^{init})$</td>
</tr>
<tr>
<td>$\alpha^{2}_{G3}^{init}$</td>
<td>Calculated using MLE from the last 3% of the high intensity data of the observed histogram</td>
</tr>
<tr>
<td>$w_{G1}^{init}$</td>
<td>Set to 3% because the proportion of the vessels in the volume ranges from 1% to 5%</td>
</tr>
<tr>
<td>$w_{G2}^{init}$</td>
<td>The area of $h(x)$ covered by $h_{res}^{init}(x)$</td>
</tr>
<tr>
<td>$w_{G3}^{init}$</td>
<td>The area of $h(x)$ covered by $h_{res}^{init}(x)$</td>
</tr>
<tr>
<td>$w_{G1}^{init}$</td>
<td>$1 - w_{G1}^{init} - w_{G2}^{init} - w_{G3}^{init}$</td>
</tr>
</tbody>
</table>
\( L = \{V,B\} \), where \( V \) denotes the vessels class and \( B \) is the background class. Therefore, given a set of observed feature vectors \( Y = y \), and the contextual information modeled by a MRF, \( P(X = x) \), the problem is to find the optimal estimate of the true labeling \( x^* \). The current trend is to combine both these steps using Bayesian formulation and then we use maximum a posteriori (MAP) method to choose the estimate \( \hat{x} \) that maximizes the posterior probability of \( p(X = \hat{x}|Y = y) \). The Bayesian formulation is given by

\[
p(X = x | Y = y) = \frac{p(Y = y | X = x)P(X = x)}{p(Y = y)}. \tag{12}
\]

The term \( p(X = x | Y = y) \) is the posterior probability of the true labeled volume given the observed one. The term \( p(Y = y | X = x) \) is the probability of the observed data given the true labels, which is assumed to be conditionally independent at each voxel in the volume. The term \( p(Y = y) \) is constant. If we take the log function of Eq. (12), we get

\[
\log p(X = x | Y = y) \propto \log p(Y = y | X = x) + \log P(X = x). \tag{13}
\]

The right hand side of this equation consists of two terms, the first term from left is the low level process which is modeled as seen in the previous section, and the second term is the high level process, which is given by the Gibbs distribution according to Hammersley–Clifford Theorem (Hammersley and Clifford, 1971)

\[
p(X = x) = \frac{e^{-U(x)}}{Z}, \quad Z = \sum_{x \in \Omega} e^{-U(x)}. \tag{14}
\]

The energy function is denoted by \( U(x) \), where the higher the energy of a configuration \( x \), the lower the probability of its occurrence. The space of all possible labeling is denoted by \( \Omega \). The denominator \( Z \), is called the partition function, which is a normalizing constant obtained by summing the numerator over all possible configuration \( x \).

### 3.2. High level model

It is known that MRF is sensitive to edges. Therefore, applying 2D MRF to each slice in the volume will destroy most of the small and medium sized vessels because blood vessels have small cross-sections and represent roughly 5% of the intracranial volume. To reduce this limitation, we employed 3D MRF to exploit the information provided by neighboring slices and hence increase the probability of correctly classifying blood vessels when most of the neighboring voxels belong to the background class. For example, consider an ideal blood vessel that passes through three slices \( S_{i-1}, S_i \) and \( S_{i+1} \) as shown in Fig. 2(a), where the voxels of the blood vessels and background signal are represented by black and white dots, respectively. Let us assume that during MRA acquisition process, noise is added to the slice \( S_i \) and turned all vessel voxels into background ones except for the middle voxel \( s \) of slice \( S_i \) as shown in Fig. 2(b). Applying any MAP optimization technique to the slice \( S_i \), using 2D neighborhood system will diminish \( s \) and classify it as a background signal because all in-plane neighboring voxels to \( s \) belong to the background class. Therefore, the probability that \( s \) belongs to the background class is higher than that of the vessels class. But, if we extend the neighborhood system to be 3D (cube) around \( s \), then the voxels of blood vessels of neighboring slices to \( S_i \) will increase, and hence the probability of correct classification to the vessels class.

The high level model is constructed as follows:

1. **Second order neighborhood of pair-wise interaction.**
2. **The neighborhood \( \eta \) is a cube of size \( 3 \times 3 \times 3 \).**
3. **The cliques are of order 2 and formed by the voxel \( s \) and its 26 nearest neighbors.**

We used the isotropic Multi-Level Logistic (MLL) (Derin and Elliott, 1987) as our MRF model. The energy function of this model is given by

\[
U(x_r) = \sum_{r \in \eta} \beta_{sr} V(x_s, x_r), \quad \forall r \in \eta, \tag{15}
\]

where the potential functions are defined as

\[
V(x_s, x_r) = \begin{cases} 
1, & x_s = x_r, \\
0, & \text{elsewhere}.
\end{cases} \tag{16}
\]

The parameter \( \beta_{sr} \) describes the strength of the interaction between pair-wise neighboring voxels. In our model, we set all \( \beta_{sr} \) to the same value such that the Markov prior model for the vessels and background classes is directly proportional to the multiplier of the number of adjacent vessels and background voxels, respectively. (4) The model parameter is estimated using MPLE method. (5) The true label of the volume is
estimated using the iterated conditional modes (ICM) algorithm (Besag, 1986).

3.3. Maximum pseudo likelihood estimator

This method was proposed by Besag (1975, 1977), where the true likelihood of the volume is approximated by the product of the conditional likelihood at each voxel. It is proved that under large lattice, pseudo likelihood converges to the true one with probability 1.0 (Geman and Graffigne, 1986), which holds for our model because we are dealing with volume rather than a slice. The pseudo likelihood is given by

$$ l(X|\theta) = \log \prod_{s \in S} p(X_s = x_s | X_{x_s} = x_{x_s}). $$

Substituting the Gibbs distribution for the conditional probability,

$$ l(X|\theta) = \sum_{x(s)} \log \frac{e^{-U(x_s)}}{e^{-U(x_s)_{\text{P}}} + e^{-U(x_s)_{\text{B}}}}. $$

3.4. High level segmentation

ICM is a computationally feasible alternative to MAP estimation (Besag, 1986). It estimates the class label of a voxel $x_s$ according to the following equation:

$$ x_s = \arg \max P(X_s = x_s | Y_s = y_s), \quad \text{where } x_s \in [V, B]. $$

ICM requires an initial estimate of the true labels, which is provided by the low level segmentation. According to Bayes' theorem

$$ P(X_s = x_s | Y_s = y_s) \propto P(Y_s = y_s | X_s = x_s)P(X_s = x_s). $$

Therefore, the posterior probability of the vessel class is given by

$$ P(X_s = V | Y_s = y_s) \propto f_{G3}(y_s) \exp \left\{ -U(V) \right\}. \tag{21} $$

And the posterior probability of the background class is given by

$$ P(X_s = B | Y_s = y_s) \propto \frac{w_R f_R(y_s) + w_{G1} f_{G1}(y_s) + w_{G2} f_{G2}(y_s)}{w_R + w_{G1} + w_{G2}} \exp \left\{ -U(B) \right\}. \tag{22} $$

ICM is applied recursively until no further changes in the labels of the vessels and background classes.

4. Validation

We may find ground truth segmentation for carotid, aneurysm, or both but not for a complete vasculature because of its complexity and the more levels of details it involves. Therefore, in order to validate our method, we created several synthetic 3D phantoms that mimic bifurcation, zero, and high curvature vessels at different spatial resolution as well as a wooden tree phantom whose ground truth is acquired using CT scan.

4.1. Experiment 1

In this experiment, a wooden tree branch which has similar topology to blood vessels is CT scanned and then manually segmented to be the ground truth (GT) of the vessels as shown in Fig. 3(a). The proportion of the voxels of the GT is set to 5% (similar to real vasculature) by adjusting the size of its containing volume. The labels of the GT are set to $\mu_{G3}$. The background signal is generated as follows: three types of voxels are randomly

![Fig. 3](image-url)(a) Ground truth. (b) Segmentation by the proposed algorithm. Undetected voxels are marked by dark color.)
Fig. 4. (a) Ground truth slice. (b) Raw data slice generated by applying our model to the slice in (a). (c) ML segmentation. (d) MRF segmentation.

Fig. 5. (a) Phantom raw data histogram, mixture fit (solid), and mixture components (dotted). (b) Segmentation improvement of MRF over ML, where the error drops from 0.19% to 0.05%.

Fig. 6. (a) Synthetic vessels with zero curvature (ground truth). (b) ML segmentation. (c) MRF segmentation. (d) Segmentation improvement of MRF over ML, where the error drops from 0.05% to 0.02%.
generated over the non-vessel labels (background volume) at a constant intensity, \( \mu_R, \mu_{G1}, \) and \( \mu_{G2} \) of proportions \( w_R, w_{G1}, \) and \( w_{G2} \), respectively. Thus, the histogram of the ground truth phantom consists of impulses at intensities \( \mu_R, \mu_{G1}, \mu_{G2}, \) and \( \mu_{G3} \). We then reshape it to form the TOF signature histogram as follows: we add three independent normal noise components of zero mean and variance \( \sigma^2_{G1} \) with proportions \( w_{G1} \) to the voxel intensities marked by \( l_{G1} \), which implies \( G_I \sim (\mu_{G1}, \sigma^2_{G1}) \forall I \in [1,3] \). We also replaced each voxel marked with \( \mu_R \) by a Rayleigh noise of mode \( \beta = I_{\text{peak}1} \) to form a Rayleigh distribution. The parameters used in the phantom design are the average values of those parameters extracted by the EM algorithm as applied to clinical data of several patients. Once the phantom is created, we apply our proposed method which includes initial estimate of the model parameters based on histogram analysis, EM parameter estimation, maximum likelihood segmentation, MRF parameter estimation, and finally MRF segmentation as shown in Fig. 3(b), where dark areas represent those voxels that are wrongly classified by our approach as background signal. The error was 0.03%.

Fig. 4(a) shows a slice from the ground truth, which is converted into raw data by applying the low level model as shown in Fig. 4(b). The ML and MRF segmentation by our model is shown in Fig. 4(c) and (d), respectively.

Fig. 5(a) shows the raw data histogram generated from the ground truth phantom after applying our low level model to it. The mixture fit and model components

![Fig. 7](image7.png)

Fig. 7. (a) Ground truth slice. (b) Raw data slice generated by applying our model to the slice of (a). (c) ML segmentation. (d) MRF segmentation.

![Fig. 8](image8.png)

Fig. 8. (a) Synthetic vessels with high curvature (ground truth). (b) Proposed segmentation. (c) Ground truth and proposed segmentation are colored in light and dark colors, respectively. (d) Segmentation improvement of MRF over ML, where the error drops from 0.16% to 0.05%.
Fig. 9. (a) Synthetic vessels bifurcation (ground truth). (b) Proposed segmentation. (c) Ground truth and proposed segmentation are colored in light and dark colors, respectively. (d) Segmentation improvement of MRF over ML, where the error drops from 0.06% to 0.03%.

Fig. 10. Our statistical model fits the clinical data accurately for different patients with 3D acquisition (a), (b), and (c) and 2D acquisition (d).
are shown in solid and dotted line, respectively. Fig. 5(b) shows the improvement of using 3D MRF segmentation over ML segmentation by computing the number of misclassified voxels for each slice in the volume when compared to the known ground truth voxels.

4.2. Experiment 2

In this experiment, we need to determine the lowest spatial resolution of vessels that our method can delineate. Therefore, we validated our method against three different phantoms representing three synthetic vessels with zero curvature, high curvature (tortuous), and bifurcation at different diameters (1, 3, 5, 7, and 11) voxels. We will follow the same procedure presented in experiment 1, where our noise model is added to the ground truth vessel, and then we segment it using the proposed segmentation method. In Fig. 6, we show the results of our method on synthetic vessels with zero curvature. Since ML segmentation takes into account voxel intensity only, it can preserve parts of the vessels down to one voxel wide as long as it has high intensity as shown in Fig. 6(b). However, it did not preserve most of the vessel voxels in the vascular regions that have

![Fig. 11](image_url) Fig. 11. Each row represents a patient: (a) MIP image; (b) segmentation by the proposed model; (c) same as (b) except that small island vessels and noise are filtered using largest connected components.
been corrupted with very low intensity noise similar to the background intensity as shown in Fig. 7(c). On the contrary, MRF segmentation preserved the voxels of such regions as shown in Fig. 7(d), and failed to preserve the one voxel wide vessel as shown in Fig. 6(c).

In Fig. 6(d), we show the segmentation improvement achieved by MRF over ML segmentation by computing the number of misclassified voxels (false-positive and false-negative) for each slice in the volume. We repeated the experiment again on a synthetic phantom of a vessel with high curvature and bifurcation as shown in Figs. 8 and 9, respectively. According to our experimental results on the different synthetic phantoms, we can conclude the following: (1) 3D MRF enhances the segmentation results over the low level process (ML segmentation) in the vascular regions where the signal is too low or has low intensity noise similar to the background intensity. (2) Although 3D MRF is still sensitive to edges, it gives better results than 2D MRF as it exploits the information of adjacent slices. (3) Our method can delineate vessels down to 3 voxel diameters.

5. Results

We have also tested our new segmentation method on several 2D/3D TOF clinical datasets that are acquired from two different 1.5 T (Picker Edge and GE) MRI scanners. The 3D datasets came in two different sizes, 512×512×93 and 512×512×63 with spatial resolution 0.43×0.43×1.0. The size of the 2D datasets is 256×256×60 with spatial resolution 0.78×0.78×1.0. In Fig. 10, we show how the proposed model provides high quality fit to the clinical data for both 2D and 3D acquisitions of different patients.

Fig. 11 shows the segmentation results of the same patients using the proposed method. Vessel surfaces are rendered in 3D using the visualization toolkit (VTK). TOF is sensitive to short T1 tissues such as subcutaneous fat, which appears in the segmented volume obscuring vessels as shown in the first row of Fig. 11(b). Therefore, to eliminate them, we filtered the volume by automatically selecting the largest connected tree structure using 3D region growing algorithm as shown in Fig. 11(c). To show the accuracy of the results, a comparison is done with the maximum intensity projection (MIP) images (Rosnick et al., 1986), as shown in Fig. 11(a). The average processing time taken by our method is ≈5 min on a single 400 MHz processor, Onyx2 SGI supercomputer.

6. Discussion and conclusion

In this paper, we have presented an automated stochastic segmentation method for extracting cerebrovascular blood vessels from TOF-MRA data. The proposed method is based on two stochastic models for the observed data. The blood vessels are modeled by one normal distribution, while the background noise is modeled by a mixture of one Rayleigh and two normal distributions. To improve the quality of segmentation achieved by the proposed low level model, a MRF is used as a high level model to adaptively adjust the local threshold during the extraction of vessel voxels from within the background noise. We chose MRF with 3D neighborhood system to exploit the vascular information provided by adjacent slices and hence increase the probability of detecting blood vessels with small cross-sections.

The parameters of the low level model are estimated by the EM algorithm. The parameters update equations of the Rayleigh distribution are derived since it is barely used in the literature. To ensure the convergence of the EM algorithm, we presented an automatic method based on residual histogram analysis for finding a good estimate of them. This method is applicable to both 2D and 3D acquisitions.

Although there are several MRF parameter estimation techniques such as the coding method (Besag, 1974) and the least square error method (Derin and Elliott, 1987), we selected maximum pseudo likelihood method because it converges to the true likelihood under large lattice, which holds in our case since we are dealing with large data volume. We used the ICM as an optimization method to MAP estimation. Although ICM is a local optimization technique and is function of the initial condition, it gave promising results, which implies that our low level model is quite adequate. For all datasets, the ICM converged in less than 8 iterations.

Our method is validated against several synthetic phantoms representing bifurcation, zero curvature, and high curvature vessels at different diameters. Experiments on those phantoms showed that our method is capable of delineating vessels down to 3 voxel diameters. In addition, the proposed high level model reduced the number of misclassified voxels when compared to the ML segmentation. Our experimental results on different clinical datasets showed that the low level model has high quality of fit to the clinical data, while the overall method provides high quality segmentation when compared with MIP images, which has been proven to be the most popular rendering algorithm for MRA although it is sensitive to high intensity noise. Finally, the only known limitation of our approach is that, it is dedicated only to MRA-TOF data.

In the future, we would like to further enhance the quality of segmentation by combining our statistical model with one of the geometrical models that takes shape into account. In addition, build a virtual angioscopy system suitable for vessel exploration and disease quantification. The components of such a system are
segmentation, centerline generation, and rendering of
the vessels internal views. We are having now an ongo-
ing research on extracting reliable centerlines suitable
for vascular fly-through applications, whose initial
results are available at http://www.cvip.uofl.edu/
skeletons.

Acknowledgments

This work has been supported by the US-Army
Grant DABT60-02-P-0063 and Norton Health Care
System Grants 97-33 and 97-72.

Appendix A. EM parameters update equations

This section shows how to derive the parameters
update equations of our statistical model in its closed
form. The posterior probabilities of the Rayleigh
and Gaussian distribution are given by

\[
 f(R|x_i) = \frac{w_{R,f}(x_i)}{f(x_i)},
\]

\[
 f(G_i|x_i) = \frac{w_{G,f,G_i}(x_i)}{f(x_i)}, \quad l \in [1, 3].
\]

The fundamental equation of the EM expectation step
for finite mixture of densities is given by

\[
 Q(\Theta^{k+1}, \Theta^k) = \sum_{i=1}^N \sum_{l=1}^M \log \left( \frac{w_{l}^{k+1} f_l(x_i|\Theta^{k+1})}{\sum_{l=1}^M w_{l}^{k+1} f_l(x_i)} \right) \frac{\log (w_{l}^{k+1} f_l(x_i|\Theta^k))}{f(l|x_i, \Theta^k)},
\]

where

\[
 f(l|x_i, \Theta^k) = \frac{w_{l}^{k} f_l(x_i)}{\sum_{l=1}^M w_{l}^{k} f_l(x_i)},
\]

where \( Q \) is the expected value of the log-likelihood function
of the mixture distribution, \( \Theta^k \) is the mixture
parameters at iteration \( k \), \( M \) is the total number of class-
es, which is four in our case, and \( N \) is the total number
of voxels in the data volume. Eq. (A.2) can be rewritten
as the sum of two independent terms:

\[
 Q = \sum_{i=1}^N \sum_{l=1}^M \log \left( w_{l}^{k+1} f_l(x_i|\Theta^{k+1}) \right) \frac{\log (w_{l}^{k} f_l(x_i|\Theta^k))}{f(l|x_i, \Theta^k)}
 + \sum_{i=1}^N \sum_{l=1}^M \log(f_l(x_i|\Theta^{k+1}) \frac{\log (w_{l}^{k+1} f_l(x_i|\Theta^k))}{f(l|x_i, \Theta^k)}.
\]

The maximization step of the EM finds the mixture
parameters by maximizing each term of Eq. (A.4)
independently. Let us maximize the term containing
\( w_{l}^{k+1} \) under the constraint \( \sum_{l=1}^M w_{l} = 1 \) using the
Lagrange multiplier and solve the following equation:

\[
 \frac{\partial}{\partial w_{l}^{k+1}} \left[ \sum_{i=1}^N \sum_{l=1}^M \log \left( w_{l}^{k+1} f_l(x_i|\Theta^k) \right) + \lambda \left( \sum_{l=1}^M w_{l}^{k+1} - 1 \right) \right] = 0
\]

or

\[
 \sum_{i=1}^N \frac{1}{w_{l}^{k+1}} f(l|x_i, \Theta^k) + \lambda = 0.
\]

Summing both sides over \( l \), we get that \( \lambda = -N \) resulting
in

\[
 w_{l}^{k+1} = \frac{1}{N} \sum_{i=1}^N f(l|x_i, \Theta^k).
\]

Maximizing the term of Eq. (A.4) that contains \( \beta^2 \),

\[
 \frac{\partial}{\partial \beta^2} \left[ \sum_{i=1}^N \log \left( \frac{x_i^2}{2\beta^2} \right) \frac{\log (w_{l}^{k+1} f_l(x_i|\Theta^k))}{f(l|x_i, \Theta^k)} \right] = 0.
\]

This yields

\[
 \left( \frac{\beta^{k+1}}{\beta^k} \right) = \frac{\sum_{i=1}^N x_i^2 f(l|x_i, \Theta^k)}{2 \sum_{i=1}^N f(l|x_i, \Theta^k)}.
\]

The update equation of the Rayleigh mode is given by

\[
 \mu_{G}^{k+1} = \frac{\sum_{i=1}^N x_i f(l|x_i, \Theta^k)}{\sum_{i=1}^N f(l|x_i, \Theta^k)}.
\]

Hence, the update equation of the mean of the Gaussian
distribution is given by

\[
 \mu_{G}^{k+1} = \frac{\sum_{i=1}^N x_i f(l|x_i, \Theta^k)}{\sum_{i=1}^N f(l|x_i, \Theta^k)}.
\]

The variance of each Gaussian distribution can be derived
by maximizing the term of Eq. (A.4) that contains \( \sigma_{G}^{k+1} \),

\[
 \frac{\partial}{\partial \sigma_{G}^{k+1}} \left[ \sum_{i=1}^N \log \left( \frac{1}{\sqrt{2\pi \sigma_{G}^{k+1}}} \frac{\log (w_{l}^{k+1} f_l(x_i|\Theta^k))}{f(l|x_i, \Theta^k)} \right) \right] = 0.
\]

Hence, the update equation of the variance of the Gaussian
distribution is given by

\[
 \sigma_{G}^{k+1} = \frac{\sum_{i=1}^N (x_i - \mu_{G}^{k+1})^2 f(l|x_i, \Theta^k)}{\sum_{i=1}^N f(l|x_i, \Theta^k)}.
\]

Notice that all of the update equations involve taking
summations over the entire number of voxels in the
volume which is computationally expensive. To reduce
the processing time, the summation is carried out over
every possible intensity in the volume by multiplying
each intensity with its histogram frequency in the volume (Wilson and Noble, 1999). Therefore,
\[
\sum_{i=1}^{N} f(l|x_i, \Theta^k) = \sum_{i=0}^{l_{\text{max}}} h(i) f(l|i, \Theta^k),
\]
where \(l_{\text{max}}\) is the maximum voxel intensity in the observed volume, and \(h(i)\) is the frequency histogram of intensity \(i\). Therefore, the update equations of the parameters becomes
\[
w_{G_i}^{k+1} = \frac{1}{N} \sum_{i=0}^{l_{\text{max}}} h(i) f(G_i|i, \Theta^k), \quad l \in [1, 3],
\]
(A.16)
\[
w_{R}^{k+1} = \frac{1}{N} \sum_{i=0}^{l_{\text{max}}} h(i) f(R|i, \Theta^k),
\]
(A.17)
\[
\mu_{G_i}^{k+1} = \frac{\sum_{i=0}^{l_{\text{max}}} h(i) f(G_i|i, \Theta^k)}{\sum_{i=0}^{l_{\text{max}}} h(i) f(G_i|i, \Theta^k)}, \quad l \in [1, 3],
\]
(A.18)
\[
(\sigma^2_{G_i})^{k+1} = \frac{\sum_{i=0}^{l_{\text{max}}} (i - \mu_{G_i}^{k+1})^2 h(i) f(G_i|i, \Theta^k)}{\sum_{i=0}^{l_{\text{max}}} h(i) f(G_i|i, \Theta^k)},
\]
(A.19)
\[
(\beta^2)^{k+1} = \frac{\sum_{i=0}^{l_{\text{max}}} \beta h(i) f(R|i, \Theta^k)}{2 \sum_{i=0}^{l_{\text{max}}} h(i) f(R|i, \Theta^k)}.
\]
(A.20)

Appendix B. MRF basic definitions

This appendix gives a quick introduction to the basic definitions and terminology of MRF models.

B.1. Basic definitions

Let \(S = \{s_1, s_2, \ldots, s_N\}\) be a set of sites (voxels), where \(N\) is the total number of voxels in the MRA volume. \(\eta = \{\eta_s|s \in S\}\) is a neighborhood system in \(S\) if:

1. A site \(s\) is not neighboring to itself: \(s \notin \eta_s\).
2. The neighboring relationship is mutual: \(r \in \eta_s \iff s \in \eta_r \forall r, s \in S\).

A clique \(c\) is a subset of \(S\) such that

1. \(c\) consists of a single site \(s\), or
2. a set of sites containing \(s\), where every pair sites are mutual neighbors.

B.2. Markov random field

Let \(X = \{X_1, X_2, \ldots, X_N\}\) be a random field representing the true label of the MRA volume \(S\) of size \(N\), where each voxel \(s\) is assigned a random variable \(X_s\). The random field \(X\) is a Markov Random Field (MRF) with respect to a neighborhood system, if and only if the following conditions are satisfied:

(1) **Positivity**
\[p(X = x) > 0,\]
(2) **Markovian**
\[p(X_s = x_s|X_s \cap X_s \cap x_s = x_s, x_s) = p(X_s = x_s|X_s \cap x_s = x_s),\]
(3) **Homogeneity**
\[p(X_s = x_s|X_s \cap x_s = x_s)\]

is the same for all sites \(s\).

The notation \(S \setminus s\) refers to the set of all \(N\) sites excluding site \(s\) itself. The notation \(\hat{s}\) refers to all sites in the neighborhood excluding site \(s\).

B.3. Gibbs random fields

A random field \(X\) is a Gibbs random field (GRF) if and only if the probability density function has the following form:
\[
p(X = x) = \frac{\exp(-U(x))}{Z}, \quad Z = \sum_{x \in \Omega} \exp(-U(x)).
\]
(B.1)

The energy function is denoted by \(U(x)\), where the higher the energy of a configuration \(x\), the lower the probability of its occurrence. The space of all possible labeling is denoted by \(\Omega\). The denominator \(Z\), is called the partition function, which is a normalizing constant obtained by summing the numerator over all possible configuration \(x\). The partition function is usually cannot be computed because it is a computationally very expensive. We can specify the energy function in terms of the potentials of the individual cliques of a neighborhood system as follows:
\[
U(x) = \sum_{c \in C} V_c(x).
\]
(B.2)

The potential function of a clique characterizes the interaction among local group of spatially neighboring voxels by assigning a large cost to configurations of voxels which are less likely to occur. Early work with MRFs was impeded because it was not known how to calculate MRF joint probability distribution \(p(X = x)\) such that it satisfies the Markovian property, which is solved later by Hammersley–Clifford theorem (Hammersley and Clifford, 1971).

B.4. Hammersley–Clifford theorem

The theorem states that, if \(X\) is a MRF defined on a lattice \(S\) of a neighborhood system \(\eta\), then it is a GRF (i.e., it can be represented by a Gibbs distribution) under the condition that the energy of the Gibbs distribution is defined in terms of the cliques which is the key to prove
the theorem. Hence, if $X$ is a MRF, its joint probability can be given by

$$P(X = x) = \frac{e^{-U(x)}}{Z}.$$  \hfill (B.3)

**Appendix C. Implementation details**

The implementation of our method is very straightforward and follows from the text. The flowchart of Fig. 12 shows the block diagram of our proposed TOF extraction method.

**C.1. Residual histogram analysis method**

Since the volume histogram $h(x)$ may be noisy, we smooth it couple of times using an average filter in order to find $I_{\text{peak1}}, I_{\text{peak2}},$ and $I_{\text{min}}$ at which $h(x)$ achieves maximum. The implementation of this method follows directly from Table 1.

**C.2. EM segmentation**

The parameters update equations of our statistical model in its closed form are derived in Appendix A. The proportion and the mode of the Rayleigh distribution are given by Eqs. (A.17) and (A.20), respectively. The proportion, mean, and variance of the normal distribution are given by Eqs. (A.16), (A.18), and (A.19), respectively. The volume is segmented using Eq. (5).

**C.3. MRF parameter estimation**

The implementation of this part follows directly from Eq. (18). We utilized the simplex method (Nedler and Mead, 1965) as a non-linear optimization method to find the parameter $\beta_{sr}$ that maximizes the conditional likelihood function.

**C.4. MRF segmentation**

The implementation follows directly from the equations of Section 3.4.

Finally, we implemented our method using MATLAB and rendered our results in 3D using the visualization toolkit (VTK).

**Appendix D. Supplementary data**


**References**


