Comparison of thyroid segmentation techniques for 3D ultrasound

T. Wunderling[a], B. Golla[a], P. Poudel[b], C. Arens[c], M. Friebe[b], and C. Hansen[a]

[a]Department of Simulation and Graphics, Otto von Guericke University Magdeburg, Germany
[b]Department of Medical Engineering, Otto von Guericke University Magdeburg, Germany
[c]Clinic of Otorhinolaryngology, University Hospital Magdeburg, Germany

ABSTRACT

The segmentation of the thyroid in ultrasound images is a field of active research. The thyroid is a gland of the endocrine system and regulates several body functions. Measuring the volume of the thyroid is regular practice of diagnosing pathological changes. In this work, we compare three approaches for semi-automatic thyroid segmentation in freehand-tracked three-dimensional ultrasound images. The approaches are based on level set, graph cut and feature classification. For validation, sixteen 3D ultrasound records were created with ground truth segmentations, which we make publicly available. The properties analyzed are the Dice coefficient when compared against the ground truth reference and the effort of required interaction. Our results show that in terms of Dice coefficient, all algorithms perform similarly. For interaction, however, each algorithm has advantages over the other. The graph cut-based approach gives the practitioner direct influence on the final segmentation. Level set and feature classifier require less interaction, but offer less control over the result. All three compared methods show promising results for future work and provide several possible extensions.

Keywords: Segmentation Methodologies, Validation, Head and Neck Ultrasound

1. INTRODUCTION

The thyroid gland is a butterfly-shaped organ, consisting of two lobes connected by an isthmus. It is located in the neck, below the Adam's apple. The thyroid is part of the endocrine system and regulates important body functions. A change in volume is a reliable indicator for pathological changes. Imaging techniques are used frequently to measure the volume, with ultrasonography being the most prominent applied method for thyroid gland diagnostics.[1]

Ultrasonography has many advantages, including its ease of use and inexpensiveness. However, it also suffers from imaging artifacts and noise. There already exist a number of publications presenting methods for freehand-tracked three-dimensional ultrasound (3DUS) thyroid volume measurement. However, to the knowledge of the authors, no publication compares methods based on practicality-oriented characteristics.

In this paper, we compare and discuss three methods of thyroid segmentation based on their accuracy and ease of use. This analysis is done to make the methods directly comparable and add evaluation relevant to practitioners. In practice, fast and easy to use methods are preferable for regular application. We compare the methods to determine which ones are viable. We only consider methods that segment the exact thyroid. We provide the sixteen 3D volume datasets including the ground truth online and make them publicly available to support further research in this domain.

The conventional method of determining thyroid volume is using 2D B-mode ultrasound and approximates simple ellipsoidal shapes for both thyroidal lobes. The volume of each lobe is calculated by multiplying its dimensions $\mathrm{length}_{\mathrm{max}}$, $\mathrm{width}_{\mathrm{max}}$ and $\mathrm{depth}_{\mathrm{max}}$ and a correction factor f. Different values for f have been recommended, but usually 0.5 is applied in clinical routine.[1,2]

Further author information: (Send correspondence to C. Hansen)
Jun.-Prof. Dr. Christian Hansen: E-mail: christian.hansen@ovgu.de, Telephone: ++49 391 67-52861
Prof. Dr. rer. medic. Michael Friebe: E-mail: michael.friebe@ovgu.de, Telephone: ++49 391 67-57025

Freesmeyer et. al. have demonstrated that 3DUS is equivalent to CT and MRI in thyroid volume estimation. They also determined that manually tracing the thyroid is superior to the ellipsoid model for both regularly shaped and deformed thyroid phantoms.[1]

Automated segmentation using image processing and computer vision techniques instead of manually tracing the thyroid's shape is desirable for practical application. Several methods have been applied to the thyroid segmentation problem in the past, including neural networks,[3] geodesic active contours[4] and support vector machines.[5]

Only a few papers comparing different 2D ultrasound thyroid segmentation methods have been published so far.[6,7] Zhao et al. introduce a new segmentation algorithm and compare seven different methods in their work.[7] Their tested algorithms also include basic segmentation methods like edge detection, thresholding, region growing and watershed segmentation. Additionally, they tested active contour, normalized graph cut and their improved normalized graph cut. They found that only their improved normalized graph cut would produce a meaningful segmentation. However, their results only show segmentation on a single image, and they provide little information about their testing procedure and the data they used.

Kaur et al. compare three algorithms: active contours without edges, localized region based active contours and distance regularized level set. They compared the algorithms on scintigraphy and ultra sound images, based on parameters like execution time, number of iterations and precision. They also provide little information about the data they used for evaluation and their results are unclear. They identified the localized region based active contour algorithm as the best out of the three.[6]

The current comparative work only focuses on strictly 2D image segmentation; none compares advanced 3DUS image segmentation methods to the authors knowledge. In summary, our contributions are:

- Development of three thyroid segmentation algorithms
- Comparison of these algorithms based on
  - Segmentation quality
  - Usability
- Publicly available data for further research

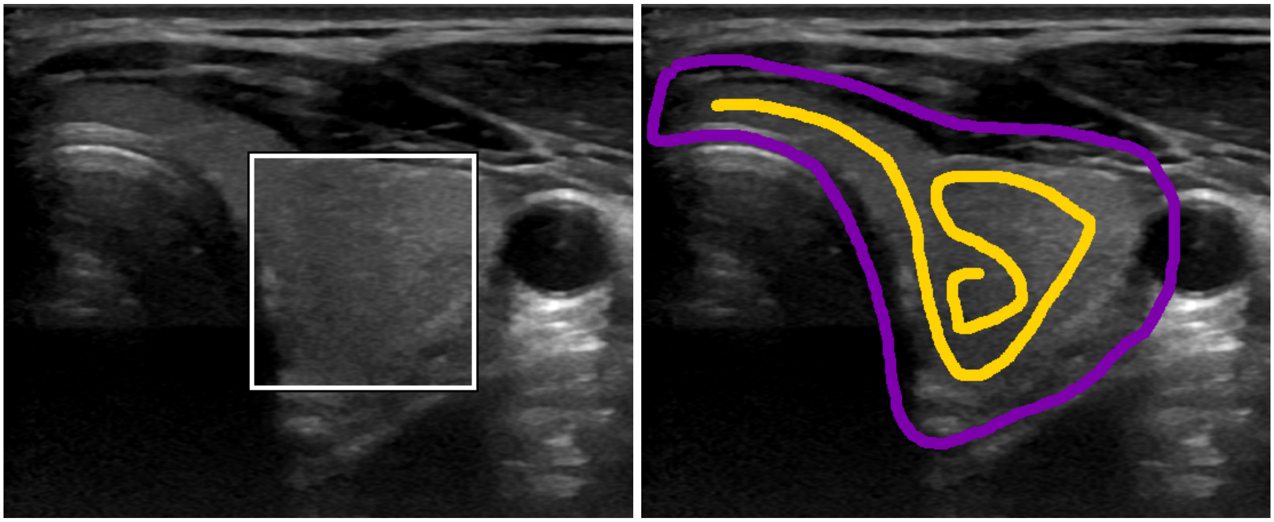
2. MATERIALS AND METHODS

Our test data consists of sixteen records of 3DUS volumes reconstructed from a tracked free-hand sweep. The volumes are stored in DICOM format and were directly obtained from the ultrasound imaging device. The device used is a GE Logiq E9 XDclear 2.0 extended with Ascension driveBay electromagnetic tracking. The probe used is the GE ML6-15. To run the 2D segmentation algorithms on the volume, it was split into individual slices. All files contain one lobe of the thyroid, but can include some parts of the isthmus and opposite lobe. It was not possible to construct a usable 3D volume from a complete sweep of both sides in one continuous motion. The ground truth contains the isthmus as part of the segmented region. It was produced by a medical expert using MeVisLab.[*] The thyroid was manually traced with Contour Segmentation Objects (CSOs) which were converted to a 3D mask afterwards via the CSOConvertoTo3DMask module. All the records were taken from healthy thyroids. The sixteen records and corresponding ground truth images are available at OpenCAS.[†]

As described in the prior section, many methods have been developed. In this section, we describe and motivate the methods we selected for comparison. The three selected methods are a level set, graph cut and a feature classifier-based approach.

[*] MeVis Medical Solutions AG, Bremen, Germany: http://www.mevislab.de/

[†] Open-CAS: http://opencas.webarchiv.kit.edu/data/thyroid.zip



(a) (b)

Figure 1. (a) Level set initialization: The user draws a rectangle to mark the size and location of the thyroid lobe. (b) Graph cut initialization: Violet scribbles mark the outside, yellow the inside of the thyroid.

2.1 Level Set

Our first approach makes use of a level set segmentation technique developed by Chan and Vese, named active contours without edges.[8] This method is based on the minimization of the Mumford-Shah functional. The contour evolution depends on region attributes instead of gradient-based edges. Hence, this method is a suitable candidate for the segmentation of noisy ultrasound images where the organ boundaries are often fuzzy and not clearly visible. Furthermore, the thyroid is visually separated from the other organs in that region, which are darker and less homogeneous.

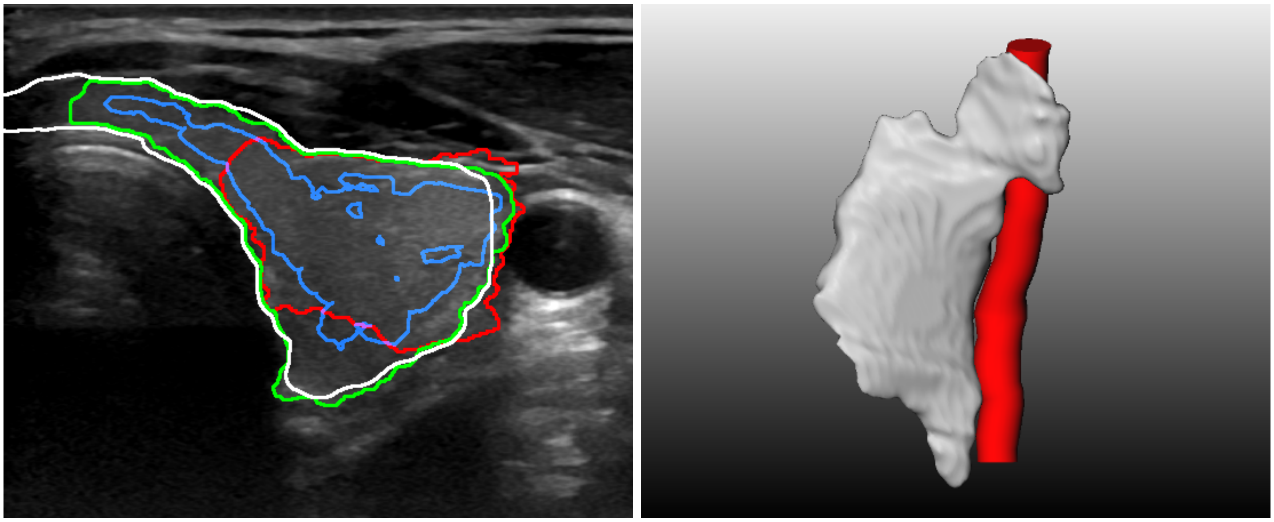
We implemented the algorithm in MATLAB to segment the thyroid in 2D images. Figure 2a shows an example segmentation in red. The main principle of this approach is to perform the segmentation task in an initial axial slice of the tracked 2D ultrasound sweep selected by the user, and then use this first result in additional segmentations of consecutive slices, by placing the initialization contour around the center of the last segmentation. This slice-wise approach was chosen because level set segmentation methods can show difficulties with narrow or elongated structures, like the isthmus.

The user is only required to select a single slice and draw a rectangle inside the thyroid lobe to initialize the segmentation process, see Fig. 1a. After this step, no more interaction is necessary, but manual corrections in form of re-initializations are possible. However, the result of the level set segmentation is very sensitive to changes in the user's initializations, making the interaction less intuitive than a graph cut-based technique.

2.2 Graph Cut

The graph cut approach also semi-automatically segments 2D ultrasonic data and combines the result to form a 3D model of the thyroid. It was realized using the implementation of the GrabCut algorithm in the OpenCV library.[9] The user marks the thyroid and surrounding regions by placing yellow and violet scribbles in some of the 2D slices, see Fig. 1b. This input is then used by the algorithm to segment the thyroid. Figure 2a shows the result in green. The output can be interactively altered by the user to correct any errors in the process.

The 2D contours of the segmented thyroid slices are combined to a 3D model. Interpolated shapes are used as an initialization for a completely automated GrabCut in other slices where the thyroid has not been fully segmented yet. The user checks the results and may correct any errors. The 3D model is updated with the new information and further automated 2D segmentations are triggered. This process repeats until a sufficiently detailed 3D model of the thyroid has been created. The accuracy of this method improves with the amount of interaction invested by the user.



(a) (b)

Figure 2. (a) Ground truth reference (white) and segmentation results: level set (red), graph cut (green) and feature-based classifier (blue). (b) Visualization showing the right lobe of the thyroid and the arteria carotis.

Thorough testing and analysis of this approach was done in cooperation with our clinical partner to find the optimum between the computationally intensive fully automated segmentation and the time-consuming user-assisted segmentation to increase accuracy while reducing computation time and interaction complexity. We suggest a compromise of allowing the user to intervene every 10 slices or every 2 mm. However, in practice the level of interaction necessary should be individually determined with each new case.

2.3 Feature-based Classifier

This segmentation method is based on the training of a decision tree, trained with three image features. The motivation for this segmentation method was the need for a fast method that allows interactive segmentation of the thyroid. Due to its speed, it is possible to implement a live visual feedback showing the segmentation to the practitioner. Based on this feedback the decision can be made whether or not the segmentation needs adjustment. With a few clicks, image regions inside and outside the thyroid can be selected such that the features from that region are used for training of the classifier. The algorithm was implemented using OpenCV.

The feature selection is based on the work of Chang et al.[3] We compute the coefficient of variation $C_V = \sigma/\mu$, with σ being the standard deviation and μ the mean, on two different sized local neighborhoods. Additionally, the mean of the smaller of the two neighborhoods is used as a feature. However, Chang et al. used additional image features as well. In separate tests, we observed that adding features only improves the segmentation marginally, while the process of feature computation takes significantly more time. The tests were only run on healthy thyroids so that the addition of features might see significant improvement in pathological cases. For the same reasons (lower computation time with small loss in accuracy) a decision tree was chosen as classifier instead of the RBF neural network used in the work of Chang et al. A drawback of this method is that no local changes are possible. Adding additional training samples via interaction changes the segmentation globally.

The segmentation obtained by the classifier itself contains some noise and regions which are not connected to the thyroid. To clean the segmentation, we performed morphological opening followed by a connected component analysis. This is only done once, after the user interaction has been completed and confirmed. Because the selected features already are defined on a local neighborhood, we did not perform pre-processing, e.g. smoothing.

3. RESULTS AND DISCUSSION

For visualization of the final segmentation results MeVisLab was utilized again: A fully automated method of segmenting the neighboring arteria carotis was implemented for this purpose using a Hessian-based vesselness

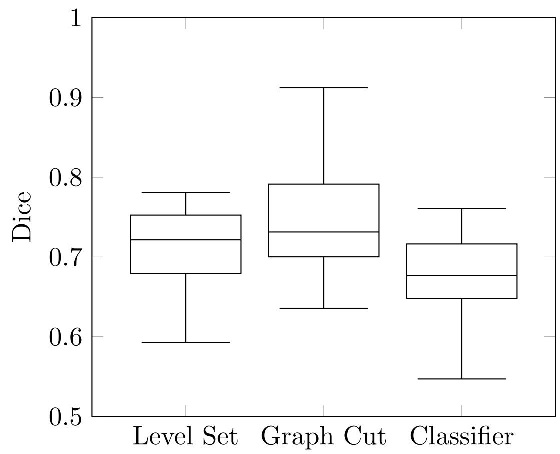


Figure 3. A box plot depicting Dice coefficients per algorithm. Whiskers represent minimum and maximum values.

filter.[10] It works reliably if the artery is completely contained within the borders of the 3D ultrasound image and recorded without any discontinuities. The result is a combined visualization, see Fig. 2b. For evaluation of the quality of the segmentation we chose the Dice coefficient. This similarity index is commonly used in medical image segmentation to compare algorithm output against reference masks.

On average, the three methods perform almost identically, with a Dice coefficient of roughly 70 percent compared to the reference segmentation. However, the range of quality differs, see Fig. 3. The feature-based classifier showed an outlier where the segmentation was particularly under-segmented (detailed overview in Figure 4). This was the case for a thyroid with a thin shape. Connected component analysis cannot prevent a leakage effect in some instances. Balancing under- and over-segmentation was also an issue for the level set approach. Limiting the number of evolution iterations to avoid leakage resulted in the contour not reaching into thin structures like the isthmus. The quality of the graph cut results scales with the amount of user interaction, which makes this method unique in this regard. This explains the greater distance between the best and worst result when using the graph cut method.

The interaction time and effort also differ between methods. The level set method requires an initial contour for segmentation. In case of over- or under-segmentation, the level set must be re-initialized by the user. Re-initialization was required 6.7 times on average. The process of initialization takes only few seconds, but the segmentation process is slow. The graph cut requires the most amount of interaction by the user. The average time spent per data set was 36 seconds. The feature-based classifier takes minimum interaction with a simple click inside and outside of the thyroid in the best case. This is mostly intuitive, but data sets with lower quality can make sample selection difficult. If the segmentation contains errors, additional clicks for adjustment are required. Sample selection took 13 seconds on average. If wrong samples were selected, the process has to be started from the beginning. This is a drawback of the implementation, but minor due to the immediate feedback and short interaction required.

4. SUMMARY AND CONCLUSION

In this paper, we compared three methods for thyroid segmentation in 3DUS. All methods aim to compute the exact object boundaries. We compared the methods based on the accuracy against a manually segmented ground truth. We also compared their usability based on duration and complexity of interaction.

No significant differences in accuracy between the compared algorithms were observed. The average Dice coefficient is similar, but varied most for the graph cut approach. In terms of interaction time the level set approach as well as the feature-based approach are preferable, whereas the graph cut method requires the most amount of interaction but offers practitioners the best control over segmentation results. All three methods show promising potential.

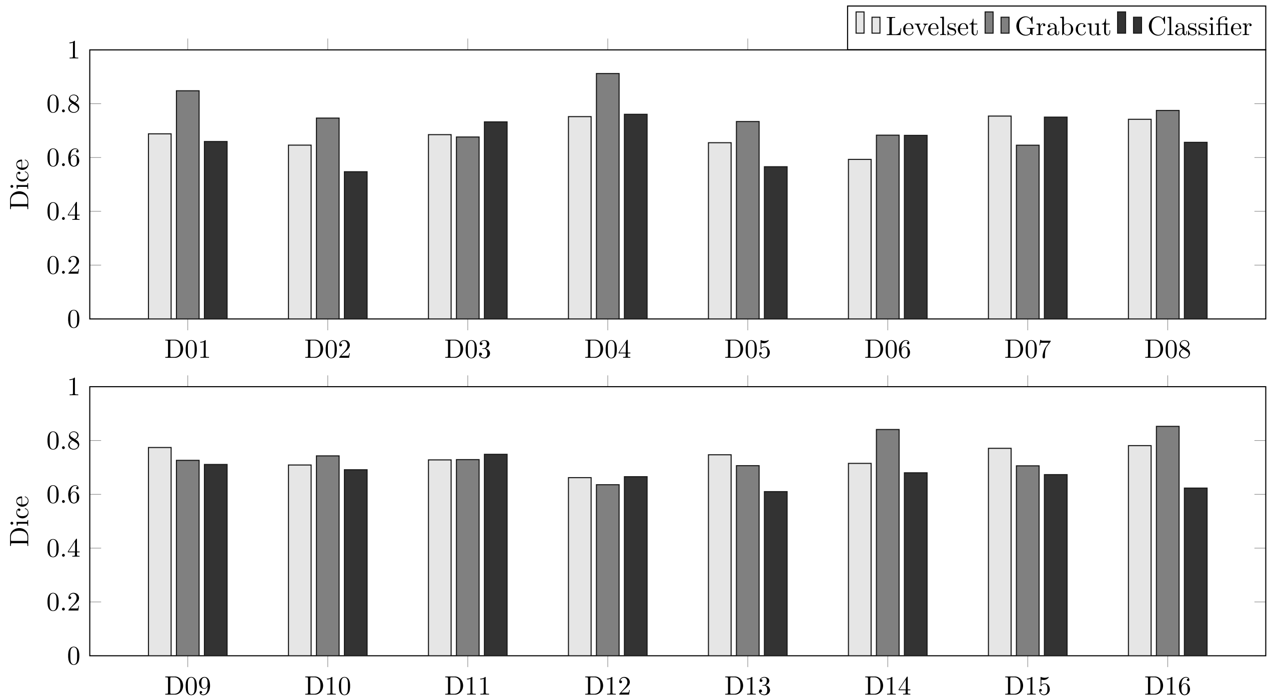


Figure 4. The Dice coefficients per data set and algorithm. Exact data can be found in Table 1.

In future work, the compared methods can be refined or combined to improve accuracy, for example offering the graph cut as an optional post-processing step after one of the other methods. Another possibility for improvement would be the automation of initialization to avoid user interaction at all. The described methods have yet to be tested on pathological datasets, which is important to evaluate the practical use and could indicate areas for further improvement.

APPENDIX A. DATA

Table 1. The Dice coefficients per data set and algorithm.

	D01	D02	D03	D04	D05	D06	D07	D08
Level Set	0.688	0.646	0.685	0.752	0.655	0.593	0.754	0.742
Graph Cut	0.848	0.746	0.676	0.912	0.734	0.683	0.646	0.775
Classifier	0.659	0.547	0.732	0.761	0.566	0.682	0.750	0.656

	D09	D10	D11	D12	D13	D14	D15	D16
Level Set	0.774	0.709	0.728	0.662	0.747	0.715	0.771	0.781
Graph Cut	0.726	0.743	0.729	0.636	0.706	0.841	0.706	0.853
Classifier	0.711	0.691	0.749	0.666	0.610	0.680	0.673	0.623

ACKNOWLEDGMENTS

We thank General Electrics, USA for providing us with financial support to carry out the research. We also thank Prof. M. Freesmeyer at the University Clinic in Jena for helping to obtain thyroid ultrasound datasets.

REFERENCES

[1] Freesmeyer, M., Wiegand, S., Schierz, J.-H., Winkens, T., and Licht, K., "Multimodal evaluation of 2-d and 3-d ultrasound, computed tomography and magnetic resonance imaging in measurements of the thyroid volume using universally applicable cross-sectional imaging software: A phantom study," *Ultrasound in Medicine and Biology* **40**(7), 1453–1462 (2014).

[2] Shabana, W., Peeters, E., and De Maeseneer, M., "Measuring thyroid gland volume: should we change the correction factor?," *American Journal of Roentgenology* **186**(1), 234–236 (2006).

[3] Chang, C. Y., Lei, Y. F., Tseng, C. H., and Shih, S. R., "Thyroid segmentation and volume estimation in ultrasound images," *IEEE Transactions on Biomedical Engineering* **57**(6), 1348–1357 (2010).

[4] Kollorz, E. N. K., Hahn, D. A., Linke, R., Goecke, T. W., Hornegger, J., and Kuwert, T., "Quantification of thyroid volume using 3-d ultrasound imaging," *IEEE Transactions on Medical Imaging* **27**(4), 457–466 (2008).

[5] Selvathi, D. and Sharnitha, V. S., "Thyroid classification and segmentation in ultrasound images using machine learning algorithms," *Proc. International Conference on Signal Processing, Communication, Computing and Networking Technologies*, 836–841 (2011).

[6] Kaur, J. and Jindal, A., "Comparison of thyroid segmentation algorithms in ultrasound and scintigraphy images," *International Journal of Computer Applications* **50**(23), 24–27 (2012).

[7] Zhao, J., Zheng, W., Zhang, L., and Tian, H., "Segmentation of ultrasound images of thyroid nodule for assisting fine needle aspiration cytology," *Health Information Science and Systems* **1**(1), 1–12 (2013).

[8] Chan, T. F. and Vese, L. A., "Active contours without edges," *IEEE Transactions on Image Processing* **10**(2), 266–277 (2001).

[9] Rother, C., Kolmogorov, V., and Blake, A., "Grabcut: Interactive foreground extraction using iterated graph cuts," *Proc. SIGGRAPH '04*, 309–314 (2004).

[10] Sato, Y., Nakajima, S., Atsumi, H., Koller, T., Gerig, G., Yoshida, S., and Kikinis, R., "3d multi-scale line filter for segmentation and visualization of curvilinear structures in medical images," *Proc. CVRMed-MRCAS'97*, 213–222 (1997).