Anatomy-informed 3D reconstruction of tracked ultrasound sweeps: a proof of concept

Alexander Furtner, Zoe Reinke, Thomas Wendler

Angaben zur Veröffentlichung / Publication details:

Furtner, Alexander, Zoe Reinke, and Thomas Wendler. 2026. "Anatomy-informed 3D reconstruction of tracked ultrasound sweeps: a proof of concept." In *Bildverarbeitung für die Medizin 2026: proceedings, German Conference on Medical Image Computing, Lübeck, March 15-17, 2026*, edited by Heinz Handels, Katharina Breininger, Thomas Deserno, Andreas Maier, Klaus Maier-Hein, Christoph Palm, and Thomas Tolxdorff, 185–92. Wiesbaden: Springer Fachmedien. https://doi.org/10.1007/978-3-658-51100-5_38.

Anatomy-informed 3D Reconstruction of Tracked Ultrasound Sweeps

A Proof of Concept

Alexander Furtner [1,2], Zoe Reinke [1,2], Thomas Wendler [1,2,3,4,5]

[1]Department of Diagnostic and Interventional Radiology and Neuroradiology, University Hospital Augsburg, Germany
[2]Digital Medicine, University Hospital Augsburg
[3]Computer-Aided Medical Procedures and Augmented Reality, Technical University of Munich, Germany
[4]Bavarian Cancer Research Center (BZKF) Augsburg, Germany
[5]Center of Advanced Analytics and Predictive Sciences, University of Augsburg, Germany
zoe.reinke@uni-a.de

Abstract. Ultrasound is a widely used imaging technique that typically produces a sequence of 2D images. However, tracked ultrasound is an increasingly popular method of converting 2D sweeps into 3D. The tracker provides data that enables subsequent reconstruction of an ultrasound volume. However, due to factors such as measurement inaccuracy, synchronization issues, and other disruptive influences, this tracking data may not accurately provide pose with exactitude. Here, we demonstrate that modifying the tracking data using anatomical knowledge can improve the compounded volume. This was achieved using a gradient descent approach to detect and correct potential errors in the tracking data. Validating against an MRI scan and comparing metrics – such as the Dice-Sørensen coefficient, Hausdorff distance, as well as volume and surface area change – demonstrates that improvements can be achieved.

1 Introduction

Ultrasound (US) is an imaging modality that is frequently used in clinics due to its low cost, non-invasiveness, and real-time visualisation capabilities. Conventional US techniques yield a series of 2D images or videos. However, it is also possible to obtain 3D volumes in real-time using special 3D US probes. Nevertheless, these systems are very cost-intensive, and are not widely available. Therefore, the utilisation of tracked US (TUS) as means of extending 2D US to 3D US is gaining popularity. Such devices combine a conventional 2D US probe with a position tracker (typically electromagnetic, optical, or inertial), whose data subsequently allows the 2D US frames ("sweep") to be reconstructed into a 3D US volume. However, tracking may contain errors due to measurement noise, synchronization issues, or interference

H. Handels et al. (Hrsg.), *Bildverarbeitung für die Medizin 2026*,
Informatik aktuell, https://doi.org/10.1007/978-3-658-51100-5_38

factors such as metallic objects for electromagnetic tracking (EMT), and occlusions for optical tracking. Consequently, here, an effort is made to optimise the obtained tracking data based on anatomical knowledge and the acquired 2D US frames.

1.1 Related work

The acquisition of 3D volumetric data from US imaging has been an active research area, driven by clinical demand for improved visualization and quantitative assessment of complex anatomical structures.

Barratt et al. provide an example of EMT device optimization in their study on carotid artery imaging. An EMT device's measurement rate and filter settings were tuned to reduce mean positional and rotational deviations, achieving sub-millimetre accuracy suitable for freehand 3D reconstruction [1]. Another significant class of work focuses on tracking and field calibration to dynamically reduce EMT errors. One method combines sensor motion models with redundant electromagnetic field measurements to perform real-time localization and field distortion mapping [2].

Beyond calibration methods, machine learning-based error compensation has gained traction. In particular, artificial neural networks have been employed to learn systematic distortion patterns from EMT data and correct them after the acquisition [3]. Moreover, advances in deep learning enable 3D freehand ultrasound acquisition without the need for external tracking systems, allowing the spatial relationship between successive frames to be inferred directly from the image data [4, 5].

Yet, to the best of our knowledge, no efforts have been made to use anatomical knowledge and the ultrasound frames to directly correct the tracked poses.

2 Materials and methods

The analysis was conducted using the SegThy dataset [6], a public dataset consisting of 3D US reconstructions of the neck with a focus on the thyroid gland, recorded in a total of 28 healthy volunteers. The US was acquired with a linear 12 MHz probe tracked using the TUS system by PIUR Imaging GmbH (Vienna, Austria). In addition to the US, the data set for each patient includes an MRI using a T1-weighted VIBE (volumetric interpolated breath-hold) sequence. In 16 of the 28 volunteers, both modalities have the structures thyroid gland, left and right carotid artery, and left and right jugular vein as segmented labels, annotated by an expert on the 3D US.

For 15 of the 16 patients, we additionally had access to two US sweeps (left and right), along with tracking data provided as a time series of 4×4 rigid transformation matrices. In this work, the tracking data was then matched to each frame using the timestamps on the frames and in the tracking data recordings.

2.1 Preprocessing of the data

The labels of the vena jugularis were excluded due to deformations by transducer pressure. As the corresponding US images exclusively display the carotid arteries in the region of the thyroid gland, the field of view of the MRI on both sides was cropped at the level of the thyroid gland.

2.2 Segmentation and reconstruction

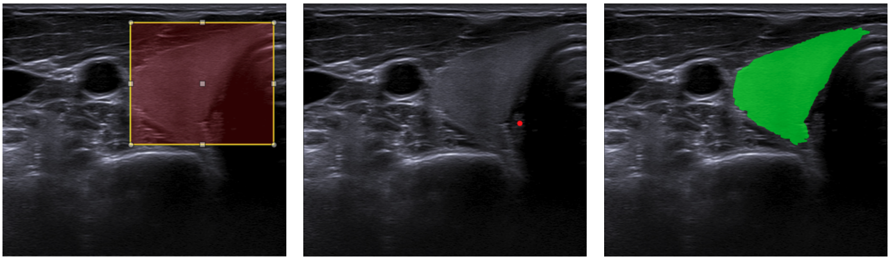
On the 2D US frames, AI-supported segmentation of the carotid arteries and thyroid lobes was performed bilaterally for each patient. To this end, a prompt-based approach employing the MedSAM2 [7, 8] model was adopted. Initially, the prompts were set on the middle frame for each US sweeps. In accordance with the provided prompts, the automatic segmentation process was executed and propagated bi-directionally. The carotid artery and thyroid lobes were both primarily segmented using bounding box prompts, but also using additional negative prompts to finetune the segmentation (Fig. 1). Notably, the prompts were manually defined for each patient to ensure optimal segmentation results, reflecting our best interpretation of the corresponding US scans.

The reconstruction of the 3D volume from transformation matrices and binary segmentation frames was performed using a voxel-based approach. First, the physical size of the volume and the corresponding bounding box were determined. For this purpose, the corner points of all US frames containing relevant image information were transformed into tracking coordinates using their respective transformation matrices.

Subsequently, a 3D bounding box was filled with voxels, employing the same voxel spacing as given in the MRI (0.65 mm $\times$ 0.65 mm $\times$ 1 mm). The frames were then transformed into their place in the volume. To minimise the gaps between the slices, a Gaussian weighting was applied in the direction of insertion. The standard deviation σ of this Gaussian was set to half the slice thickness, ensuring smooth interpolation between adjacent frames.

2.3 Optimisation of the tracking data

The optimisation of tracking data refers to the modification of individual transformation matrices targeting a more precise mapping of the anatomy.



(a) Bounding box prompt to specify that the largest coherent structure within this area is to be segmented.

(b) Additional negative prompt to ensure that this area is excluded from the segmentation.

(c) Predicted segmentation mask on the prompted frame.

Fig. 1. Example of MedSAM2 prompts used for segmentation of the right thyroid lobe and the resulting segmentation mask on the center frame of a US sweep.

Alg. 1 Optimisation of pose parameters for US reconstruction.

procedure OPTIMISATION(frames, $\text{poses}_{\text{base}}$, λ_R, λ_t, iterations, lr, ϵ)
    $\delta_{\text{rot}} \leftarrow \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^\top$ ▷ identity quaternion
    $\delta_t \leftarrow \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^\top$ ▷ zero translation vector
    Init ADAM with learning rate lr
    for $i = 1$ to iterations **do**
        Reset gradients
        $\Delta T \leftarrow$ convert δ_{rot} and δ_t to 4x4 transformation matrices
        $\text{poses}_{\text{new}} \leftarrow \text{poses}_{\text{base}} \times \Delta T$ ▷ Apply error transformation
        Reconstruct 3D volume from frames and $\text{poses}_{\text{new}}$
        Compute SDF from volume
        Compute loss L_{total}
        Backpropagate L_{total}
        Update parameters using ADAM
        Normalize δ_{rot} quaternions
    end for
    $\Delta T \leftarrow$ convert δ_{rot} and δ_t to 4x4 transformation matrices
    $\text{poses}_{\text{final}} \leftarrow \text{poses}_{\text{base}} \times \Delta T$
    return $\text{poses}_{\text{final}}$
end procedure

As demonstrated in the pseudocode (Alg. 1), this is achieved through the utilisation of gradient descent, in conjunction with a differentiable loss function that penalises anatomically non-plausible edges from the segmentation masks and "jumps" in the rotation and translation of the tracking data. The optimiser was ADAM [9]. In particular, two vectors are optimised: a quaternion rotation vector and a translation vector.

After optimisation, the quaternion vector is converted into a rotation matrix and combined with the translation vector to form a transformation matrix, as in the original tracking data. The resulting matrix is then multiplied by the actual transformation matrix, thus functioning as an error correction matrix. In each iteration of the optimisation process, the volume is reconstructed using the adjusted data. Converting the volume into a signed distance field (SDF) emphasizes the morphology of the thyroid and carotids, revealing edges and discontinuities, while simultaneously providing a representation of the surface geometry of the anatomical structures.

The specific loss function employed in this optimisation is a Charbonnier loss, which is applied to the gradients of the SDF in the x-, y-, and z-directions. This approach is designed to promote spatial smoothness in the reconstructed 3D volume by penalising substantial local alterations in the SDF.

Similar loss functions, including L1, L2, and Huber loss, were considered but showed no significant advantages.

The loss is given by

Tab. 1. Aggregated mean ± standard deviation per modality, averaged bilaterally, for ΔSurface and ΔVolume relative to MRI, with Mann-Whitney U-test p-values relative to the optimized tracking method.

Modality	ΔSurface [%]	p-value	ΔVolume [%]	p-value
Original	28.987 ± 18.117	0.202	33.126 ± 23.619	0.106
Optimised	36.475 ± 20.607	–	18.980 ± 21.736	–
Label	-2.352 ± 8.759	<0.001	1.258 ± 11.083	0.01

Tab. 2. Aggregated mean ± standard deviation per modality, averaged bilaterally, for DSC and HD95 relative to MRI, with Mann-Whitney U-test p-values relative to the optimized tracking method.

Modality	DSC	p-value	HD95 [mm]	p-value
Original	0.619 ± 0.084	0.775	6.191 ± 2.291	0.902
Optimised	0.608 ± 0.106	–	6.748 ± 3.863	–
Label	0.717 ± 0.066	0.001	3.889 ± 1.473	0.007

$$L_{\text{Charb}} = \frac{1}{N} \sum_{x,y,z} \sqrt{\left(\frac{\partial \text{SDF}}{\partial x}\right)^2 + \left(\frac{\partial \text{SDF}}{\partial y}\right)^2 + \left(\frac{\partial \text{SDF}}{\partial z}\right)^2 + \epsilon^2} \tag{1}$$

where N is the total number of voxels, used for normalization, and ϵ is a constant for numerical stability.

Furthermore, the optimisation of translation and rotation parameters is regularised by two weighting terms, λ_R and λ_t, with the objective of preventing excessive deviation from the initial tracking data. This is achieved by incorporating quadratic penalty terms into the overall loss calculation.

The total composite loss is

$$L_{\text{total}} = L_{\text{Charb}} + \lambda_{\text{rot}} \,\|\mathbf{q}_{\text{vec}}\|^2 + \lambda_{\text{t}} \,\|\mathbf{t}\|^2 \tag{2}$$

where $\mathbf{q}_{\text{vec}} = (x, y, z)$ is the vector part of the quaternion representing the rotation correction, $\mathbf{t} = (t_x, t_y, t_z)$ is the translation correction vector, λ_{R} and λ_{t} are weighting factors both fixed to 10^{-5} following a grid search between 10^{-6} and 0.5.

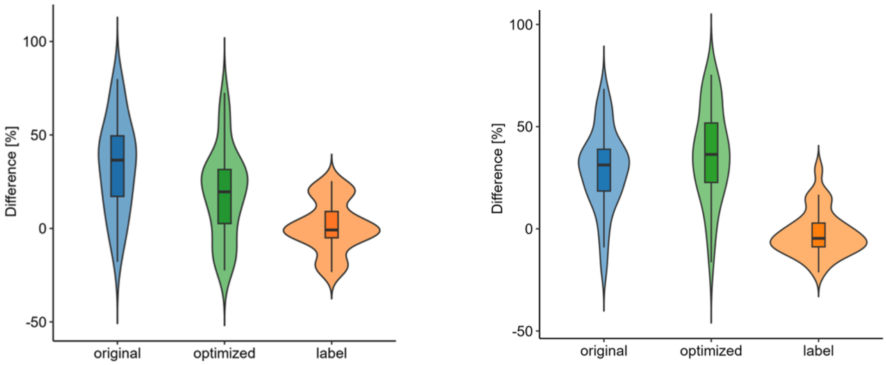
2.4 Validation

To evaluate the effect of the optimisation on the tracking data, the results were validated against MRI in a bilateral analysis. The right thyroid lobe and carotid artery were treated as a single volume, with the left side assessed accordingly. For this purpose, ImFusion Suite (Imfusion GmbH, Munich, Germany) is utilised to rigidly register the individual structure (as a labelmap) against its equivalent from the MRI and calculate the dice-sørensen coefficients (DSC), the 95th percentile of the Hausdorff distance (HD95 [mm]), and the percentage change in volume and surface area relative to the MRI (ΔV [%]). The metrics were calculated for the following three US volumes: (a) reconstruction with original tracking data, (b) reconstruction with optimised tracking data, and (c) US labels provided within SegThy – which were labeled manually in the compounded 3D US without any correction.

3 Results

The structures were evaluated bilaterally for all 15 patients, and the mean of both sides was used to obtain a patient-level quantification. The proposed tracking optimisation improved ΔV (Tab. 1, Fig. 2) compared to reconstruction using the original tracking data. Yet, the difference was not statistically significant. However, no improvement was observed for the remaining metrics using our approach (Tab. 2, Fig. 3).

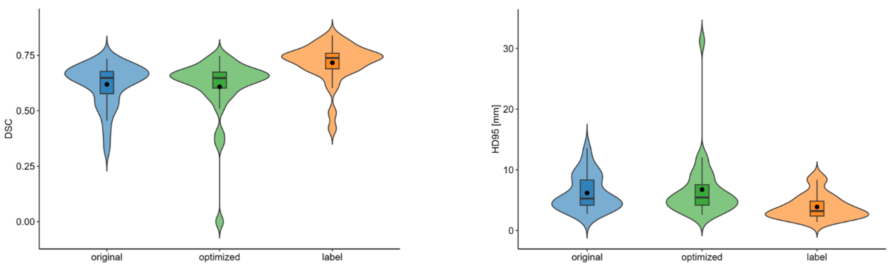
When compared to the 3D US label maps, the optimized images showed statistically significant differences in all metrics.



(a) Violin plots of percentage differences in volume relative to MRI.

(b) Violin plots of percentage differences in surface area relative to MRI.

Fig. 2. Violin plots of reconstruction accuracy using ΔVolume and ΔSurface.



(a) Violin plots of DSC relative to MRI.

(b) Violin plots of HD95 relative to MRI.

Fig. 3. Violin plots of reconstruction accuracy using DSC and HD95.

Qualitatively, the reconstructed images look smoother and lack the aforementioned anatomically non-plausible "jumps" (Fig. 4). Visual comparison with MRI masks is also satisfactory.

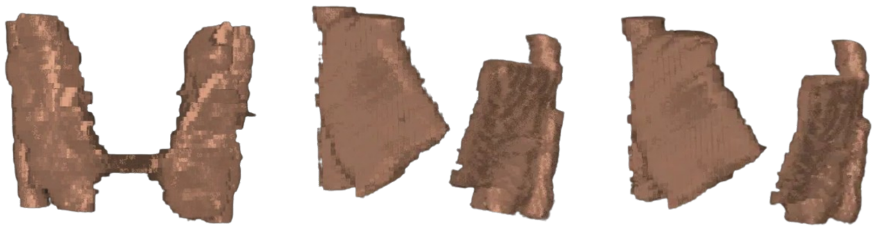
4 Discussion

In this work, we introduced a method for correcting tracking data in TUS guided by anatomical plausibility. We evaluated our approach on a fifteen-patient subset of the SegThy dataset by comparing segmentation masks of the carotid arteries and thyroid lobes between 3D US and MRI.

Our results indicate that improving the recorded tracking data by incorporating anatomical knowledge can enhance the correspondence between 3D US and MRI. Although the observed improvements were not statistically significant, the validation results point towards a potential positive change, providing an indicative proof of concept. The tracking data exhibited minimal noise and no outliers, which may explain why the optimisation had only a limited effect. The discrepancies observed between the publicly provided SegThy masks and our reconstructions suggest possible inaccuracies in the public dataset, or alternatively, that the segmentation masks of 3D US and MRI collaboratively generated by an expert and MedSAM2 are inherently more consistent.

The statistically significant difference between the proposed approach and the label maps provided by SegThy may point at differences in the interpretation of the US images by the experts in the current evaluation and the ones labeling the SegThy dataset. Potentially, the annotators of SegThy "corrected" errors in tracking, resulting in smoother structures that deviate from the underlying images.

The study is limited by the small number of evaluated patients. Moreover, the employed validation is error-prone, as MRI masks may contain inaccuracies due to the low contrast of the thyroid in the used sequences. MRI was employed as a surrogate for the ground truth, although its accuracy may be inherently limited. Also,



(a) Segmentation of the MRI volume of the thyroid and carotid arteries.

(b) Non-optimised reconstruction of the thyroid and carotid arteries.

(c) Optimised reconstruction of the thyroid and carotid arteries.

Fig. 4. Exemplary reconstructions of segmentation masks for thyroid and carotid for different modalities, including the optimised result.

perfect registration is infeasible due to motion and swallowing artifacts in both 3D US and MRI, as well as the non-rigid deformations that were not modeled here.

Future extensions of this work could focus on improving segmentation quality and incorporating deformable registration techniques to achieve better alignment between 3D US and MRI. Robustness to noise typical for EMT tracking could also be investigated to validate the optimisation for noisy data.

Extension of the evaluation to as many patients of SegThy as possible and inclusion of more organs (muscles, trachea, etc) would also be possible, as well as validation using other public datasets.

In summary, this study provides a proof-of-concept that anatomically informed correction of tracking data can enhance the reliability and internal consistency of tracked ultrasound, paving the way towards more robust and reproducible 3D US imaging.

Acknowledgement. This research was partially funded by the Bavarian Ministry of Economic Affairs, Regional Development and Energy (StMWi) under grant number LSM-2403-0017.

References

1. Barratt DC, Davies AH, Hughes AD, Thom SA, Humphries KN. Optimisation and evaluation of an electromagnetic tracking device for high-accuracy three-dimensional ultrasound imaging of the carotid arteries. Ultrasound Med Biol. 2001;27(7):957–68.
2. Sadjadi H, Hashtrudi-Zaad K, Fichtinger G. Simultaneous electromagnetic tracking and calibration for dynamic field distortion compensation. IEEE Trans Biomed Eng. 2016;63(8):1771–81.
3. Krumb H, Hofmann S, Kügler D, Ghazy A, Dorweiler B, Bredemann J et al. Leveraging spatial uncertainty for online error compensation in EMT. Int J Comput Assist Radiol Surg. 2020;15(6):1043–51.
4. Luo M, Yang X, Wang H, Du L, Ni D. Deep motion network for freehand 3D ultrasound reconstruction. Proc MICCAI. 2022.
5. Prevost R, Salehi M, Jagoda S, Kumar N, Sprung J, Ladikos A et al. 3D freehand ultrasound without external tracking using deep learning. Med Image Anal. 2018;48:187–202.
6. Krönke M, Eilers C, Dimova D, Köhler M, Buschner G, Schweiger L et al. Tracked 3D ultrasound and deep neural network-based thyroid segmentation reduce interobserver variability in thyroid volumetry. PLoS One. 2022;17(7):e0268550.
7. Ma J, Yang Z, Kim S, Chen B, Baharoon M, Fallahpour A et al. MedSAM2: segment anything in 3D medical images and videos. arXiv: 2504.03600. 2025.
8. Ravi N, Gabeur V, Hu YT, Hu R, Ryali C, Ma T et al. SAM 2: segment anything in images and videos. Proc ICLR. 2025.
9. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv: 1412.6980. 2017.