Vessel segmentation in MRI using a variational image subtraction approach

Ayşe Nurdan SARAN1,*, Fatih NAR2, Murat SARAN1
1Computer Engineering Department, Çankaya University, Yenimahalle, Ankara, Turkey
2Space and Defence Technologies, ODTÜ Teknokent, Ankara, Turkey

Received: 05.06.2012 ● Accepted: 08.11.2012 ● Published Online: 17.01.2014 ● Printed: 14.02.2014

Abstract: Vessel segmentation is important for many clinical applications, such as the diagnosis of vascular diseases, the planning of surgery, or the monitoring of the progress of disease. Although various approaches have been proposed to segment vessel structures from 3-dimensional medical images, to the best of our knowledge, there has been no known technique that uses magnetic resonance imaging (MRI) as prior information within the vessel segmentation of magnetic resonance angiography (MRA) or magnetic resonance venography (MRV) images. In this study, we propose a novel method that uses MRI images as an atlas, assuming that the patient has an MRI image in addition to MRA/MRV images. The proposed approach intends to increase vessel segmentation accuracy by using the available MRI image as prior information. We use a rigid mutual information registration of the MRA/MRV to the MRI, which provides subvoxel accurate multimodal image registration. On the other hand, vessel segmentation methods tend to mostly suffer from imaging artifacts, such as Rician noise, radio frequency (RF) inhomogeneity, or partial volume effects that are generated by imaging devices. Therefore, this proposed method aims to extract all of the vascular structures from MRA/MRI or MRV/MRI pairs at the same time, while minimizing the combined effects of noise and RF inhomogeneity. Our method is validated both quantitatively and visually using BrainWeb phantom images and clinical MRI, MRA, and MRV images. Comparison and observer studies are also realized using the BrainWeb database and clinical images. The computation time is markedly reduced by developing a parallel implementation using the Nvidia compute unified device architecture and OpenMP frameworks in order to allow the use of the method in clinical settings.

Key words: Magnetic resonance imaging, magnetic resonance angiography, magnetic resonance venography, vessel segmentation, total variation, parallel processing, compute unified device architecture

1. Introduction
Patient-specific modeling of vessel structures can be useful to investigate the human anatomy and in vascular disease characterization and assessment. Assessing vessel characteristics from magnetic resonance imaging (MRI) enables quantitative pathological and clinical evaluations, where MRI provides indispensable information. In order to acquire useful information from vessel structures, the measurements of the vessel’s width, color, reflectivity, tortuosity, and abnormal branching and the occurrence of vessels of a certain width are required. The complexity and large number of vessels for the especially large number of images make manual delineation of the vessels tedious to a large extent. Therefore, the main clinical goals of surgical planning and quantitative monitoring of disease progression require accurate automatic vessel segmentation methods with high reproducibility due to the limited number of images available per patient [1].

The performance of segmentation methods directly impacts the detection and target definition, as well

*Correspondence: buz@cankaya.edu.tr
as the monitoring of disease progression. However, vessel segmentation is a challenging task due to the complex nature of vessel structures. Because of the nature of the imaging technology and of human tissue, difficulties arise from medical image segmentation in MRI images, including noise and radio frequency (RF) inhomogeneity [2,3]. Although improvements in scanner technology have reduced these artifacts to some extent, noise and inhomogeneity remain problematic, particularly in MRI. Furthermore, noise and RF inhomogeneity within an image may change for new image acquisitions, more so if different scanners or imaging protocols are used. Dealing with noise without losing accuracy is another challenging task for medical image segmentation, since the differentiation of thin vessel structures and noise is a challenging problem. If noise removal were to be used as a preprocessing step for MRI and magnetic resonance angiography (MRA)/magnetic resonance venography (MRV) images, then the success of the noise removal step would affect the performance of any subsequent steps, such as image registration or vessel segmentation. Likewise, if intensity inhomogeneity removal were to be used as a preprocessing step for an MRI image, then the success of this step would affect the performance of further steps. The correction of intensity inhomogeneity in MRI was studied in several papers, such as [4–6], but to the best of our knowledge, there is no effective method for removing intensity inhomogeneity from MRA or MRV images, which eliminates the possibility of removing this imaging artifact as a preprocessing step.

There are several comprehensive reviews on vessel segmentation [7–9] and a vast number of research papers on the medical image segmentation problem [10–15], as either general segmentation methods or the segmentation of specific biological structures. Segmentation methods may be grouped into many different categories, such as histogram- and thresholding-based, clustering-based, edge-detection-based, region-growing-based, split-and-merge-based, graph-based, and partial differential equation-based [7]. For example, Sekiguchi et al. proposed a branch-based local region-growing scheme, where region growing is allowed to continue only towards a single branch at one time [16]. Atlas-based approaches have also been used in combination with region growing [17]. Kass et al. initially proposed an active contour (snake) algorithm [18], and it was first used in the brain segmentation of MR images by Atkins and Mackiewich [19]. The snake algorithm uses an analytical description of the segmented geometry rather than a discrete set of voxels. The snake is modeled as an evolving contour represented by a series of points. By iteratively moving the snake points, it aims to minimize the total energy to enable the snake to fit the image features well. The total energy is the weighted sum of the energy of internal and external forces, such as the image gradient fitted to vessel structures [20]. The snake algorithm is extremely sensitive to the initialization and noise, as well as to the concave parts of image contour attracting the segmentation. Holtzman-Gazit et al. used a variational method, which employs a function including an edge-based term, a minimal variance term, and a geodesic active contour (GAC) term [21]. The GAC term was implemented for regularization and the level set formulation was used for surface evolution. The level set method starts from a seed location or an initial shape and evolves to a discrete set of labeled voxels according to the image information such as image gradient and internal constraints (e.g., smoothness of the resulting segmented surface) [22,23]. Strzelecki et al. analyzed level set segmentation methods in simulated vessel structures with differing noise levels and concluded that segmentation results are good even for noisy images [24]. Gooya et al. introduced a level set-based geometric regularization method based on GAC, which is a modification to the curve regularization [25]. Chan and Vese [26] proposed a powerful level set image segmentation method based on an energy minimization problem. Xingce et al. used a maximum intensity projection (MIP) algorithm to remove tissues, which requires a time-consuming preprocessing method. In the preprocessing step, MRI was denoised, and then the Laplacian operator was used to sharpen the image and the Robert operator was used to realize the edge detection [27].
In this study, a registration-based vessel segmentation approach is developed that uses an MRI image as an atlas for MRA and MRV images. Vessels are found as changes between the MRI and MRA/MRV images, which are registered to each other, where inherent difficulties of change detection [28] are solved by unifying image subtraction, RF field estimation, and noise removal problems as a variational subtraction model. It is obvious that in order to use 2 modalities to extract vessels, the fundamental task is to register these images with high accuracy. For that reason, we use rigid mutual information (MI) image registration to register the MRI with either MRA or MRV. In the literature, there exist registration-based segmentation algorithms [2], yet we have not encountered any algorithm using MRI/MRA or MRI/MRV image pairs. These algorithms use a model or an atlas that is built to represent prior knowledge, such as white or gray matter. However, atlas-based vessel segmentation is not a feasible approach for vessel segmentation, since anatomic variability in vascular structures is large. In addition to all of this, building an atlas is a challenging problem in and of itself. Using prior information about the shape and topology of vessels, such as vessels being tubular, vessels being bright, vessel surface being smooth, or vessels having tree-like topology, succeeds to a certain level; however, especially for vascular structures with anomalies, all such assumptions may fail. In our study, we use MRI as prior information and our only assumption is that of the vessels being bright when 2 images are subtracted in the case that the bias field and noise are eliminated. Our assumption and used prior information have no such limitations, therefore allowing the segmentation of a large set of vessel structures. In [29], the skin surfaces between MRI-CT images were segmented. In [30], the authors presented 2 clustering algorithms for segmenting images. In [31], nerve cells between the original image and the constructed image were segmented. There are also some segmentation algorithms based on registration in the video frames, but these studies do not belong in the medical image analysis domain. To the best of our knowledge, this paper is the first paper that uses registration-based vessel segmentation in medical image analysis.

The outline of the paper is as follows. In the next section, the details of the proposed method and its subcomponents are explained. In Section 3, parallel implementations in OpenMP and Nvidia compute unified device architecture (CUDA) are explained, and the obtained speed-ups are demonstrated. In Section 4, our proposed method is validated using various quantitative tests on synthetic images and the results of an observer study are given. In Section 5, visual results on clinical images and comparison with segmentation using combined multiscale vessel enhancing and the level set algorithm are given. Finally, the conclusion is given in Section 6.

2. Method

The proposed method suggests the segmentation of vessels by subtracting registered MRI images from MRV/MRA images, which is not straightforward due to Rician noise [32,33] and RF field inhomogeneity [34]. Therefore, we formulate our optimization model as a simultaneous search of vessels and the normalization field, which equalizes the RF field of registered MRI images, while also handling noise, as shown in Eq. (1) below:

\[ T = (O - V^2)N, \]  

where \( T \) (target) is the MRI image that has no vessels, \( O \) (object) is the MRA/MRV image with vessels that is registered to \( T \) (as seen in Figure 1a), \( N \) is the smooth normalization field for equalizing RF inhomogeneity in \( O \) (as seen in Figure 1b) to RF inhomogeneity in \( T \), and \( V \) is the vessels to be determined. \( V \) is squared to make a positivity constraint on the vessels (as seen in Figure 1c). Figure 1d is prepared by multiplying \( O - V^2 \) (as seen in Figure 1e) by \( N \) (as seen Figure 1f), which looks very similar to \( T \) (MRI), where \( V^2 \) and \( N \) are the proposed method’s results.
Our cost function is defined as in Eq. (2), in which we seek to minimize $N$ and $V$. These values are constrained to be smooth by the corresponding regularization terms:

$$J(N, V) = |T - (O - V^2)N|^2 + \lambda_1 \|D(N)\|^2 + \lambda_2 \|D(V)\|^k,$$

where $\lambda_1$ and $\lambda_2$ are 2 scalar parameters controlling the smoothness of the $N$ and $V$, and $D$ is the deviation from the flatness operator, defined as:

$$D(I) = \sqrt{(\nabla^2_x I)^2 + (\nabla^2_y I)^2 + (\nabla^2_z I)^2},$$

where $\nabla^2_x I$ denotes the second-order partial derivative of $I$ with respect to $x$. Therefore, Eq. (2) can also be represented as follows.

$$J(N, V) = |T - (O - V^2)N|^2 + \lambda_1 \sqrt{(\nabla^2_x N)^2 + (\nabla^2_y N)^2 + (\nabla^2_z N)^2)} + \lambda_2 \sqrt{(\nabla^2_x V)^2 + (\nabla^2_y V)^2 + (\nabla^2_z V)^2}},$$

(3)

We use a gradient-based minimization approach in this study. Therefore, we use smooth approximation to $l_k$-norm for $k \leq 1$, as given in [35], to avoid problems due to the nondifferentiability of the $l_k$-norm:

$$\|z\|_k^k \approx \sum_{i=1}^n \left(|z_i|^2 + \frac{2}{k}\varepsilon\right)^{\frac{k}{2}},$$

for $k \leq 1$, and $\varepsilon \geq 0$ is a small constant, where $n$ is the length of the real-valued vector $z$ and $z_i$ is the $i$th element of $z$.

Next, the final form of the cost function $J(N, V)$ becomes:

$$J(N, V) = \sum_{x, y, z} J_{x, y, z}(N, V) = \sum_{x, y, z} (F_{x, y, z} + \lambda_1 \tilde{N}_{x, y, z} + \lambda_2 \tilde{V}_{x, y, z}),$$

(4)
where \( F_{x,y,z} = (T_{x,y,z} - (O_{x,y,z} - V^2_{x,y,z})N_{x,y,z})^2 \) is the fidelity term,
\[
\hat{N}_{x,y,z} = \left( \nabla^2 N_{x,y,z} \right)^2 + \left( \nabla^2 N_{y,x,z} \right)^2 + \left( \nabla^2 N_{z,x,y} \right)^2
\]
is the regularization term for \( N \),
\[
\hat{V}_{x,y,z} = \left( \nabla^2 V_{x,y,z} \right)^2 + \left( \nabla^2 V_{y,x,z} \right)^2 + \left( \nabla^2 V_{z,x,y} \right)^2 + \varepsilon
\]
is the regularization term for \( V \), and
\[
\nabla^2 f_{x,y,z} = \frac{f_{x+h,y,z} - 2f_{x,y,z} + f_{x-h,y,z}}{h^2}, \quad h = 1
\]
is the central difference approximation for second-order partial derivatives.

The first term in the objective function is a quadratic data fidelity term, which forces \((O - V^2)N\) to be similar to \( T \). The second term, which is also a quadratic term, implies the generation of a smooth \( N \) field in a Tikhonov regularization [36] manner, which is a commonly used method for the regularization of ill-posed problems. For the third term, which is nonquadratic, we use \( k \leq 1 \), since using a smaller \( l_k \)-norm implies fewer penalties on large intensity differences between neighbor voxels, which forms a smooth yet edge-preserving \( V \) field [37,38]. Our experiments show that \( k \approx 1 \) produces better results for sharp edges. In most cases, \( V \) has a smooth curvature, but in some cases, it has sharp-edged vascular structures. Therefore, we use one of the curvature models named ‘deviation from flatness’, which is a useful way of measuring local unflatness. The regularization terms also aim to generate noise-free \( N \) and \( V \) fields. We estimate \( N \) and \( V \) by minimizing cost function \( J(N,V) \) with respect to \( N \) and \( V \):
\[
\hat{J}(N,V) = \arg\min_{N,V} J(N,V).
\]

The flow chart of our method is given in Figure 2. The sole purpose of the steps in the given flowchart is the minimization of Eq. (4) to obtain vessels. Since a gradient-based numerical optimization method is used to minimize Eq. (4), a suitable initial \( N \) and \( V \) are necessary, namely \( N_0 \) and \( V_0 \). To this end, minimization of Eq. (4) is realized at step 7 and the necessary data, such as \( T \), \( O \), \( V_0 \), and \( N_0 \), are prepared between steps 1 and 6, where all of the voxels are normalized to 1. Minimization of Eq. (4) yields a vessel image, which may still be contaminated by noise, especially for very noisy MRI images, which consequently is cleaned in step 8. Each step in our method is explained in detail below.

**Step 1:** Registration of the images.

MI has had a wide range of interest for medical image registration since it was introduced by Collignon and Maes [39] and Viola and Wells [40] simultaneously. It is based on the widely utilized mutual information concept in information theory and provides a quantity that measures the mutual dependence of 2 images. We use normalized MI (NMI), proposed by Studholme et al. [41], as it is less sensitive to changes in the overlap; the formulation is given in Eq. (6). Its robustness and subvoxel accuracy for the registration of 3D multimodal medical images of various organs have been demonstrated in various research papers. We suggest the detailed reviews in [42,43] for the interested reader. Our MRI images and MRV/MRA images are registered using NMI, which is defined as:
\[
NMI(A,B) = \frac{H(A) + H(B)}{H(A,B)},
\]
where \( H(A) \) is the Shannon entropy of image \( A \), and \( H(A,B) \) denotes the amount of information \( B \) contains about \( A \).
Steps 2–4: Applying the brain mask to MRI and MRA/MRV.

We are only interested in the vessels in the brain and in the region between the brain and the skull. For that reason, we use the Brain Extraction Tool (BET; www.fmrib.ox.ac.uk/analysis/research/bet/), which is a widely used tool to separate brain structures from other brain structures in MRI images. BET only extracts brain structures; therefore, we use an adaptive dilation scheme to expand the extracted brain, as in Figure 3 (Figure 3a is the MRI image, Figure 3b is the extracted image, Figure 3c is the dilated brain, and Figure 3d is the mask generated). At this step, dilation is terminated when a voxel reaches the skull, where the skull threshold is set as an average intensity value of the extracted brain. As a result, the potential influence of the skull and background on the estimation of $N$ and $V$ is eliminated, and the computational domain is reduced. A generated brain mask is applied to the MRI image and registered to a vessel image such as MRA or MRV. Hence, the brain mask-applied MRI image is labeled $T$ and the brain mask-applied MRA/MRV images are labeled $O$, as used in Eq. (2).

![Figure 3. Generation of the brain mask.](image)

Steps 5 and 6: Creating $V_0$ and $N_0$.

The nonlinear objective function may have a large number of local minima and maxima. Optimization methods such as steepest descent, Newton’s method, or the Goldfeld, Quant, and Trotter (GQT) method may be applied to find these points. A good initial point is important to reach a global optimum or at least a
sufficiently good solution. In any case, a good initial point may also help the algorithm to converge faster. Likewise, our optimization method, GQT, requires an initial point, namely $V_0$ and $N_0$. To find $V_0$, we use the method proposed by Alonso-Montes et al., where vessels are preestimated by blurring the original image and subtracting the blurred image from the original image [44]. $V_0$ does not need to contain the fine details of the vessels and it may even contain structures that are not vessels, since the proposed cost function and minimization procedure produce fine vessel structures from this rough and even slightly erroneous initial vessel structure. A small value, $10^{-3}$, is added to the results of the square root of the Alonso-Montes method in only the brain region, so that the method can find vessels for all of the voxels. The square root of the Alonso-Montes method is used since $V$ is squared in the fidelity term in the cost function. A rough $N_0$ is sufficient for our proposed method to produce a fine and correct $N$. The selection procedure of $N_0$ to find this constant adaptively is the following:

$$N_0 = \frac{\mu_T}{\mu_O-V^2},$$

where $\mu_T$ is the mean of $T$, and $\mu_{O-V^2}$ is the mean of $O - V^2$. The extraction of $V_0$ and $N_0$ from the MRI image and MRA/MRV image is presented in Figure 4, where the MRI image is labeled as $T$ (Figure 4a). $N_0$ (Figure 4b) is the initial $N$. The estimated $N$ (Figure 4c) is the output of the proposed method. $N_0$ is quite similar to the true $N$ (Figure 4d). Brain mask-applied MRA/MRV images are labeled $O$ (Figure 4e) and, starting with the initial point $V_0$ (Figure 4f), $V$ (Figure 4g) is the estimated vessel. $V_0$ is somewhat similar to the true $V$ (Figure 4h); however, there are still a number of differences.

**Figure 4.** Extraction of $V_0$ and $N_0$ from the MRI image and MRA/MRV image.

**Step 7:** Minimizing $J$ with respect to $N$ and $V$.

There is no closed-form solution for Eq. (5); hence, a numerical optimization technique is needed. The steepest descent method may be used to minimize Eq. (5), using partial derivatives with respect to $N$ and $V$. However, the steepest descent is a first-order method and works rather slowly. For our case, we can calculate
the Hessian matrix analytically; therefore, Newton’s method [Eq. (7)], which is a second-order method, may be employed in order to obtain faster convergence. In order to be able to use Newton’s method, the initial point must be close to the optimum point and the cost function must be quadratic around that optimum point. Notice that in Eq. (4), the cost value for each voxel only depends on neighboring voxels due to the deviation of the flatness operator. Therefore, Newton’s method may be applied separately to each voxel, whereas Eq. (7) must be applied to all of the voxels for the $n$th iteration before proceeding to the next iteration.

$$\begin{bmatrix} N_{x,y,z} \\ V_{x,y,z} \end{bmatrix}^{(n+1)} = \begin{bmatrix} N_{x,y,z} \\ V_{x,y,z} \end{bmatrix}^{(n)} - \gamma_{x,y,z} [H(J_{x,y,z}(N,V))]^{-1} \nabla J_{x,y,z}(N,V)$$ (7)

Here, $\gamma_{x,y,z}$ is the step size, $H(J_{x,y,z}(N,V))$ is the Hessian matrix, and $\nabla J_{x,y,z}(N,V)$ is the gradient:

$$H(J_{x,y,z}(N,V)) = \begin{bmatrix} \frac{\partial^2 J_{x,y,z}(N,V)}{\partial N \partial N} & \frac{\partial J_{x,y,z}(N,V)}{\partial N} \\ \frac{\partial J_{x,y,z}(N,V)}{\partial V} & \frac{\partial^2 J_{x,y,z}(N,V)}{\partial V \partial V} \end{bmatrix}$$ and $\nabla J_{x,y,z}(N,V) = \begin{bmatrix} \frac{\partial J_{x,y,z}(N,V)}{\partial N} \\ \frac{\partial J_{x,y,z}(N,V)}{\partial V} \end{bmatrix}$.

However, classical numerical optimization techniques, such as Newton’s method, have been shown to perform poorly due to the presence of nonquadratic constraints, such as the third term in our optimization model [27]. Since Eq. (4) contains nonquadratic terms, the Hessian matrix is not always positive definite, which prevents the direct use of Newton’s method. Goldfeld et al. [45,46] proposed an optimization method that forces the Hessian matrix to remain positive definite. Therefore, the GQT method does not require a cost function to be concave. It estimates the step size using the quadratic approximation of the function in a given point.

$$\hat{H}(J_{x,y,z}(N,V)) = \frac{H(J_{x,y,z}(N,V)) + \beta_{x,y,z} I_{2 \times 2}}{1 + \beta_{x,y,z}}$$ (8)

For a positive definite $H$ matrix, $B_{x,y,z}$ is set as 0, and thus $\hat{H}$ becomes $H$. If $H$ is not a positive definite matrix, then $B_{x,y,z}$ is incremented by 1 until $\hat{H}$ becomes a positive definite matrix. If the determinant of $\hat{H}$ is greater than 1, then $\hat{H}$ is accepted as positive definite, since any matrix becomes ill-conditioned as its determinant approaches zero [45]. For example, for $\beta_{x,y,z}$ equaling 0, Eq. (7) turns into Newton’s method and for a large $\beta_{x,y,z}$ value, Eq. (7) turns into the steepest descent method. As a result, for each voxel, a $2 \times 2$ positive definite Hessian matrix is obtained, which is also analytically invertible so that Eq. (7) may be executed in the most efficient manner. GQT contains extra computations such as the calculation of the Hessian matrix and a search for the best $\beta_{x,y,z}$ value. However, in our method, we observe that the GQT method is an order of magnitude faster in comparison to the steepest descent method.

The step size is determined using the determinant of the Hessian matrix and $\beta_{x,y,z}$ value:

$$\gamma_{x,y,z} = \frac{1}{\beta_{x,y,z} \det(H)}$$ (9)

The convergence is determined by analyzing the change in the result of Eq. (4) in subsequent iterations. To eliminate the difference in the cost value for different cases, $J(N,V)$ is normalized:

$$\tilde{J}^{(n)}(N,V) = \sqrt{\frac{J^{(n)}(N,V)}{J^{(0)}(N,V)}}$$ (10)
In Figure 5, the value of the normalized cost function with respect to the iterations during the minimization procedure is shown (BrainWeb MRIs with 20 dB of noise, 20% RF, $\lambda_1 = 10^{-2}$, $\lambda_2 = 10^{-3}$, $\varepsilon = 10^{-4}$). Since the cost value is monotonically decreasing, the maximum cost value is obtained in the 0th iteration. Therefore, the normalized cost function is calculated using the cost value at the 0th iteration.

![Figure 5. Normalized cost function during the minimization of the cost function.](image)

The convergence tolerance value is evaluated by analyzing the change in the normalized cost function using Eq. (11):

$$C = \frac{\tilde{j}^{(n)}(N,V) - \tilde{j}^{(n-p)}(N,V)}{p},$$

where $p$ is the length of the convergence check period with the default value of 10, and $C$ is the change in the normalized cost function value between subsequent iterations. If $C$ is less than the convergence tolerance ($10^{-8} \leq C \leq 10^{-4}$), then the algorithm is stopped. The tolerance value may be increased to $10^{-4}$ for a decreasing convergence time by sacrificing segmentation accuracy on a small scale, less than 2%. The segmentation accuracy changes less than 1% for $C$ between $10^{-8}$ and $10^{-5}$, but for $C$ between $10^{-8}$ and $10^{-7}$, the computation time increases significantly. In this study, the default convergence tolerance value is set as $10^{-6}$, since this achieves a good balance between the accuracy and computation time at that convergence tolerance value.

**Step 8:** Postprocessing on $V$.

After step 7 is completed, $V_{\text{estimated}}$ is taken as the square of the final $V$ since it is squared in the fidelity term of the objective function, where $V_{\text{estimated}}$ corresponds to the vessel that we are trying to estimate. In $V_{\text{estimated}}$, regions having an element with a value higher than the threshold, 0.05, are all found and set as vessels. The remaining voxels are set as 0, which corresponds to nonvessel tissue.

### 3. Implementation

One way to increase the speed of the algorithm is to utilize parallel processing using multicore central processing units (CPUs). On the other hand, in recent years, the Nvidia CUDA technology offers a better way to increase massive parallel processing power using graphical processing units (GPUs). While the GPU consists of hundreds of smaller cores, a CPU may consist of 4 to 8 CPU cores. This parallel architecture gives the GPU its high computation performance and massive parallelism.

One of the advantages of our cost function and optimization method is that it is constructed based on the logic of parallel programming. We also consider the memory bandwidth limitation of GPUs and design our cost function and optimization method so that minimum memory space is required. Our function contains quadratic and nonquadratic terms. Therefore, efficient half-quadratic regularization is applicable where a new cost function with the same minimum as the original nonquadratic cost function is introduced and solved using
linear algebraic methods \[38\]. For instance, efficient modification of half-quadratic regularization was proposed by Çetin and Karl, where an approximate Hessian matrix is defined and the conjugate gradient (CG) method is used to solve the linear system instead of inverting this Hessian matrix \[35,37\]. This approach is efficient and the solution is obtained in only a few iterations, but within each iteration, 12 temporary arrays of the same size as the MRI volume and a CG execution are needed. Since our method tries to find \( N \) and \( V \), the required number of arrays will be even larger. This large amount of memory usage causes degradation in parallelism due to the memory bandwidth limitation of GPUs. However, our approach causes more iterations with more computation, but only with the usage of an array for \( N \) and an array for \( V \), which makes it more suitable for GPU architectures. The execution time of the steps in the proposed method is given in Table 1 (BrainWeb 20 MRIs with a size of 181 \( \times \) 217 \( \times \) 181, a 20-dB noise level, and 20\% RF, \( \lambda_1 = 10^{-2}, \lambda_2 = 10^{-3}, \varepsilon = 10^{-4} \)).

<table>
<thead>
<tr>
<th></th>
<th>CPU 1-thread</th>
<th>CPU 2-threads</th>
<th>CPU 4-threads</th>
<th>CPU 8-threads</th>
<th>GPU</th>
<th>Fastest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generate brain mask</td>
<td>1.08</td>
<td>0.51</td>
<td>0.34</td>
<td>0.31</td>
<td>0.31</td>
<td>0.31</td>
</tr>
<tr>
<td>Estimate ( V_0 )</td>
<td>1.42</td>
<td>1.42</td>
<td>1.42</td>
<td>1.42</td>
<td>1.42</td>
<td>1.42</td>
</tr>
<tr>
<td>Minimize Eq. (4)</td>
<td>1751.78</td>
<td>1131.31</td>
<td>813.35</td>
<td>297.78</td>
<td>8.53</td>
<td>8.53</td>
</tr>
</tbody>
</table>

The GPU-based MI registration method of Saxena et al. \[47\] was used for fast image registration. Execution times for GPU-based MI are quite similar to those of GPU-based NMI, since MI and NMI have equal computational complexity \[48\]. The computation times for the MI registration and BET are:

- MI : 9 s (for 181 \( \times \) 217 \( \times \) 181 voxels) on a GPU,
- BET : 3 s on a CPU (for 1 thread).

In addition, since MI is executed on a GPU and BET is executed on a CPU, they are executed simultaneously to introduce the task level parallelism.

The generation of the brain mask and estimation of \( V_0 \) and \( N_0 \) are quite fast using multicore CPU implementation. Applying the brain mask and estimating \( N_0 \) are executed in a negligible period. Essentially, the most computationally intensive part in the proposed algorithm is the minimization of Eq. (4). Therefore, this part is parallelized using CPU/OpenMP and GPU/CUDA, where the GPU implementation is 35 times faster in comparison to the 8-thread CPU implementation. Single precision is used for both the CPU and GPU implementation. The total computation time of the proposed method can be given as:

- Steps 1–4: 9 s (MI and brain mask),
- Steps 5 and 6: 2 s (creation of \( V_0 \) and \( N_0 \)),
- Steps 7 and 8: 8 s (minimization of Eq. 4 postprocessing),

Total: 20 s.

4. Validation
In this study, the coefficients of the regularization terms, \( \lambda_1 \) and \( \lambda_2 \), are determined empirically. These parameters must be adjusted for different datasets, yet we end up with quite similar parameters for BrainWeb
phantom images, clinical MRA images, and clinical MRV images. Moreover, we also observe that tuning the $\lambda_1$ and $\lambda_2$ parameters is quite easy and requires only a few attempts.

4.1. Experiments with phantom images

We demonstrate the effectiveness of our method on 20 cases from the BrainWeb database. The vessel class in the BrainWeb datasets is used to generate a MRA/MRV image using the real MRI image taken from MRICro software (http://www.mccauslandcenter.sc.edu/mricro/mricro/index.html). This MRI is obtained by averaging multiple acquisitions of the same patient; hence, it is almost noise-free. Rician noise is added to the noise-free MRI image by taking the norm of the subsequent addition of independent and identically distributed additive Gaussian noise into the noise-free MRI image [49,48]. RF inhomogeneity is also added to the MRI and MRA/MRV images, where 3 samples of RF inhomogeneity in the BrainWeb database are used for that purpose.

The signal-to-noise ratio for the image deteriorated by Rician noise is defined as:

$$\text{SNR} = 10 \log_{10} \frac{I}{\hat{I}},$$

where $I$ is the noise-free image and $\hat{I}$ is the added noise. A quantitative analysis of the segmentation success is shown using the precision, recall, and Jaccard index (similarity coefficient), calculated as:

$$\text{Precision} = \frac{tp}{tp + fp}, \quad \text{Recall} = \frac{tp}{tp + fn}, \quad J(V, V_{\text{estimated}}) = \frac{|V \cap V_{\text{estimated}}|}{|V \cup V_{\text{estimated}}|},$$

where $tp$ is the true positive (correct result), $fp$ is the false positive (unexpected result), $fn$ is the false negative (missing result), $V$ and $V_{\text{estimated}}$ are in binarized form, and $J(V, V_{\text{estimated}})$ is the Jaccard index and is calculated as the size of the intersection of $V$ and $V_{\text{estimated}}$ divided by their union. Performance values are only evaluated in the brain mask since the proposed method tries to extract vessels in that region. Typical noise and RF levels in the MRI images are simulated on the BrainWeb database so that the performance of the proposed method can be evaluated for clinical settings (see Table 2 for BrainWeb images with different noise and RF levels). The effects of the noise and RF inhomogeneity are separately analyzed in the BrainWeb database and the results are shown in Figures 6 and 7 ($\lambda_1 = 10^{-2}$, $\lambda_2 = 10^{-3}$, and $\varepsilon = 10^{-4}$) may be evaluated for clinical settings (see Table 2 for BrainWeb images with different noise and RF levels). The effects of the noise and RF inhomogeneity are separately analyzed in the BrainWeb database and the results are shown in Figures 6 and 7 ($\lambda_1 = 10^{-2}$, $\lambda_2 = 10^{-3}$, and $\varepsilon = 10^{-4}$). As seen in Figure 6, the proposed method is very robust to RF inhomogeneity. Even for a high level of RF inhomogeneity, the performance only decreases to 86.48%, which shows the effectiveness of the regularization term on the $N$ field. As seen in Figure 7, the proposed method is also robust to Rician noise. Even for a high level of Rician noise (Figure 8), the performance decreases to 89.80%, which shows the effectiveness of the regularization term on the $V$ field. The combined effect of Rician noise and RF inhomogeneity can be seen in Table 2, where the common level of Rician noise and RF inhomogeneity average performance is 96.07%, and for a high level of Rician noise and RF inhomogeneity, the average performance is 92.13%. All of these analyses show that the proposed method performs well, even for fixed $\lambda_1$, $\lambda_2$, $\varepsilon$, and convergence tolerance parameters. As seen in Table 1, the execution time of the proposed method is also fast. Therefore, the proposed method is suitable for clinical use due to its speed, accuracy, and robustness to Rician noise and RF inhomogeneity.
Figure 6. Analysis of the RF inhomogeneity on the segmentation success.

Figure 7. Analysis of the Rician noise on the segmentation success.

Figure 8. Different Rician noise levels.

Table 2. BrainWeb images with different noise and RF levels.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Zero noise, no RF field</th>
<th>20 dB of noise, 20% RF field</th>
<th>15 dB of noise, 30% RF field</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
<td>Jaccard</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>99.89</td>
<td>99.89</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>99.86</td>
<td>99.86</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>99.95</td>
<td>99.95</td>
</tr>
<tr>
<td>18</td>
<td>100</td>
<td>99.88</td>
<td>99.88</td>
</tr>
<tr>
<td>20</td>
<td>100</td>
<td>99.93</td>
<td>99.93</td>
</tr>
<tr>
<td>38</td>
<td>100</td>
<td>99.84</td>
<td>99.84</td>
</tr>
<tr>
<td>41</td>
<td>100</td>
<td>99.93</td>
<td>99.93</td>
</tr>
<tr>
<td>42</td>
<td>100</td>
<td>99.95</td>
<td>99.95</td>
</tr>
<tr>
<td>43</td>
<td>100</td>
<td>99.92</td>
<td>99.92</td>
</tr>
<tr>
<td>44</td>
<td>100</td>
<td>99.92</td>
<td>99.92</td>
</tr>
<tr>
<td>45</td>
<td>100</td>
<td>99.94</td>
<td>99.94</td>
</tr>
<tr>
<td>46</td>
<td>100</td>
<td>99.90</td>
<td>99.90</td>
</tr>
<tr>
<td>47</td>
<td>100</td>
<td>99.93</td>
<td>99.93</td>
</tr>
<tr>
<td>48</td>
<td>100</td>
<td>99.89</td>
<td>99.89</td>
</tr>
<tr>
<td>49</td>
<td>100</td>
<td>99.90</td>
<td>99.90</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>99.93</td>
<td>99.93</td>
</tr>
<tr>
<td>51</td>
<td>100</td>
<td>99.92</td>
<td>99.92</td>
</tr>
<tr>
<td>52</td>
<td>100</td>
<td>99.94</td>
<td>99.94</td>
</tr>
<tr>
<td>53</td>
<td>100</td>
<td>99.89</td>
<td>99.89</td>
</tr>
<tr>
<td>54</td>
<td>100</td>
<td>99.91</td>
<td>99.91</td>
</tr>
<tr>
<td>Average</td>
<td>100</td>
<td>99.91</td>
<td>99.91</td>
</tr>
<tr>
<td>Min.</td>
<td>100</td>
<td>99.84</td>
<td>99.84</td>
</tr>
<tr>
<td>Max.</td>
<td>100</td>
<td>99.95</td>
<td>99.95</td>
</tr>
</tbody>
</table>
4.2. Experiments with clinical images

To obtain a medical expert’s point of view, 9 clinical MRA segmentation results of the proposed method were shown to a pediatric neurologist. Table 3 shows the results of our observer’s evaluation. The observer states that, for patient 1, “although the patient has sagittal sinus in the MRA, the vessels are segmented correctly”. In general, “there are some thin vessels that the algorithm fails to segment; however, the error is too small to be taken into account”.

<table>
<thead>
<tr>
<th>Patient 1</th>
<th>Thick</th>
<th>Thin</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>98%</td>
<td>99%</td>
<td>99%</td>
<td>100%</td>
</tr>
<tr>
<td>Patient 2</td>
<td>100%</td>
<td>96%</td>
<td>98%</td>
<td>99%</td>
</tr>
<tr>
<td>Patient 3</td>
<td>95%</td>
<td>99%</td>
<td>97%</td>
<td>100%</td>
</tr>
<tr>
<td>Patient 4</td>
<td>100%</td>
<td>92%</td>
<td>96%</td>
<td>96%</td>
</tr>
<tr>
<td>Patient 5</td>
<td>92%</td>
<td>96%</td>
<td>94%</td>
<td>100%</td>
</tr>
<tr>
<td>Patient 6</td>
<td>99%</td>
<td>86%</td>
<td>96%</td>
<td>100%</td>
</tr>
<tr>
<td>Patient 7</td>
<td>99%</td>
<td>98%</td>
<td>98%</td>
<td>100%</td>
</tr>
<tr>
<td>Patient 8</td>
<td>90%</td>
<td>97%</td>
<td>96%</td>
<td>98%</td>
</tr>
<tr>
<td>Patient 9</td>
<td>92%</td>
<td>96%</td>
<td>92%</td>
<td>96%</td>
</tr>
<tr>
<td>Overall</td>
<td>96.1%</td>
<td>95.4%</td>
<td>96.2%</td>
<td>98.8%</td>
</tr>
</tbody>
</table>

5. Results and comparison

Segmentation results for the BrainWeb images (parameters: TR = 22 ms, TE = 9.2 ms, FA = 30°, 1.00 mm isotropic) for \( \lambda_1 = 10^{-2}, \lambda_2 = 10^{-3}, \) and \( \varepsilon = 10^{-4} \) are shown in Figure 9. Figures 9a and 9b show the MRA/MRV image and MIP MRA/MRV for various RF and noise levels. Figure 9c is the estimated vessel and Figure 9d is the ground truth.

The vessel segmentation (\( \lambda_1 = 10^{-2}, \lambda_2 = 10^{-3}, \text{and} \ \varepsilon = 10^{-4} \)) results for the clinical MRA (parameters for the 1.5 T Siemens scanner: TR = 23 ms, TE = 7 ms, FA = 25°, 12 bits, 0.35 mm \( \times \) 0.35 mm \( \times \) 0.70 mm) and MRV (parameters for the 1.5 T Siemens scanner: TR = 25 ms, TE = 7 ms, FA = 60°, 12 bits, 0.98 mm \( \times \) 0.98 mm \( \times \) 1.34 mm) are shown in Figure 10. The clinical MRI (parameters for the 1.5 T Siemens scanner: TR = 20 ms, TE = 3.1 ms, FA = 15°, 12 bits, 0.48 mm \( \times \) 0.48 mm \( \times \) 1.00 mm) (Figure 10a) and MRA (Figure 10b) pairs are registered using NMI. Six degrees of freedom within NMI rigid registration are used: 3 \( \times \) translate and 3 \( \times \) rotate (scale factors are calculated using pixel resolution information in the MRI DICOM header). Figures 10c and 10d show the estimated vessels and 3D-rendered vessels, respectively. Similarly, the clinical MRI (Figure 10e) and MRV (Figure 10f) pairs are registered using NMI. Figures 10g and 10h show the estimated vessels and 3D-rendered vessels, respectively.

The proposed method is compared with the Frangi multiscale vessel enhancement filtering method [50] followed by GAC segmentation [51] in the BrainWeb database. The Frangi method and level set method are combined to increase the vessel segmentation performance, as suggested in the literature. The parameters of the Frangi method, GAC, and their combination are hand-optimized to produce the best results for the BrainWeb database. However, the results (Table 4) show that the proposed method performs considerably better compared to the combined segmentation of the Frangi method and GAC level set.
Figure 9. BrainWeb vessel segmentation results.

Figure 10. Visual vessel segmentation results for the clinical MRA and MRV.
Table 4. Segmentation results of the combined Frangi and GAC methods.

<table>
<thead>
<tr>
<th>Patient</th>
<th>Precision</th>
<th>Recall</th>
<th>Jaccard</th>
<th>Patient</th>
<th>Precision</th>
<th>Recall</th>
<th>Jaccard</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>90.57</td>
<td>85.70</td>
<td>78.68</td>
<td>45</td>
<td>92.49</td>
<td>87.02</td>
<td>81.27</td>
</tr>
<tr>
<td>5</td>
<td>91.63</td>
<td>86.69</td>
<td>80.33</td>
<td>46</td>
<td>92.26</td>
<td>89.40</td>
<td>83.16</td>
</tr>
<tr>
<td>6</td>
<td>93.34</td>
<td>87.95</td>
<td>82.76</td>
<td>47</td>
<td>92.42</td>
<td>89.09</td>
<td>83.02</td>
</tr>
<tr>
<td>18</td>
<td>91.59</td>
<td>86.72</td>
<td>80.32</td>
<td>48</td>
<td>91.92</td>
<td>87.70</td>
<td>81.42</td>
</tr>
<tr>
<td>20</td>
<td>91.70</td>
<td>88.77</td>
<td>82.17</td>
<td>49</td>
<td>92.69</td>
<td>87.57</td>
<td>81.92</td>
</tr>
<tr>
<td>38</td>
<td>91.92</td>
<td>86.22</td>
<td>80.15</td>
<td>50</td>
<td>91.35</td>
<td>88.75</td>
<td>81.87</td>
</tr>
<tr>
<td>41</td>
<td>92.42</td>
<td>86.77</td>
<td>81.00</td>
<td>51</td>
<td>93.58</td>
<td>87.06</td>
<td>82.16</td>
</tr>
<tr>
<td>42</td>
<td>91.29</td>
<td>88.61</td>
<td>81.70</td>
<td>52</td>
<td>91.62</td>
<td>84.74</td>
<td>78.64</td>
</tr>
<tr>
<td>43</td>
<td>90.16</td>
<td>88.27</td>
<td>80.52</td>
<td>53</td>
<td>92.22</td>
<td>88.35</td>
<td>82.22</td>
</tr>
<tr>
<td>44</td>
<td>90.91</td>
<td>88.81</td>
<td>81.57</td>
<td>54</td>
<td>93.31</td>
<td>85.85</td>
<td>80.87</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Combined Frangi and GAC methods</th>
<th>Proposed method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>average</td>
<td>91.97</td>
</tr>
</tbody>
</table>

6. Conclusion

In this study, a novel registration-based vessel segmentation approach is proposed, where an MRI image is registered to either an MRA or MRV image. The proposed approach intends to increase vessel segmentation accuracy using the available MRI image as prior information. Any unfavorable effects of Rician noise and RF inhomogeneity in the MRI, MRA, and MRV images during vessel segmentation are eliminated using a subtraction schema, which simultaneously seeks out vessel field $V$ and RF equalization field $N$. The proposed subtraction schema utilizes a GQT method for the minimization of the proposed cost function, which contains a quadratic fidelity term in addition to quadratic and nonquadratic regularization terms. The proposed method is validated quantitatively using real MRI and BrainWeb phantom images. Our method is also visually analyzed using clinical MRI, MRA, and MRV images. As seen in the Sections 4 and 5, the proposed method is robust to Rician noise and RF inhomogeneity and the execution speed is also fast, therefore making it suitable for clinical use.

Parallelization is taken as a priority during the development of our cost function and choice of minimization method. Our experiments show that the computation time is markedly decreased using the Nvidia CUDA and OpenMP frameworks. This will allow easy use of the method in clinical settings.

Acknowledgment

The authors would like to thank Prof Dr Kivlicm Güçyüener for her valuable time in evaluating our results; Asst Prof Dr Didem Gökçay for providing the clinical MRI, MRA, and MRV images; and Osman Erman Okman, Alireza Mazloumi, and Atilla Özgür for their valuable suggestions.

References


