

Comparison of Vessel Segmentations using STAPLE

Julien Jomier, Vincent LeDigarcher, and Stephen R. Aylward

Computer-Aided Diagnosis and Display Lab
The University of North Carolina at Chapel Hill, Department of Radiology
27510 Chapel Hill, USA
{jomier, aylward}@unc.edu

Abstract. We propose a novel method for the validation of vascular segmentations. Our technique combines morphological operators and the STAPLE algorithm to obtain ground truth of centerline extractions as well as a measure of accuracy of the methods to be compared. Moreover, our method can be extended to the validation of any open-curves. We also present a comparison study of three vascular segmentation methods: ridge traversal, statistical and curves level set. They are compared with manual segmentations from five experts.

1 Introduction

Blood vessels and their branches vary considerably but are often critical in planning and performing neurosurgical and interventional procedures. In planning the feeding and drain vessels of a lesion must be defined. During surgery the vessels serve as landmarks and guidelines to the lesion [1]. The more precise the vascular segmentation is, the more accurate the plan and navigation.

Segmentation of tubular structures and more specifically vascular segmentation has been of high interest in medical imaging, and several excellent techniques have been developed, incorporating a variety of different approaches. For instance, the curve evolution algorithm [4] produces accurate vascular segmentations by combining the modified curvature diffusion equation (MCDE) with a level-set based technique. On the other hand, Aylward et al. [2] use a ridge traversal technique with width estimation to extract vascular centerline and estimated radius at each point along blood vessels. Wink et al. [6] have developed a front propagation method for the extraction of blood vessels. Given a starting point, a front wave is moved to fit the tubular structure.

Surveys have been conducted to contrast the vascular segmentation algorithms [3], however, no direct comparison has been done on the accuracy of these techniques. In this paper, we propose a measure for vessel segmentation comparison and we apply it to the analysis of vascular segmentation algorithms from 3-dimensional images of the liver (CT and MR combined). The rest of this paper is structured as follow: first the STAPLE algorithm and its extension to open curves is described, second the segmentation methods to be compared and the results of the comparison are presented.

2 Method

In this section we present the STAPLE algorithm [5] and detail how we extend its use for open curves validation.

2.1 STAPLE algorithm

The Simultaneous Truth and Performance Level Estimation (STAPLE) algorithm generates ground truth volumes from a set of binary expert segmentations as well as a simultaneous accuracy assessment of each expert.

STAPLE works as follow: considering $\mathbf{p} = (p_1, p_2, \dots, p_r)^T$ a column vector of R elements, with each element a sensitivity parameter characterising one of the R segmentations, and $\mathbf{q} = (q_1, q_2, \dots, q_r)^T$ a column vector representing the specificity parameter of one of R segmentations. Let \mathbf{D} be an $N \times R$ matrix describing the binary decisions made for each segmentation at each voxel of the image and \mathbf{T} be an indicator vector of N elements representing the hidden binary true segmentation. The complete data can be written as (\mathbf{D}, \mathbf{T}) and the probability mass function of these data $f(\mathbf{D}, \mathbf{T} | \mathbf{p}, \mathbf{q})$. An expectation-maximization algorithm then estimates the performance level of the experts characterized as a tuple (p, q) , where p represents the sensitivity (“true positive fraction”) and q the specificity (“true negative fraction”), which maximizes the complete data log likelihood function

$$(\hat{\mathbf{p}}, \hat{\mathbf{q}}) = \operatorname{argmax}_{\mathbf{p}, \mathbf{q}} \ln f(\mathbf{D}, \mathbf{T} | \mathbf{p}, \mathbf{q}) \quad (1)$$

The STAPLE algorithm treats segmentation as a pixelwise classification, which leads to an averaging scheme that accounts for systematic biases in the behavior of experts in order to generate a fuzzy ground truth volume and simultaneous accuracy assessment of each expert. One can notice that for STAPLE to work, the set of binary segmentations should overlap which is often not the case for open curves. In order to overcome this issue, we propose an iterative scheme to construct ground truth of open curve segmentations such as blood vessels. First, each centerline extraction is discretized to form a binary image volume - having an intensity of one on the centerline and zero outside. Second, our method iteratively creates new morphologically dilated segmentations from the initial volume. The STAPLE algorithm is then used to compute the level of each experts as well as the ground truth at each stage of the dilation process. The resulting output is a set of probability maps as well as sensitivity and specificity levels of the experts for each dilation factor.

2.2 Toy example

We illustrate the behavior of our technique using a simple example composed of three segmentations (one per expert) of an horizontal straight line. Figure 1 shows the three original segmentations as well as the resulting probability map. The sensitivity p and specificity q of the experts are shown in figure 2.

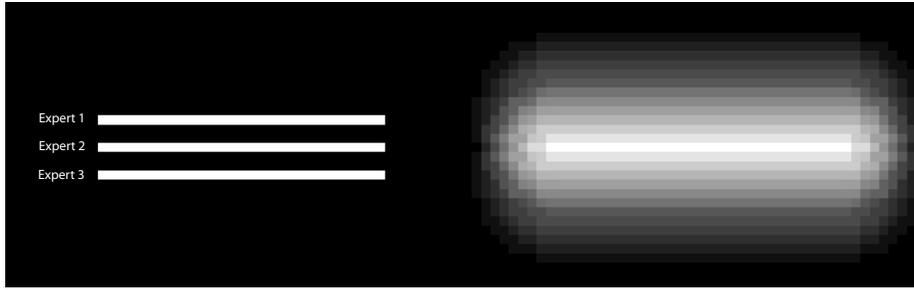


Fig. 1. Simple example composed of three segments representing segmentation of the same structure by three experts (**left**) and the resulting probability map obtained by averaging the ground truth probability volumes (**right**).

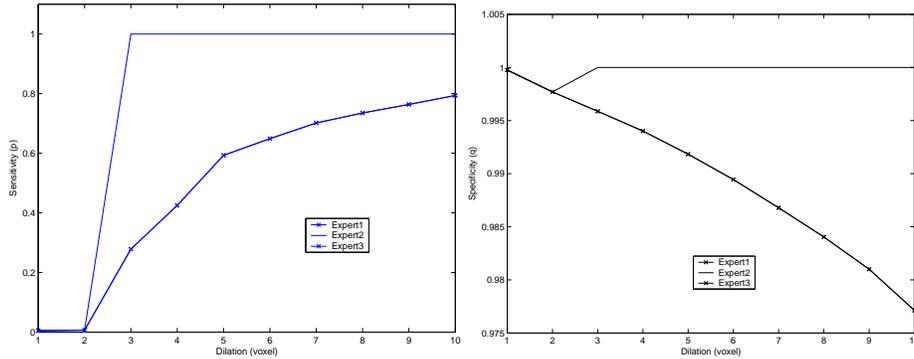


Fig. 2. Sensitivity and specificity of the experts in the simple example of three straight line segmentations. Expert 1 and Expert 3 are confused.

In this example, the three initial segmentation were separated by three pixels. As we can predict the expert levels graph shows high specificity and very low sensitivity as long as the three binary segmentations do not overlap ($dilation < 3$). As soon as the dilated segmentation start to overlap, the sensitivity increases and the specificity decreases. One can notice that the sensitivity of the second expert is higher than the two others which is the intended behavior of the algorithm since the second expert has extracted the real structure. The sensitivity will increase logarithmically since in the limit ($dilation \rightarrow \infty$) the sensitivity will be one for all experts and specificity is undefined. The specificity graph shows that all experts are following the same specificity until segmentations overlap, then, once again, the second expert outperforms the other two as expected.

2.3 Open curves validation

In order to validate our technique, we created simulated data composed of single spiral going down the Z axis. A binary mask (high intensity inside the spiral, low

intensity outside) was computed from it. After a gaussian blurring of the binary volume ($\sigma = 2$), six experts extracted the spiral using a manual (point and click) segmentation technique. We then applied our validation method from $n = 2$ experts to $n = 6$ experts to create five normalized ($I \in [0, 1]$) probability maps of the manually segmented spiral. We then superimposed the original spiral with each probability map and computed the mean probability along the centerline. A mean probability of 1.0 would mean that the segmented spiral is the exact representation of the original. Figure 3 shows the results of the experiment.

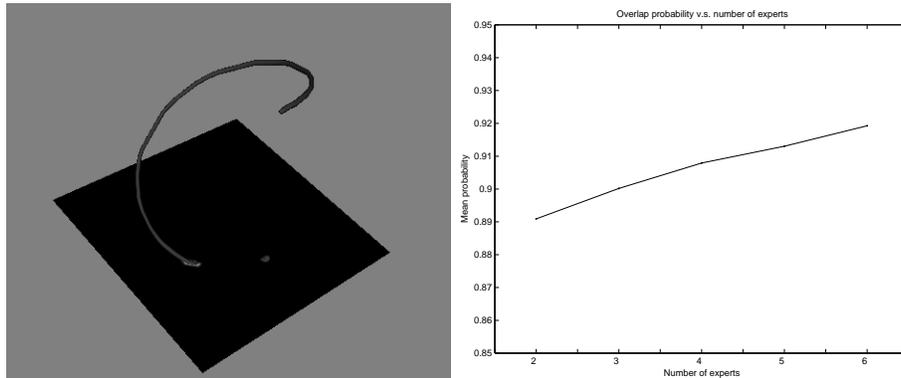


Fig. 3. Simulated open-curve: a spiral (**left**) and the mean probability along the centerline of the spiral given an increasing number of experts (**right**).

As one can see the segmentation keeps improving with the number of experts. This is due to the STAPLE algorithm which limits the contribution of “bad” experts. From the probability map we can then reconstruct the mean spiral as shown in figure 4.



Fig. 4. Volume rendering of the output probability map (**left**) and the reconstructed spiral from the set of 6 experts (**right**).

3 Comparison of segmentation techniques

In this section we compared three vascular segmentation techniques on seven liver datasets (CT and MR). CT volumes are contrast enhanced $1 \times 1 \times 3mm^3$ voxels and MRA volumes are time-of-flight data with $1 \times 1 \times 1mm^3$ voxels. In order to establish truth, five experts have segmented vascular centerlines manually by point and click. We then compared the expert segmentations with three other techniques: (a) a ridge traversal and width estimation method, (b) a statistical segmentation technique using connected component and (c) a region growing via level set method using the Curves algorithm. Next we present these three techniques in details.

3.1 Ridge traversal and width estimation

This vessel extraction method extracts blood vessels from 3-dimensional images using a scale space technique. The algorithm traverses a ridge in an intensity function F , using the Hessian at a point x . Let's define α and β as ascending-ordered eigenvalues of the Hessian at x , \mathbf{u} and \mathbf{v} as the corresponding eigenvectors of the Hessian, and \mathbf{P} and \mathbf{Q} as the directional derivatives: $\mathbf{P} = \mathbf{u} \cdot \mathbf{F}$ and $\mathbf{Q} = \mathbf{v} \cdot \mathbf{F}$. Therefore, if x is exactly in the middle of the ridge the following conditions must hold: $\alpha < 0$ and $\mathbf{P} = 0$. Given an initial starting point close to the ridge, the intensity ridge is computed to minimize \mathbf{P} using a direction search with respect to the Hessian. The line search is performed from x in the direction \mathbf{u} to find the local minimum of \mathbf{P} . If the resulting minimum is not within a specified tolerance a new initial point is required. This vascular segmentation method has shown sub-voxel accuracy but has never been compared with other techniques on physical datasets.

3.2 Statistical Segmentation

The second segmentation technique to be compared uses statistics on the intensity of the blood vessel as well as a region growing technique. First the user specifies a set of points inside a blood vessel, then statistics are computed to define a gaussian probability class corresponding to the blood vessel likelihood. Second, a connected component algorithm is performed based on the previously defined class. Third and last, morphological operators are used to remove unwanted structures and obtain a smooth segmentation.

Skeletization is then done using a ridge finding technique. Basically the binary image volume is blurred by an amount proportional to the expected radius value of the tubular structure and the ridge is tracked.

3.3 Segmentation via levelset

The third segmentation method is based on the Curves algorithm [4], an extension of geodesic active contours based on a level set implementation. The

evolution of the level set is driven by the curvature equation. In the case of a 1-dimensional curve in a 3-dimensional space, the following equation should be minimized

$$\oint_1^0 g(|\nabla I(c(p))|)|C'(p)|dp \quad (2)$$

where $C(p)$ is the 1-dimensional curve, I the image, and g a strictly decreasing function such that $g(r) \rightarrow 0$ as $r \rightarrow \infty$. By computing the Euler-Lagrange equations, the curve evolution equation can be formulated as

$$C_t = kN - \frac{g'}{g} \prod \left(H \frac{\nabla I}{|\nabla I|} \right) \quad (3)$$

where H is the Hessian of the intensity function. Therefore the equation for embedding space is

$$\nu_t = \lambda (\nabla \nu(x, t), \nabla^2 \nu(x, t)) + \frac{g'}{g} \nabla \nu(x, t) \cdot H \frac{\nabla I}{|\nabla I|} \quad (4)$$

The output of the Curves algorithm is a binary image obtained from the zero level set. Therefore, the same skeletization technique as in 3.2 is used to obtain the centerlines.

3.4 Results

Given the centerline segmentations (5 experts and 3 automated) for the seven datasets we applied our algorithm until a maximum dilation of 7 voxels. We are interesting in the sensitivity and specificity of each methods; these two values characterize the confidence level of a method and allows for comparisons. The mean results are presented in figure 5.

One can notice that the ridge traversal technique outperforms the statistical and curves method in both terms of sensitivity and specificity. Moreover, the sensitivity of the ridge traversal and curves technique are close to the sensitivity of the experts. However, the statistical method performs poorly compared to the experts and the other automated techniques.

From the specificity graph, one can notice that the three automated method have a smaller specificity than the experts. There are two reasons for this. First, the length of the segmented centerlines appear to be shorter for all the experts segmentation. This can be explained by the noise present in the image which makes small structures invisible to the human eye but the automated algorithms are able to extend further the tubular structure. Second, Curves and statistical methods tends to “leak” outside of the structure making the skeleton less reliable.

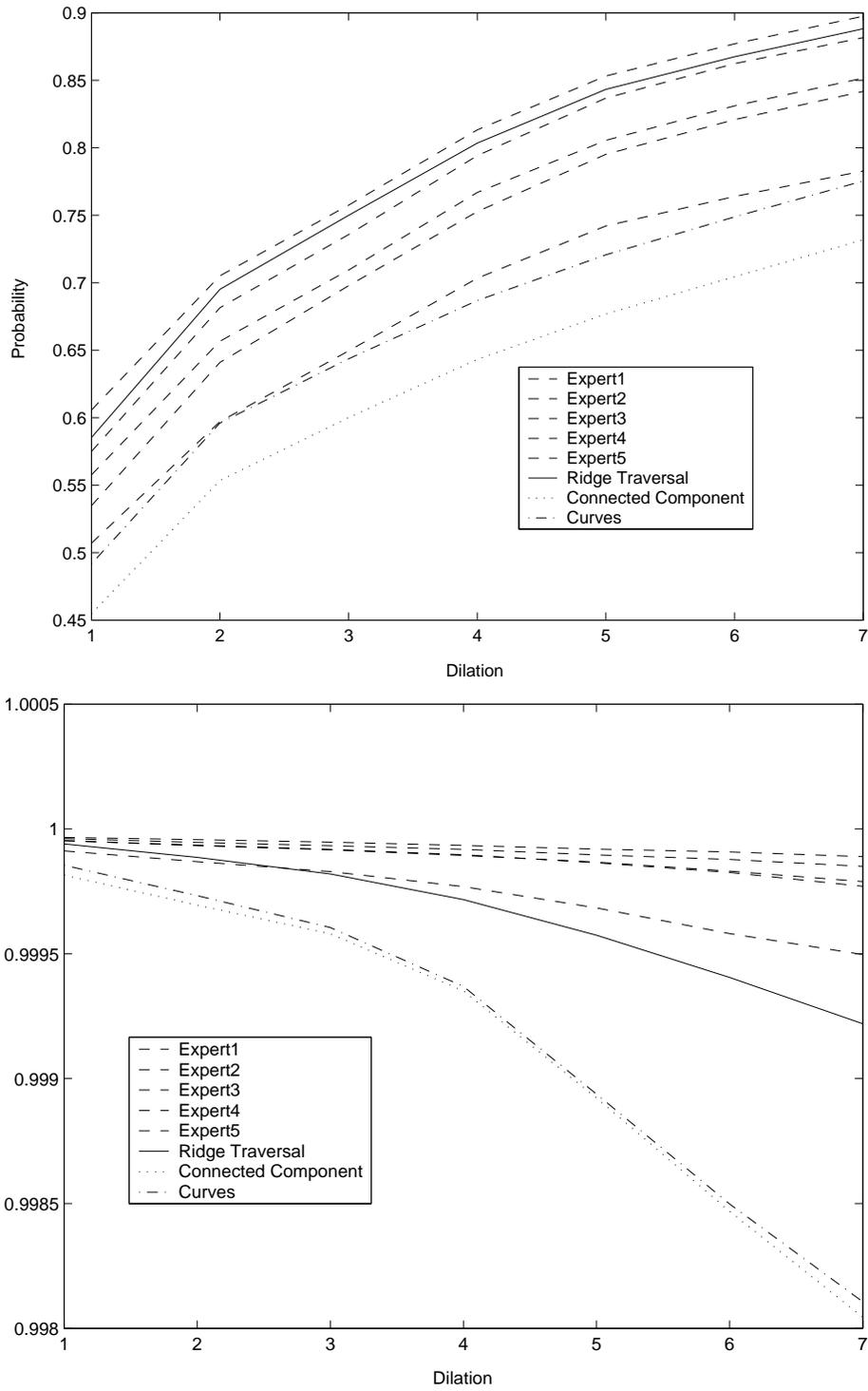


Fig. 5. Sensitivity (left) and specificity (right) of the 5 experts and the 3 segmentation methods to be compared.

4 Discussion and Conclusions

We have presented a novel algorithm for validation of open-curve segmentation and applied it to evaluating vessel segmentation methods. Our technique quantifies the accuracy of different segmentation techniques and also produces an accurate probability map of the ground truth segmentation. We have shown that ridge traversal techniques can outperform level sets and statistical method in localizing the centerline of a tubular structure.

The main drawback of our approach is that it discretizes centerlines to the pixel level. Some segmentations can have a sub-voxel accuracy, but this technique cannot be used to discern that accuracy

We are currently extending our method to estimate the radius of segmented tubular structures assuming a constant radius at each centerline point.

This work is funded in part by the Whitaker Foundation (TF-04-0008).

References

1. Aylward, S., Jomier, J., Weeks, S., Bullitt, E.: Registration of Vascular Images International Journal of Computer Vision, November 2003, pages 15
2. Aylward, S., Bullitt, E.: Initialization, Noise, Singularities, and Scale in Height-Ridge Traversal for Tubular Object Centerline Extraction IEEE Transactions on Medical Imaging, Feb, 2002, Pages 61-75
3. Kirbas C., Quek Francis K. H.: Vessel Extraction Techniques and Algorithms: A Survey Third IEEE Symposium on BioInformatics and BioEngineering (BIBE'03)
4. Lorigo L., Faugeras O., Grimson W.E.L., Keriven R., Kikinis R., Nabavi A. and Westin C.-F.: Curves: Curve evolution for vessel segmentation. Medical Image Analysis, 5:195-206, 2001.
5. Warfield S., Zou K., Wells W.: Validation of image segmentation and expert quality with an expectation-maximization algorithm MICCAI 2002: Fifth International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer-Verlag, Heidelberg, Germany, 2002, pp. 298-306
6. Wink O., Niessen W. J., Verdonck Bert, Viergever M. A.: Vessel Axis Determination Using Wave Front Propagation Analysis. MICCAI 2001: Fourth International Conference on Medical Image Computing and Computer-Assisted Intervention. 845-853