Evaluation of Low-Cost Hardware Alternatives for 3D Freehand Ultrasound Reconstruction in Image-Guided Neurosurgery

Chapter · September 2021

DOI: 10.1007/978-3-030-87583-1_11

5 authors, including:

Étienne Léger
Brigham and Women's Hospital

Houssem-Eddine Gueziri
McGill University

Louis Collins
McGill University

Tiberiu Popa
Concordia University Montreal

Evaluation of low-cost hardware alternatives for 3D freehand ultrasound reconstruction in image-guided neurosurgery

Étienne Léger[1] Houssem Eddine Gueziri[2], D. Louis Collins[2], Tiberiu Popa[1,3], and Marta Kersten-Oertel[1,3]

[1] Department of Computer Science and Software Engineering, Gina Cody School of Engineering and Computer Science, Concordia University, Montréal, QC, Canada
etienne.leger@mail.concordia.ca
[2] McConnell Brain Imaging Centre, Montreal Neurological Institute and Hospital, McGill University, Montréal, QC, Canada
[3] PERFORM Center, Concordia University, QC, Canada

Abstract. The evolution of consumer-grade hardware components (*e.g.,* trackers, portable ultrasound probes) has opened the door for the development of low cost systems. We evaluated different low-cost tracking alternatives on the accuracy of 3D freehand ultrasound reconstruction in the context of image-guided neurosurgery. Specifically, we compared two low-cost tracking options, an Intel RealSense depth camera setup and the OptiTrack camera to a standard commercial infrared optical tracking system, the Atracsys FusionTrack 500. In addition to the tracking systems, we investigated the impact of ultrasound imaging on 3D reconstruction. We compared two ultrasound systems: a low-cost handheld ultrasound system and a high-resolution ultrasound mobile station. Ten acquisitions were made with each tracker and probe pair. Our results showed no statistically significant difference between the two probes and no difference between high and low-end optical trackers. The findings suggest that low cost hardware may offer a solution in the operating room or environments where commercial hardware systems are not available without compromising on the accuracy and usability of US image-guidance.

Keywords: Low Cost · 3D Freehand Reconstruction · Ultrasound · Neurosurgery.

1 Introduction

Image-guided neurosurgery (IGNS) systems have shown positive impacts on tailoring craniotomies, reducing interventional errors, increasing tumour resection percentages and improving patient survival rates. However, these systems suffer from accuracy degradation as a procedure progresses and the patient to image alignment computed at the beginning of surgery gets invalidated by the movement and deformation of the brain, or *brain shift*. To reregister the patient intraoperatively, updated images can be acquired, using either intraoperative magnetic resonance images (iMRI) (*e.g.,* Clatz *et al.* [4]), intraoperative computed tomography (iCT) (*e.g.,* Riva *et al.* [12]) or intraoperative ultrasound (iUS) (*e.g.,* Reinertsen *et al.* [11]). The latter is much less expensive, has a smaller footprint in the OR, and has shown usefulness in neurosurgery, for intraoperative image-based registration correction to account for brain shift [15].

Although more affordable than MRI and CT solutions, the price range of US systems still varies significantly from a low-cost handheld system (∼2-8k USD) to a high-resolution station (between 50 - 250k USD). In addition to the intraoperative imaging modality, another hardware component used in IGNS to perform 3D freehand ultrasound reconstruction is the tracking system. This component, usually an optical tracking system, accounts for a substantial portion of the hardware costs of open-source IGNS systems. In this work, we compared the accuracy of ultrasound reconstruction obtained with different hardware setups at a broad range of price points. For the ultrasound transducer, two options were compared: a ∼250k USD mobile station and a ∼7k USD handheld system. For tracking four options are compared: a ∼25k USD high-end optical tracker, a ∼3k USD lower-end optical tracker, a sensor fusion hybrid tracking method which uses a ∼200 USD depth camera and a ∼20 USD equivalent RGB camera. Using these different setups we aimed to answer the following questions: Can compromises be made on some of the components without sacrificing too much on accuracy of the 3D freehand US reconstruction? If so which ones and how is a given budget best invested between these components?

2 Previous Work

Cenni *et al.* [3] looked at the effect of using different hardware setups on 3D freehand US reconstruction quality. They tested their method with two different optical tracking systems (exact models not disclosed) and report having found no noticeable difference in the reconstruction quality with the two systems. However, separate acquisitions were made independently with each tracker, making the comparison less robust.

The low cost alternative tracking method tested in our study is similar to that presented by Asselin *et al.* [1]. In their work, Asselin *et al.* developed a sensor fusion tracking method that uses a depth camera and an RGB camera to detect an ArUco marker in the RGB image to determine the x and y position in space and the depth camera to determine its z position. They found that the method worked very well, much better than when using an ArUco marker alone and would be suitable for intraoperative tool tracking. The present study thus extends their work in assessing if that low-cost hardware and method can be used for 3D freehand reconstruction. The final tracking method tested is that of using an ArUco marker alone. It serves as our baseline, similarly to Asselin *et al.*'s study.

A number of studies have investigated the reconstruction quality of different 3D freehand reconstruction algorithms. The interested reader is referred to Solberg *et al.* [14] and Rohling *et al.* [13] for a review in this area. A review of

US probe calibration in the context of 3D freehand US reconstruction is also available in [9].

3 Methodology

We designed an experiment to simultaneously test the impact of hardware on the accuracy of 3D US reconstructed volumes, with both dimensional distortion as well as shape (angular) distortion in all three dimensions.

3.1 Hardware Setups

The acquisitions were made using two ultrasound probes in combination with four tracking systems. The tested ultrasound probes are: (1) the MicrUs MC4-2R20S-3 probe (TELEMED, Vilnius, Lithuania); and (2) the BK3500 14L3 probe (BK Medical, Peabody, MA, USA). The tracking systems that were tested are: (1) ArUco markers [5] captured with the RealSense RGB camera; (2) the RealSense D435 (Intel Corporation, Santa Clara, CA, USA); (3) the Optitrack V120:Duo (NaturalPoint Inc., Corvallis, OR, USA); and (4) the Atracsys FusionTrack 500 (Atracsys LLC, Puidoux, Switzerland). All combinations of ultrasound and tracker were used to capture ultrasound acquisitions.

3.2 Phantoms and Marker Construction

To enable the most precise and fair assessment possible, a phantom and markers were designed for the experiment. A wire phantom built from Lego™ with eight wires pulled tautly through between Lego bricks was constructed (Figure 1a.). The wires form a cuboidal shape of precisely known dimensions. All wires cross perpendicularly, thus angles between line segments are precisely known. Lego bricks themselves are accurate to within 0.04 mm [7] and the wires were carefully pulled between them, which translates to a very accurate phantom. The cuboid measures 11.20 mm by 9.60 mm by 19.00 mm in x, y and z respectively. The phantom was immersed in water for US acquisition.

A custom marker, similar to that of Asselin *et al.* [1], was designed to enable all trackers (*i.e.,* both RGB camera and optical) to capture the position of the probe in the same coordinate frame (see Figure 1b.). The marker pivot (3D position) for all tracking methods was defined to be a common point at the center of the construction (the center of the ArUco marker, which corresponds to the centroid of the reflective sphere positions). The marker was 3D printed on a Raise 3D Pro2 printer (Raise 3D Technologies, Inc., Irvine, CA, USA) using a 0.1 mm layer height. A rigid probe attachment bracket was also designed and printed with the same printer settings. A similar marker was designed as a reference and attached to the phantom. This custom design and precise alignment of the tracked position for all trackers was done in order to reduce potential bias in the comparison.

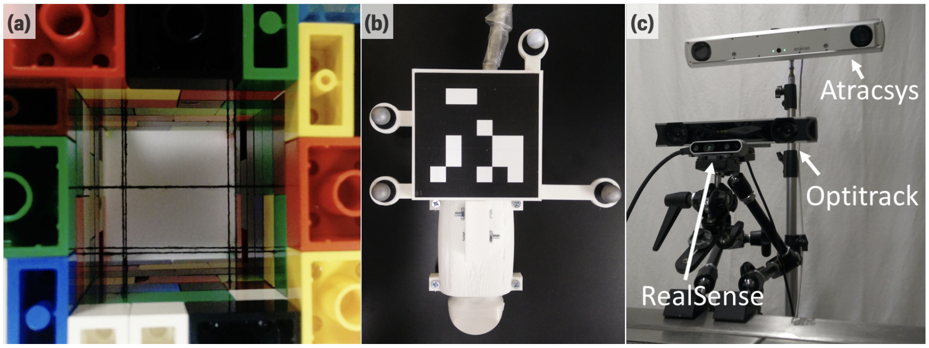

Fig. 1. Experimental setup: (a) Lego-wire phantom, (b) custom hybrid markers and probe attachment and (c) tracker arrangement for data acquisition.

3.3 Experiments

Tracking was captured simultaneously with all tracking systems for each US acquisition. Trackers were each placed at their optimal working distance from the scene to mimic a real-world scenario. Trackers were also each placed at as close as possible to the same viewing angle with respect to the scene, thus minimizing measurement volume as a confounding factor. The trackers were all aligned with the phantom so that the axes of the tracking volume would match that of the phantom (see Figure 1c.). To simplify the setup the live camera feed used for detecting the ArUco markers was that of the RealSense. This enabled us to have only three physical devices in the test setup while allowing testing with four tracking methods. The resolution of the RealSense RGB camera is 848 by 480 pixels, which is low for modern hardware. So, even though it was captured on a more expensive device, it could be achieve just as well with a $20 webcam.

The ultrasound probes were calibrated, both temporally and spatially, using fCal from PLUS (Public software Library for UltraSound imaging research) toolkit [6] (version 2.8). fCal implements the 3 N-wires calibration procedure [2], a method that was previously shown to be reliable and accurate [8]. This calibration was computed for each ultrasound probe with the tracker corresponding to the high-end of its price bracket (for the BK imaging system, the FusionTrack 500 tracker and for the Telemed imaging system, the V120:Duo tracker). The reasoning behind this was that using similarly priced devices in a system might be a more common use case.

Ten sweeps of the phantom were acquired with each ultrasound probe. For each acquisition, the tracking data was recorded simultaneously with all tracker systems. All sweeps were done in one linear motion done along the z axis. Independent reconstructions were then computed from each sweep and hardware combination, using the PLUS reconstruction [6]. Thus, from the 20 acquired sweeps, a total of 80 volume reconstructions were computed.

On each of these reconstructed volumes, the eight lines corresponding to all wires were manually segmented using 3D Slicer [10] version 4.11.20210226. The order in which segmentation was performed for each trial was randomized between conditions to reduce potential operator bias. All segmentations were performed by the same operator. The intersection of the eight line pairs corresponding to the corners of the reconstructed cuboid were computed in a least-square sense. These eight constructed points were then used in all further analysis. The distance between these points were used to compute the dimensions of the cuboid, or the dimensional distortion (DD) along each axis and angles between the line segments were used to compute angular distortion (AD) around each axis. All metrics were averaged over each axis. This means that for the dimensions, all four segments spawning from the connections of points along that axis are averaged. As well, for angles, all eight angles corresponding to rotation around that axis are averaged. Averaging is more robust and reduces the effect of uncertainty associated with segmentation. Finally, the total cuboid volume was computed, which allows easier comparison with previous studies which computed the error as a percentage of volume.

4 Results

The image to probe calibration reprojection error for the Telemed system was 0.87 mm and for the BK system 1.26 mm. The temporal calibration yielded a 38 ms latency for the Telemed and a 48 ms one for the BK.

4.1 Reconstruction Quality Results

For both DD and AD, the absolute value of the error is used in analysis as both a negative or positive error would have similarly undesirable effects on the usability of the resulting reconstruction. Table 1 shows the DD results for all combinations of hardware. Table 2 shows the AD for all combination of hardware. DD is reported as a percentage error of the supposed length value and AD is reported as an angle difference from the supposed angle (90°). Results in both tables show the mean value for each setup with the standard deviation in parentheses.

We found that the probe used had little impact on the overall accuracy of the reconstruction. A two-way ANOVA revealed that the BK and Telemed reconstructions were not statistically significantly different from one another on neither dimensional nor angular error on almost any axis. They were only different in the x dimension, where the BK was worse than the Telemed ($p = 0.0275$), for all other metrics they were not statistically different. For that reason, data for both probes was bundled in Figure 3. Reconstructions done with the Atracsys and Optitrack trackers were also not significantly different from one another on any metric and any dimension. All differences that were statistically different from the null hypothesis are labelled with stars in Figure 3.

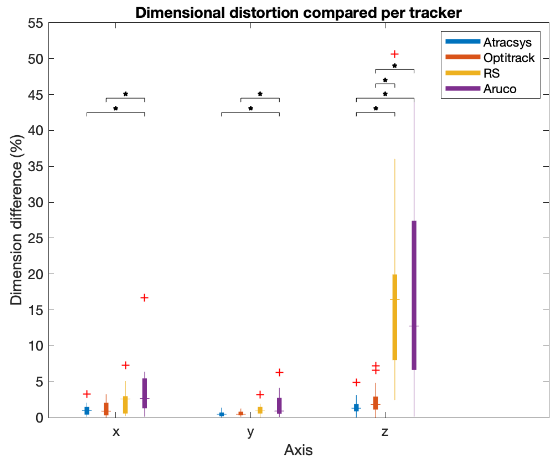

Fig. 2. Boxplots of dimensional distortions compared per tracker. Relationships marked with a star ($\star$) are those where the difference between group means are statistically significant to within $p < 0.05$. Devices are ordered in decreasing order of cost.

4.2 Qualitative Results

When visually inspecting the 3D ultrasound reconstructions a number of things can be seen. First, the Telemed and BK are clearly different in terms of image quality (Figure 4). The wires appear more fuzzy in the images acquired with the Telemed. Second, there was very noticeable visual differences in reconstruction quality between volumes obtained with either the Atracsys or Optitrack and those obtained with ArUco or RealSense. In those from the ArUco alone or RealSense the wires are much less clearly defined (those of the ArUco alone being slightly worse). This lower visual quality of the reconstruction translated

Table 1. Mean dimensional distortion per axis for each combination of hardware as well as total volumetric error. Devices are ordered in decreasing order of cost.

Probe	Tracker	Dimensional distortion (%)			Volume (%)
		x-axis	y-axis	z-axis	
BK	Atracsys	1.37 ± 0.96	0.59 ± 0.33	1.77 ± 1.29	2.54 ± 2.16
	Optitrack	1.40 ± 1.05	0.66 ± 0.38	2.13 ± 2.16	3.30 ± 2.77
	RS	2.79 ± 2.32	0.82 ± 0.43	12.28 ± 6.53	12.75 ± 9.27
	ArUco	5.06 ± 4.80	1.90 ± 2.16	19.02 ± 10.63	20.73 ± 14.94
Telemed	Atracsys	0.78 ± 0.37	0.41 ± 0.39	1.28 ± 0.85	1.14 ± 0.89
	Optitrack	1.08 ± 1.09	0.47 ± 0.28	2.70 ± 1.80	3.44 ± 2.71
	RS	1.85 ± 1.22	1.36 ± 0.86	21.65 ± 13.72	22.53 ± 13.23
	ArUco	2.64 ± 1.96	1.50 ± 1.20	13.74 ± 13.87	13.83 ± 14.66

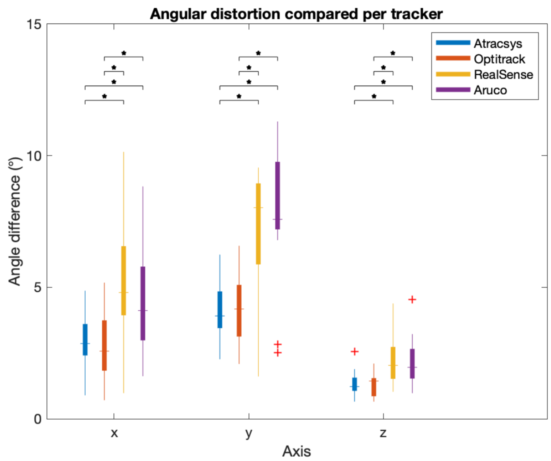

Fig. 3. Boxplots of angular distortions compared per tracker. Relationships marked with a star ($\star$) are those where the difference between group means are statistically significant to within $p < 0.05$. Devices are ordered in decreasing order of cost.

very strongly when doing the manual segmentation. Wires in the ArUco and RealSense acquired volumes were more difficult to segment. They are much noisier and jagged, which made the segmentation process more error prone. Picking the center of those wires was harder on those reconstructions than those obtained with the other two systems.

5 Discussion

The fact that reconstructions made with the BK and the Telemed probe were visually different is not very surprising, image resolution and probe frequency of the BK are significantly higher, 728×892 and 12 MHz compared to 512×512 and 4 MHz for the Telemed. However, this difference did not translate into a measurable difference in reconstruction volume quality, meaning that the wires appeared more diffuse but their position corresponded. Even though images are noisier with a low cost probe, the reconstructed volume is still accurate, which leads us to believe it would perform reasonably well in brain-shift correction or for visualizing tool trajectories (*e.g.,* catheter, ventricular drain or needle) with appropriate user training. At the same time the jaggedness of the reconstructed edges on lower cost hardware might impact intraoperative registration given that the noise introduces artificial gradients. This will be explored in future work.

The fact that cheap hardware (both probe and optical tracker) works similarly to more expensive hardware hints that errors arising from other sources (*e.g.,* calibration, reconstruction, unevenness in the sweep acquisition movement) are higher than that of the measurements for all devices, even lower cost ones.

Table 2. Mean angular distortion per axis for each combination of hardware. Devices are ordered in decreasing order of cost.

Probe	Tracker	Angular distortion (°)		
		x-axis (pitch)	y-axis (yaw)	z-axis (roll)
BK	Atracsys	2.79 ± 0.95	3.31 ± 0.56	1.36 ± 0.57
	Optitrack	3.24 ± 1.16	3.36 ± 0.96	1.31 ± 0.48
	RS	5.24 ± 1.38	7.75 ± 2.44	2.29 ± 0.94
	ArUco	5.03 ± 2.12	9.65 ± 1.37	2.16 ± 0.65
Telemed	Atracsys	3.12 ± 1.17	4.82 ± 0.70	1.19 ± 0.35
	Optitrack	2.30 ± 1.31	5.00 ± 1.12	1.36 ± 0.44
	RS	5.00 ± 2.98	6.30 ± 2.38	2.13 ± 0.90
	ArUco	4.26 ± 1.87	6.53 ± 2.18	2.16 ± 1.09

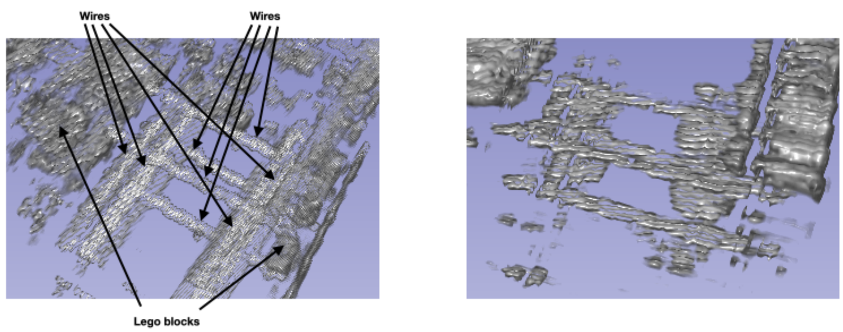

Fig. 4. Side-by-side comparison of typical reconstruction results obtained with each US acquisition systems. Both acquisitions depicted were acquired with the Atracsys tracking system. Left: BK imaging system; Right: Telemed imaging system.

In our experiment and in general in 3D freehand US reconstruction, the time difference between image timestamps and tracking timestamps is assumed to be fixed. The temporal calibration done prior to acquisition enables computing this time difference, which can then be compensated for in software upon data streams arrivals. However, it was observed that the BK latency fluctuated over time. For this reason, the BK was perhaps at a bit of a disadvantage. In our particular setup this fluctuation could have arisen from many sources: US system software, network card drivers, operating system or other receiving software. Latency should be considered with great care in this application and efforts should be made to ensure that the latency is not only as low as possible, but also, and very importantly, that it remains as constant as possible throughout an acquisition.

The difficulty described in the previous section in doing the manual segmentation on the cheapest tracking hardware has consequences beyond just the segmentation process itself. Not only is the process more time consuming for these acquisitions as the viewer takes longer to understand the US images but more importantly this leads to less accurate segmentation. This less accurate segmentation might be what causes both ArUco and RealSense to be indistinguishable statistically. Values are quite different for the z-axis, but the standard deviation on both samples is also large. There is a possibility that a genuine statistical difference between the two might be obfuscated by this segmentation difficulty due to low quality of the reconstructed volumes.

We found that all systems, even higher end ones, performed significantly worse in the z direction. This was expected, as all tracking methods tested, be it the commercial optical trackers or the experimental sensor-fusion method, are vision-based, meaning that they measure distances in images. They are therefore more accurate in the image plane than perpendicular to it. Although, and while all system suffer from this, the marker-based (ArUco and RealSense) were much more affected.

Finally, it is worth noting an important limitation in the design of our experiment. Manual wire segmentation, as performed in the experiment, allowed us to compensate for discontinuous data, especially when the quality of the reconstructed volume was low. Although this approach allows for capturing of the overall dimensional and angular distortions, local artifacts such as deformations and mis-reconstructions were attenuated. The effect of these artifacts on the outcome of an IGNS application need to be investigated.

6 Conclusion and Future Work

In this study, a phantom and protocol to measure 3D freehand US reconstruction distortion was presented. The wire and Lego phantom is easy and cheap to build and the protocol is easy to replicate. This allows for a more standardized comparison of reconstruction methods and tracking methods in the future. The protocol was used in a study to gain insight into the impact of different hardware components' cost on reconstruction accuracy. Four tracking systems were compared, whose cost were an order of magnitude apart from one another, as well as two US imaging systems that were roughly two orders of magnitude apart in price. We found that the cheapest US imaging system didn't yield reconstructions that were measurably worse than the high-end system. This thus suggests that for this application a cheap US imaging system may be used to reduce overall system cost. For tracking, the cheapest optical tracking system performed statistically the same as the high-end optical tracker. This shows the feasibility of using low cost hardware. However, the camera-based tracking methods performed significantly worse. To improve on the sensor fusion method, the depth camera could be used to track the shape of a marker in space, and this will be explored in future work. In future work as well, a second series of experiments will be performed to test more specifically how volume registration is impacted by the varying quality of reconstructions obtained with different hardware components.

References

1. Asselin, M., Lasso, A., Ungi, T., Fichtinger, G.: Towards webcam-based tracking for interventional navigation. In: Fei, B., III, R.J.W. (eds.) Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling. vol. 10576, pp. 534 – 543. International Society for Optics and Photonics, SPIE (2018). https://doi.org/10.1117/12.2293904, `https://doi.org/10.1117/12.2293904`
2. Carbajal, G., Lasso, A., Gómez, Á., Fichtinger, G.: Improving N-wire phantom-based freehand ultrasound calibration. International Journal of Computer Assisted Radiology and Surgery **8**(6), 1063–1072 (2013). https://doi.org/10.1007/s11548-013-0904-9
3. Cenni, F., Monari, D., Desloovere, K., Aertbeliën, E., Schless, S.H., Bruyninckx, H.: The reliability and validity of a clinical 3D freehand ultrasound system. Computer Methods and Programs in Biomedicine **136**, 179–187 (2016). https://doi.org/10.1016/j.cmpb.2016.09.001
4. Clatz, O., Delingette, H., Talos, I.F., Golby, A.J., Kikinis, R., Jolesz, F.A., Ayache, N., Warfield, S.K.: Robust nonrigid registration to capture brain shift from intraoperative MRI. IEEE Transactions on Medical Imaging **24**(11), 1417–1427 (2005). https://doi.org/10.1109/TMI.2005.856734
5. Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J., Marín-Jiménez, M.J.: Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognition **47**(6), 2280–2292 (2014). https://doi.org/10.1016/j.patcog.2014.01.005
6. Lasso, A., Heffter, T., Rankin, A., Pinter, C., Ungi, T., Fichtinger, G.: PLUS: Open-source toolkit for ultrasound-guided intervention systems. IEEE Transactions on Biomedical Engineering **61**(10), 2527–2537 (2014). https://doi.org/10.1109/TBME.2014.2322864
7. Lemes, S.: Comparison of similar injection moulded parts by a coordinate measuring machine. SN Applied Sciences **1**(2), 1–8 (2019). https://doi.org/10.1007/s42452-019-0191-3
8. Mercier, L., Del Maestro, R.F., Petrecca, K., Kochanowska, A., Drouin, S., Yan, C.X.B., Janke, A.L., Chen, S.J.S., Collins, D.L.: New prototype neuronavigation system based on preoperative imaging and intraoperative freehand ultrasound: system description and validation. International Journal of Computer Assisted Radiology and Surgery **6**(4), 507–522 (2011). https://doi.org/10.1007/s11548-010-0535-3
9. Mercier, L., Langø, T., Lindseth, F., Collins, D.L.: A review of calibration techniques for freehand 3-D ultrasound systems. Ultrasound in Medicine and Biology **31**(4), 449–471 (2005). https://doi.org/10.1016/j.ultrasmedbio.2004.11.015
10. Pieper, S., Halle, M., Kikinis, R.: 3D Slicer. 2004 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano **1**(March 2015), 632–635 (2004). https://doi.org/10.1109/isbi.2004.1398617
11. Reinertsen, I., Lindseth, F., Askeland, C., Iversen, D.H., Unsgård, G.: Intraoperative correction of brain-shift. Acta Neurochirurgica **156**(7), 1301–1310 (2014). https://doi.org/10.1007/s00701-014-2052-6
12. Riva, M., Hiepe, P., Frommert, M., Divenuto, I., Gay, L.G., Sciortino, T., Nibali, M.C., Rossi, M., Pessina, F., Bello, L.: Intraoperative computed tomography and finite element modelling for multimodal image fusion in brain surgery. Operative Neurosurgery **18**(5), 531–541 (2019). https://doi.org/10.1093/ons/opz196

13. Rohling, R., Gee, A., Berman, L.: A comparison of freehand three-dimensional ultrasound reconstruction techniques. Medical Image Analysis **3**(4), 339–359 (1999). https://doi.org/10.1016/S1361-8415(99)80028-0
14. Solberg, O.V., Lindseth, F., Torp, H., Blake, R.E., Nagelhus Hernes, T.A.: Freehand 3D Ultrasound Reconstruction Algorithms-A Review. Ultrasound in Medicine and Biology **33**(7), 991–1009 (2007). https://doi.org/10.1016/j.ultrasmedbio.2007.02.015
15. Unsgaard, G., Rygh, O.M., Selbekk, T., Müller, T.B., Kolstad, F., Lindseth, F., Hernes, T.A.N.: Intra-operative 3D ultrasound in neurosurgery. Acta Neurochirurgica **148**(3), 235–253 (2006). https://doi.org/10.1007/s00701-005-0688-y