

9-21-2021 2:00 PM

Mixed-reality visualization environments to facilitate ultrasound-guided vascular access

Leah Groves, *The University of Western Ontario*

Supervisor: Terry, Peters M., *The University of Western Ontario*

Co-Supervisor: Elvis, Chen C. S., *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Biomedical Engineering

© Leah Groves 2021

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Other Biomedical Engineering and Bioengineering Commons](#), and the [Systems and Integrative Engineering Commons](#)

Recommended Citation

Groves, Leah, "Mixed-reality visualization environments to facilitate ultrasound-guided vascular access" (2021). *Electronic Thesis and Dissertation Repository*. 8151.
<https://ir.lib.uwo.ca/etd/8151>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

Ultrasound-guided needle insertions at the site of the internal jugular vein (IJV) are routinely performed to access the central venous system. Ultrasound-guided insertions maintain high rates of carotid artery puncture, as clinicians rely on 2D information to perform a 3D procedure. The limitations of 2D ultrasound-guidance motivated the research question: “Do 3D ultrasound-based environments improve IJV needle insertion accuracy?”. We addressed this by developing advanced surgical navigation systems based on tracked surgical tools and ultrasound with various visualizations. The point-to-line ultrasound calibration is used to enable the use of tracked ultrasound. We automated the fiducial localization required for this calibration method such that fiducials can be automatically localized within 0.25 mm of the manual equivalent. The point-to-line calibration obtained with both manual and automatic localizations produced average normalized distance errors less than 1.5 mm from point targets. Another calibration method was developed that registers an optical tracking system and the VIVE Pro head-mounted display (HMD) tracking system with sub-millimetre and sub-degree accuracy compared to ground truth values. This co-calibration enabled the development of an HMD needle navigation system, in which the calibrated ultrasound image and tracked models of the needle, needle trajectory, and probe were visualized in the HMD. In a phantom experiment, 31 clinicians had a 94 % success rate using the HMD system compared to 70 % for the ultrasound-only approach ($p=0.018$). We developed a machine-learning-based vascular reconstruction pipeline that automatically returns accurate 3D reconstructions of the carotid artery and IJV given sequential tracked ultrasound images. This reconstruction pipeline was used to develop a surgical navigation system, where tracked models of the needle, needle trajectory, and the 3D Z-buffered vasculature from a phantom were visualized in a common coordinate system on a screen. This system improved the insertion accuracy and resulted in 100 % success rates compared to 70 % under ultrasound-guidance ($p=0.041$) across 20 clinicians during the phantom experiment. Overall, accurate calibrations and machine learning algorithms enable the development of advanced 3D ultrasound systems for needle navigation, both in an immersive first-person perspective and on a screen, illustrating that 3D US environments outperformed 2D ultrasound-guidance used clinically.

Summary for Lay Audience

Central line insertions are used to access the central venous system require inserting a needle into a vein, such as the internal jugular vein (IJV) on the neck. Ultrasound-(US)-guidance is used to navigate the needle into the vein using 2D US images, but still results in high complication rates, such as carotid artery (CA) puncture, as clinicians rely on 2D information to guide a 3D procedure. The use of 3D US environments for surgical navigation is an active research area and this thesis focuses on the development of 3D US-based systems for needle navigation into the IJV. A surgical navigation system was developed that employed a spatial tracking system and provides tracked US images aligned with tracked models of the needle, needle trajectory, and US probe. Tracked US requires US probe calibration that enables the US image to be positioned and scaled to the field-of-view of the US beam, such that the interactions between the image and surgical tools can be visualized for the user. This system can be rendered on a monitor or within a head-mounted display (HMD) by means of a co-calibration apparatus that places the surgical navigation system described above in a first-person perspective in the HMD. The components required to develop this system were evaluated and found to be highly accurate. The surgical navigation system displayed on the monitor and in the HMD were compared to the US-guidance technique, and the HMD system significantly improved the needle insertion accuracy in a phantom study with 31 experienced clinicians compared to the US-guidance approach. As clinicians have a preference for monitor-based systems, a second surgical navigation system was developed, based on a method to automatically obtain 3D models of the CA and IJV given a tracked US recording that was developed in this work. In a phantom study with 20 experienced clinical practitioners a system that employed models of the 3D vessel models and surgical tools, delivered significantly improved the needle insertion accuracy, compared to the system that employed US-only guidance. Overall, 3D US environments outperformed US-guidance for needle navigation.

Keywords: Surgical navigation, ultrasound, tracking, reconstruction, needle-based interventions, machine learning

Co-Authorship Statement

This thesis integrates several publications that are published or in preparation for submission. Details regarding the author's contributions to these manuscripts are provided below.

Chapter 2 Authors: L. A. Groves, B. B. VanBerlo, T. M. Peters, and E. C. S. Chen

Status: Published

Reference: Groves, L. A., VanBerlo, B., Peters, T. M., & Chen, E. C. (2019). Deep learning approach for automatic out-of-plane needle localisation for semi-automatic ultrasound probe calibration. *Healthcare technology letters*, 6(6), 204.

Contributions: There are significant differences between the work published in this paper and the work presented in this thesis. For the original publication, my contributions were data collection and manual annotation, guidance and assistance with machine learning development, integration of the machine learning algorithm into the calibration algorithm, study methodology, data analysis, and manuscript preparation. B. VanBerlo developed and trained the machine-learning algorithm and assisted with manuscript preparation and editing. The work was performed under the supervision of both T. M. Peters and E. C. S. Chen, and both helped in reviewing and editing the manuscript. For this publication my contribution was 80 %. However, the work presented in the thesis chapter varies from the publication. For this chapter, I altered machine-learning algorithm, performed a four-fold train-test analysis, expanded the results to include calibration accuracy analysis, and made significant changes to the methods, results, and discussion sections. My contributions to this thesis chapter are 90 %.

Chapter 3 Authors: L. A. Groves, P. Carnahan, D. R. Allen, A. Rankin, T. M. Peters, E. C. S. Chen.

Status: Published

Reference: Groves, L. A., Carnahan, P., Allen, D. R., Rankin, Adam., Peters, T. M., & Chen, E. C. S. (2019). Accuracy assessment for the co-registration between optical and VIVE head-mounted display tracking. *International journal of computer assisted radiology and surgery*, 14(7), 1207-1215.

Contributions: My contributions to this work include: assistance with system development, study design, data collection, data analysis, and manuscript preparation. P. Carnahan: Assisted with data collection, CAD design, and data analysis, and manuscript preparation. D. R. Allen: Assisted with data collection and analysis. AR: conceptual and CAD design for the co-calibration apparatus. The work was performed under the supervision of both T. M. Peters and E. C. S. Chen, and both helped in reviewing and editing the manuscript. E. C. S. Chen also assisted with the study design. My contributions to this work were 75%.

Chapter 4 Authors: L. A. Groves, N. Li, T. M. Peters, and E. C. S. Chen

Status: Published

Reference: Groves, L., Li, N., Peters, T. M., & Chen, E. C. (2019, October). Towards a mixed-reality first-person point of view needle navigation system. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 245-253). Springer, Cham.

Contributions: My contributions include the system and phantom development, study design, data collection and analysis, and manuscript preparation and writing. N. Li. assisted with the phantom development, data collection and data analysis. E. C. S. Chen contributed to the technical development of the systems and the study design. The work was performed under the supervision of both T. M. Peters and E. C. S. Chen. All authors helped in reviewing and editing the manuscript. My contribution to this work was 90%.

Chapter 5 Authors: L. A. Groves, B. VanBerlo, N. Veinberg, A. Alboog, T. M. Peters, and E. C. S. Chen.

Status: Published

Reference: Groves, L. A., VanBerlo, B., Veinberg, N., Alboog, A., Peters, T. M., & Chen, E. C. S. (2020). Automatic segmentation of the carotid artery and internal jugular vein from 2D ultrasound images for 3D vascular reconstruction. International Journal of Computer Assisted Radiology and Surgery, 15(11), 1835-1846.

Contributions: My contributions to this work was all data collection, development of the reconstruction pipeline, the research methods and data analysis, and writing. B. VanBerlo trained the Mask R-CNN and U-Net algorithms and provided the trained models to me for analysis and use within the reconstruction pipeline. N. Veinberg performed all the manual segmentations of the vasculature. A. Alboog performed the US scans. This work was performed under the supervision of both T. M. Peters and E. C. S. Chen. All authors helped in reviewing and editing the manuscript. My contributions to this work was 80%.

Chapter 6 Authors: L. A. Groves, N. Li, B. VanBerlo, N. Veinberg, A. Alboog, T. M. Peters, and E. C. S. Chen.

Status: Published

Reference: Groves, L. A., Li, N., VanBerlo, B., Veinberg, N., Peters, T. M., & Chen, E. C. (2020). Improving central line needle insertions using *in-situ* vascular reconstructions. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 1-7.

Contributions: My contributions include system developed, study design, data collection and

analysis, and manuscript preparation and writing. N. Li assisted with phantom develop and data collection. B VanBerlo trained the machine learning algorithm on phantom data. N. Veinberg manually segmented the phantom data for training. This work was performed under the supervision of both T. M. Peters and E. C. S. Chen. All authors helped in reviewing and editing the manuscript. My contributions to this work was 90%.

Acknowledgements

I would like to acknowledge both Dr. Terry Peters and Dr. Elvis Chen for their mentorship, guidance and supervision during my time at Western. I am incredibly grateful to have had the opportunity to work with both of these amazing scientists and for all the opportunities they have provided me during my time in their lab. I would like to thank Dr. Peters for always allowing his students to investigate research areas that peak their interest. Dr. Peters was always encouraging us to have a full graduate student experience, supporting students with extracurricular involvement, engagements with other researchers through conference attendance, and never forgetting to remind us to get outside on nice days. I would like to express my gratitude to Dr. Chen, for his patience, valuable discussions, and enthusiasm about research. Dr. Chen always provided his opinions to me as both a mentor and a friend, which helped me navigate my research and graduate school path. From both Dr. Peters and Dr. Chen I have learned what amazing mentorship looks like, and I will carry this with me for the rest of my life.

I would like to express my gratitude to my advisory committee members Dr. Daniel Bainbridge and Dr. Jody Culham for their feedback and insights. I would like to thank Adam Rankin for his technical expertise, teaching me the ropes of 3D Slicer and his continuous mentorship and honest opinions. I would like to thank Hilda for assistance with recruitment of clinical practitioners and assistance with study logistics. I would also like to recognize all of the members of the VASST Lab. It has been an honour to work along side such a talented, kind, hardworking, supportive, and collaborative team.

I would like to express my gratitude to the many friends I have made along this journey who have supported me and enriched my graduate student experience. To the biomedical graduate student committee for brightening my life and the lives of every BME graduate student. Thank you for all you do for the program, and for supporting every endeavor I initiated. A special thanks to Salma Dammak who was always willing to be involved and take on responsibilities outside of the scope of her position, many my contributions to the committee would not have been possible without her hard work.

To my high school and undergraduate friends who supported me, listened when I needed an outside ear to talk to, and maintained our friendships despite my ever busy and changing schedule. To my family, thank you for supporting me, encouraging me, and having the utmost faith in my abilities. A special thank you to my mom, who patiently listened to endless ramblings and frustrations and was always there when I needed her. To my Omi who did not make it to the end of this journey but whose love, support, and pride I will carry with me forever. To my amazing partner Em, who loved and supported me through all the difficult days and who was there to celebrate every success, I could not have done this without you.

Contents

Abstract	i
Summary for Lay Audience	ii
Co-Authorship Statement	iii
Acknowledgements	vi
List of Figures	xi
List of Tables	xviii
List of Appendices	xx
1 Introduction	1
1.1 Clinical Procedure of Interest	1
1.1.1 Relevant Anatomy	1
Current clinical approaches and complications	2
1.2 Ultrasound for Interventional Guidance	3
1.2.1 Overview of Ultrasound principles	3
1.2.2 Physical principles of US as they relate to interventional imaging ad- vantages and limitations	7
1.2.3 Ultrasound guidance	8
1.3 Tracking Technology	10
1.3.1 Spatial Tracking	10
1.4 Ultrasound Registration and Calibration	15
1.4.1 Registration approaches	17
1.4.2 Tool Calibration and Visualization	19
1.5 Ultrasound Probe Calibration	23
1.5.1 Free-hand Tracked Ultrasound	26
1.5.2 Temporal Calibration	26

1.6	Registration Errors and Ultrasound Calibration Evaluation	27
1.6.1	Sources of Error in Fiducial-based Registrations	27
1.6.2	Calibration Quality Assessment	29
1.7	Ultrasound segmentation	30
1.7.1	Image processing-based segmentation	31
1.7.2	Machine learning-based segmentation	31
1.7.3	Image-based tracking	34
1.8	Ultrasound Reconstruction Methods	35
1.9	US Visualization	37
1.9.1	Image Fusion	38
1.10	Display Technology	39
1.10.1	Mixed-reality Spectrum	39
1.10.2	Ergonomic displays	40
1.10.3	Wearable Technology	41
1.10.4	Overview of Current Devices	42
1.10.5	Psychophysical aspects	45
1.10.6	Surgical Navigation Systems	46
1.11	Limitations of current systems	46
1.12	Thesis Outline	47
1.12.1	Point-to-Line Ultrasound Probe Calibration with automatic fiducial localization	47
1.12.2	Accuracy Assessment for the Co-registration Between Optical and VIVE Head-Mounted Display Tracking	48
1.12.3	First-Person Perspective Ultrasound-Guided Central Line Insertions	48
1.12.4	Accuracy Assessment for the 3D Reconstruction of the Neck Vasculature from Transverse 2D Ultrasound Images	48
1.12.5	Effect of Using 3D Anatomical Information on a 2D Monitor on Complication Rates During Central Line Insertions	49
1.12.6	Conclusions	49
2	Point-to-Line Ultrasound Probe Calibration with automatic fiducial localization	50
2.1	Introduction	50
2.2	Methods	54
2.2.1	Equipment and Data Collection	55
2.2.2	Machine learning-based needle segmentation	57
2.2.3	US calibration with automatic fiducial localization	58

2.2.4	Quality Assessment	59
2.3	Results	64
2.3.1	Automatic Needle Localization Accuracy	64
2.3.2	Quality Assessment	65
2.4	Discussion	69
2.5	Conclusions	73
3	Accuracy Assessment for the Co-registration Between Optical and VIVE Head Mounted Display Tracking	74
3.1	Introduction	74
3.2	Materials and Methods	76
3.2.1	Equipment	76
3.2.2	Co-calibration Process	78
3.2.3	Accuracy Analysis	79
3.2.4	Positional accuracy assessment	79
	Rotational accuracy assessment	80
3.3	Results	82
3.4	Discussion	83
3.5	Conclusions	86
4	Towards a mixed-reality first-person point of view needle navigation system for central line insertions	87
4.1	Introduction	87
4.2	Materials and Methods	89
4.2.1	Statistical Analysis	92
4.3	Results	93
4.4	Discussion	95
4.5	Conclusion	98
5	Automatic segmentation of the carotid artery and internal jugular vein from 2D ultrasound images for 3D vascular reconstruction	99
5.1	Introduction	99
5.2	Materials and Methods	101
5.2.1	Data Collection	101
5.2.2	Deep Learning Segmentation	103
5.3	Vessel reconstruction	105
5.4	Results	106

5.5	Discussion	111
5.6	Conclusions	114
6	Improving central line needle insertions using in-situ vascular reconstructions	115
6.1	Introduction	115
6.2	Materials and methods	116
6.2.1	System development	116
6.2.2	Experimental Methods	121
6.3	Results	123
6.4	Discussion	124
6.5	Conclusions	126
7	Conclusions and Future Directions	127
7.1	Contributions of the Thesis	127
7.2	Future Directions	130
	Bibliography	133
	A Questionnaires	151
	B Research Ethics Approval	154
	C Reprint Permissions	158
	Curriculum Vitae	163

List of Figures

1.1	Vascular anatomy in the neck, where the direction of blood flow is indicated with black arrows	2
1.2	Visual depiction of the landmark-based central line insertion technique	3
1.3	Visual depiction of the three ways an US wave may be redirected after making contact with a material boundary, illustrating a) a reflected wave, b) a refracted wave, and c) a scattered wave.	5
1.4	Diagram illustrating the axial, lateral, and elevation resolution. This diagram also depicts the 3D beam profile from a single crystal element.	6
1.5	Two representative visualizations of US-guided central line insertions. Figure a) depicts the US visualization provided for out-of-plane needle insertions and figure b) depicts an example of the spatial configuration of the needle, the US probe and image, and the vessel during the out-of-plane insertion. Figure c) provides an US visualization for an in-plane needle insertion, and figure d) provides the spatial configuration of the needle, US probe and image, and vessel.	9
1.6	Collinear camera stereo reconstruction (triangulation) diagram illustrating how a point can be reconstructed in 3D using two cameras. The diagram shows the Left and Right cameras, the distance between them (b), the focal length (f), the pixel coordinate of the object on the image U_L and U_R , and the reconstructed point $P(X_R, Y_R, Z_R)$	11
1.7	Polaris tracker FoV, showing the front, side and top view with all dimensions labelled. The diagram has the optimal zone labelled on the side and top view, denoted as O.Z., which extends from the vertical line in the diagram.	12
1.8	Aurora tabletop tracker and FoV labelled with measurements.	13
1.9	Graphic demonstrating the 5 and 6 DoF tracking of a stylus, courtesy of NDI https://www.ndigital.com/technology/6dof-explained/	14
1.10	Two instances of tool tracking, with graphic a) illustrating tracking a sensor with respect to the origin of the tracking system, and graphic b) illustrating tracking a sensor with respect to another sensor denoted as the “reference”. . .	15

1.11	Sample of the transformations associated with tracking a calibrated object, with objectToObjectSensor being the calibration transform that when concatenated with objectSensorToTracker can directly provide the transform objectToTracker.	20
1.12	Sample of the transformations associated with tracking a stylus including the raw transform stylusSensorToTracker, the calibration StylusTipToStylusSensor, and the output transform stylusTipToTracker.	21
1.13	Sample of the transformations associated with tracking an object that was calibrated using a fiducial-based method, where objectToObjectSensor is the output calibration transformation which can be concatenated with objectSensorToTracker to produce objectToTracker.	22
1.14	Sample of the transformations associated with a template-based calibration. Depicting the transformation chain toolTipToTemplate, templateToTemplateSensor, templateSensorToToolSensor, which provides the output transform toolTipToToolSensor.	22
1.15	a) depicting the relationship between the calibrated US image and the probe sensors and b) provides an example of a calibrated US image interacting with a virtual models tracked needle, US probe and spine model	23
1.16	Various visualizations of prostate imaging are presented in this Figure. Image a) depicts an US slice from a 3D volume registered and overlaid on an MRI from the same patient, Image b) shows a prostate segmentation from the US overlaid onto the MRI image, figure c) shows a MPR visualization with three MRI slices intersecting the prostate segmentation, and figure d) shows a MPR visualization where a front-facing US image and two slices from the MRI are intersecting the segmentation.	38
1.17	The MR spectrum, where the green portion on the left represents AR, yellow represents VR on the right end and blue indicates MR. Below the MR spectrum is the spectrum of MR devices, with optical see-through head-mounted displays on the left, video pass head-mounted displays through in the middle, and VR head-mounted displays on the right.	39
2.1	The transformations required for two US calibration methods, where a) depicts the required transforms for a cross-wire calibration that requires phantom-calibration and b) illustrated the transforms required for the point-to-line calibration demonstrating the reduction in transforms that may propagate error.	51

2.2	This figure depicts how the point-to-line calibration approach mitigates issues with mid-plane alignment, and the FLE in the Z direction and the mid-plane are labelled on the diagram.	53
2.3	Both graphics depict 30 calibrated image planes where sub-figure a) shows the variation in calibrations obtained by 30 first-time users, and sub-figure b) depicts the variation in the 30 calibrations obtained by an expert user. Less variations between the calibrated image planes represents more precise calibrations.	54
2.4	Figure a) illustrates the configuration used for calibration, where the tracked US probe is clamped and submerged in water and the tracked pre-calibrated needle is inserted at an oblique angle into the US beam stabilized with a mechanical arm. Figure b) is a representative image of the needle reflections produced using this set-up.	55
2.5	This figure depicts the needle path used for data collection, with graphic a) illustrating the path from the top to bottom of the image with the initial needle orientation, graphic b) shows the needle at a similar angle to the needle graphic a) with a bottom to top outlined path, graphic c) has a similar path to the needle in graphic a) but the needle is inserted at the opposite angle, and graphic d) shows the needle path from the bottom to top of the image maintaining a similar angle to the needle in graphic c).	56
2.6	Architecture of the neural network model used to predict the coordinates of the centroid of an US needle reflection.	57
2.7	The cross-wire phantom used within the point reconstruction analysis with a) illustrating the CAD model of the phantom overlaid with the wires and point target, and b) depicting the real phantom with the cross-wire and tracking sensor.	61
2.8	Sample images from the scanning protocol, where figure a) shows the US image with the two wire reflections prior to the cross-wire junction, figure b) shows the cross-wire junction, and figure c) shows the two wire reflections subsequent to the cross-wire junction.	63
2.9	Four sample US images overlaid with the automatically localized centroid positions obtained from our CNN trained on set C.	65
2.10	Visual representation of the calibration precision by applying all transformations obtained during calibrations with both manual and automatic calibrations to a plane the size of the US image and overlaying them in space.	66

2.11	The calibration reproducibility across the 75 mean shifted reconstructed pixel locations obtained for both calibration approaches, where a larger spread between samples indicates less precision. The 95 % prediction interval bounds are provided for the X, Y, and Z axes as an ellipsoid.	67
2.12	Visual representation of the cross-wire accuracy analysis for one calibration per imaging depth. Each row of images arises from the same imaging depth. The first column (a, d, e) shows the phantom target overlaid on the US image, while the second (b, e, h) shows the calibrated US image in a common coordinate system, with the cross-wire phantom aligned with the target point, and the third column (c, f, i) shows the calibrated US image in a common coordinate system with the cross-wire phantom aligned with the two cross-wires.	68
2.13	The absolute error between the three ground-truth targets and US reconstructed positions of the cross-wire target plotted across the 15 calibrations obtained per imaging depth, with the 95 % prediction interval provided as an ellipsoid. . . .	69
3.1	Schematic of the experimental setup. The black and blue region indicates the optimal tracking volume of the optical/VIVE Pro's tracking systems, respectively. A grid of locations was programmed in CNC coordinate system and is located in the overlapping, optimal, tracking volume of both systems.	77
3.2	calibration device with optical DRF and hemispherical divots.	78
3.3	(a) Visual depiction of the point-based registration, (b) Visual depiction of the transformation chain used for the hybrid system.	79
3.4	(a) Dual tracked validation tool outfitted with a stylus, (b) Position accuracy validation set-up with arrows to highlight the data collection path.	80
3.5	Hemisphere phantom for rotation validation. Image a) depicts the 3D printed hemisphere phantom with cylindrical openings and Image b) shows the CT scan of the hemisphere phantom and the rotation vectors for each cylindrical opening.	81
3.6	Positional (a) and rotational (b) accuracy summarized in box plots	82
3.7	Examples of first-person Point-of-View MR visualization for surgery: (a) a VR system for spine needle intervention where streaming ultrasound and real-time needle advancements are visualized using HMD, b) VR representation of this procedure as viewed within the HMD, c) AV scene with streaming ultrasound and the edge mapping feature for the stereo camera video input, d) AV scene with streaming ultrasound and the translucent feature for the stereo camera video input	84

4.1	Comparison of guidance techniques for CVC a) landmark guidance b) US-only guidance and c) AV guidance.	88
4.2	a) Phantom development, b) Phantom under US, c) and d) health human neck vasculature under US. Images c) and d) courtesy of healthy volunteers.	90
4.3	a) The co-calibration apparatus tracked by the VIVE controller and magnetic pose sensor b) the calibration apparatus for the syringe, and c) Visual representation of an example of tracked tools registered in the HMD's coordinate system, where the alignment between the models of the spine and needle and their reflections in the US image indicated the total system accuracy	91
4.4	Visual representation of each mode of visualization where a) is the US-only system, b) is the monitor system, and c) is the HMD system. Images b) and c) comprise models of the US probe, needle, needle trajectory and the calibrated US image.	91
4.5	Distance metrics presented with respect to system where a) and c) summarize the distances from the final needle tip position to the closest point on the centre line of the vessel for experts and novices respectively and b) and d) summarize the distances from the final needle tip position to the closest point on the vessel wall for experts and novices respectively	93
5.1	The Mask R-CNN architecture depicting the CNN backbone, the Region Proposal Network, and the RoIAlign Layer. The "box head" is a series of Fully Connected layers that outputs the predicted class and bounding box for each object. The "mask head" outputs binary segmentation masks for each object instance.	104
5.2	Visual depiction of the reconstruction process. The left-most figure shows a calibrated US image positioned and scaled to the true field-of-view of the US beam. The second graphic demonstrates a segmented and calibrated US image, where the CA and IJV have been delineated. The third panel depicts a vascular skeleton where each image in the tracked scan has been segmented and spatially positioned using the calibration to form a skeleton. The final figure depicts the closed surface reconstructed vessels after the application of binary morphological hole filling and Gaussian blur smoothing.	106
5.3	Average Dice, Recall, and Precision for the CA from each of the four-folds. These results are reported separately for normal and patient data.	107
5.4	Average Dice, Recall, and Precision for the IJV from each of the four-folds. These results are reported separately for normal and patient data.	108

5.5	Four sample images with their respective outputs from the Mask R-CNN and U-Net with and without processing. Row a) shows an example of a small cluster of misclassified pixels from the U-Net. Row b) provides an example of a large group of pixels that have been misclassified as CA when they should be IJV in the U-Net output. Row c) depicts a vessel-like structure that has been misclassified as the IJV and the post-processing algorithm selecting this false segment as the IJV. Row d) depicts an image that has accurate outputs across all 3 algorithms. Despite the U-Net output from images a-c containing erroneous segmentation, the Mask R-CNN produced accurate segmentations across all sample images.	109
5.6	IJV surface-to-surface distances between the reconstructed US and the ground truth CT for all four-folds. The colour progresses to warm colours as distances increase.	110
5.7	CA surface-to-surface distances between the reconstructed US and the ground truth CT. The colour progresses to warm colours as distances increase.	110
5.8	Each letter (a, b, c, d) represents a unique human subject who was not used to train the algorithm used to produce the segmentations.	111
6.1	a) depicting the phantom mould, b) depicting an US scan of the phantom vasculature and b) depicting an US of human vasculature	117
6.2	a) and b) depict vascular reconstructions from human US, with a needle superimposed on the scene to provide an example of the system for human vasculature. c) and d) provide example vascular segmentations, where image d) also contains a markup that is representative of the needle tip pose from images a) and b).	118
6.3	Visual comparison between using a wireframe and a solid but translucent model to visualize the IJV. This figure compares visualizations with: the needle and trajectory in front of the IJV, the trajectory intercepting the IJV, the needle and trajectory behind the IJV, and the needle and trajectory intercepting the IJV. . .	120
6.4	Visual depiction of how to position the needle at the centre of the vessel, where a) and b) depict the needle trajectory at the periphery and c) depicting the trajectory centred about the vessel through optimizing the amount of needle trajectory that is rendered blue.	120
6.5	Histogram of the surface distance between the CT and US reconstructions of the phantom CA (a) and IJV (b)	121

6.6	Experimental setup for a) the 3D system and b) US-only guidance. Where A is the tracked syringe and its tracked calibration block, and B is the tracked and calibrated US probe. This figure is drawn from the point of view of the user. . .	122
A.1	User questionnaire used in Chapter 4, where each scale is 10 cm long and has three anchors with written descriptions at 0, 5 and 10 cm	152
A.2	User questionnaire used in Chapter 6, where each scale is 10 cm long and has three anchors with written descriptions at 0, 5 and 10 cm	153

List of Tables

1.1	Summary of technical specifications for current commercially available HMDs.	44
2.1	The number of pixels that comprise the X and Y axes of the US image for 4, 6 and 8 cm, as well as the mean and standard deviation for the scaling factors extracted from all calibrations produced for 4, 6 and 8 cm.	61
2.2	The absolute pixel distance between the manual and automatically localized needle centroids obtained from the four trained networks for the X- and Y-axis, respectively. This table reports the mean and standard deviation, for the absolute error across the test and validation sets for each fold and the average across the four-folds.	64
2.3	The absolute distance between manual and automatically localized centroid coordinates obtained during calibration. For each imaging depth, the mean, standard deviation, 95 % confidence interval, and the maximum and minimum absolute distances were calculated in millimetres across the 225 point pairs. . .	65
2.4	The standard deviations are reported for the 75 mean shifted reconstructed pixel locations obtained for each imaging depth and both calibration approaches. The p-value from the Z-test that was performed to compare variances between the two distributions produced using manual and automatic localizations are also included.	66
2.5	The mean and standard deviation for the absolute distance between the ground-truth cross-wire position and the US reconstructed position calculated using the 15 calibrations obtained per imaging depth across the 3 point targets. The absolute error for X, Y, and Z axes, as well as the normalized absolute distance, are summarized below.	69
3.1	The average, standard deviation, and 95 % CI calculated using the absolute distances between each collected sample and its corresponding CNC location. .	82

3.2	The average, standard deviation, and 95 % CI calculated using absolute angular differences between each collected sample and its corresponding ground truth angle extracted from the cylindrical phantom CT scan.	83
4.1	Post Hoc test results following repeated-measures ANOVA analysis for the distance from the final needle tip position to the closest point on the centre line of the vessel for both experts and novices.	94
4.2	Post Hoc test results following repeated-measures ANOVA analysis for the distance from the final needle tip position to the closest point on the vessel wall for both experts and novices	94
4.3	Summary of the success rates by system and the associated statistics	94
4.4	Summary of insertion time and number of users who punctured the CA under each mode of visualization.	95
4.5	User questionnaire results	95
5.1	Summary of number of images allocated to each data set used to train and evaluate the networks. The percentage of normal control images in the training, test and validation sets are denoted in brackets. The number of normal and patient images for each fold are the images that were used to evaluate the networks. . .	103
5.2	Summary of the Dice, Recall, and Precision averaged across the patient and control data from all four-folds for the raw U-Net, post-processed U-Net, and Mask R-CNN. These data are presented for the IJV and CA separately. . . .	108
5.3	Summary of the SA and volume ratio between the metrics produced from the US reconstructions from the four trained networks to the metrics extracted from the CT segmented vasculature.	111
6.1	Vessel targeting results for the US-only and 3D guidance systems across time, distance from the centre line (smaller is better), distance from vessel wall (larger is better), and number of successful insertions, all with respect to the US and CT reconstructions. Bold faced values are considered the better result. .	123
6.2	Summary of user questionnaire result (scored out of 10)	124

List of Appendices

Appendix A: Study Questionnaires	151
Appendix B: Ethics approval	154
Appendix C: Copyright Transfers and Reprint Permissions	158

List of Abbreviations

2D	Two-dimensional
3D	Three-dimensional
AC	Alternating current
A-mode	Amplitude mode
AOPA	Anisotropic orthogonal Procrustean analysis
AR	Augmented-reality
AV	Augmented-virtuality
B-mode	Brightness mode
CA	Carotid artery
CNC	Computer numerical control
CI	Confidence interval
CNN	Convolutional neural network
CT	Computed tomography
CVC	Central venous catheterization
DoF	Degrees-of-freedom
DC	Direct current
DRF	Dynamic reference frame
DFVW	Distance from vessel wall
FLE	Fiducial localization error
FRE	Fiducial registration error
FoV	Field-of-view

HMD	Head-mounted display
IJV	Internal jugular vein
IMU	Inertial measurement unit
IR	Infrared
JV	Jugular vein
LED	Light-emitting diode
MRI	Magnetic resonance imaging
MAE	Mean absolute error
MR	Mixed-reality
M-mode	Motion mode
MPR	multi-planar reformatting
OST-HMD	Optical see-through head-mounted display
OTS	Optical tracking system
PSF	Point spread function
POSE	Position and orientation
PVAc	Poly (vinyl alcohol) cryogel
ReLU	Rectified linear unit
RMSE	Root mean square error
RoI	Region of interest
RPN	Regional proposal network
SI	Successful insertion
SA	Surface area
SLAM	Simultaneous localization and mapping
SVD	Singular value decomposition
TRE	Target registration error
US	Ultrasound
VPT-HMD	Video pass-through head-mounted display

VR	Virtual-reality
VR-HMD	Virtual-reality head-mounted display

Chapter 1

Introduction

1.1 Clinical Procedure of Interest

Blood vessels play vital roles in the circulation of blood to and from the central venous system. Many interventional procedures routinely access the central venous system using central venous catheterization (CVC), also referred to as central line insertion, with over 6 million insertions performed annually between the United States and Europe [126, 90, 13]. Central lines may be required for hemodynamic monitoring, delivery of drugs, hemodialysis, aspiration of air emboli, or insertion of transcutaneous pacing leads [16, 99, 126, 106]. Catheters are inserted following a percutaneous needle puncture into a vein, which provides the path for the catheter (line) to enter the vein, such that it can be navigated to the central venous system [17]. The three insertion sites that are commonly used for central venous access are the internal jugular vein (IJV), the subclavian vein, and the femoral vein [6]. Performing the insertion at the IJV is preferred as it provides direct access to the superior Vena Cava, which allows for easier navigation to the central venous system [17, 6]. While the IJV is the most common insertion site [17, 6], it also results in high rates of arterial puncture due to the close proximity between the IJV and the carotid artery (CA) [17], as seen in Fig. 1.1. This thesis focuses on central line insertion at the IJV due to the high prevalence of this insertion and the risks involved.

1.1.1 Relevant Anatomy

Jugular veins (JV) and carotid arteries are major vessels required for the circulation of blood to and from the organs superior to the heart [66]. The CA supplies the face, neck, and brain with oxygenated blood [66], and the JV receives and returns the de-oxygenated blood from these areas to the heart to be oxygenated and redistributed [66]. Humans have a CA and JV on the left and right sides of the neck [66]. Both vessels bifurcate as the common trunk of each

blood vessel is divided into internal and external branches [66]. Typically, the JV is positioned 1.5 cm below the skin and is proximal to the CA in the anterior and lateral directions. However, the relationship between the CA and JV is extremely variable across humans [49, 148]. The internal branch is positioned medially compared to the external branch for both the CA and JV [66]. The relevant anatomy and direction of blood flow in the neck are depicted in Fig. 1.1.

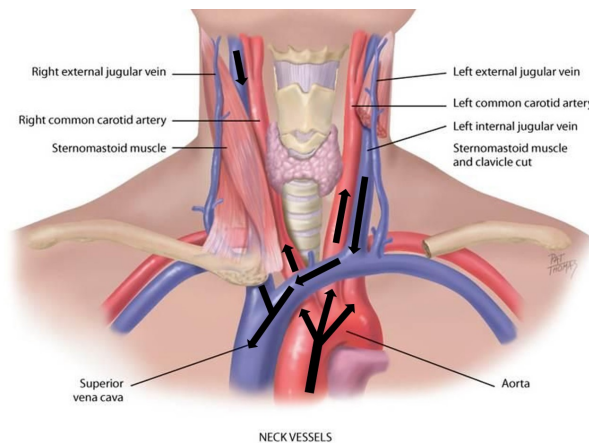


Figure 1.1: Vascular anatomy in the neck, where the direction of blood flow is indicated with black arrows

Current clinical approaches and complications

The Seldinger technique is the most commonly used method to cannulate the vein during the central line insertion procedure, regardless of the insertion site [6]. First, a needle is inserted into the vein percutaneously [6], and a guidewire is advanced through the needle into the vein. Following extraction of the needle, a catheter is inserted into the vein along the path formed by the guide wire [6]. The guidewire is then extracted and the catheter is navigated from the insertion vein to the Vena Cava providing access to the central venous system [6]. As the IJV is also deformable in humans the Trendelenburg position is commonly employed for IJV puncture during central venous catheterization [157]. Trendelenburg position requires the patient to be supine on the table with their head declined below their feet at an angle of roughly 16° , which allows the vessel to distend reducing the amount of deformation during scanning [157].

The central line needle insertion location was originally selected using the landmark guidance technique, where clinicians employed landmarks that could be identified visually or via palpation to determine the location of needle insertion [90]. To perform the insertion at the IJV, clinicians relied on the knowledge that the IJV is located at the apex of the triangle formed by the heads of the sternocleidomastoid muscle and the clavicle [90], as depicted in Fig. 1.2.

Complications that occur during central line insertions at the IJV are classified as either

infectious, thrombotic, or mechanical [17]. The complication of highest clinical importance is arterial puncture, especially at the site of the IJV where the CA is at risk of puncture and is classified as a mechanical complication [17]. CA punctures have been found to occur in 10.6 % of landmark-guided cases [69, 17] and can result in adverse outcomes such as hematoma, hemorrhage, pseudo-aneurysm, and stroke [17]. The high complication rates associated with this approach motivated a shift from landmark-based guidance to image-guidance. Ultrasound (US) imaging was introduced to the central line insertion procedure to improve insertion accuracy by providing visual anatomic information to the clinicians. Studies reported in the literature have demonstrated that US-guided central line insertions significantly reduced the number of mechanical complications, compared to the landmark-based approach [93, 11, 98].

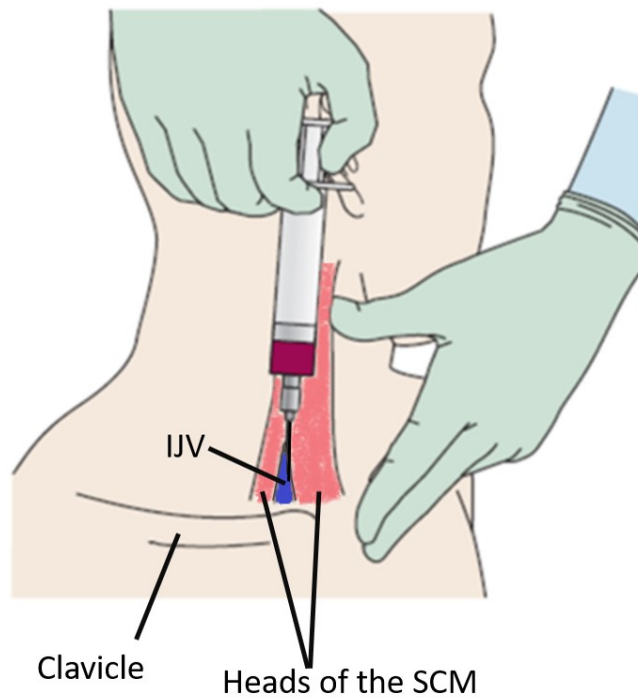


Figure 1.2: Visual depiction of the landmark-based central line insertion technique

1.2 Ultrasound for Interventional Guidance

1.2.1 Overview of Ultrasound principles

US images are formed by the transmission of sound waves between 2 MHz and 12 MHz through a medium [1]. The velocity or speed of sound of an US wave is related to the frequency and wavelength by equation 1.1 [1].

$$velocity_{sound}(m/s) = Frequency(Hz) * wavelength(m) \quad (1.1)$$

The wavelength varies depending on the density and elasticity of the material through which the sound wave is travelling, and thus the velocity is also material dependent [1]. To account for the variance in US velocity, US systems assume the velocity of sound to be 1540 m s^{-1} , which is an average associated with various soft tissues found in the human body [1]. The frequency of the sound wave interpreted by the US imaging system is either set by the operator or is inherent to the US probe [1].

US probes are designed to transmit and receive US signals, where the transmitted signal is referred to as a pulse and the received signal as an echo. US pulses are transmitted by the elements in the US probe and interact with the tissues through which they pass. As the US wave traverses through tissue, the intensity of the signal is attenuated as it can be reflected, refracted, or absorbed. At a tissue interface, the US wave may be reflected or refracted, where the amount of signal that is reflected depends on the difference in acoustic impedance between the two tissues [64]. The impedance (Z) of a material is determined by the density and speed of sound as per equation 1.2 [64]. Since the speed of sound of an US pulse is largely consistent amongst soft tissues, the difference in tissue impedance is proportional to the difference in density [64]. The reflected echoes return to the transducer and are used within the image formation process.

$$Z = density * velocity_{sound} \quad (1.2)$$

US waves may also be refracted when the tissue boundary is at an oblique angle to the direction of the wave [64]. Refracted waves are typically not received by the transducer [64]. US waves are also refracted if they traverse a tissue boundary between two media with different acoustic impedances. This large difference in tissue impedances produces a large change in the velocity of the wave [64]. The velocity change results in the redirection of the sound wave, where the angle of the refracted wave is quantified by Snell's law [64] as defined in equation 1.3. These refracted signals may be received by any element in the US probe or may miss the probe altogether, resulting in artifacts [1]. US waves are also scattered, which refers to their redirection as they interact with small, rough, or uneven structures [64]. This redirection results in the specular artifact typical of US images as the signals are being received by a different element from which the signal was produced or miss the transducer entirely [64]. US waves are attenuated as they traverse tissue as they are converted into heat and absorbed by the tissue [64]. Fig. 1.3 illustrates the redirection of the US wave through reflection, refraction, and scatter when it comes into contact with the material boundary.

$$\frac{\sin(\theta_2)}{\sin(\theta_1)} = \frac{V_2}{V_1} \quad (1.3)$$

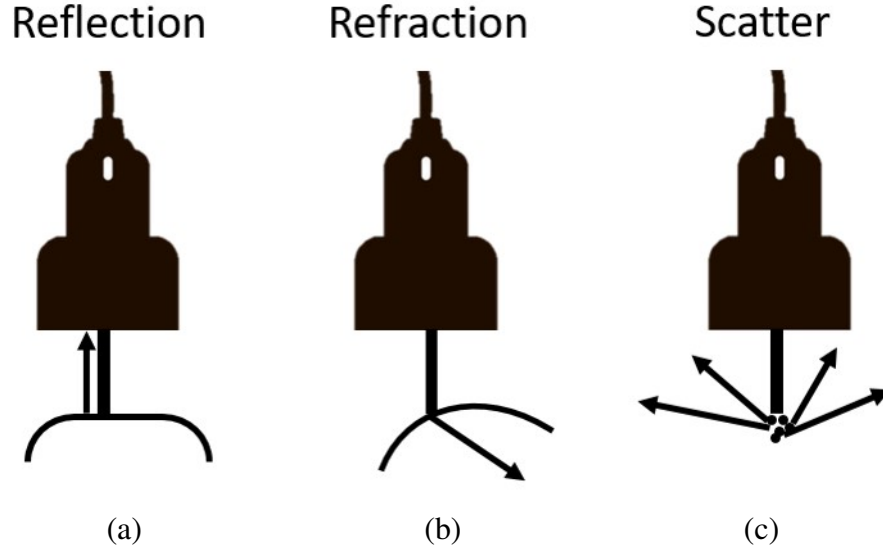


Figure 1.3: Visual depiction of the three ways an US wave may be redirected after making contact with a material boundary, illustrating a) a reflected wave, b) a refracted wave, and c) a scattered wave.

The sound waves received by the elements in the probe are converted to electrical signals where the amplitude of the voltage is proportional to the strength of the received echo [14]. Amplitude Mode (A-mode) displays a plot of the voltage amplitudes over time on the monitor [14], it is not commonly used for clinical interventions due to its lack of spatial information. Brightness Mode (B-mode) provides a 2D visualization of the US information as the received echoes are plotted on a 2D pixel grid, where signals with higher amplitudes result in brighter pixels in the image [14]. The echo is mapped to a pixel and the lateral position depends on the element receiving the echo and the axial position is determined by the time delay between the transmitted pulse and the received echo [14]. B-mode is commonly used for clinical applications, as it provides an intuitive 2D image of the anatomy scanned by the probe. Motion Mode (M-mode) is another way to use US signals, where a single B-mode scan line is selected and the system displays the change in echo intensity as a function of time [14]. M-mode is typically used to map the motion associated with the beating heart, such as the motion of a valve [14]. The M-mode graphs are typically used to supplement B-mode images. Doppler signals offer another method to use US information. Doppler physics relates changes in frequency to velocity, thus the change in frequency between the pulse and echo can be used to calculate the velocity of blood flow [14]. The velocity of a scan line can be plotted over time and visualized

as a graph. Alternatively, power Doppler or duplex US imaging expands on B-mode scanning through the inclusion of a visual representation of the blood flow overlaid onto a B-mode image [14]. If the blood is flowing away from the transducer it is depicted in blue, and if blood flows in the opposite direction, it is displayed in red [14].

The resolution of a B-mode US image differs in the axial, lateral, and elevation directions as the nature of the US image in each of these directions is determined by independent factors, these three resolutions are depicted in Fig. 1.4. The lateral resolution (X-axis) determines the ability to distinguish two objects perpendicular to the direction of the US beam [14]. The lateral resolution is determined by the beam width, if the two structures are smaller than the width of the beam they are visualized as one object [14]. The axial resolution (Y-axis) determines the ability to distinguish two objects that are aligned with the US beam [14]. The axial resolution is determined by the wavelength and pulse length of the US pulse, where the smaller the wavelength and pulse length the better the resolution [14]. The elevation resolution (Z-axis) is the ability to distinguish two structures that are aligned perpendicularly to the image plane and is determined by the elevation beam width [14]. The elevation and axial resolutions change throughout the beam and are typically best at the focal point [14].

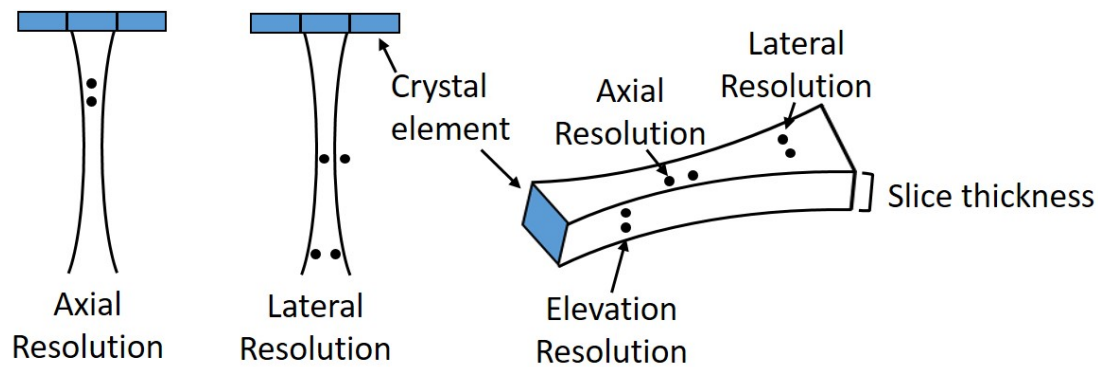


Figure 1.4: Diagram illustrating the axial, lateral, and elevation resolution. This diagram also depicts the 3D beam profile from a single crystal element.

B-mode US imaging can be used for a wide range of applications, and there are ways to optimize images for different applications. Many different types of US probes are used clinically with linear, curvilinear, and phased array transducers being the most common [14]. Linear arrays have the elements arranged in a straight line producing a rectangular image through parallel scan lines that are perpendicular to the transducer face [14]. Curvilinear transducers have the elements arranged in a convex curve to produce a fan-shaped image where the image grows progressively wider in the lateral direction as depth increases [14]. Phased arrays can produce various US waveforms by strategically timing the firing of the elements in the

array [14]. These delays cause purposeful interference between the US signals generated by the elements, thereby changing the US beam profile and the shape of the resulting image [14]. This allows for the focal point and scanning direction of the beam to be altered depending on the application and imaging depth.

1.2.2 Physical principles of US as they relate to interventional imaging advantages and limitations

An image artifact is something that can be seen in the image that is not present in reality. There are a few artifacts common to US images due to the way that US waves interact with tissues. Acoustic shadowing occurs when the US beam passes through an interface with a large difference in acoustic impedance between materials, such as an interface between soft tissue and air, bone, or metal [1]. Due to the difference in impedance, all or most of the US signal is reflected back to the transducer or absorbed at the interface [1]. Therefore, the area behind these structures appears black because almost no signal reaches this area, resulting in almost no reflections [1]. Acoustic enhancement occurs when the US wave passes through a fluid (as opposed to the transmission through an attenuating medium) because the US signal has almost no attenuation under these conditions [1]. The structures distal to the fluid appear hyper-echoic as the signal is stronger than expected for the depth due to the lack of attenuation of the wave when travelling through the liquid [1]. Such artifacts can be reduced by using the time gain compensation control (see below) that enables signals reflected from various depths to be visualized with equal brightness [1]. In general, these artifacts are a limitation to US, as issues such as shadowing or enhancement may affect the ability to produce accurate diagnoses or clinically relevant metrics.

The general attenuation of US waves as they travel through a medium also affects the appearance of the output image. The amount of wave attenuation depends on depth and frequency, as quantified by equation 1.4 where α is the attenuation coefficient, d is the depth, and f is the frequency.

$$Attenuation = \alpha \left[\frac{dB}{MHz * cm} \right] * d[cm] * f[MHz] \quad (1.4)$$

The Sonographer has control over the frequency of the signal and the imaging depth. Higher frequency signals produce images with better axial resolution but will cause more attenuation [1]. Thus, there is a trade-off between frequency and imaging depth and it is recommended to use the highest frequency that is capable of penetrating the required depth, while still producing acceptable image quality. Since US signals attenuate as they interact with tissues, signals received from deep structures have inherently lower amplitudes compared to the

echoes that return from closer objects. Thus, a direct mapping of amplitude would result in pixels with increasingly lower grey levels as depth increases regardless of the tissues that are being imaged [1]. To account for this issue, the time gain compensation control is implemented on the system, where the operator can adjust the amplitude of returning echoes at set depths [1]. The time gain compensation can be used to compensate for the attenuated signals and allow for deep structures to be more visible [1]. Time-gain compensation is also used to minimize the effects of artifacts such as acoustic enhancement and shadowing [1]. The ability for the operator to alter the appearance of the US image, while imaging the structure of interest, can be seen as an advantage as well as a disadvantage. While the sonographer can use the controls to alter the image appearance based on the application, this may result in them developing their own style and habits, which could lead to differences in the appearance of images collected by different operators. Most US scanners have preset configurations that intend to optimize the US parameters for different applications.

1.2.3 Ultrasound guidance

US can be used for both diagnosis and real-time image guidance. Due to its portable, real-time characteristics, and lack of ionizing radiation, US is commonly used for interventional image-guidance, including needle navigation [110]. US is an ideal imaging modality for surgical navigation as the operator can relate the position of the US probe to the visual information of the anatomic target and surgical tool displayed within the US image that is being updated in real-time [110]. Some applications that use US needle guidance include biopsy, regional anesthesia (including epidural or other injections), drainage, aspiration, and venous catheterization [110]. As venous catheterization at the IJV is the focus of this thesis, I describe US-guided CVC in more detail. US-guided central line insertions with percutaneous access at the IJV employ two common US image orientations. The out-of-plane approach provides a cross-sectional view of the CA and IJV [157, 135], as depicted in 1.5 a) and b). In this approach, the needle is inserted out-of-plane or perpendicular to the image plane, such that a bright point echo formed by the needle is seen within the IJV in the US image [135], as shown in Fig 1.5 a). Alternatively, the in-plane method can be used, providing a view of the IJV along its long axis (Fig. 1.5 d)). In the in-plane approach, the needle is inserted at an oblique angle parallel to the image plane, allowing the entirety of the needle to be visualized within the image [157, 135], (Fig. 1.5 d)). Regardless of the approach, once the clinician believes the needle to be within the vessel lumen, they attempt to aspirate blood into the syringe to verify the needle placement [135]. While US-guidance has improved the accuracy of the needle insertion required for central venous catheterization, clinical studies have found rates of CA puncture to be approximately 7 %

for US-guided cases [93].

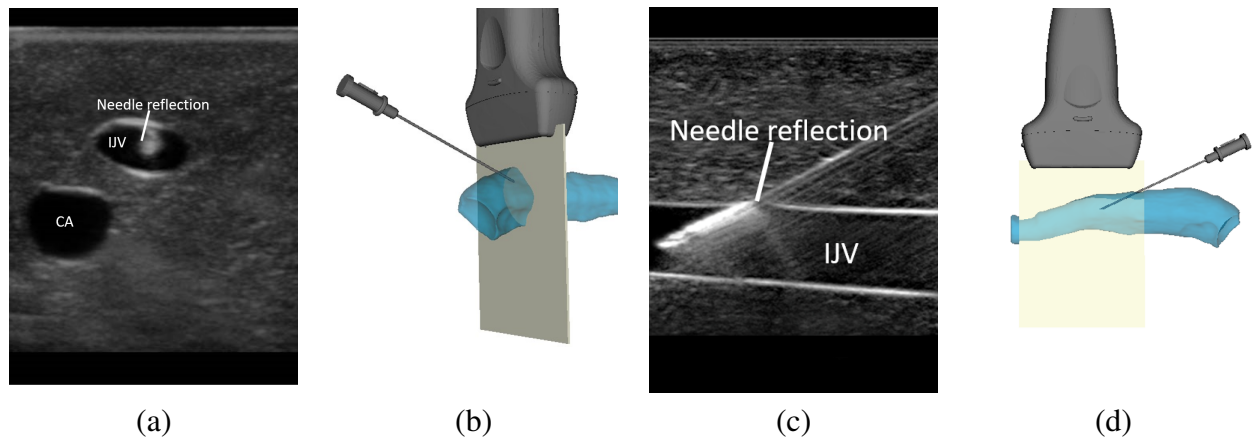


Figure 1.5: Two representative visualizations of US-guided central line insertions. Figure a) depicts the US visualization provided for out-of-plane needle insertions and figure b) depicts an example of the spatial configuration of the needle, the US probe and image, and the vessel during the out-of-plane insertion. Figure c) provides an US visualization for an in-plane needle insertion, and figure d) provides the spatial configuration of the needle, US probe and image, and vessel.

Both the in-plane and out-of-plane approaches have their own benefits and disadvantages. One factor that is believed to influence the targeting accuracy of out-of-plane US-guided needle insertions, is the inability to distinguish the needle tip from any other point along the needle shaft [9, 157]. For out-of-plane needle insertions, the probe scans the cross-section of the needle and the resultant reflection is a bright hypo-echoic point in the US image. This approach may result in users losing line-of-sight with the needle tip, as the reflections produced when imaging the needle tip and shaft are very similar [135]. The visualization of this needle makes it difficult for clinicians to properly identify whether the needle tip has inadvertently punctured the posterior wall of the vessel [9, 157]. The inability to visualize the needle tip precisely can result in complications such as inadvertent arterial and posterior wall punctures [9, 157]. The in-plane approach provides the ability to see the entire shaft of the needle including the tip [9, 157]. However, it can be difficult to keep the needle within the thin image plane, as even small movements can result in losing sight of the needle within the US image [9, 157]. Another limitation of this method is that it provides no information about the spatial relationship between the CA and IJV [9, 157]. Both of these factors can increase the number of inadvertent punctures that occur during the out-of-plane approach [9, 157]. The limitations of these two methods can be summarized as the inability to simultaneously visualize all relevant anatomy and the entirety of the needle including the needle tip due to the use of 2D US-guidance for needle navigation. The current state-of-the-art US-based navigation systems typically employs

external spatial tracking systems, US registration or fusion, 3D US or anatomic reconstruction, advanced visualizations [30].

1.3 Tracking Technology

1.3.1 Spatial Tracking

External spatial tracking technologies provide the pose (position and orientation) of associated sensors with respect to the tracker's coordinate system [42]. For surgical applications, these sensors are typically rigidly integrated into or fixed onto surgical tools or the patient [30, 143]. In general, spatial tracking systems comprise one or more sources that generate the signal, one or more sensors that receive the signal(s), and a data acquisition and signal processing unit [143]. The two main tracking systems used for surgical navigation are optical and magnetic tracking, each having advantages and disadvantages with respect to the other [143].

The most commonly used optical tracking categories include videometric or visible light-based systems and infrared (IR) light-based system [143]. In both methods, one or more camera(s) act as the sensor that receives the signal. IR-based optical tracking markers can be active or passive. Passive IR-based systems contain light-emitting diodes (LEDs) that flood the field-of-view (FoV) with IR light, which in turn is reflected by the passive reflective markers, to be detected by the IR camera(s) [143]. Active IR-based systems use sensors containing embedded IR LEDs and directly emit IR light that is detected by the IR camera(s) [143]. In videometric systems, the sensor contains one or more calibrated video cameras that are used to localize checkered-type markers, with distinct patterns similar to those found on a QR code, that act as the signal source [143]. Videometric systems are inherently passive, as the checkered patterned markers act as a "source" that provides the signal by reflecting visible light to the camera, rather than generating a signal of a specific type. The video camera acts as the sensor that receives the visible-light signals reflected by the markers [143]. To localize optical tracking markers with six degrees-of-freedom (DoF) both types of systems require that at least three "source" markers that generate or reflect the desired light waves, be arranged in a rigid and known non-collinear geometric pattern [143]. These distinct groupings of markers are called dynamic reference frames (DRF) [143]. Multiple DRFs can be simultaneously localized by the camera(s) if they have mutually unique patterns [143]. Mathematical triangulation is used to estimate the pose of the DRFs in 3D.

Stereo reconstruction or stereo vision is the process of using two cameras with known configurations that simultaneously capture the same scene from different vantage points, to estimate the 3D pose of an object [143]. Stereo reconstruction is commonly implemented in optical

tracking systems, as optical markers can be localized with superior accuracy to mono-camera triangulation [143]. Stereo vision mimics the process whereby humans and other animals simultaneously use the visual information from both eyes to interpret the position of objects in 3D. The stereo cameras capture left and right images of the same scene at the same time [143]. As the viewpoint of each camera is different, the two images contain depth-dependent disparities [143]. Stereo vision relies on the known intrinsics of the camera, including the focal length, the known horizontal distance between the cameras, and the pose of the camera [143]. In addition, stereo vision also relies on the known pixel correspondences between left and right images and the knowledge that the light rays that generate the pixel positions must pass through the camera's focus [143]. The stereo vision problem is commonly used to reconstruct a 3D scene given the left and right images. This problem is simplified for the case of estimating the 3D pose of an optical marker, as optical tracking systems have image-processing methods to localize the pixel coordinates that represent the tracking markers [143]. It is for this reason that IR-based systems have superior accuracy to visible light-based systems, as the IR markers appear as bright reflections on a black background, making localization trivial [143]. These known parameters allow for simple trigonometry to be used to solve for the 3D X, Y, and Z coordinate of the optical marker [143]. As the 3D pose of the optical sensor is derived using 2D images with known X and Y pixel locations, the tracking will be the least accurate in the Z-direction and thus the accuracy is not isotropic. A diagram of the collinear camera stereo reconstruction problem can be seen in Fig. 1.6.

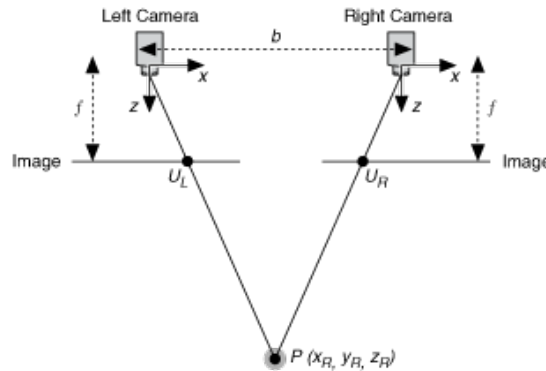


Figure 1.6: Collinear camera stereo reconstruction (triangulation) diagram illustrating how a point can be reconstructed in 3D using two cameras. The diagram shows the Left and Right cameras, the distance between them (b), the focal length (f), the pixel coordinate of the object on the image U_L and U_R , and the reconstructed point $P(X_R, Y_R, Z_R)$

The specific optical tracking system that is used throughout this work is the Polaris Spectra with passive spherical markers (NDI, Canada). This tracking system has a pyramid-shaped FoV where the tracking has a 0.3 mm root mean square error (RMSE) in the optimal tracking

area and 0.3 mm RMSE in the non-optimal zone [103]. The tracking FoV is depicted in Fig. 1.7 [103].

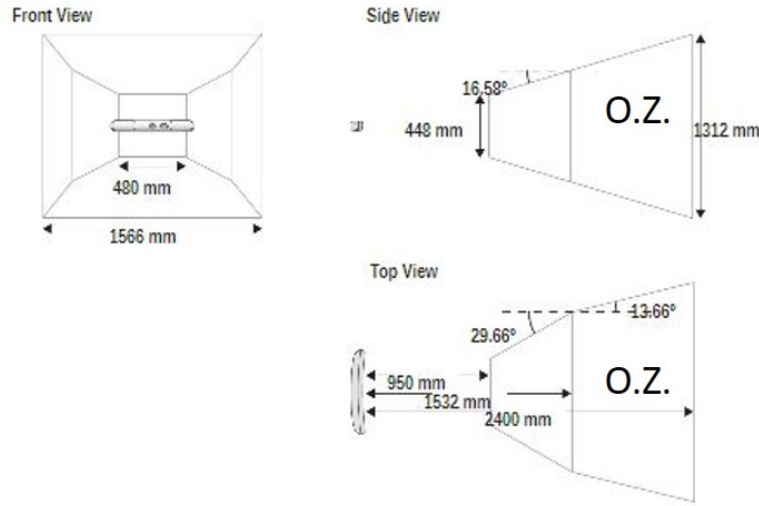


Figure 1.7: Polaris tracker FoV, showing the front, side and top view with all dimensions labelled. The diagram has the optimal zone labelled on the side and top view, denoted as O.Z., which extends from the vertical line in the diagram.

Magnetic (electromagnetic) tracking systems comprise a magnetic field generator, magnetic tracking sensors, and a control unit. The field generator contains solenoids that generate a pulsed magnetic field and magnetic sensors that interact with the changing magnetic field to generate current pulses [143]. The solenoids are arranged in a specific configuration, whose geometry and nature of the excitation electric currents determine the properties of the magnetic field that defines the tracking coordinate system [143]. The control unit aids in the generation of the magnetic field, the data acquisition, and processing required to localize magnetic sensors [143]. There are three common types of magnetic tracking systems including alternating current (AC) driven systems, quasi-static direct current (DC) driven systems, or passive systems that use permanent magnets or transponders that emit a radio-frequency signal [143]. Passive systems are the only type of magnetic tracker that can perform wireless tracking, but due to the reduction of accuracy in passive systems, the majority of the magnetic tracking systems used for medical applications are AC- or DC-driven systems. Magnetic tracking sensors contain small coil(s) [143]. When the sensor intercepts the magnetic field produced by the field generator, there are local changes in the field strength that induces current flow in the coils windings [143]. The magnitude of the currents is related to the position of the sensor within the range of influence of the field generator [143]. The two commonly used sensing coils are search coils and fluxgate sensors. Search coils measure the magnetic flux over time using inductors and require an AC-driven magnetic field [42]. Fluxgate coils contain two inversely arranged

inductors that measure the second harmonic Fourier component of the magnetic field and thus are used to measure static or low-frequency alternating fields such as those provided by DC-driven systems [42]. AC-driven systems are susceptible to local distortions caused by eddy currents [143] and pulse-driven DC systems are highly affected by the presence of metallic and magnetic materials [143].

The magnetic tracking system used throughout this work is the NDI Aurora Desktop Magnetic Tracker (NDI, Canada) that is capable of tracking 5 and 6 degree-of-freedom (DoF) sensors (NDI, Canada). The Aurora Desktop Tracker is an AC driven system with a dome volume a position and rotational accuracy of 0.7 mm (RMSE) and 0.3° (RMSE), for 6 DoF sensors [102].

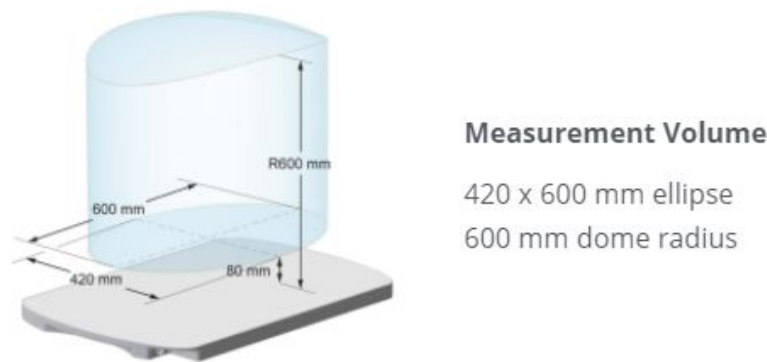


Figure 1.8: Aurora tabletop tracker and FoV labelled with measurements.

Both optical and magnetic tracking have advantages and disadvantages, thus the system selection is dependent on the required application. Optical tracking offers a larger FoV [120] and superior tracking accuracy compared to magnetic tracking [41, 159]. However, optical tracking requires a clear line of sight between the camera and sensors, which can be difficult to ensure in crowded surgical environments [41, 159]. Magnetic trackers do not require line of sight and the tracking sensors are significantly smaller than optical sensors allowing for the integration of sensors into surgical tools. One of the major limitations of the magnetic tracking system is that the accuracy may become compromised when a ferromagnetic material is present [42, 89]. Active magnetic tracking also requires a wired connection between the tool and the control unit, whereas optical tracking is wire-less [41, 159].

The NDI optical and magnetic tracking systems represent the pose of a tracking sensor as a 4×4 homogeneous transform provided in Eq. 1.5. The DoF associated with tracking sensors describes the number of axes in which the system can capture the movement of the sensor in 3D space. Most tracking sensors can be tracked with 5 DoF (translation in X, Y, and Z, and rotation about the Y-axis (pitch) and Y-axis (yaw)) or 6 DoF (translation in X, Y, and

Z, and rotation about the X-axis (roll), pitch, and yaw), as illustrated in Fig. 1.9. The values of each element in the inner rotation matrix (R_{ab}) within the 4×4 matrix in Eq. 1.5 depend on the DoF of the tracking sensor as transforms produced by 5 DoF sensors exclude the roll. The homologous transform in Eq. 1.5 can be applied to a 3D coordinate in space ($[X,Y,Z,1]$) through matrix multiplication to produce a transformed 3D coordinate ($[X',Y',Z',1]$).

$$\begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1.5)$$



Figure 1.9: Graphic demonstrating the 5 and 6 DoF tracking of a stylus, courtesy of NDI <https://www.ndigital.com/technology/6dof-explained/>

The pose of a tracking sensor can be provided with respect to the origin of the coordinate system defined by the tracker (SensorToTracker), or with respect to another pose sensor tracked by the same system, which is denoted as a reference sensor (SensorToReference). Reference sensors may be fixed to a patient to provide the pose of the tools with respect to a patient coordinate system. A visual example of tracking a sensor with respect to the tracker's coordinate system and a reference sensor are provided in Fig. 1.10 a) and b), respectively. More information on the use of virtual models to visualize the motion of a tracked tool is provided in Section. 1.4.2

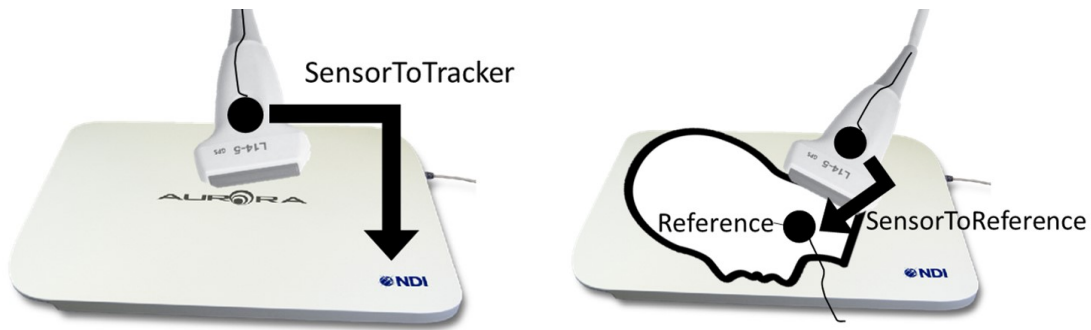


Figure 1.10: Two instances of tool tracking, with graphic a) illustrating tracking a sensor with respect to the origin of the tracking system, and graphic b) illustrating tracking a sensor with respect to another sensor denoted as the “reference”.

1.4 Ultrasound Registration and Calibration

Registration algorithms are used to find the geometric transform that maps one coordinate system to another to align them in space [19]. US registration is required for US calibration and registration between US and other medical images. US calibration registers the US image to the coordinate system defined by the external spatial tracking system being used, through a tracking sensor fixed to the US probe [63]. US-image registration aims to solve for the transformation matrix that best aligns the two images or volumes in space, where the coordinate system is defined by the pixel grid [19]. Real-time US images may be registered with medical images from other modalities such as MRI or CT scans [19]. Alternatively, US images may be registered with other US images to detect changes over time, relate 2D to 3D US images, or compare images across patients [19].

Image registration pipelines require five main components, i) the input to the pipeline, ii) an initial estimate of the transform, iii) a registration or similarity metric, iv) an interpolation algorithm, and v) an optimization algorithm. A registration or similarity metric provides a way to judge the accuracy of the registration [19]. The optimizer algorithm employs the similarity metric to comprehensively search for the transformation that maximizes or minimizes this parameter [19]. Optimization algorithms for registration are typically iterative processes. The input to a registration algorithm is either images, volumes, or some corresponding information extracted with respect to both coordinate systems [19]. Registration pipelines may require a pre-processing step to extract useful information, which can aid in the selection of the similarity metric and thus the optimization algorithm [19]. One coordinate system is designated as the fixed coordinate system, while the other coordinate system is designated as the moving coordinate system implying that the transformation matrix derived by the optimization algorithm is applied to the moving coordinate system to register it to the fixed coordinate system [19].

This is more commonly referred to as the moving image, as in the case of most US registration applications, the 2D US image is typically designated as the coordinate system that is adjusted to align with a second coordinate system, whether it is the external tracker's coordinate system, or a 2D or 3D medical image [19]. For image registration, after the registration transform is applied to compute the coordinates of the output (transformed) image, the pixel intensities of the output image need to be resampled from the original image [19]. To avoid artifacts, this is typically performed with inverse mapping by applying the inverse transformation to map the output image pixel coordinate back to the original image space to compute the output pixel intensity [19]. However, the pixel positions determined during the inverse mapping may not lie within the bounds of the original image or may not be integers indices (as pixels are discrete), thus the pixel intensity can not be directly resampled. In these cases, interpolation is required to estimate the intensity value of the transformed pixel [19]. The pixel intensity can be interpolated using a nearest-neighbour approach where the output pixel intensity is the same as the closest neighbouring pixel to the coordinate produced during inverse mapping [19]. Alternatively, the intensities of the pixels surrounding the inverse transformed position can be averaged using bilinear or cubic interpolation to produce the transformed pixel intensity [19]. The interpolation step is a consistent error source in all registration problems as the new pixel intensities are estimates. Once the moving image has been interpolated, the similarity metric is calculated to measure the accuracy of the derived transformation matrix [19]. Depending on the resultant similarity metric, the optimization algorithm adjusts the transformation matrix and the process is repeated until the similarity metric has been optimized [19]. The algorithm typically ends when the change in the similarity metric is within a certain tolerance, or it has run for a set number of iterations [19]. Registrations can be performed in both 2D in the case of image-to-image registration, or 3D which is required for registration where at least one of the two coordinate systems is in 3D, such as an image volume or the tracker's coordinate system. In the case where a 2D image is registered to a 3D coordinate system, the image is represented as a 3D volume with a single pixel-width plane in the Z-dimension.

There are different types of geometric transformations, broadly categorized as linear (matrix) and elastic that can be solved to align two coordinate systems in space [19]. Linear transformations uniformly transform the entire image (or coordinate system) such that all pixels (coordinates) are manipulated in the same manner. Elastic transforms may apply both global and local changes to the moving image and there is no consistency between how the pixels are manipulated. This section focuses on linear transforms as they are used extensively throughout this thesis. Linear transformations can be sub-categorized as rigid, similarity, and affine. These transformations are grouped together as they are all typically modelled using matrices and share common characteristics. Linear transforms between 2D coordinate systems require

a 3×3 transformation matrix [19]. For linear registrations where one or more of the coordinate systems is in 3D, a 4×4 transformation matrix is required [19]. These transforms can be grouped based on how they alter the moving image and are often described in terms of degrees of Freedom (DoF), which is the number of independent ways that the transformation can alter the coordinate system to which it is applied. Rigid transforms preserve the relative distances and angles of the original moving image coordinate system by only altering the translation and rotation parameters [19]. For 2D registration, rigid transforms have 3 DoF, which includes X and Y translation and rotation about the Z-axis [19]. In 3D rigid transforms contain 6 DoF - rotation about the X, Y, and Z axes and translation in the X, Y, and Z directions [19]. Rigid registrations are commonly used to register two images collected from the same system at different time points, as they are typically of the same size. Similarity transforms can change the rotation, translation, and scaling, but the scaling factor is consistent in all directions, maintaining the original aspect ratio of the coordinate system but increasing or decreasing it by some factor [19]. In 2D, similarity transforms have 4 DoF which includes rotation about the Z-axis, X and Y translation, and scaling. In 3D, similarity transforms have 7 DoF including translation in the X, Y and, Z directions, rotation about the X, Y, and Z axes, and scaling. Similarity transforms are common for the registration of images with isotropic pixel size. Affine transforms alter the translation, rotation, shear and scaling while preserving lines and parallelism [19]. Unlike similarity transforms, the scaling values can be unique for each dimension for affine transforms. 2D affine transforms have 6 DoF, - rotation about the Z-axis, X and Y translation, X and Y scaling, and shear about the Z-axis. 3D affine transforms have 12 DoF including rotation about the X, Y and Z axes, X, Y and Z translation, X, Y and Z scaling, and shear about the X, Y, and Z axes. Affine registrations are commonly used for registration between different modalities and US Probe calibration.

1.4.1 Registration approaches

Numerous registration approaches have been developed for various image registration applications. Broadly they can be classified as intensity- or feature-based methods. Intensity-based methods implement a similarity function that measures the correlation between the pixel intensities within the two images, and the higher the correlation the more accurate the registration. Intensity-based similarity metrics are used to quantify how similar two overlapping images are based on the values of the pixels in the image, and include Mutual Information, Normalized Cross-Correlation, and Correlation Coefficient [19]. Gradient descent or ascent is a well-used optimization algorithm for intensity-based image registration [19]. A gradient measures how much the similarity function changes based on small changes to the transformation matrix [19].

The algorithm aims to change the transformation matrix such that the resultant gradient is smaller than in the previous iteration. This process continues until the applied transformation results in a local minimum of the gradient where the similarity metric is minimized [19]. Gradient ascent algorithms output the transformation matrix that results in a local maximum and a maximized similarity function [19]. Intensity-based methods are commonly used for multi-modal registration and to register images from the same patient over time, but are not employed in the work described in this thesis. Thus, this section focuses on feature- and landmark-based registration methods.

Landmark-based registration requires the identification of a set of homologous points in both coordinate systems, which are used to solve for the transformation matrix that best aligns the landmarks and thus the coordinate systems [111]. The landmarks used for registration are commonly referred to as fiducials and can be automatically or manually identified, and can be anatomical structures, features common in both images, or physical markers. Pair-wise landmark-based registration requires corresponding pairs of coordinates in 2D or 3D, where both point clouds comprise the same number of points [111]. A least-squares solution can be applied to pair-wise landmark registration to solve for the transform that minimizes the distance error between the fixed point set (Y_i) and transformed moving point set (X_i) [5], as per equation 1.6, where R is the rotation matrix, t is the translation vector, and N_p is the number of points used in the registration.

$$E = \sum_{i=1}^{N_p} \|RX_i + t - Y_i\|^2 \quad (1.6)$$

In the most trivial case, the least-squares approach [5] solves for the rotation and translation to transform each point p_i in the moving set to p_i' as per the equation $p_i' = Rp_i + t$. To solve for the translation values, the centre of mass (centroid) of both point clouds are calculated [5]. The translation value is obtained by calculating the difference between the centroid values between the two point clouds across the X, Y and Z (for 3D) axes independently [5]. Both point clouds are shifted by their respective centroid value such that the centroid is at the origin (0,0,0), allowing for the rotation to be solved independently from the translation [5]. To solve the rotation the point clouds are represented by a $3 \times N_p$ matrix, where N_p is the number of points [5]. A 3×3 cross-covariance matrix (W) can be calculated by performing matrix multiplication between the $3 \times n$ matrix of points from one cloud and the transposed matrix of points ($n \times 3$ matrix) from the other cloud [5]. Performing singular value decomposition (SVD) on this 3×3 cross-covariance matrix allows W to be decomposed into $W = UDV^T$, where U and V are 3×3 basis vectors [5]. The 3×3 rotation matrix that aligns the two point clouds in space is solved for using $R = UV^T$ (T is the transpose) [5]. Alternatively, applying Eigen de-

composition to the cross-covariance matrix can be used to derive the rotation matrix, where the eigenvector with the smallest eigenvalue provides the rotation in quaternion form, which can be converted to a 4x4 matrix [62]. Fiducial-based pair-wise registrations are used throughout this thesis to register phantoms to tracker or image space.

Iterative closest point (ICP) registration is required if the correspondences between points are not known, or if one point cloud contains more points than the other. In the case where there is an unequal number of points, the smaller point cloud should be registered to the larger one. Due to the lack of known correspondence between points, an additional step is added into the method to find the corresponding points [8]. This is typically done by calculating the Euclidean distance between each point and the closest point from the other point cloud [8]. Once the correspondence has been estimated, the same process outlined above is implemented such that the translation and rotation values can be obtained. An error metric $E(R,t)$ is then computed, typically by calculating the Euclidean distance between points, and the algorithm continues to iterate until the error is less than some specified threshold or the algorithm has run for a set number of iterations [8].

In this thesis, a point-to-line registration method that solves for the translation, rotation, and anisotropic scaling is used extensively to solve for US calibration [35]. This problem is formatted as an Anisotropic Orthogonal Procrustean Analysis (AOPA) and is implemented as an iterative approach to solve for the relationship between corresponding point and line pairs [35]. This method is similar to fiducial-based and ICP registrations with isotropic scaling but allows for the scaling to be unique across the X, Y, and Z axes [35]. Additionally, as the input data are point and line pairs, there is an additional step where the shortest distance between each point and its corresponding line is computed allowing for a coordinate on each line to be extracted to reformat the problem as a point-to-point registration [35]. The use of this procedure in the context of US calibration is outlined below.

1.4.2 Tool Calibration and Visualization

To virtually represent a tracked instrument, the object must be tracked and a transformation matrix providing the relationship between the coordinate system of the tracked element, and that of the instrument must be determined via a calibration process [25]. The output of the calibration is a homogeneous transformation matrix that registers the coordinate system of the object and the coordinate system defined by the spatial tracker fixed to the object [25]. This process allows for: i) the position of landmarks on the object to be known with respect to the pose sensor, ii) the motion of the virtual object to replicate the motion of the true object, and iii) the interactions between virtual models of tracked objects to replicate real interactions. Tool

calibrations are typically used within a transformation chain that concatenates all the transformation matrices within that chain. An example is provided in Fig. 1.11 depicting the calibration transform (ObjectToObjectSensor) which provides the relationship between the object geometry and the object sensor that would be concatenated with the tracking of the object sensor (ObjectSensorToTracker) to produce the output transform ObjectToTracker.

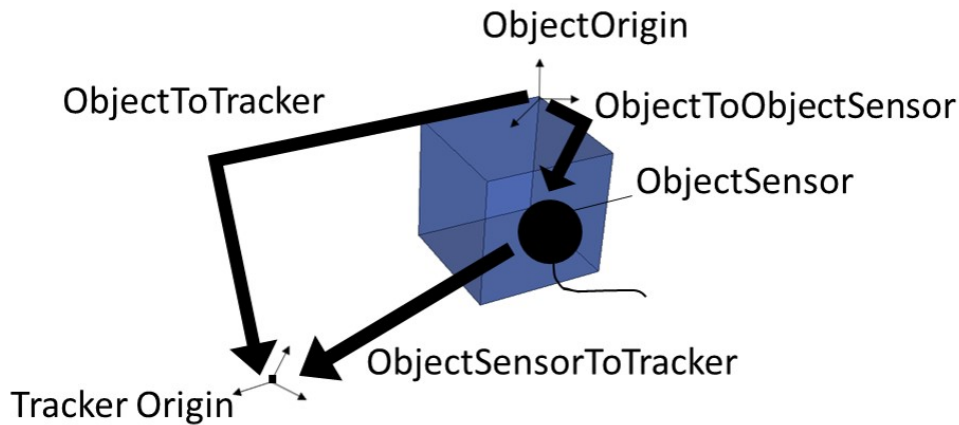


Figure 1.11: Sample of the transformations associated with tracking a calibrated object, with `objectToObjectSensor` being the calibration transform that when concatenated with `objectSensorToTracker` can directly provide the transform `objectToTracker`.

Pivot-based calibrations are commonly used to calibrate a ball-tipped stylus, where the relevant information is the position of the stylus tip [25]. Pivot calibration is the process of pivoting the tool around its tip, such that the motion of the tracking sensor forms the surface of a sphere about the stylus tip [25]. Sphere fitting can be used to solve for the relationship between the collected data points and the centre of the sphere using a least-squares approach [163]. Pivot calibration can only determine the tip location but not the mechanical axis of the stylus [25]. If the anatomical axis of the stylus does not align with the mechanical axis of the sensor spin calibration is required to provide the correct axis of rotation [152]. The output of this process is the transform that relates the stylus tip to the stylus sensor (`StylusTipToStylusSensor`). Thus, the pose of the stylus tip can be tracked with respect to the tracking system through concatenating the transforms `StylusTipToStylusSensor` and `StylusSensorToTracker`, which provides the transform `StylusTipToTracker`. A visual example of these transformations and their use for tracking a calibrated stylus is provided in Fig. 1.12.

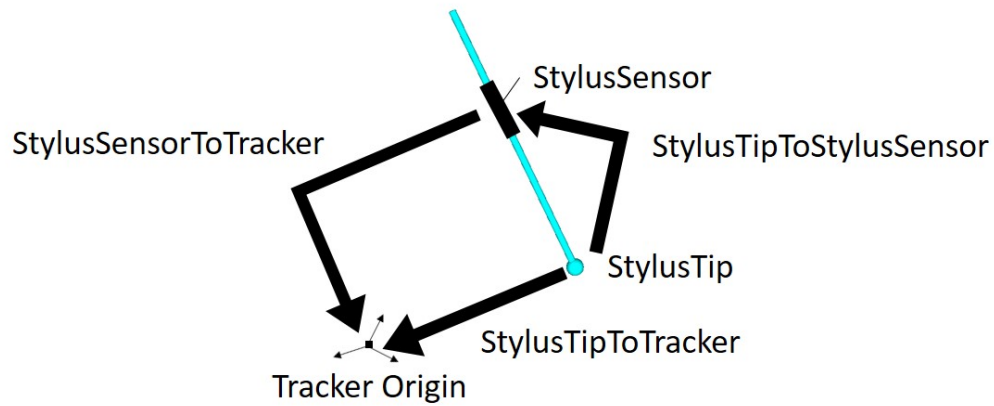


Figure 1.12: Sample of the transformations associated with tracking a stylus including the raw transform $\text{stylusSensorToTracker}$, the calibration $\text{StylusTipToStylusSensor}$, and the output transform $\text{stylusTipToTracker}$.

Fiducial-based calibrations require an object that is equipped with a tracking sensor and has calibration fiducials that are fixed to or engraved into the object [25]. The positions of these fiducials are known with respect to the geometry of the object, either through the CAD model or a CT scan of the object. The model (or volume) of the object has its own coordinate system such that its geometry, and that of the fiducials, are known with respect to the origin of the model. The position of the fiducials are extracted from the model or volume of the object by segmentation or by design and are obtained with respect to the origin of the model [25]. The positions of the fiducials are also digitized, such that their positions are collected with respect to the tracking sensor fixed to the object using a tracked and calibrated ball-tip stylus [25]. A paired fiducial-registration approach can be applied to solve the relationship between the geometry of the tool and the position of the tracking sensor fixed to the object ($\text{objectToObjectSensor}$) [25]. An example of the transformations used to position the model or volume of the object with respect to the optical tracking sensor is provided in Fig. 1.13, depicting the transforms $\text{objectToObjectSensor}$ and $\text{objectSensorToTracker}$ which are concatenated to produce objectToTracker . Thus, if a tracked model of a calibrated stylus ($\text{stylusTipToTracker}$) and a tracked model of the calibrated object (objectToTracker) are both visualized, their interactions would replicate the interactions between the real objects. This provides a method to visually validate a calibration, as the interactions and relative position between the two virtual objects should mimic the relationship between the true objects. This process is commonly required within phantom-based US calibrations, as described in Section 1.5.

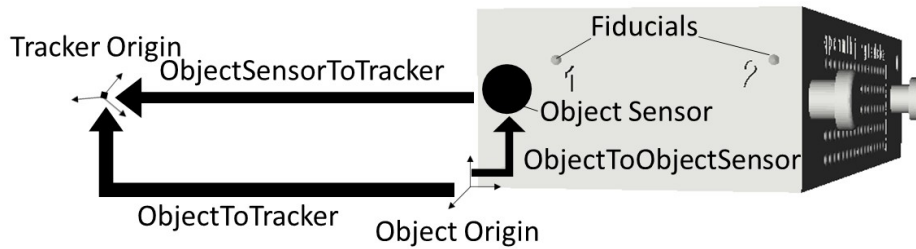


Figure 1.13: Sample of the transformations associated with tracking an object that was calibrated using a fiducial-based method, where objectToSensor is the output calibration transformation which can be concatenated with $\text{objectSensorToTracker}$ to produce objectToTracker .

Template-based calibrations provide another method to solve for the relationship between a tooltip, such as a needle, and the sensor fixed to the tool [25]. The calibration template is an additional tracked device that has the negative imprint of the surgical instrument, a tracking sensor, and calibration fiducials [25]. The position of these fiducials and the geometry of the tool are known with respect to the geometry of the imprint of the tool from the CAD model or a CT scan of the template. The known geometry of the phantom is used to extract the transform that relates the tip of the tool to the origin of the template geometry templateToToolTip [25]. The positions of these fiducials are measured with respect to the tracking sensor fixed to the template [25]. The transformation matrix that registers these two paired fiducial sets provides the relationship $\text{templateSensorToTemplate}$ [25], between the template and its tracking sensor. The relationship between the sensors on the tool and the template in which the tool is inserted is provided by the tracking system through the transform $\text{toolSensorToTemplateSensor}$ [25]. Thus, the calibration can be solved by concatenating the transforms $\text{toolSensorToTemplateSensor}$, $\text{templateSensorToTemplate}$, templateToToolTip to provide the relationship between the tooltip and the tool sensor ($\text{toolTipToToolSensor}$), as depicted in Fig. 1.14 [25].

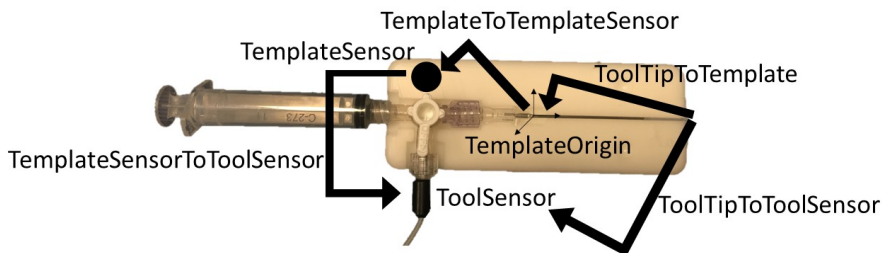


Figure 1.14: Sample of the transformations associated with a template-based calibration. Depicting the transformation chain toolTipToTemplate , $\text{templateToTemplateSensor}$, $\text{templateSensorToToolSensor}$, which provides the output transform $\text{toolTipToToolSensor}$.

1.5 Ultrasound Probe Calibration

US probe calibration is a main application of US registration and is a specific case where the US image is registered to an external spatial tracking system such as an optical or magnetic tracking system [63]. All US probe calibration methods require some form of tracked marker (optical or magnetic) to be rigidly attached to the US probe, whose location is determined with respect to the tracker's coordinate system via the tracking operation [94]. The tracking system can provide the pose of the probe sensor with respect to the tracker's coordinate system or another reference sensor that may be fixed to the patient. Through the concatenation of transforms, the US beam coordinates can ultimately be placed within the patient's coordinate system by means of the reference sensor(s). In general, the resulting 3D transformation matrix is a 4×4 rigid or affine matrix that can consist of any combination of scaling, rotation and translation in up to three directions X, Y, and Z [63]. Assuming the US image lies on the X-Y plane, this matrix can be applied to each pixel location $[X, Y, 0, 1]$ within the image to produce a new location $[X', Y', Z', 1]$ within the tracker or patient coordinate system. Thus, the images obtained from a calibrated US probe are positioned and scaled to the true FoV of the US beam, allowing for pixel spacing to be related to standard units of measurement [63]. US probe calibration allows the relationship between the US image and virtual models of tracked tools to be visualized and enables free-hand 3D US reconstruction [63]. Fig. 1.15 a) displays the relationship between the calibrated US image and the probe sensor, and Fig. 1.15 b) provides an example of a calibrated US image and its interactions with other tracked objects. The various methods used to perform US probe calibration are summarized below [63].

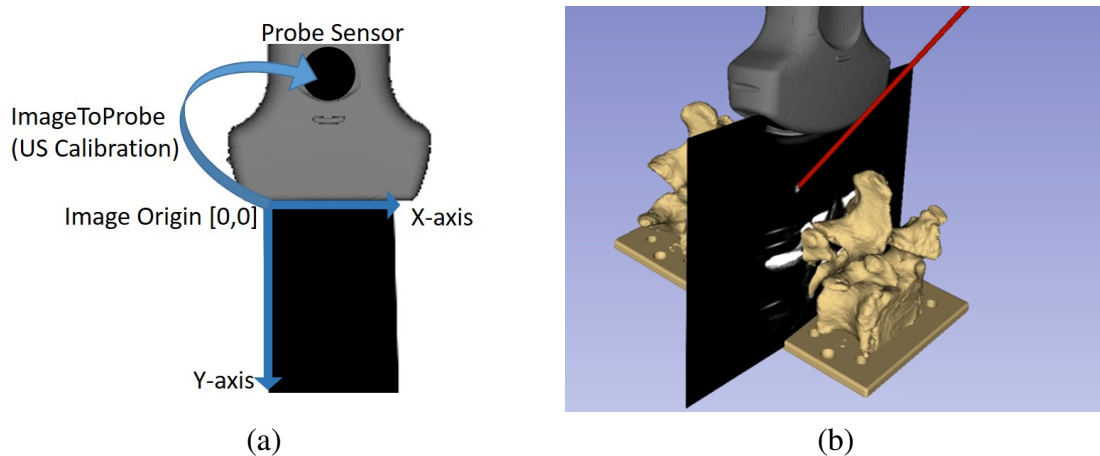


Figure 1.15: a) depicting the relationship between the calibrated US image and the probe sensors and b) provides an example of a calibrated US image interacting with a virtual models tracked needle, US probe and spine model

Free-hand Ultrasound Probe Calibration Methods

Fiducial-based US calibration methods rely on imaging some form of phantom containing landmarks whose positions are known with respect to the probe sensor in the tracker's coordinate system, such that the corresponding landmarks can be automatically or manually localized in the image coordinate system [63]. Fiducial-based US-calibration approaches are solved using landmark-based registrations to provide the relationship between the homologous point sets known in the tracker's coordinate system with respect to the probe sensor and within the US image, which defines the relationship between the US image and the probe sensor [63]. Fiducial-based US calibrations are grouped into different categories based on the type of calibration phantom being used, although the specifics of each method within the category may differ based on the calibration approach. These categories include single-point target phantoms, multiple cross-wire phantoms, multiple-point target phantoms, three-wire phantoms, Z-fiducial phantoms, and needle phantoms. Fiducial-based calibrations are used throughout this thesis and thus are the focus of this section.

The first type of phantom to be implemented for US probe calibration was a single point 4 mm bead target phantom [144]. This concept was extended to the cross-wire phantom, where the phantom is composed of two intersecting wires producing a point within the image [144]. The point target or cross-wire phantoms are aligned with the US image plane and are scanned from several viewing angles as the probe moves freely. These approaches are known as point-based methods, as they consist of mapping a single point from image space to phantom space, and then from phantom space to the tracker's coordinate system. The concepts behind the point target and single cross wire approaches were extended to phantoms that include multiple point targets or cross-wires [75]. The principals of these approaches remained the same, however through increasing the number of points, the registration algorithm has more information to use as inputs [75]. The major limitation of point-based approaches is the requirement to place the phantom in the mid-plane of the US beam. If the phantom does not lie within the mid-plane of the US beam the fiducial within the image won't represent its true location and error is introduced into the calibration [94].

Three-wire phantoms involve three orthogonal wires that are scanned sequentially along their length [15]. For this method, the origin of the phantom coordinate system is located at the intersection of the three wires and then each wire is assigned to one axis [15]. When the phantom is scanned the wire appears as a dot within the US image and again can be input into the point-based registration algorithm. This method eliminates the issues associated with the correct location within the US beam as it is intrinsically easier to scan the length of a wire than a single point [120]. The three-wire method is still limited by the requirement of a precisely designed phantom to insure perfect orthogonality of the wires and the requirement that the user

must keep track of which wire they are imaging.

Z-fiducial (also known as N-fiducial) phantoms are perhaps the most commonly used for probe calibration. The Z-phantom consists of three wires that are organized in a Z formation, this phantom is tracked and requires calibration to the tracker's coordinate system [120]. When scanned, the three collinear points that correspond to the geometry of the phantom are visible within the image, and thus can be used as corresponding point-pairs for registration [120]. This method has been extended to include multiple Z-phantoms to improve the accuracy of the registration. This method is limited by the requirement of a precisely designed phantom for that specific US probe and a phantom calibration step, which may introduce additional errors.

Tracked stylus or needle-based approaches utilize a tracked and pre-calibrated stylus or needle as the calibration phantom, allowing it to be employed universally across all US probes without any precise design requirements [21]. For the point-to-point implementation, the US probe is fixed while imaging a water bath. The operator inserts the needle into the US beam until they believe the tip of the needle to be aligned with the beam's mid-plane [63]. The stylus tip location, and the segmented centroid of the associated reflection, are used as input to a point-to-point registration algorithm [63]. As previously mentioned, optimally positioning an object in the mid-plane of the US beam is non-trivial and error-prone. Thus, line-based calibration methods were developed to address this limitation. This approach does not require positioning the needle in the mid-plane of the US beam, but rather inserting it such that the shaft of the needle intercepts the entire US-beam [21]. The operator still selects the centroid of the reflection produced by the needle, which provides a 3D $([X, Y, 0])$ image coordinate that is used as one of the inputs to the algorithm. The secondary algorithm input is the line formed by the needle, which is defined using the stylus tip location as the origin, along with the direction of the needle, which can be extracted from the needle's transformation matrix delivered by the tracker [21]. In both approaches, the needle is positioned in various locations (and poses for the point-to-line approach) throughout the US image, forming corresponding sets of points or points and lines which can be used to iteratively solve the calibration.

The point-to-line US calibration approach is consistently employed throughout the work presented in this thesis. For an optimal calibration, two scans of the needle are collected in each corner of the image, one with the far end of the needle tilted to the left and the other tilted to the right, with respect to the needle tip [21]. Next, the number of scans with the needle in varying locations and orientations throughout the centre portion of the image are collected. The pixel location, the line formed by the needle, and the location of the tracker on the US probe are input to the closest-point Procrustean point-to-line registration algorithm that iteratively solves for the transformation matrix [21]. In this iterative process, for each new measurement, the algorithm solves for a new transformation matrix that builds upon the previous solution [21].

This approach provides the relationship between the US image and the sensor fixed to the US Probe which is represented as a 4×4 transformation matrix comprised of translation, rotation, and anisotropic scaling. The use of anisotropic scaling is necessary for calibration as it allows for the calibrated US image to be accurately scaled to the imaging depth being used. The number of pixels that comprise an US image differs in the axial and lateral directions. The lateral pixel scaling is related to the number of transducer elements, whereas the axial spacing is related to the speed of sound. For example, if the estimated speed of sound is 1540 m s^{-1} but the true speed of sound through the medium is 1450 m s^{-1} , it takes longer for the US wave to reach the same distance. To compensate for this difference the image can be scaled such that the same number of pixels can represent the actual depth of the image. However, due to the difference in how the lateral and axial pixel spacing is obtained, errors occur if the scaling is isotropic particularly for deep images.

1.5.1 Free-hand Tracked Ultrasound

US probe calibration enables the use of free-hand tracked US, where the position of the US image is tracked and updated in real-time. Tracked free-hand US allows for the US image to be visualized in a common coordinate with surgical tools, such that the virtual model of the tool is aligned with the reflection of the tool within the US image [63], as depicted in Fig. 1.15 b). Due to this capability, tracked free-hand US forms the foundation for many US-guided surgical navigation systems [30]. Free-hand tracked US also enables free-hand US reconstructions, as the pose of the US image is known in 3D. Free-hand US reconstruction is described in more detail in Section 1.8. Last, as free-hand tracked US scans are scaled to the true FoV of the US beam, the pixel spacing can be related to real units of measurements, which allows for clinical measurements and could assist with diagnosis.

1.5.2 Temporal Calibration

Temporal calibration solves for the time offset between data streams acquired by different devices, such as the US and tracking streams. Temporal misalignment between the US and tracking streams affects the accuracy of the calibration if any of the components are moving during data collection, as the reflections produced by the calibration phantom are not representative of the pose of the tool at the same point in time. The method used in this thesis to solve the temporal calibration is a feature in the Plus Toolkit [77]. This method aims to collect the same signal from both data streams. In the case of performing temporal calibration between the US and tracking streams, the US probe outfitted with a tracking sensor is used to image the bottom of a water bath. The user attempts to move the probe in a sinusoidal fashion by smoothly

moving the transducer face closer and farther from the bottom of the water basin. Imaging the bottom of a water basin under US produces a straight line within the US image, where the line is closer to the top of the US image the closer the transducer is to the bottom of the basin. The Plus Toolkit automatically segments this line from the US image and maps the position over time in a plot that resembles a sinusoidal curve. The pose of the tracking sensor fixed to probe is also plotted against time, and also resembles a sinusoidal curve. The two curves have the same appearance but are shifted from one another by some time delay. Performing cross-correlation on the two signals solves for the time offset (lag) that maximally aligns the plots. This offset is the temporal calibration, which can be used to ensure the data streams are synchronous.

1.6 Registration Errors and Ultrasound Calibration Evaluation

1.6.1 Sources of Error in Fiducial-based Registrations

All US-based registrations, whether they are performed with an external tracking system or a secondary image or volume, have some associated errors. For landmark-based registration, the corresponding landmark coordinates must be localized with respect to the US image and the coordinate system to which it is being registered. Fiducial localization error (FLE) is defined as the error associated with localizing the fiducial or landmark, which is an important consideration in fiducial-based registrations as the accuracy of localizing US fiducials can be affected by many sources of error [30]. First, US image generation requires an assumption for the speed of sound, and typically fiducials used for image registration are considerably denser than tissue, and thus the assumed sound speed is far lower than its true velocity when traveling through the fiducial material. This phenomenon may displace the fiducial within the US image [30]. Moreover, most US probe calibration methods require the calibration phantom to be submerged in water, which results in less attenuation of the US beam compared to human tissue [30]. The misalignment with the true and assumed speeds of sound may result in reflections being erroneously mapped within the US image [63]. Additionally, US systems produce a 2D image using a beam that has a 3D profile. Therefore, if the user is attempting to localize the centroid of the fiducial they need to position it in the mid-plane of the US beam [63]. This process is difficult and there is no method to verify that the image was collected with the fiducial optimally placed with respect to the beam geometry. Once the fiducial has been captured in the US image, it must be localized, either manually or automatically [63]. Typically, a single-pixel representing the centroid of the fiducial is selected. The ability to accurately

interpret and localize the fiducial in the US image directly relates to the resolution and point spread function (PSF) of the US image [3]. The PSF is the response of a system when imaging a single point target, and measures the degree of blur or spread of the point in the image [3]. The PSF can be used to characterize the spatial resolution at the location of the point fiducial. PSF in US images can extend by several pixels in both the axial and lateral dimensions but are typically anisotropic with a larger spread in the lateral dimension. The PSF and resolution in the axial dimension differs throughout the image due to US beam characteristics [3]. Thus, selecting the pixel that represents the centroid of a fiducial in US images can be difficult and often localizations are erroneous by a few pixels, as the fiducial's appearance in the US image is anisotropically blurred with varying degrees based on the axial position of the fiducial [3]. Accurate fiducial segmentation can also be related to the operator's knowledge of US physics and imaging, as more experience results in easier interpretation of reflections in US images.

For US calibration, the fiducials must be localized (or digitized) with respect to the tracker's coordinate system. External tracking systems have an inherent error associated with them. While tracking systems used for clinical tasks require sub-millimetre and sub-degree accuracy, these small errors affect the ability to accurately digitize fiducials, which propagate into the FLE. Many calibration phantoms require a phantom calibration step such that the position of the landmarks are known with respect to the phantom sensor and thus the tracker's coordinate system, as described in Section 1.4.2. Any errors in the calibration of this phantom will erroneously position the landmark in the tracker's coordinate system, resulting in FLE [30, 96]. Moreover, if the tracking and image streams are not temporally calibrated, the recorded tracking data may have differences between the true location of the tracking data and the recorded location for its corresponding image containing the reflection produced by the calibration phantom. Any errors associated with localizing the fiducial propagate into fiducial registration error (FRE). FRE is the error between the corresponding landmarks after registration. Target registration error (TRE) is defined as the error between two sets of non-fiducial points after registration [30] and is potentially the most meaningful registration error. This metric is meaningful as the FRE over or underestimates the accuracy of the registration [30]. The relationship between FLE and TRE has been quantified by equation 1.7, where n is the number of collected fiducials [96]. This model works under the assumptions that the FLEs are isotropic and are independent and identically distributed randomly with a zero-mean Gaussian distribution [96]. The relationship between FLE and TRE is defined by equation 1.8, where p is the point target and d_k^2 and f_k^2 are the squared distance between the target point and the fiducial points and the k th principle axis of the fiducial configuration, respectively. Equation 1.8 follows the same assumptions as equation 1.1.

$$E(FRE^2) \approx \frac{(n-2)E(FLE^2)}{n} \quad (1.7)$$

$$E(TRE^2(p)) \approx \frac{E(FLE^2)}{n-2} \left(1 + \frac{1}{3} \sum_{k=1}^3 \frac{d_k^2}{f_k^2} \right) \quad (1.8)$$

1.6.2 Calibration Quality Assessment

The accuracy of an US calibration is difficult to quantify, as there is no way to produce a ground truth relationship between the probe tracking sensor and the US image [63]. Thus, there is no standard for assessing calibration quality. However, there are common methods used to measure the precision and trueness to validate the calibration [63]. Precision is the spread between repeated measurements or samples in some coordinate system, and trueness is how close each sample is to some known ground truth [63]. Other important factors to consider when assessing an US calibration approach are the equipment, space, and time required for the calibration approach.

A common precision method used is calibration reproducibility, whereby a single-pixel $[X, Y, 0, 1]$ in image space is transformed by the calibration to a new coordinate $[X', Y', Z', 1]$ in 3D space [63]. This process is repeated using multiple calibrations obtained using the same calibration approach [63]. The variation of this reconstructed point cloud represents the precision of the calibration. The selected pixel coordinate has an effect on the precision obtained, as selecting points near the centre of the image may be less affected by rotational errors [63]. Thus, it is recommended to test the calibration reproducibility using multiple reconstruction targets, such as all four corners as well as a few samples close to the centre of the image [63]. Calibration precision is also quantified by the variation across each DoF within the calibration matrix (translation and rotation in the X, Y, and Z axes) using multiple calibrations obtained using the same calibration method [63]. Both of these processes measure the consistency of the calibration approach and can be used in junction to quantify the calibration precision.

It is near impossible to measure calibration accuracy as the ground-truth calibration is not known, since if there was a method to obtain the optimal calibration this technique would replace the need to perform US calibration [63]. As quantifying the accuracy of the components within an image-guidance system is vital to their success many groups have attempted to quantify the accuracy of their US calibration methods. Point reconstruction accuracy uses a point-based phantom, where the point has a known position in the tracker or world coordinate system to provide the ground-truth [63]. This phantom is imaged by the calibrated US probe and the position of its reflection in the calibrated image is compared to the ground-truth location [63]. This process is repeated many times such that a cloud of reconstructed points can

be compared to the ground-truth [63]. Many groups have also performed point reconstruction accuracy using phantoms with multiple point sources, typically produced using wire-crossings in a Z or cube pattern [63]. These points can be reconstructed in 2D by extracting their position from a 2D image or in 3D by extracting their locations from a reconstructed volume [63]. Similar to the 3D point reconstruction accuracy, many groups employ 3D distance reconstruction accuracy to assess their calibrations, where a phantom with multiple point sources is reconstructed in 3D by the calibrated US probe [63]. The distances between the point sources are compared to the known configuration of the phantom [63]. A common phantom used in this approach is a 4×2 matrix of small spheres. However, phantoms with a large number of point sources have also been used [63].

While point-reconstruction analysis provides a method to assess calibration accuracy, many sources of error are introduced into the analysis. The centre location of the point source has to be extracted by manual or automatic localization, which may affect the accuracy results produced [63]. In the case of 2D point reconstruction, the point fiducial has to be imaged at the mid-plane of the US beam. 3D point reconstruction resolves the issue of alignment in the mid-plane. However, the accuracy of the 3D reconstruction depends on the sampling rate and reconstruction approach, and artifacts can be introduced during this process [63]. Thus, any errors within the reconstruction or fiducial localization affect the overall accuracy measure obtained for the calibration [63]. Moreover, even comparing the same accuracy metric reported from two independent groups, can be difficult as the accuracy of the calibration is affected by the imaging depth, type of tracking device, quality of equipment, the experience of the operator, and there may be additional factors depending on the calibration approach used [63].

1.7 Ultrasound segmentation

Medical image segmentation is the process of delineating or outlining structures present in medical images such as relevant anatomy or surgical tools. For US-based applications, structures may be segmented to assist with diagnosis and treatment planning or to improve the visualization of tools or targets for US-guided applications [109]. US images may be segmented using traditional image-processing-based techniques, or machine learning-based approaches that require some form of training phase where the algorithm learns how to perform the segmentation [112]. Many different image-processing and machine learning-based methods for US segmentation have been developed for applications within cardiology, oncology, and vascular disease diagnosis [112]. Another common application of US image segmentation is image-based tool tracking.

1.7.1 Image processing-based segmentation

Traditional image processing-based segmentation relies on hand-crafted or known features, such as intensity, intensity distributions, or edges [109]. Edge detectors are an example of a known feature, as they are designed to detect abrupt changes in the intensities within an image by applying some form of derivative across neighbouring pixel values to quantify rates of change, where areas with high rates of change are edges [109]. Edge detection is typically performed using a filter to detect horizontal, vertical, diagonal or all edges in the image. Some common edge detection filters include Sobel, Canny, and Laplacian [112]. Segmentation based on pixel similarities are also designed features, for example, pixels with similar intensities can be grouped as the intensity of a pixel in an US image relates to the tissue properties. Some commonly used intensity-based segmentation approaches include threshold-based, region-based, region growing [112]. It can be difficult to develop a robust image processing-based algorithm that accurately segments structures across images collected from various patient groups, different US scanners, and different sonographers.

1.7.2 Machine learning-based segmentation

Machine learning algorithms are commonly used for US segmentation applications, as they are capable of performing segmentation on diverse data sets. Supervised learning algorithms require both images and ground-truth segmentation examples as training inputs [33]. The goal of a supervised machine learning algorithm is to try and develop a function that aims to replicate the corresponding ground-truth labels when given a training input image [33]. Supervised learning algorithms for segmentation are referred to as classification algorithms, as pixels are grouped into classes such as disease or non-diseased, or some anatomy and background [33]. Unsupervised learning only requires images as the training examples with no training labels, as pixels are grouped into classes based on statistical trends. Unsupervised learning algorithms for segmentation are referred to as clustering algorithms that group pixels into classes based on similar properties or features [112]. There are also semi-supervised approaches that combine supervised and unsupervised learning [33]. As this thesis only implements supervised machine learning methods, the following section focuses on this approach. One of the most commonly used machine learning models is a convolutional neural network (CNN). CNNs comprise several components such as convolutional layers, pooling layers, and fully connected layers, where the architecture of the CNN depends on the order and frequency of these components [162]. Convolution layers perform feature extraction using kernels or filters, which are matrices of values usually significantly smaller than the image, such as a 3X3 grid, where the depth of the image and the filter are the same [162]. The values within these filters are

called weights, and determine what kind of feature the kernel is trying to detect [162]. The convolution operator produces the sum between the element-wise kernel and the underlying image matrix values [162]. The magnitude of the result is a measure of the similarity between the elements of the kernel and the image at the position in the image matrix overlaid by the kernel [162]. If the value of the convolution is high the feature is present in that portion of the image, if the value is low the feature is not present [162]. A feature map is generated by passing the filter across the entire image by shifting the kernel by some set stride amount and performing the convolutions. The output convolution value is recorded at the pixel coordinate at the centre of the kernel where the convolution was applied [162]. After the feature map is produced it is summed with a bias term and then an activation function or transfer function is applied that defines some consistent method of mapping of the values in the feature map to a new value [162]. This altered feature map or activation map next undergoes a pooling step to produce a pooled feature map [162]. One of the most common pooling steps is max-pooling, which extracts sub-grids of pixels from the feature map finds the maximum value in the sub-grid and discards the rest [162]. Depending on the architecture of the network, there may be many consecutive convolution, activation, and pooling layers, these layers are typically referred to as hidden layers [162]. After the final pooling layer, the resultant pooled feature map sized $n \times m$ is flattened or reshaped to a one-dimension feature vector of size $1 \times n * m$ [162]. The flattened layer is then input into one or more fully connected layer(s) in which every input is connected to every output by a learnable weight [162]. There are usually a few fully connected layers, where each layer is followed by an activation function. The last fully connected layer is input into an activation function that is usually different from the others used.

The softmax function is a commonly used activation function for multi-class classification as it produces class probabilities in the range 0 to 1 [162]. The softmax function (σ) is a normalized exponential function as defined in Eq. 1.9, where \vec{z} is the input vector, e^{z_i} is the standard exponential function for the input vector, K is the number of classes in the multi-class classifier, and e^{z_j} is the standard exponential function for the output vector [162]. Two other common activation functions used within machine learning architectures are the Rectified Linear Unit (ReLU) and the leaky ReLU. The ReLU activation function is defined as $y = \max(0, x)$, such that all x -values less than zero follow the function $y = 0$ and all x -values greater than zero follow the function $y = x$ [86]. The leaky ReLU function is the same as the traditional ReLU for x -values greater than zero, but all x -values less than zero are input to the function $y = \alpha x$ where α is a small value such as 0.01 [86]. The leaky ReLU improves upon the traditional ReLU as there are no zero-slope components, and typically converges faster during training [86]. Numerous CNN architectures that have been applied to medical image segmentation problems. Two pre-developed architectures that are used in this thesis

include U-Net and Mask R-CNN.

$$\sigma(\vec{Z})_i = \frac{e^{Z_i}}{\sum_{j=1}^K e^{Z_j}} \quad (1.9)$$

U-Net is one of the main architectures used for medical image segmentation [131]. The U-Net algorithm has three phases, contraction, bottleneck, and expansion, and the model resembles the shape of a U [131]. The contraction phase consists of multiple convolution blocks, where one block performs two consecutive 3X3 convolutions. Each convolution block is followed by the ReLU activation function and a single 2X2 max-pooling step with stride 2 [131]. The number of feature channels doubles for each convolution block [131]. The bottleneck phase links the contraction and expansion portions by performing 2 3X3 convolutions followed by an up-sampling convolution layer that performs the reverse operation of a max-pooling layer [131]. The expansion phase also uses convolution blocks, where 2 3X3 convolutions occur sequentially [131]. Each convolution block is followed by a ReLU and an up-sampling layer. To maintain symmetry the number of feature channels is halved for each convolution block in the expansion portion of the model [131]. There is also a crop and copy connection between the output of each contraction layer and the corresponding expansion layer, to append the input with the feature maps produced in the contraction phase [131]. There is one final 3X3 convolution in the expansion phase, where the number of feature maps is equal to the number of desired segmentations [131]. The U-Net architecture has been successfully used to segment nerves, breast lesions, arterial wall, cardiovascular structures, thyroid and other structures from US images [138].

Mask R-CNN is another machine learning algorithm that has been applied to medical image segmentation problems. Mask R-CNN was inspired by Faster R-CNN [58] for object detection [129] and consists of two stages. In the first, a Region Proposal Network (RPN) determines possible bounding boxes that may contain objects of interest. In the second, two components execute in parallel, receiving the region proposals from the RPN as input. The first component, inspired by Faster R-CNN, predicts object class and bounding box localization, while the second predicts pixel segmentation (masks) for each RoI [58]. Mask R-CNN has been successfully applied to US segmentation of breast tumors [29], femoral cartilage [74], and other anatomy.

Once the architecture of the CNN has been developed, a model is trained using the training images and associated labels. During training, the algorithm runs for a set number of epochs [162]. A small dataset (validation dataset), which is independent from the training dataset, will be used during training to monitor the performance of the trained model [162]. A cost (i.e. loss) function quantifies the correctness of the algorithm's predictions when com-

pared to corresponding labels across the validation set [162]. Optimization algorithms, such as gradient descent or the Adam Optimizer, search for the values of the model's learnable parameters that minimize the cost function using back propagation [162]. Early stopping may be applied based on the performance on the validation dataset to prevent over-fitting [162].

1.7.3 Image-based tracking

Image-based tracking can be used as an alternative to external tracking systems, where the pose or location of a surgical tool is tracked solely based on information extracted from images. Tools can be tracked in image space, where the pixels representing the tool on the US image are identified and labelled in real-time to improve the visualization for the user [136, 7]. Needle localization in US images is a common application, where the tip or entirety of the needle is tracked in the US image to assist with needle identification and visualization for navigation [136, 7]. When imaging a needle under US, it appears as a bright hypoechoic reflection compared to the background. Thus, pixel characteristics can be used to separate pixels belonging to the needle and background tissue [136, 7]. The majority of needle-tip identification algorithms rely on the assumption that the needle comprises the brightest pixels in the image and that the needle is a linear structure [136, 7]. These assumptions allow for traditional image processing-based algorithms to detect the needle, such as the Randon and Hough transforms, to identify linear structures, and filters such as the Gabor filter to differentiate between light and dark pixels [136, 7]. Advancements in machine learning algorithms have promoted the development of needle detection algorithms based solely on machine learning, or some combination of machine learning and image-processing, to localize the needle in US images. In general, these machine learning algorithms are built upon and out-perform traditional image-processing approaches for needle localization [56]. Classification is commonly used to label all pixels as either background or needle [56]. Regardless of the method used to identify the needle, post-processing steps are often used to obtain needle trajectory or identify the needle-tip [55]. Various methods have been developed that can localize the needle with less than 0.5 mm of error [56, 55].

More advanced image-based tracking algorithms use the information within images to provide the pose of surgical tools or anatomy in 3D with respect to a world coordinate system. This process differs from 2D US-based needle tracking where the needle pose is provided with respect to the image coordinate system. 3D US images can be used to estimate the needle pose in 3D, where the relationship between real and pixel spacing is already known. Both image processing [34] and machine learning-based [117, 4] solutions have been developed for needle tracking from 3D US with various approaches achieving sub-millimetre accuracy [34, 117, 4].

While these approaches are highly promising, there are limitations to using 3D US volumes to obtain the needle pose in 3D. The formation of 3D US volumes without specialized 3D array transducers requires some form of tracking system, typically mechanical or magnetic tracking, to accurately position each 2D slice within the 3D volume [40]. Due to the methods used to obtain 3D US volumes, they cannot be updated in real-time. In the case of mechanical tracking the user no longer has free-hand control of the US probe. While estimating the pose of tools or anatomy in 3D from 2D US images is a process that is more prone to error due to the extrapolation of information to obtain the third dimension, the user can nevertheless maintain free-hand control over the US probe and the US images can be updated in real-time.

Another common application of image-based tracked is US probe pose estimation for sensorless free-hand US reconstruction. Sensorless 3D US reconstruction uses sequential US images and feature detection, matching and tracking to estimate the pose of the US probe and thus image in 3D [119]. While this is a major application of image-based tracking, it is more relevant to US reconstruction and is expanded upon in Section 1.8.

1.8 Ultrasound Reconstruction Methods

The benefits of 3D imaging have motivated the development and use of 3D US volumes, which are reconstructed using a series of sequential 2D US images [119]. In general, the pose of each US slice within the scan is obtained using external tracking components or estimated using image-based tracking. Using the spatial position of the 2D scans the 3D volume is populated by either resampling and interpolating the information from the 2D images into a rectangular volume or visualizing the information as sequential 2D slices across a 3D volume [119]. The process of reconstructing the 3D volume by resampling the image data trades off an increased time required to obtain the 3D volume with the improved visualization that closer resembles a CT or MRI volume.

The commonly used methods for 3D US reconstruction include: specially designed 2D array probes, mechanically sweeping the transducer, free-hand with external tracking, and free-hand without external tracking [119]. An US probe with a 2D array of elements can obtain 3D volumes by strategically timing the firing of pulses from the various elements. However, the requirement to have a specific type of non-traditional US probe is a major limitation. For methods that require a mechanically sweeping transducer, an US probe is outfitted with a mechanical steering device that moves the probe in a known trajectory to collect 2D US scans covering a predetermined 3D volume [119, 40]. 3D US reconstruction using a free-hand approach with external tracking requires an US probe with a rigidly fixed tracking sensor, typically optical or magnetic [119]. This US probe must be spatially calibrated such that the pose of the image is

known with respect to the tacker's coordinate system. The calibrated probe provides the pose of each image in the scan such that they can be spatially positioned to form a 3D volume [119]. Free-hand US reconstruction can also be performed without the use of an external tracking sensor. Some methods to produce freehand 3D reconstruction replace the external tracking systems with other components, such as inertial sensors and cameras, that provide information that can be used during the reconstruction [119, 123]. Alternatively, entirely image-based tracking methods can be implemented that estimate the pose of each US slice without any additional equipment.

The first sensorless freehand reconstruction algorithms relied on tracking the speckle component of US images. Speckle is an inherent US artifact caused by the redirection of US waves as they come in contact with small, rough or uneven surfaces resulting in the signals being received by a different element than which the signal was produced. Speckle is always present in US images and differs from noise as it is not random. Speckle tracking can be used to determine the velocity and direction of scanning by following speckle patterns between slices [119]. Speckle decorrelation techniques used the statistical similarity between the speckle in two images to determine the motion between them, it was observed that the higher the speckle correlation the lower the distance between neighbouring slices [119]. Substantial research has been performed to improve the accuracy of speckle-based approaches for reconstruction, with the majority of work focused on speckle decorrelation methods [119]. While the relationship between speckle and relative image motion has been well developed and quantified, the US reconstructions produced using only speckle decorrelation have been found to not meet clinical requirements [124]. Recently, machine learning methods have been developed to automatically quantify the relationship between the speckle and the pose of the image with respect to the other images in the scan using trained statistical models [119, 124].

Prevost *et al.* [124] were the first formulated US reconstruction as a machine learning problem and demonstrated that a convolutional neural network (CNN) can automatically estimate motion between consecutive US frames with reasonable accuracy. This 2D CNN takes two sequential US frames and estimates the relative rotation and translation between slices using an optical flow field [124]. Guo *et al.* [54] expanded upon this solution by developing a deep contextual learning network with a novel case-wise correlation loss that takes multiple consecutive frames as input and focuses on speckle-rich portions of the images. These sensorless machine learning approaches that provide are promising but they are much more prone to distance and drift errors compared to approaches that use additional equipment and spatial information.

1.9 US Visualization

The standard method used to visualize medical imaging and surgical navigation information is via a 2D monitor [139]. A major advantage of this display is the ease of clinical integration, as operating rooms and US machines are already equipped with monitors [139]. As monitors are already used to transmit information in a clinical setting, clinicians do not have to adapt to both a new system and a new display. However, displaying surgical navigation information on a monitor results in disparities between clinician's visual and motor fields, as clinicians focus is on the screen, not the surgical site [139, 73]. 2D monitors can also contribute to perceptual targeting issues when displaying 3D information on a 2D monitor [158]. Furthermore, systems have been developed that promote the use of 3D volumes for surgical planning or diagnostic measurements [158, 128].

Visualization techniques can be used to improve the clinician's ability to interpret 3D information on a 2D monitor. One way to simplify the visual information is using the multi-planar reformatting (MPR) approach [39]. In the MPR method, the volume is resliced and used to render three intersecting and orthogonal planes. Alternatively, volume rendering can be used to display the entire 3D volume or to project the 3D volume onto a 2D plane [39]. While texture mapping by means of transfer functions can be applied to US volumes or MPR images to make features of interest stand out, transfer functions can be difficult to optimize and developing a generalizable function that works across various systems as clinical cases is a non-trivial task [39]. Segmentation is one method that is commonly used to extract information that is pertinent to the clinical application and improve visualization, by automatically or manually delineating structures of interest [39]. Anatomical reconstruction can be obtained by first performing a 3D US reconstruction, and then segmenting each slice within the volume or segmenting the entire 3D volume to extract the anatomy in 3D [39]. Alternatively, 2D US images can be segmented first and the resulting label maps can be positioned within a 3D volume if the pose of the US slices is known [151, 133]. 3D anatomical segmentations are typically collected using a mechanically sweeping transducer or a free-hand transducer calibrated to an external spatial tracking system [39]. After segmentation, opacity, Z-buffering, colour, and other properties can be manipulated to enhance and augment medical image visualization. Typically, these segmented structures are visualized with respect to a resliced image that intercepts the 3D models. The two main applications for US-derived anatomical reconstructions include diagnosis and guidance [153]. Performing reconstruction using an external tracking system and a calibrated US probe allows the 3D anatomy and the surgical tools tracked by the spatial tracking system to be in a common coordinate system. This provides a method to visualize virtual models of the surgical tools, the anatomy and the interactions between them.

US-based anatomical reconstruction has been applied to scoliosis measurements [151], prostate brachytherapy [39], and plaque detection and quantification [133].

1.9.1 Image Fusion

Image fusion is the combination of information from two or more image sources into a single image or displaying the co-registered images side by side [37]. This process requires registration between the images or volumes, followed by resampling to optimally co-register the images or volumes [37]. US image fusion can be performed using static and real-time images. Static US image fusion is the combination of unchanging US images or volumes with other static image volumes such as pre-operative MRI or CT [37]. Real-time US image fusion is the combination of real-time streaming US with other static or real-time images [37]. This process allows for pre-operative images to be related to or overlaid on the true anatomy of the patient during the surgical intervention as captured by the streaming US [37]. Real-time image fusion requires real-time and continuous registration between volumes. US image fusion with CT or MRI has been successfully applied to brain [118, 76] and prostate [83, 142] interventions. A secondary application of real-time US image fusion is the combination of US images and video feeds from one or more camera(s) [26]. US images can be fused with surgical video feeds, with the most common application being the fusion of stereoscopic and US feeds (surface or laparoscopic) to provide more optimal visualization to the clinician [27, 26]. Additionally, US video feeds can be fused with stereo video feeds from VPT-HMDs such that the US and real world information can be simultaneously visualized.

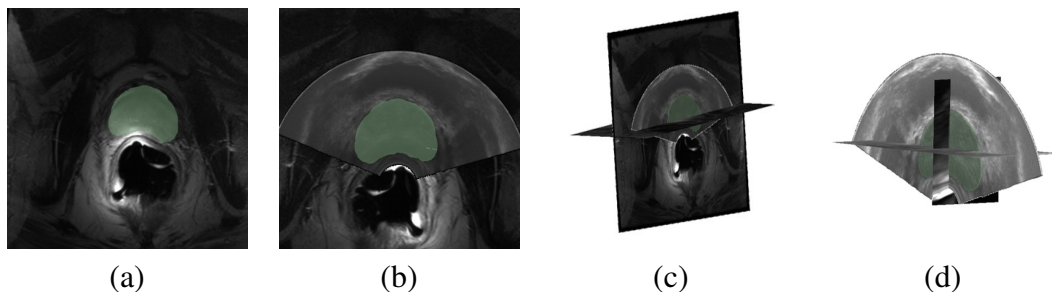


Figure 1.16: Various visualizations of prostate imaging are presented in this Figure. Image a) depicts an US slice from a 3D volume registered and overlaid on an MRI from the same patient, Image b) shows a prostate segmentation from the US overlaid onto the MRI image, figure c) shows a MPR visualization with three MRI slices intersecting the prostate segmentation, and figure d) shows a MPR visualization where a front-facing US image and two slices from the MRI are intersecting the segmentation.

An example of various US visualizations and image fusion are depicted in Fig. 1.16. The data used to make these images were downloaded from the Sample Data Module that is a part

of the 3D Slicer Platform [70]. This figure depicts an MRI image with prostate segmentation, the US image registered and overlaid on the MRI with the segmentation and two MPR views one using three slices from the MRI and the other using two MRI views and one US view that intersect the prostate segmentation.

1.10 Display Technology

US images are traditionally displayed on the 2D monitor affiliated with the US console, but advancements in 3D visualization methods to display US-derived information have motivated the integration of advanced displays into US-based surgical navigation systems. These displays include 2D monitors with advanced visualizations, handheld devices with a screen (phone or tablet), 3D monitors or 3D capable monitors, and wearable technology. These various displays span the mixed-reality (MR) spectrum.

1.10.1 Mixed-reality Spectrum

MR is the overarching term used to describe an environment with any combination of real and virtual information [97], as depicted in Fig. 1.17. The MR spectrum ranges from physical reality to digital reality [97]. Terms such as augmented- and virtual-reality are used to describe specific combinations of physical and digital reality [97]. Augmented-reality (AR) is the inclusion of some computer-generated graphics overlaid on to information from physical reality [97]. Virtual-reality (VR) exists at the other end of the spectrum and represents an environment that is comprised only of computer-generated graphics [97]. MR systems are commonly developed for image-guided surgical navigation or assistance with clinical diagnoses [30].

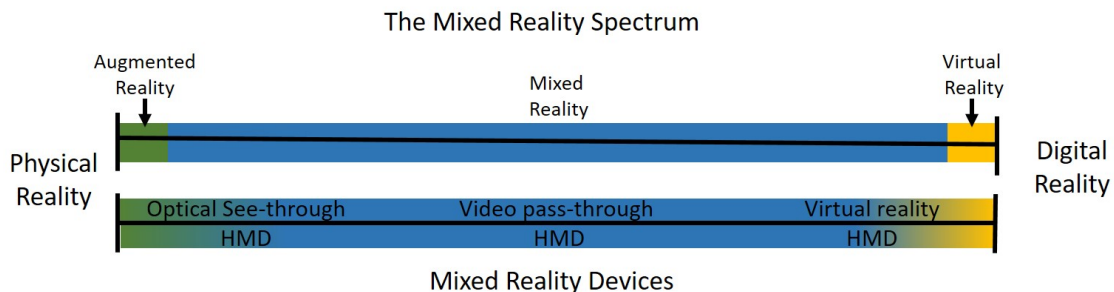


Figure 1.17: The MR spectrum, where the green portion on the left represents AR, yellow represents VR on the right end and blue indicates MR. Below the MR spectrum is the spectrum of MR devices, with optical see-through head-mounted displays on the left, video pass head-mounted displays through in the middle, and VR head-mounted displays on the right.

1.10.2 Ergonomic displays

Ergonomic displays such as high-resolution monitors, hand-held or mobile screens such as a tablet, and 3D capable displays, can provide visual information spanning the MR spectrum. High-resolution monitors can visualize US images and other information such as virtual models of tracked tools or anatomy in a common coordinate system. 2D external monitors allow for more control over the visualization compared to the monitors fixed to US consoles, as many US systems only allow for images rendered by the company's proprietary software to be rendered on the screen. These displays can be mobile to provide a more ergonomic viewing position without the limitation of the length of the US probe cable or the limited mobility of the display fixed to the console. A high-resolution monitor trivially enables VR scenes to be rendered, as monitors are designed to render computer-generated graphics. AR-based systems can display their output on a 2D monitor, and by incorporating a calibrated video camera into the system, the real world information can be visualized on the monitor by rendering some or all of the video-feed from this camera [97]. High-resolution monitors can be limited for guidance or navigation as they divide the user's attention between the visual and motor fields as the procedure is performed in one location, while the information about the procedure is provided in another location. The user is required to mentally transform the information being visualized on the monitor to the site of the procedure, resulting in a high cognitive load. Moreover, the user is provided information on a 2D monitor to perform a 3D procedure. While the information presented on the monitor can be three-dimensional, the ability of the human visual system to perceive 3D information from a 2D context can be limited [146]. This may require systems to be designed with 3D visualizations that are easy for the user to interpret in 3D. Mobile or hand-held devices have similar capabilities to high-resolution monitors but have the added benefit of being able to be positioned closer to the patient to mitigate the visual and motor field disparities. Many mobile handheld devices include a camera, making it easier to develop AR systems as the visualization can comprise computer graphics and real world information from the camera's video feed, which can ultimately be overlaid on the patient [146]. Tablet displays share the same 2D perceptual limitations as monitors. The last type of ergonomic display is a 3D capable monitor such as a 3D TV. These displays maintain the visual and motor disparities but attempt to improve perception issues by visualizing the graphics in 3D [146]. Commercially available 3D-capable or stereoscopic displays simultaneously provide two slightly different views of a scene. Various solutions have been implemented such that each eye only receives one of the two images [91]. The majority of the commercially available stereoscopic displays have associated glasses that use mechanisms such as shutters, polarizing filters, or colour filters to separate the information going into each eye [91]. The human eye can fuse the information obtained by each eye, where the disparities between the images obtained by the left and right

eye enable 3D perception. Stereoscopic displays have been successful used for laparoscopic surgery and to visualize 3D US images for diagnostic purposes [61], but there is little to no research pertaining to the use of stereoscopic displays for US-guidance.

1.10.3 Wearable Technology

Head-mounted displays (HMDs) have been proposed for use in surgical navigation systems to improve 3D perception. An HMD is a wearable technology that allows users to visualize information directly in their FoV in 3D [115, 139]. Monocular HMDs contain one screen that is simultaneously viewed by both eyes. Binocular HMDs have two separate screens one for each eye. Binocular HMDs are the focus of this section as they provide superior perception compared to monocular devices [113]. There are three main types of HMDs, which include: optical see-through head-mounted display (OST-HMD), video pass-through head-mounted displays (VPT-HMD), and virtual-reality head-mounted display (VR-HMD) [115, 139]. OST-HMDs are the most common HMD used for surgical navigation systems, as the translucent display allows the clinician to maintain the line of sight with the patient [125, 68, 23]. OST displays are inherently AR environments as computer-generated graphics are projected onto the real world by means of the transparent display. While the ability to visualize the real world is beneficial for surgical navigation, OST-HMDs have limitations. These HMDs utilize inside-out tracking that does not require external tracking components. Inside-out tracking suffers from lower accuracy compared to its outside-in counterpart that makes use of external tracking components [115]. Furthermore, OST-HMDs have a smaller FoV compared to VPT and V-HMDs, which can affect the user's sense of immersion [115]. Additionally, OST-HMDs are prone to perceptual issues as the semi-translucent nature of the holographic projections affects the user's depth perception, resulting in projections that appear to be floating in the foreground [139].

VPT-HMDs are outfitted with stereo cameras, which allows real world information to be visualized in the HMD. VPT-HMDs are capable of spanning the entire MR spectrum as they can render an entirely virtual scene or include increasing amounts of real world information by integrating the video stream from the stereo cameras. The video feeds from the current commercial VPT-HMDs are low resolution and are not perfectly spatially calibrated to the tracking of the HMD. These factors result in registration errors between and surgical navigation information and the real world video stream. VPT-HMDs more often employ outside-in tracking, which requires external tracking components but results in more accurate tracking [115]. VPT-HMDs have large FOVs and address the depth perception issue associated with OST-HMDs [115].

The last class of HMDs is VR-HMDs, which were the precursor to VPT-HMDs. While these HMDs have all the same benefits as VPT systems, they are only capable of render-

ing scenes generated entirely from computer graphics [92]. VR-HMDs can be classified as computer- or phone-dependent [92]. Computer-dependent VR-HMDs rely on connections to an external system such as a desktop computer to render the graphics and stream the data to the HMD [92]. Computer-dependent HMDs are more technologically advanced, provide software, have superior tracking, and have more realistic rendering compared to phone-dependent systems [92]. The trade-off of computer-dependent HMDs is the increased cost and hardware required. Lower-cost phone-dependent HMDs have been developed to provide a low-cost alternative to computer-dependent systems [104]. These companies typically offer kits that can be self-assembled to form a light weight HMD with a place to secure your phone. The phone acts as a VR display. Most of these companies provide development platforms and packages such that the user can design their own software and visualize it using their phone. These systems are the lowest cost, but require the most development by the user [104]. Phone-dependent systems may be limited by the phone being used and may not be suitable for medical-based interventions.

HMDs have been used for a variety of clinical scenarios, and have been successfully applied to surgical navigation for pedicle screw navigation and spinal needle insertions [80, 24, 60]. They have also been used to view 3D images for surgical planning and diagnostic measurements [158, 128]. There are now many commercially available HMDs, which has enabled more rapid development of HMD-based surgical navigation systems.

1.10.4 Overview of Current Devices

Various commercially available HMDs have been developed for the entertainment industry. These HMDs fall under the three categories: OST-, VR-, and VPT-HMDs as outlined above. This section covers the most popular commercially available HMDs. The technical specifications of a large portion of the commercially available HMDs are summarized in Table 1.1

The Microsoft HoloLens (Microsoft, USA), is a commonly used HMD for medical applications and both the HoloLens 1 and 2 are commercially available. Both models of the HoloLens are OST-HMDs are untethered and do not require connections to power, a computer, or external tracking components. All tracking is performed using simultaneous localization and mapping (SLAM) and inertial measurement units (IMU), where the tracking coordinate system is defined with respect to the HMD. The HoloLens 2 is the current state-of-the-art OST-HMD and the technical specifications are summarized in Tab. 1.1. The major differences between the HoloLens 1 and 2 are the increased FoV, improved resolution, implementation of eye-tracking and two hand tracking (HoloLens 1 only tracked one hand), rendering using eye-tracking, improved processor unit, Bluetooth capabilities, more memory, a USB connection, and a larger

gesture recognition library. One additional feature that is clinically useful is the HoloLens 2 has a display that can flip up to allow the user to choose when the display is needed. The HoloLens 1 was used to develop an FDA-approved AR surgical planning system called OpenSight (Novorad Corporation, USA). The HoloLens has been successfully used for pre-operative planning [137, 114], surgical navigation [80, 43, 121], and training [79, 149]. The HoloLens is the most expensive commercial HMD due to its untethered nature and see-through display.

The HTC VIVE (HTC, Taiwan ROC) and HTC VIVE Pro (HTC, Taiwan ROC) (second edition) are commercially available and commonly used for gaming and entertainment. Both models of the VIVE are VPT-HMDs and can render an entirely virtual scene or include input from video streams from the camera(s) fixed to the outside of the HMD. The VIVE Pro has two stereo cameras used to stream the video input which improves the visualization compared to the VIVE that has one camera centred on the display. The VIVE Pro can display the input from the stereo camera in the form of coloured video with varying opacity or edge mapping. Both models come with two associated controllers, the controller's design remained consistent between the VIVE and VIVE Pro and can be used with either system. The VIVE Pro also has an audio input, a larger FoV, improved tracking accuracy, and better resolution compared to the first edition, the technical specifications for the VIVE Pro are summarized in Tab. 1.1. The VIVE tracking system uses Lighthouse technology, where two base stations send out alternating sweeps of infrared (IR) light that span 120° in both the horizontal and vertical directions [107]. Photodiodes on the surfaces of the HMD and controllers detect the IR light and calculate the position and orientation (pose) by comparing the time differences between the detected and the known locations of each photodiode on the HMD [107]. This tracking is also supplemented with inertial measurement units to add additional tracking information [107]. Both the VIVE and VIVE Pro can also be used as VR-HMDs by not using the data from the camera(s).

Strictly VR-HMDs are the most common commercially available system, due to the increased control over the environment, VR-HMDs are typically easier to design. Some commercially available VR-HMDs include the Oculus Rift (Facebook, USA), Playstation VR (Playstation, USA), FOVE (FOVE, Japan), and the Avegant Glyph (Avegant, USA). These HMDs allow for only virtual graphics to be displayed to the user. The major differences in these various VR-HMDs are the resolution and FoV. As all of these products are developed by different companies they differ in the development platform and what developers can and cannot do. Both the HMD design and development capabilities are important to consider when selecting an HMD for clinical applications. VR-HMDs have a lower price point compared to VPT- and OST-HMDs. Commercially available phone-dependent HMDs have been released at an even lower price point. The first phone-dependent HMD was the Google Cardboard

(Google, USA), but since other companies have released similar phone-dependent systems such as Zeiss VR One (Zeiss, Germany), GearVR (Samsung, Japan and Facebook, USA), and Immerse VR (Thumbs Up!, USA).

A summary of all the aforementioned HMDs and their respective technical specifications can be seen below in Tab. 1.1.

Table 1.1: Summary of technical specifications for current commercially available HMDs.

Parameter	Display Resolution	Video Resolution	Diagonal FoV	Max Refresh Rate	Weight (g)
OST-HMDs					
Hololens 1	1280*720	HD video	34°	60 Hz	579
Hololens 2	2048*1080	1080p30	52°	60 Hz	566
Google Glass 2	640*360	720p	30°	N/A	46
Magic Leap 1	1280*960	1080p	50°	60 Hz	316
VPT-HMDs					
VIVE Pro	1440*1600	640*480	110°	90 Hz	470
Valve Index	1440*1600	960*960	130°	144Hz	809
VR-HMDs					
Oculus Quest 2	1832*1920	N/A	100°	90Hz	503
PlayStation VR	1920*1080	N/A	100°	120 Hz	600
FOVE	2560*1440	N/A	90°	70 Hz	520
Parameter	Tracking method	Tracking DoF	Controllers	Untethered capabilities	Price for Dev (USD)
OST-HMDs					
Hololens 1	Inside-out	6 Dof	N	Y	3000
Hololens 2	Inside-out	6 Dof	N	Y	3500
Google Glass 2	Inside-out	6 Dof	N	Y	999
Magic Leap 1	Inside-out	6 Dof	Y	Y	2300
VPT-HMDs					
VIVE Pro	Outside-in	6 Dof	Y	Y	799
Valve Index	Outside-in	6 Dof	Y	N	999
VR-HMDs					
Oculus Quest 2	Inside-out	6 Dof	Y	Y	500
PlayStation VR	Inside-out	6 Dof	Y	N	300
FOVE	Inside-out	6 Dof	N	Y	600

1.10.5 Psychophysical aspects

Psychophysical aspects of HMDs describe the relationship between the physical stimuli produced by the HMD and its environment and the sensation and perception experienced by the user. These psychophysical effects are a product of the difference between how we normally perceive the real world and how the HMD simulates the 3D environment. The human visual perception system uses various methods and relies on cues to orient oneself and objects in 3D. These cues must be effectively simulated in the HMD such that the user can accurately determine the size, depth, and speed of an object in the virtual scene. The human visual system relies on binocular-, monocular-, eye-motion-, and object-motion-based depth cues to interpret how far objects are positioned in space [10]. The user's perception in the HMD is dependent on how accurately the system can simulate these cues. Patterson *et al.* summarized the various perceptual issues associated with HMDs such as binocular rivalry, restricted FoVs, and accommodation-vergence mismatch in their overview [113].

The FoV of the human eye is the area in which we can detect objects and movement. Humans have a slightly larger than 210° forward-facing horizontal arch-shaped FoV, where 120° of this field is binocular vision, where the eyes simultaneously obtain information from an overlapping FoV [10]. The remainder of the visual field is monocular, where the information from this region of each eye is not combined with information from the other eye [10]. HMDs typically have independent inputs into each eye, which disrupts the binocular processing that assumes an overlapping FoV and results in binocular rivalry where the eyes are in competition and one eye inhibits the visual processing of the other. Binocular rivalry results in optical misalignment, image distortion, unstable visual processing, and object or change blindness [113]. Binocular rivalry is less common in OST-HMDs and can be mitigated by implementing features that encourage binocular fusion. The restricted FoV common in HMDs typically only covers the binocular region, meaning little monocular information is processed resulting in a reduction in the processing of spatial information. Tab. 1.1 illustrates how the various commercial HMDs have a range of FOVs, which are typically less than 210° , especially for OST-HMDs. The larger the FoV the more natural the 3D scene seems to the user, as users have a greater sense of immersion and adequate visual function. Larger FoVs result in increased accuracy during simulated tasks, improved interpretation of object's spatial orientations, improved time per task, and less simulator sickness[113].

3D virtual scenes are rendered within the HMD by emitting light on flat surfaces (screens) at fixed distances. Accommodation is the process where the lens of the eye changes shape to alter the focal length and allow the eye to focus on objects at varied distances. Vergence is the change in pupil location to optimize binocular vision, when focusing on a close object the pupils rotate towards each other resulting in convergence, and when focusing on a far object

they rotate away from each other resulting in divergence. These two mechanisms typically work in junction to focus the eyes on an object. For VR-HMDs these screens are positioned very close to the eyes, resulting in a distance conflict between where the light is generated and where it intends to be simulated. This is known as the vergence-accommodation or distance conflict as the eyes accommodate to the screen distance while converging to the distance of the simulated object, which can affect the user's ability to perceive the size, depth, and distance of an object [113].

1.10.6 Surgical Navigation Systems

The use of HMDs to render surgical navigation information is not a novel idea and dates back to the 1960s. With the increased availability of commercially available HMDs (Tab. 1.1) it is more trivial than ever to develop systems that display the surgical navigation information into the HMD. The Hololens OST-HMD has been used to visualize systems that guide both percutaneous sacroiliac[155] and pedicle [44] screw insertions. The Hololens was also used to render the US-guidance information for fine-needle aspiration [68] and central line insertions [67] in the neck by displaying the US image just above the patient. Displaying the US image in the OST-HMD did not result in any significant differences as compared to traditional US-guidance [67]. Various other surgical applications of HMD-based systems are summarized in the review paper by Rahman *et al.* [127].

1.11 Limitations of current systems

As mentioned earlier in this chapter the current clinical standard for central line insertions is 2D US guidance. As this method maintains high rates of arterial punctures, more advanced surgical navigation systems have been developed. Ameri *et al.* [2] developed an augmented-virtuality (AV) monitor-based surgical navigation system for needle guidance. This system employed magnetic tracking to render tracked virtual models of the US probe, needle and needle trajectory, onto a front-facing US image. This system did not demonstrate significant improvement in the complications associated with the needle insertion compared to the US-only technique for expert users [2]. Chew *et al.* implemented a similar system, in which a magnetically tracked needle was used to superimpose the trajectory of the needle and marker's representing the actual needle position onto the front-facing 2D US image [28]. This system was compared to US-only guidance with 50 clinical insertions obtained using each approach [28]. The system developed by Chew *et al.* achieved the same success rates as the US-only system but had slower insertion times [28]. Two potential factors that may have influenced the success of

these systems were the fixed face-on 2D perspective provided to the user and the discrepancy between the visual and motor fields, as the user had to rely on a monitor exterior to the visual field of the phantom [2]. However, 3D US-based environments have been successful for other needle navigation procedures such as biopsy and aspiration [132, 68, 116]. Therefore, this thesis investigates the use of 3D US environments for the application of vascular access at the site of the IJV.

1.12 Thesis Outline

This thesis investigates the general research question “Do 3D ultrasound-based environments improve IJV needle insertion accuracy compare to the clinical standard of US-guidance?”. One focus of this work is the development and evaluation of components used within advanced 3D US-based environments for needle navigation. The second focus of this research is to validate the use of 3D advanced US-based surgical navigation systems for needle insertions at the IJV. A short summary of each chapter is provided.

1.12.1 Point-to-Line Ultrasound Probe Calibration with automatic fiducial localization

In this chapter, we developed and evaluated a point-to-line US probe calibration that used a CNN to localize the calibration fiducials (needle reflections) required for this calibration approach. The needle localization algorithm is capable of selecting the needle centroid within 0.25 mm of the manually labelled ground truth for a single US probe across 4-8 cm imaging depths. The point-to-line approach with automatic localization and the original manual formulations were found to have an average normalized point reconstruction absolute distance error less than 1.5 mm. The point-to-line calibration with both manual and automatic localizations were both found to be precise and accurate, with 1.18 ± 0.45 mm and 1.23 ± 0.47 mm being the largest normalized point reconstruction absolute distance errors, for the manual and automatic approaches respectively. The point-to-line calibration with automatic localizations produces calibrations with the same precision and trueness as the manual localization equivalent. This work improves the usability of the point-to-line calibration approach and provides the ability to visualize calibration US images in a common coordinate system with virtual models of tracked tools or anatomy.

1.12.2 Accuracy Assessment for the Co-registration Between Optical and VIVE Head-Mounted Display Tracking

The second chapter of this thesis focuses on the development and evaluation of a method to co-calibrate an OTS and the HTC VIVE Pro HMD. This chapter establishes and validates the co-calibration required to use the HTC VIVE Pro as the platform to visualize surgical navigation information. The goal is to develop and validate a method to register tracked surgical tools into an HMD with clinically acceptable accuracy. This co-calibration provides the ability for advanced surgical navigation systems to utilize this type of HMD as a visualization platform. It is shown that by using a co-calibration apparatus that is simultaneously tracked by the optical spatial tracker and controller associated with the HMD system, we can register spatially tracked tools into the HMD with sub-millimetre (0.3 ± 0.25 mm) and sub-degree ($0.06 \pm 0.11^\circ$) accuracy.

1.12.3 First-Person Perspective Ultrasound-Guided Central Line Insertions

The third chapter investigates how the disparity between the clinician's visual and motor fields during the central line insertion affects complication rates. A surgical navigation system was developed for this chapter, comprising a calibrated US image and models of the tracked surgical needle, needle trajectory and US probe. This system was displayed on a monitor in a 3D perspective, and in a first-person perspective in the HTC VIVE Pro HMD by means of the co-calibration developed in Chapter 3. The purpose of this chapter is to investigate if visualizing surgical navigation information in a first-person perspective during the central line needle insertion process provides superior targeting to the traditional US-guided insertion and an advanced monitor-based system. The HMD system provided superior targeting accuracy compared to the US-only approach for both expert users and medical students and produced insertions with higher success rates (94 %) than the US-guidance alternative (70 %).

1.12.4 Accuracy Assessment for the 3D Reconstruction of the Neck Vasculature from Transverse 2D Ultrasound Images

The objective of the fourth chapter is to develop and validate a method to obtain surface reconstructions of the IJV and CA in 3D. The purpose of this chapter is to verify the accuracy of the reconstruction, such that the 3D vasculature can be used in a clinical context. Our machine learning algorithm can delineate the CA and IJV from transverse US scans with a Dice Score of 0.88 and 0.90, respectively. Machine learning and tracking information are used to obtain a surface reconstruction of the neck from 2D US.

1.12.5 Effect of Using 3D Anatomical Information on a 2D Monitor on Complication Rates During Central Line Insertions

The objective of the fifth chapter is to develop and evaluate an advanced surgical navigation system that comprises the US-based vascular reconstructions and spatially tracked tools in a common coordinate system. The IJV is visualized using a Z-buffer to provide a more intuitive 3D navigation. The purpose of this work is to understand if proving this 3D visualization of the anatomy promotes superior navigation compared to the US-only approach while maintaining the display technology. It was shown that the advanced US reconstruction-based surgical navigation system had superior targeting accuracy and success rates compared to the US-only approach with a clinically feasible system.

1.12.6 Conclusions

The conclusions chapter summarizes the contributions of this thesis. This chapter also discusses future research directions relating to the various areas explored in this thesis and the field as a whole. matintai

Chapter 2

Point-to-Line Ultrasound Probe Calibration with automatic fiducial localization

This chapter is adapted from the following manuscript:

- Groves, L. A., VanBerlo, B., Peters, T. M., & Chen, E. C. (2019). Deep learning approach for automatic out-of-plane needle localisation for semi-automatic ultrasound probe calibration. *Healthcare technology letters*, 6(6), 204.

2.1 Introduction

Ultrasound (US) imaging is commonly used to guide procedures such as percutaneous needle insertions [30] as it is a portable imaging modality that provides non-invasive real-time imaging with the capability to depict deep tissues without using ionizing radiation. Spatial tracking systems have been used to enhance 2D US-guidance by integrating visual representations of tracked surgical instruments into the US environment [30]. Spatial tracking systems can also be used to obtain 3D US volume reconstructions to assist with surgical navigation [30]. Generating 3D US reconstructions and integrating externally tracked tools into US environments rely on accurate US probe calibration to establish the spatial relationship between the US image and a tracking sensor fixed to the transducer [63]. A calibrated US probe registers the US image to the tracker's coordinate system positioned and scaled to the true field of view of the US beam [63]. In general, US calibrations are modelled using a homogeneous 4×4 rigid or affine transformation matrix that comprise of any combination of scaling, rotation and translation in up to three directions (X, Y, and Z) [63]. Assuming the 2D US image lies on the X-Y plane (i.e.

$Z=0$) in its local coordinate system, this matrix can be applied to each pixel within the image, $[X, Y, 0]$ to produce a new location $[X', Y', Z']$ within the tracker's coordinate system. US probe calibration relates US images and other tracked tools in a common frame of reference [63].

The majority of US calibrations rely on fiducial-based methods and some common approaches use point-target, cross-wire, multiple-wire, or needle-based phantoms as summarized in Section 1.5. These phantoms are equipped with a tracking pose sensor and require phantom calibration such that the position(s) of the calibration fiducial(s) are known with respect to the phantom sensor [63]. This phantom calibration is typically achieved using a fiducial-based approach that requires the extraction of homologous fiducial marker pairs from a phantom CAD model or CT volume and from the physical phantom using a tracked stylus to digitize the markers with respect to the phantom sensor, such that the paired fiducials can be used to achieve a fiducial-based registration that provides the relationship between the phantom sensor and the phantom geometry [63]. The output of the phantom calibration process is the transform $\text{PhantomModelToPhantom}$, which can be concatenated with the transform phantomToProbe , such that the phantom geometry can be directly tracked with respect to the probe sensor ($\text{phantomModelToProbe}$), as depicted in Fig. 2.1 a). To perform the US calibration, the fiducials within the calibration phantom are imaged by the US probe. The positions of the fiducials are known with respect to the US probe sensor via the phantom calibration process described above, and the reflections of the fiducials are extracted from the US image resulting in corresponding point pairs. A second fiducial-based registration is applied to solve the calibration and provide the relationship between the US probe sensor and US image [63].

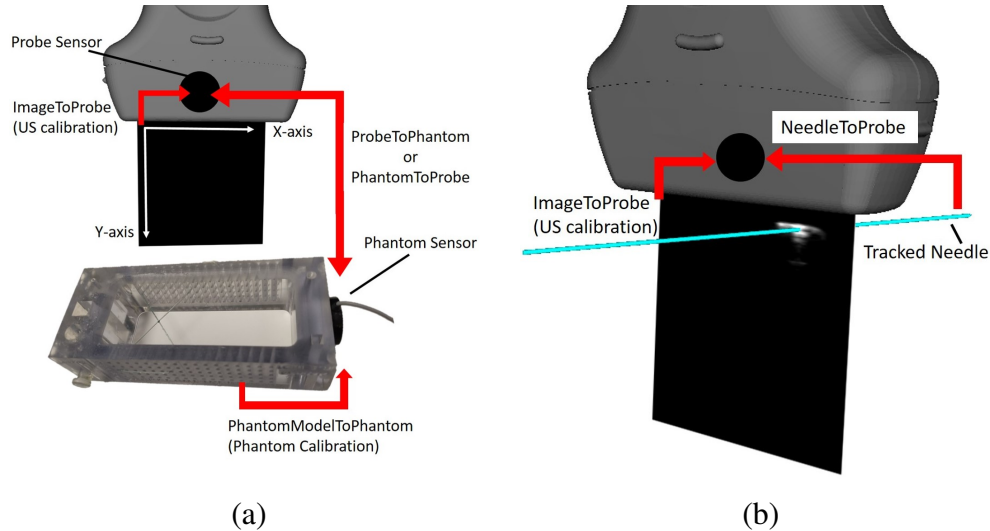


Figure 2.1: The transformations required for two US calibration methods, where a) depicts the required transforms for a cross-wire calibration that requires phantom-calibration and b) illustrated the transforms required for the point-to-line calibration demonstrating the reduction in transforms that may propagate error.

The fiducial-based registrations required to calibrate the phantom and perform the US calibration have associated errors that affect the accuracy of the US calibration. Fiducial-based registrations are affected by fiducial localization error (FLE), which is defined as the error associated with localizing a fiducial or landmark in the tracker or US coordinate systems [30]. Fiducial registration error (FRE) is a consequence of FLE and is defined as the error between the corresponding landmarks after fiducial-registration [30]. Within the phantom calibration, the FLE is quantified by the difference between the localized fiducial coordinates extracted from the phantom model and digitized with respect to the phantom sensor and the true positions of these fiducials, which affects the accuracy of the fiducial-based phantom calibration and results in FRE. The FRE affects how accurately the phantom calibration fiducials are positioned in the tracker's coordinate system, and thus the positions of these fiducials have an associated FLE. The fiducial reflections within the US image are also localized, either automatically or manually, and FLE in the X and Y directions are calculated based on the differences between the selected and true pixel coordinates representing the centre of the landmark [63]. Since the US beam also varies in thickness in the Z-direction as depicted in Fig. 2.2, the FLE can also have a Z component. Calibration fiducials are often small targets, and to accurately capture their positions in the US image, they must be imaged at the mid-plane of the Z-axis of the 3D beam [22]. The elevation resolution associated with US images is low relative to the axial and lateral directions, and thus US images have a fairly large point spread function in the Z-direction, as discussed in Section 1.6. When an object enters the Z-plane of the 3D US beam it may appear in the image despite not being at the mid-plane [22]. Thus, any difference between the true position of the object and the position of the mid-plane of the US image introduces FLE in the Z-direction. To achieve an accurate US calibration all sources of error that propagate into the calibration should be minimized.

Chen *et al.* [21] developed a point-to-line US probe calibration approach that aimed to minimize the accumulation of errors that affect the final calibration accuracy, by using a tracked needle as the calibration phantom, as depicted in Fig. 2.1 b). The 18-Gauge tracked needles (NDI, Canada) employed here incorporates a tracking sensor into the tool-tip by design, and thus does not require a phantom calibration, minimizing the FLE associated with the position of the calibration fiducials collected with respect to the probe sensor. The tracked needle is inserted into the US-beam such that it intercepts the entire US-beam, eliminating the requirement for precise placement of the phantom in the mid-plane, and improving the FLE in the Z-axis (Fig. 2.2). The centroids of the reflections produced from the needle are manually segmented, which provides a 3D image coordinate (point) which is used as one of the inputs to the registration algorithm [21]. The second input is the line formed by the needle, defined using the tip location as the origin and the direction of the needle which can be extracted from the needle's

transformation matrix from the tracker [21]. The pixel locations and the lines obtained from the needle tracking with respect to the tracker on the US probe are input to the Anisotropic Orthogonal Procrustean Analysis (AOPA) formulation developed by Chen *et al.* [21] to solve the point-to-line registration. AOPA is an iterative solution, where for each measurement (point and line pair) input to the algorithm, a new transformation matrix built upon the previous solution is obtained [21]. The output of this approach is a calibration represented as a 4×4 transformation matrix comprised of anisotropic scaling, rotation, and translation [21]. Chen *et al.* [21] found that this method converges to a consistent solution after collecting 12 to 15 measurements, with the majority of data collected in the four corners of the image. The point-to-line registration further minimizes the FRE by solving for anisotropic scaling, as for US images the pixel scaling in the X and Y directions are independent of each other. The accuracy of the point-to-line calibration approach has been validated using a theoretical estimate of target registration error (TRE). The requirement for users to manually localize the fiducials in the US image results in the propagation of FLE into the US calibration. Due to the effort to minimize the remainder of errors associated with the calibration, the accuracy of the calibration is highly dependent on the user's ability to localize the needle reflections in the US image.

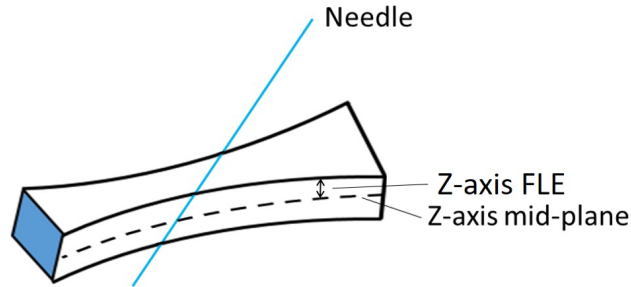


Figure 2.2: This figure depicts how the point-to-line calibration approach mitigates issues with mid-plane alignment, and the FLE in the Z direction and the mid-plane are labelled on the diagram.

The precision of the AOPA point-to-line calibration approach was assessed across 30 novice users and an expert for a single linear US probe at 6 cm depth [51]. A novice user was defined as someone without any experience in US calibration as they performed their first US calibration during this study. The expert user was someone who performed over 100 calibrations using this approach. The novice users obtained 3 calibrations each and the expert obtained 30 calibrations [51]. On average the novice users completed their point-to-line calibrations in 3.2 ± 0.9 minutes [51]. Each calibration obtained by the novice user's on their third attempt and the repeated calibrations obtained by the expert were applied to a plane the size of the US image to transform it in 3D space. These calibrated planes were overlaid with the other

calibrations obtained by the same group (novice and expert) to provide a visual example of the precision of the point-to-line calibration, as depicted in Fig. 2.3 a) and b), respectively. In this graphic a larger spread between image planes indicates less precision between calibration matrices, illustrating that the expert user produces more precise calibrations compared to the novice cohort [51]. The increased variation across the novice user calibration results is likely a product of an increased FLE compared to the expert, as the point-to-line calibration was designed to minimize the propagation of the other errors that affect the calibration. The requirement for users to manually select the calibration fiducial produced by the needle is a limitation and one of the remaining sources of error that can be minimized in this point-to-line calibration approach.

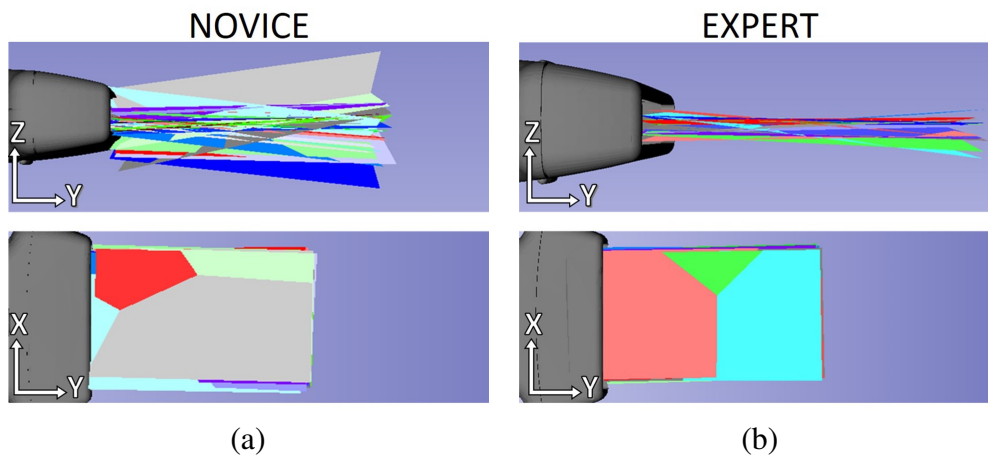


Figure 2.3: Both graphics depict 30 calibrated image planes where sub-figure a) shows the variation in calibrations obtained by 30 first-time users, and sub-figure b) depicts the variation in the 30 calibrations obtained by an expert user. Less variations between the calibrated image planes represents more precise calibrations.

To further improve the accuracy of the point-to-line calibration approach, the first objective of this work is to develop a machine learning algorithm to automatically localize the centroids of the needle reflections used within the calibration, while the second is to integrate the automatic fiducial localization algorithm into the point-to-line US calibration approach. The final objective is to validate the trueness and precision of the point-to-line calibration, with both manual and automatic fiducial localizations, using real data rather than theoretical estimates.

2.2 Methods

Towards providing a consistent and automatic fiducial segmentation for the point-to-line US probe calibration we developed a convolutional neural network (CNN) to automatically local-

ize the needle reflections used for this calibration approach for a single linear US probe. We validated this algorithm using test and validation sets partitioned from the data collected for training, as well as on images used to obtain US calibrations using the point-to-line approach. The point-to-line US calibrations obtained using both the manual and automatic fiducial localizations were analyzed for trueness and precision.

2.2.1 Equipment and Data Collection

The equipment used for this work included the Ultrasonix scanner (BK Ultrasound, US) with the L14-5 probe, the NDI Aurora Desktop magnetic tracking system (Northern Digital Inc., CA), one NDI Aurora 6 degree-of-freedom (DoF) flat disk reference sensor, and an NDI Aurora 5 DoF 18-gauge needle. The 6 DoF reference sensor was affixed to the US probe. The sensor of the tracked needle was embedded directly into the tool-tip and the needle-tip was tracked with respect to this 6 DoF reference sensor fixed to the US probe (NeedleToProbe), as depicted in Fig. 2.1 b). For all data collection, a mechanical arm with a clamp was used to stabilize the tracked US probe within a water bath, as depicted in Fig. 6.6. The Plus Server [78] was used to stream the tracking and image information into 3D Slicer [70] at 60 frames per second.

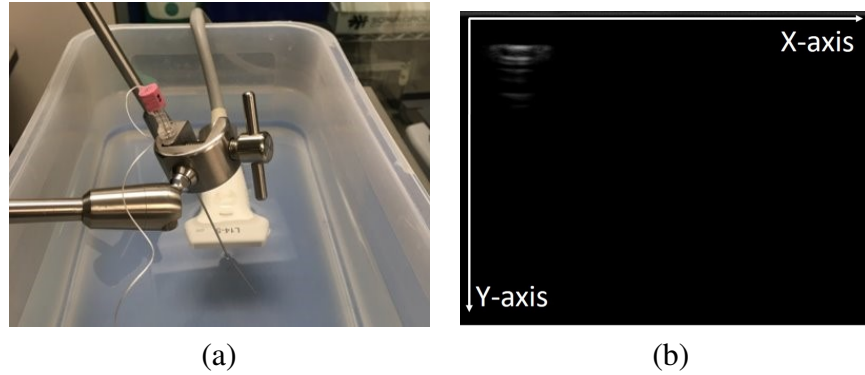


Figure 2.4: Figure a) illustrates the configuration used for calibration, where the tracked US probe is clamped and submerged in water and the tracked pre-calibrated needle is inserted at an oblique angle into the US beam stabilized with a mechanical arm. Figure b) is a representative image of the needle reflections produced using this set-up.

The data used to train this algorithm were derived from 10 tracked recordings in total, with 2 tracked recordings collected for each imaging depth in the 4 cm to 8 cm range. For each recording, the needle was inserted into the US beam at an oblique angle with respect to the US image plane using a mechanical arm as depicted in Fig. 2.4. The user traversed the mechanical arm horizontally and the clamp maintained the angle of the needle. Once the needle reached a horizontal image boundary, the user shifted the needle down to produce a reflection deeper

within the image and traversed the needle in the opposite horizontal direction (Fig. 2.5 a). The user repeated this process until they reached the bottom of the image and then repeated the process moving the needle from the bottom of the image back to the top. The entire process was then repeated with the needle inserted at the opposite oblique angle (Fig. 2.5 b).

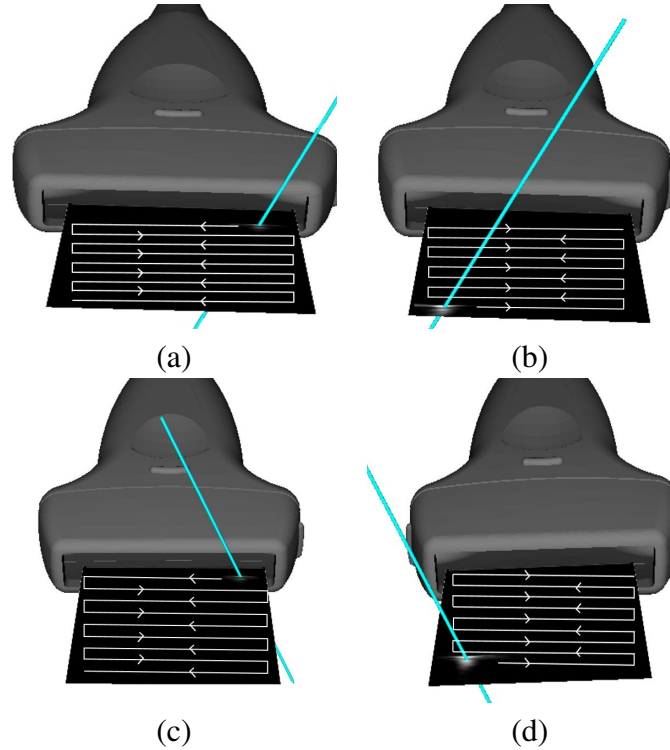


Figure 2.5: This figure depicts the needle path used for data collection, with graphic a) illustrating the path from the top to bottom of the image with the initial needle orientation, graphic b) shows the needle at a similar angle to the needle graphic a) with a bottom to top outlined path, graphic c) has a similar path to the needle in graphic a) but the needle is inserted at the opposite angle, and graphic d) shows the needle path from the bottom to top of the image maintaining a similar angle to the needle in graphic c).

For each image, the pixel coordinate that represented the centroid of the needle reflection was manually localized by an expert. The data sets were cleaned up to remove any images that did not contain the needle, such that there were 1800 images along with their respective centroid positions for each depth (9000 in total). All of the images within the array were normalized such that the pixel intensities were represented on a continuous $[0, 1]$ range, by dividing each image by the largest grey level within the image. The centroid locations were also normalized such that they were on a continuous $[0, 1]$ range.

2.2.2 Machine learning-based needle segmentation

Automatically localizing the centroid of the US needle reflections used for the point-to-line calibration (Fig. 2.4 b) is a type of key-point localization problem. A facial key-point detection algorithm [84] motivated the architecture presented in this paper. A CNN was designed to receive an US image as input and predict coordinates over a continuous $[0, 1]$ range. An iterative training and testing process was implemented to assess which architecture worked best for the required task. The final network consists of 5 alternating convolutional and max-pooling layers, followed by 5 fully connected layers. The output layer is a fully-connected layer with 2 units, which outputs the regressed coordinates of the centroid of the US needle on the range $[0, 1]$. The depth of the convolutional layers increases with each successive layer. The number of filters in the first layer is 16, and this number doubles with each successive convolutional layer. All convolutional layers use filters of size 3×3 and a stride of 1. The initial parameters of the convolutional layer were determined using Glorot uniform initialization [48]. All max-pooling layers use a pool size of 2×2 and a stride of 1. The activation function used in all pre-output layers is the Leaky Rectified Linear Unit (ReLU) function with an alpha of 0.1, which was chosen over ReLU because it converges faster during training [86]. The output layer predicts the centroid of the needle in the image, whose values are bounded by $[0, 1]$. The output values ($[X, Y]$) are reconverted to pixel units by scaling the first and second output values by the number of pixels in the X and Y dimensions, respectively. A diagram of the model's specific architecture is depicted in Fig. 2.6.

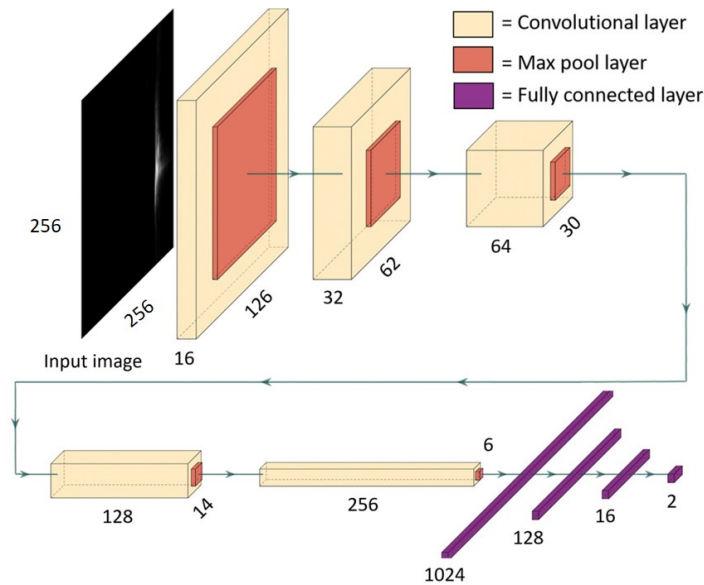


Figure 2.6: Architecture of the neural network model used to predict the coordinates of the centroid of an US needle reflection.

To validate this machine learning model, a 4-fold cross-validation analysis was performed. We collected 9000 images across imaging depths 4 cm to 8 cm at 1 cm increments, such that there were 1800 images for each depth. Manual localization of the needle reflections were performed by an expert for every image in the set ¹. The imaging depth set for the scan has a direct effect on the image size, and therefore all images were reshaped to 256×256 for consistency. To form the test set, 900 randomly sampled images were extracted from the whole data set. These images and their centroid pixel coordinates were set aside and not used within any training or validation sets. From the 8100 remaining data points, 900 were randomly selected to form the validation set. The remaining 7200 images were used as the training set. This process was repeated 4 times to obtain four training sets with entirely unique validation sets, as once an image was used within a validation set it was not allocated to another. After obtaining the unique training, test, and validation sets, all images were reflected about the horizontal centre of the image, which doubled the number of images in all of the sets. The training sets, denoted A, B, C, and D, were used to obtain 4 trained networks using the four unique training sets to obtain a four-fold cross validation analysis. For all training, the batch size was 32 and the model was trained for 150 epochs to minimize the mean absolute error loss function. We employed the Adam optimization algorithm [72], with $\alpha = 1 \times 10^{-4}$. To reduce overfitting, L2 regularization with $\lambda = 1 \times 10^{-5}$ was applied in all convolutional layers. These hyperparameters were determined in an iterative heuristic manner. One hyperparameter was varied at a time while all others were held constant.

The output of the algorithm produces two values $[X, Y]$ on a continuous $[0, 1]$ range, which provides the X and Y locations of the centroids following re-scaling to the image dimensions. After training, the validation set that corresponded to the trained model and the test set which is consistent for all folds were evaluated using the mean and standard deviation of the absolute error, where the absolute error is the absolute distance between the manual and automatic centroid positions. The absolute error in the X and Y directions for each test and validation set and the average value across the tests and validation sets produced from all four-folds are summarized in Table. 2.2.

2.2.3 US calibration with automatic fiducial localization

The trained CNN automatically localizes the centroid of the needle reflection given an US image, and this algorithm replaced the manual needle segmentation required for the original point-to-line calibration [21]. The trained CNN model that was integrated into the point-to-line calibration algorithm, was the model trained on set C as this set produced the best accuracy

¹An expert user in this calibration approach was defined as someone who had performed over one hundred point-to-line US calibrations

from the four-fold analysis. Thus, we refer to our approach as point-to-line calibration with automatic localization, and the manual approach is referred to as point-to-line calibration with manual localization. The point-to-line calibration approach requires a recorded or live tracked US scan of a tracked needle in a water bath. The user selects the input image, which also provides the associated needle tracking transform, by pausing the US stream. The user then clicks a button to apply the trained CNN to the US image to obtain the centroid. The centroid coordinates obtained from the CNN, a 3×1 vector representing the origin of the needle, and a 3×1 vector representing the direction of the needle extracted are input to the AOPA point-to-line registration algorithm developed by Chen *et al.* [21]. The user repeats this process 12 to 15 times to obtain the calibration, and can visually validate their calibration by assessing the alignment between the needle model and the needle reflection in the calibrated US image.

2.2.4 Quality Assessment

To assess the quality of the original point-to-line calibration method with both manual and automatic localizations, three separate tracked US recordings were collected for 4, 6 and 8 cm depths (9 scans in total), using the same data collection approach outlined in Fig. 2.5. These imaging depths were selected to show the results across 3 of the 5 depths used for training. These recordings were not used to train the needle detection algorithm. Prior to collecting the quality assessment recordings, the US stream and tracking data were temporally calibrated using the temporal calibration feature available through the Plus Toolkit [78], as described in Section. 1.5.2. The following methods were performed for all three imaging depths. For each collected scan, 5 calibrations were obtained using the point-to-line calibration approach with both manual and automatic localizations.

Chen *et al.* recommends that the majority of the US calibration fiducials used within the point-to-line calibration are collected in the four corners of the image. For each calibration, the expert user selected eight images that contain needle reflections close to the corners, and for each corner two images were selected with the needle at opposite oblique angles. The remaining images were collected with the needle reflection positioned throughout the centre of the image. Once an image was used within a calibration, the time stamp was recorded and that image was not used within the other calibrations obtained from the same scan, such that the 15 calibrations obtained for each imaging depth use 15 unique images. For each calibration, the same 15 images and their respective tracking transforms were used to obtain calibrations with manual and automatic localizations. For each image, the expert user manually selected the centroid of the needle reflection and in the background, the CNN automatically selected the centroid for the same image. Thus, the only difference in the calibration process was centroid

positions obtained by manual or automatic localization. The automatic and manual centroid positions and the final transformation matrix obtained using the manual and automatic centroid positions were recorded. This process allowed for the centroid positions and transformation matrices to be directly compared between calibrations obtained with automatic and manual localizations.

The manual and automatic centroid positions obtained during the calibrations resulted in 225 point pairs per depth. The mean, standard deviation, 95 % confidence interval, maximum, and minimum values for the absolute distance between the manual and automatic centroid coordinates were calculated and are summarized in Table. 2.3. The manual and automatic centroid positions were obtained in pixel coordinates and were converted to millimetres. To obtain the scaling values to perform this conversion, the X and Y scaling parameters were extracted from the 30 calibrations (manual and automatic) obtained for each imaging depth. From each calibration matrix, the inner 3×3 rotation and scaling matrix was isolated by removing the fourth row and column. The scaling factors were extracted by applying singular value decomposition to this 3×3 matrix, where the singular values are the anisotropic scaling factors. The average X and Y scaling values were calculated across all the scaling factors extracted from the 30 calibrations obtained for each imaging depth, and the mean and standard deviations for the scaling values from each imaging depth are summarized in Tab. 2.1. These scaling values were applied to the pixel-based coordinates to convert them to millimetre units.

The precision of the point-to-line calibration algorithm was quantified using calibration reproducibility, which is a standard method for measuring calibration precision [63]. Calibration reproducibility is measured by applying N calibrations obtained using the same calibration method to one or more pixel coordinates, and it is recommended to use multiple pixel locations such as the four corners of the image as targets [63]. The calibration reproducibility was measured independently for calibrations obtained using the manual and automatic localizations across the three imaging depths. All calibrations obtained using the same localization method and imaging depth were applied to the pixel coordinates that represent the four corners and the centre of the images produced for that particular imaging depth. The image shapes in pixels are provided for these imaging depths for the X and Y axes independently in Table 2.1. Applying the 4×4 homogeneous calibration matrix to a pixel coordinate represented as $[X, Y, 0, 1]$ produced a new coordinate $[X', Y', Z', 1]$, and repeating this process for the 5 pixel targets across the 15 calibrations obtained for each imaging depth and localization approach produced 5 3D point clouds, where the spread of the point cloud represents the precision of the calibration approach [63]. To right-hand the spread across all points, the 5 point clouds were shifted by their means, such that all points are centred about the $[0, 0, 0]$ origin. This process produced point clouds with 75 samples for both the automatic and manual localization methods for each

imaging depth. The standard deviations across the 75 mean shifted points for each approach were calculated to measure the calibration reproducibility. A two-sided F-test was applied to compare the variances across the 75 mean shifted points centred about $[0,0,0]$ that were obtained using the manual and automatic localization methods to right-hand if the automatic localization was capable of producing calibrations with the same precision as expert manual localization. The standard deviation of the reconstructed point clouds and the p-value associated with the F-test are summarized in Table 2.5. The 3D mean shifted point clouds were also plotted graphically in Fig. 2.11, these plots also include the 95 % prediction interval represented as an ellipsoid. Additionally, a visual representation of the precision is also depicted in Fig. 2.10, the graphics in this figure are produced by applying all the calibrations obtained for each approach and each imaging depth to a plane the size of the US image produced for that imaging depth. The output of this process is 15 calibrated image planes that provide a visual representation of the variation of point-to-line calibrations, where a greater spread amongst planes indicates lower precision in the calibration approach.

Table 2.1: The number of pixels that comprise the X and Y axes of the US image for 4, 6 and 8 cm, as well as the mean and standard deviation for the scaling factors extracted from all calibrations produced for 4, 6 and 8 cm.

Depth	Pixel # X-axis	Pixel # Y-axis	Scaling X-axis	Scaling Y-axis
4 cm	568	598	$0.068 \pm 9.43 \times 10^{-4}$	$0.066 \pm 5.44 \times 10^{-4}$
6 cm	374	589	$0.10 \pm 8.52 \times 10^{-4}$	$0.10 \pm 7.50 \times 10^{-4}$
8 cm	280	589	$0.14 \pm 2.40 \times 10^{-3}$	$0.13 \pm 1.10 \times 10^{-3}$

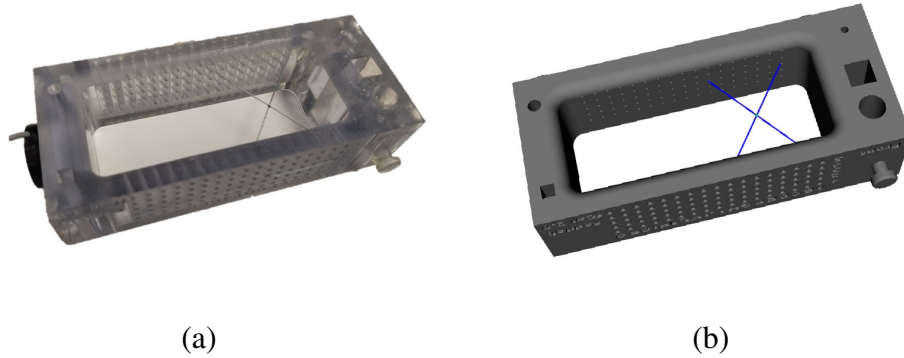


Figure 2.7: The cross-wire phantom used within the point reconstruction analysis with a) illustrating the CAD model of the phantom overlaid with the wires and point target, and b) depicting the real phantom with the cross-wire and tracking sensor.

The calibration quality was assessed using point reconstruction accuracy, such that trueness and precision were quantified independently across the 15 calibration matrices obtained by the two localization approaches for each imaging depth. Proper quantification of the accuracy of a calibration approach requires a known ground-truth, yet there can never be a ground-truth for calibration, as if we knew there was a method to obtain an accurate calibration, all other methods would be obsolete, and that method would be used for calibration. One commonly used method to assess calibration accuracy is point reconstruction in 2D and 3D, which is used to quantify the target localization error (TLE) [63]. Point reconstruction accuracy analysis requires imaging one or more point-source(s) with known positions in the spatial tracker's coordinate system, such that the point fiducial is the target. When the point source is reconstructed using a calibrated US probe, the difference between the reconstructed location and the known location in the tracker's coordinate system can be used to quantify the TLE associated with the calibration. To validate the accuracy of the calibration approach, a cross-wire phantom was used as the point source, the CAD model and the real phantom are depicted in Figure. 2.7. The cross-wire phantom contained 8 carefully milled 6.35 mm divots that were digitized with respect to the sensor fixed to the phantom using a pivot calibrated stylus [163] with a conforming 6.35 mm diameter tip. The positions of the divots were extracted from the CAD design, and the error was minimized by carefully milling the divots to mitigate any errors within the 3D printing process. A fiducial-registration was applied to solve for the relationship between the phantom's and the tracker's coordinate systems using a least-squares solution [5]. This process achieved an RMS error (FRE) of 0.35 mm, and can consistently demonstrate errors less than 0.5 mm. The position of the cross-wire junction is known within the phantom space by the construction of the phantom. The cross-wire point target position was transformed to tracker space by applying the phantom calibration to provide the ground truth.

The cross-wire phantom was imaged under US using the following scanning protocol for the 4, 6 and 8 cm imaging depths. First, the phantom was clamped vertically in a water bath, such that the tracking sensor was not submerged. The US images, and tracking transforms for the probe and phantom, were streamed and recorded at 60 frames per second using the Plus Server [78]. The cross-wire was imaged by aligning the probe with the location where the wires were anchored to the phantom, such that they appear as distinct and separate reflections. The US probe was slowly traversed across the wires, such that the reflections of the wire become progressively closer, overlap at the junction, and then become progressively wider (Fig. 2.8). To extract the cross-wire position from the US images, each recording was replayed frame-by-frame until the frame where two wires overlap was identified. The pixel coordinate representing the cross-wire centroid was manually extracted from the 2D US image from that frame. This scanning protocol and manual segmentation process were repeated three times per

imaging depth.

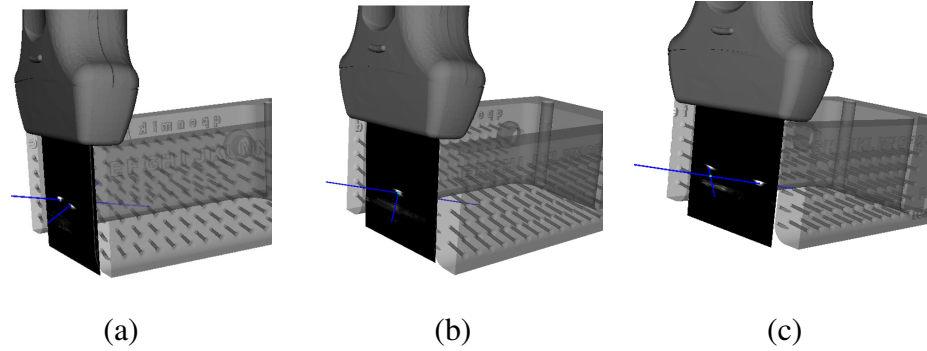


Figure 2.8: Sample images from the scanning protocol, where figure a) shows the US image with the two wire reflections prior to the cross-wire junction, figure b) shows the cross-wire junction, and figure c) shows the two wire reflections subsequent to the cross-wire junction.

The phantom target is known with respect to the tracker's coordinate system. The pixel coordinates representing the cross-wire junction were transformed to obtain their 3D position with respect to the tracker's coordinate system by concatenating the pixel coordinate, an US calibration transform, and the tracking transform that provides the position of the US probe with respect to the tracker's coordinate system. Thus, the US reconstructed point and phantom point target are registered to a common coordinate system. For each imaging depth, the cross-wire was imaged so that the reflection appeared in three different vertical locations (top, middle and bottom of the US image) with all of the reflections obtained near the lateral centre. For each image of the cross-wire, the pixel representing the centre of the cross-wire in the US image was extracted, along with the associated probe to tracker transform. The cross-wire pixel coordinates were transformed using each of the 15 US calibrations obtained for the corresponding imaging depth using both the manual and automatic localization methods and their associated probe to tracker transforms. For each pixel coordinate, the probe tracking transform was held constant such that only the US calibration changed. For each pixel coordinate 15 US reconstructed points were obtained for both calibration approaches (manual and automatic localization), producing 45 reconstructed points per imaging depth per calibration method. The absolute distance between each US reconstructed point and the corresponding ground-truth phantom coordinate were calculated. The mean and standard deviation for the X, Y, and Z axes, and the normalized distance ($\sqrt{X^2 + Y^2 + Z^2}$), as well as the maximum and minimum normalized distances for the manual and automatic calibration methods, were calculated independently to quantify the TLE. A two-sample paired T-test was used to compare the mean absolute error values for the point reconstruction error distributions obtained using the point-to-line calibration with manual and automatic localization. These metrics were confirmed to be normally distributed using the Anderson-Darling test. The 45 absolute error values obtained

for both the manual and automatic localizations were plotted in 3D along with the 95 % prediction interval represented as an ellipsoid for each imaging depth are depicted in Fig. 2.13. Fig. 2.12 provides a sample graphic illustrating the alignment between the cross-wire phantom and calibrated US images for the three imaging depths.

2.3 Results

2.3.1 Automatic Needle Localization Accuracy

Tables 2.2 and 2.3 contain the accuracy results for the CNN developed to automatically localize the needle centroid. Table 2.2 contains the results from the four-fold analysis and provides the mean and standard deviation for the absolute distance error between the automatic and manual pixel coordinates, for the X and Y axes independently. Table 2.3 presents the absolute difference between the manual and automatic pixel coordinates obtained during the 15 calibrations obtained per imaging depth, resulting in 225 point sets per imaging depth. This provides an assessment of how accurate the automatic centroid localization is on new scans collected for calibration independently for 4, 6 and 8 cm imaging depths. Sample centroid localizations obtained using the CNN are depicted in Fig. 2.9.

Table 2.2: The absolute pixel distance between the manual and automatically localized needle centroids obtained from the four trained networks for the X- and Y-axis, respectively. This table reports the mean and standard deviation, for the absolute error across the test and validation sets for each fold and the average across the four-folds.

Absolute Error (pixels)	Validation Set	Validation Set	Test Set	Test Set
	X-axis	Y-axis	X-axis	Y-axis
	Mean \pm std	Mean \pm std	Mean \pm std	Mean \pm std
Set A	1.7 \pm 1.9	1.2 \pm 1.9	1.8 \pm 1.4	1.2 \pm 1.3
Set B	1.6 \pm 2.1	1.7 \pm 2.2	1.7 \pm 1.7	1.7 \pm 1.3
Set C	1.3 \pm 1.7	1.1 \pm 2.2	1.2 \pm 2.1	1.1 \pm 2.4
Set D	1.4 \pm 1.9	1.6 \pm 2.1	1.7 \pm 1.5	1.6 \pm 1.6
Average	1.5 \pm 1.9	1.4 \pm 2.1	1.6 \pm 1.7	1.3 \pm 1.7

Table 2.3: The absolute distance between manual and automatically localized centroid coordinates obtained during calibration. For each imaging depth, the mean, standard deviation, 95 % confidence interval, and the maximum and minimum absolute distances were calculated in millimetres across the 225 point pairs.

X-axis Absolute Error (mm)	Mean \pm std	CI (95 %)	Maximum	Minimum
4 cm	0.10 \pm 0.06	[0.08, 0.10]	0.2	>0.1
6 cm	0.15 \pm 0.08	[0.13, 0.18]	0.3	>0.1
8 cm	0.21 \pm 0.12	[0.18, 0.23]	0.4	>0.1
Y-axis Absolute Error (mm)	Mean \pm std	CI (95 %)	Maximum	Minimum
4 cm	0.10 \pm 0.06	[0.11, 0.13]	0.2	>0.1
6 cm	0.15 \pm 0.09	[0.12, 0.17]	0.3	>0.1
8 cm	0.22 \pm 0.11	[0.18,0.25]	0.4	>0.1

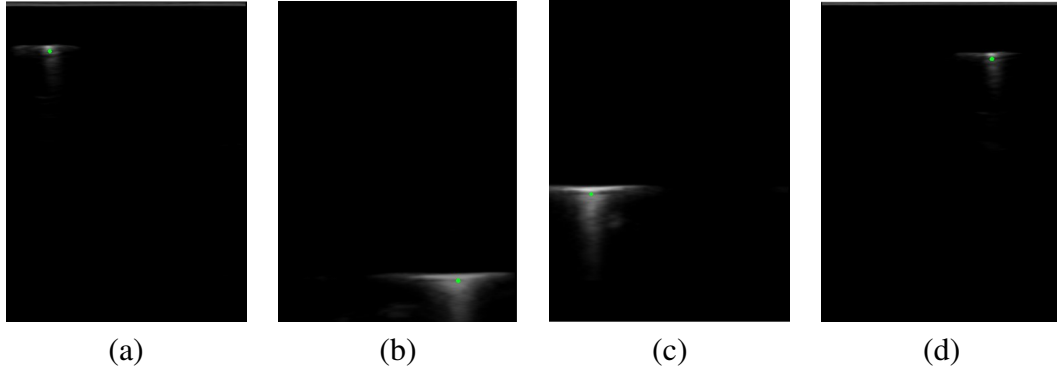


Figure 2.9: Four sample US images overlaid with the automatically localized centroid positions obtained from our CNN trained on set C.

2.3.2 Quality Assessment

Precision Analysis

Calibration reproducibility was used to quantify the precision, where the 15 calibrations obtained using the point-to-line calibration method with both automatic and manual localization were applied to the five pixel coordinates representing the four corners and centre for each imaging depth, as described in Section. 2.2.4. The standard deviation across the mean shifted point cloud and the Z-test results that compared the variance of the manual and automatic point clouds are summarized in Table 2.4. These results are also provided graphically in Fig. 2.11 using a 3D scatter plot along with the 95 % prediction interval represented as an ellipsoid. The prediction interval was calculated for the X, Y and Z axes independently. The bounds of the

ellipse are defined by the resulting prediction intervals. It should be noted that the ellipse overlaid on the graphic is a statistical measure of the range in which future samples are likely to be within and not a model that fits the data. A visual representation of the precision is also provided in Fig. 2.10, where the resultant calibrations were applied to a plane with the same dimensions as the US image produced for the specific imaging depths to show the spread of calibrated image planes.

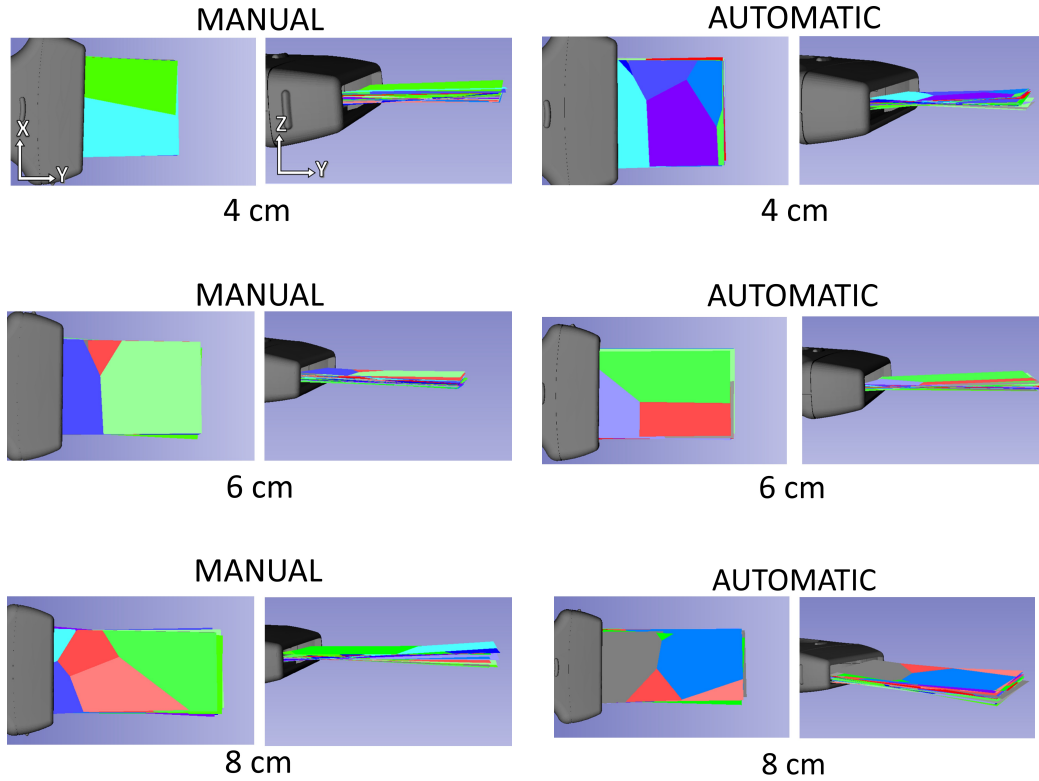


Figure 2.10: Visual representation of the calibration precision by applying all transformations obtained during calibrations with both manual and automatic calibrations to a plane the size of the US image and overlaying them in space.

Table 2.4: The standard deviations are reported for the 75 mean shifted reconstructed pixel locations obtained for each imaging depth and both calibration approaches. The p-value from the Z-test that was performed to compare variances between the two distributions produced using manual and automatic localizations are also included.

Std	X-axis (mm)			Y-axis (mm)			Z-axis (mm)			Norm (mm)		
Depth	Man	Auto	p-val	Man	Auto	p-val	Man	Auto	p-val	Man	Auto	p-val
4 cm	0.57	0.52	0.87	0.48	0.52	0.47	0.90	0.91	0.85	0.56	0.59	0.65
6 cm	0.32	0.29	0.45	0.36	0.38	0.61	0.84	0.88	0.69	0.51	0.53	0.81
8 cm	0.44	0.47	0.53	0.87	0.80	0.48	1.18	1.17	0.95	0.85	0.93	0.48

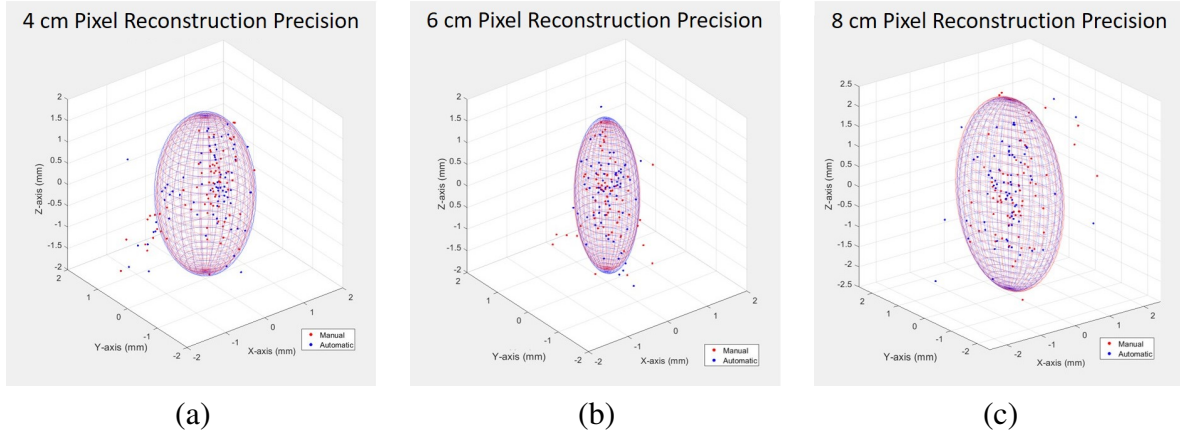


Figure 2.11: The calibration reproducibility across the 75 mean shifted reconstructed pixel locations obtained for both calibration approaches, where a larger spread between samples indicates less precision. The 95 % prediction interval bounds are provided for the X, Y, and Z axes as an ellipsoid.

Point Reconstruction Accuracy

Calibration accuracy was analyzed using a cross-wire point reconstruction approach as described in Section 2.2.4, where three points were compared per calibration. The absolute distance between the ground-truth point and the reconstructed point obtained using the calibration were calculated. The mean, standard deviation were calculated for the absolute distance in the X, Y and Z directions, and the mean, standard deviation, maximum and minimum were calculated for the normalized absolute distance, these results are summarized in Table 2.5. These results are also presented graphically in Fig. 2.13, where the absolute errors were plotted along with the 95 % predictions interval represented as an ellipsoid. Fig. 2.12 provides visual representations of the point reconstruction analysis for all three imaging depths, showing the phantom target overlaid on the US images, the calibrated US images in a common coordinate system with the cross-wire and aligned with the point target, and the calibrated US images in a common coordinate system aligned with the two cross-wires.

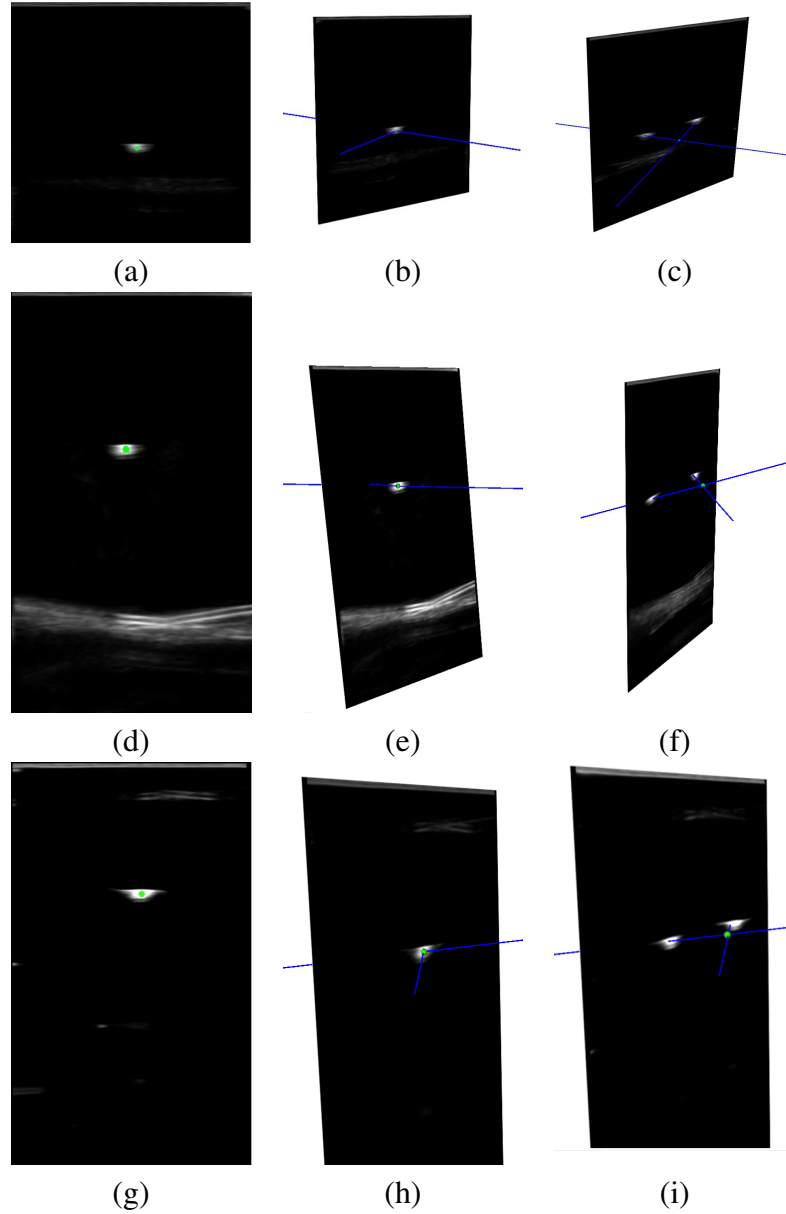


Figure 2.12: Visual representation of the cross-wire accuracy analysis for one calibration per imaging depth. Each row of images arises from the same imaging depth. The first column (a, d, e) shows the phantom target overlaid on the US image, while the second (b, e, h) shows the calibrated US image in a common coordinate system, with the cross-wire phantom aligned with the target point, and the third column (c, f, i) shows the calibrated US image in a common coordinate system with the cross-wire phantom aligned with the two cross-wires.

Table 2.5: The mean and standard deviation for the absolute distance between the ground-truth cross-wire position and the US reconstructed position calculated using the 15 calibrations obtained per imaging depth across the 3 point targets. The absolute error for X, Y, and Z axes, as well as the normalized absolute distance, are summarized below.

Absolute Error (mm)	Manual	Automatic	p-value
4 cm X-axis	0.59 ± 0.46	0.50 ± 0.38	0.21
4 cm Y-axis	0.60 ± 0.23	0.59 ± 0.20	0.96
4 cm Z-axis	0.82 ± 0.56	0.79 ± 0.55	0.77
4 cm Normalized	0.72 ± 0.58	0.70 ± 0.53	0.75
6 cm X-axis	0.70 ± 0.37	0.74 ± 0.36	0.43
6 cm Y-axis	0.24 ± 0.27	0.29 ± 0.21	0.72
6 cm Z-axis	0.61 ± 0.53	0.60 ± 0.44	0.92
6 cm Normalized	0.72 ± 0.55	0.63 ± 0.42	0.36
8 cm X-axis	0.56 ± 0.63	0.49 ± 0.46	0.57
8 cm Y-axis	0.19 ± 0.13	0.22 ± 0.14	0.42
8 cm Z-axis	1.15 ± 0.59	1.28 ± 0.57	0.31
8 cm Normalized	1.18 ± 0.45	1.23 ± 0.47	0.60

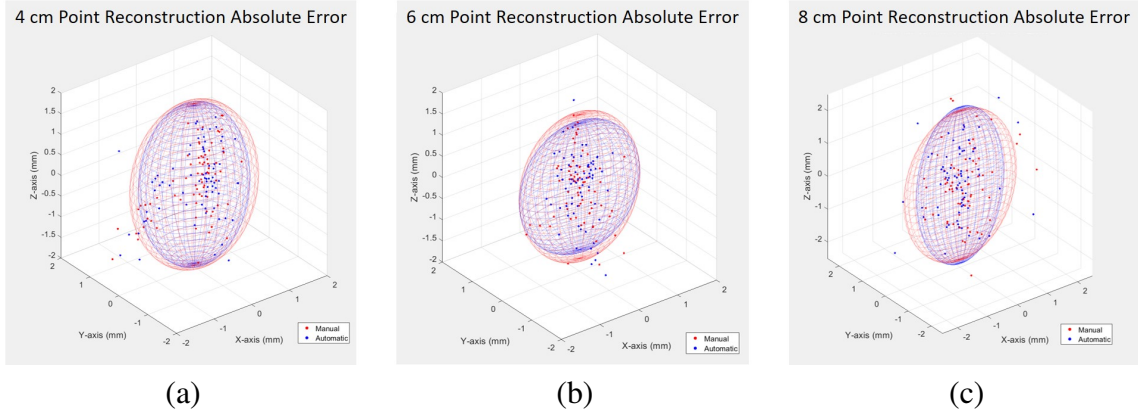


Figure 2.13: The absolute error between the three ground-truth targets and US reconstructed positions of the cross-wire target plotted across the 15 calibrations obtained per imaging depth, with the 95 % prediction interval provided as an ellipsoid.

2.4 Discussion

The objectives of this work are three-fold. First, we aimed to develop an algorithm capable of localizing the US fiducials required for the point-to-line US calibration method with accuracy equivalent to expert manual localization. The second objective was to integrate the trained machine learning model into the point-to-line calibration approach to enable automatic fiducial

localization, while the third goal was to validate the trueness and precision of the point-to-line calibration, with both manual and automatic fiducial localization approaches, using real data rather than theoretical estimates. Overall, we demonstrate that the point-to-line US probe calibration algorithm is capable of achieving high-quality calibration for both the manual and automatic fiducial localization variations for a single US probe across 4-8 cm imaging depths.

We developed a machine learning algorithm capable of localizing the centroid of the needle reflection (fiducial) within 2 pixels or 0.25 mm of an expert manual segmentation. All test and validation sets had a mean absolute error of less than 2 pixels and a standard deviation of less than 2.5 pixels. As pixels are a unit-less measurement, the error results from the 225 point pairs obtained during calibrations were obtained in millimetres by applying the average scaling factors obtained from all calibrations for each imaging depth, as summarized in Tab. 2.1. The largest absolute error obtained was 0.4 mm, and on average the absolute error was less than 0.25 mm between the automatic and manually localized fiducials. Thus, for a single linear US probe, we developed a machine learning algorithm capable of accurately localizing the US fiducials required for the point-to-line US calibration. While this algorithm produces accurate localizations for the US probe and imaging depths provided during training, it would unlikely provide the same level of accuracy if applied to images captured from a different US system or a different type of US probe from the same system. Using a known probe calibration obtained with careful manual localization, the coordinates representing the centroid of the needle can be automatically obtained to provide ground-truth centroid positions removing the need for time-consuming manual segmentation. This allows for this network to be easily trained for other US probes and US manufacturers. Additionally, this machine learning architecture could be retrained to localize fiducials for different US calibration methods, such as the Z-bar method. This automatic fiducial localization algorithm was integrated into the point-to-line calibration, removing one of the last required manual components. Thus, through the development of an algorithm to select the best images for the calibration, the automatic localization could enable a video-based fully automatic US probe calibration algorithm.

The precision results presented in Table 2.4, demonstrate that the point-to-line calibration approach is precise with both manual and automatic fiducial localizations. For each imaging depth, our analysis produced two point clouds with 75 points centred about $[0, 0, 0]$, one for calibrations obtained with manual localizations and one for the automatic equivalent. The standard deviation was calculated across the 75 samples in each distribution to quantify the precision of both calibration approaches for each imaging depth. On average all of the standard deviations were less than 1 mm regardless of imaging depth or localization approach, illustrating that the point-to-line calibration method is capable of producing precise calibrations. The results of the Z-test applied to compare variances between the two groups demonstrates

that there are no significant differences between the variance of the calibrations obtained using the automatic or manual localizations. Therefore, the point-to-line calibration with automatic fiducial localization is capable of producing calibrations with the same precision as an expert user. The precision is also presented graphically in Fig. 2.11 and a visual representation of the differences in output calibrations is provided in Fig. 2.10 such that the variation between the calibrations obtained using each approach can be understood in a qualitative context.

The accuracy (trueness and precision) of the system was quantified using a point reconstruction approach, where the point target was a cross-wire phantom. The point target is known by the construction of the phantom and is transformed into the tracker's coordinate system by applying the fiducial-based registration between the phantom and tracker space. For each imaging depth, the phantom was imaged under US such that the cross-wire reflection appeared in three distinct positions in the US image. The pixel coordinates representing the centres of the cross-wire junction in the US image were manually localized. The target position is reconstructed by applying the US calibration transformation matrices to the manually localized pixel coordinates. The difference in the ground-truth and US reconstructed positions represents the trueness and the spread of the values quantifies precision. For all imaging depths, the average normalized absolute errors were less than 1.5 mm and all standard deviations were less than 1 mm, as summarized in Tab. 2.5. For each imaging depth, the mean absolute error between the ground truth and reconstructed points obtained using the automatic and manual localizations were analyzed using T-tests. The results of these tests show that there is no significant difference between the mean absolute error produced by the manual and automatic calibrations. Indicating that the point-to-line calibration with automatic fiducial localization is capable of producing calibrations with the same level of precision and trueness as the manual localization equivalent. These results are also presented graphically in Fig. 2.13. The Z-axis has the largest error values compared to the X and Y axes for the same imaging depth as presented in Tab. 2.5. Fig. 2.13 provides a clear visual that the error is largest in the Z-direction. This is an expected result as US calibration is the least accurate in the Z-direction as the calibration fiducials extracted from US images do not contain any information about the Z-dimension. There error in the Z-axis increases as the imaging depth increases, which likely has to do with changes in the US beam profile as depth increases. As seen in Fig. 2.9 the reflections produced deeper in the image have a lower resolution compared to the reflections produced close to the transducer face, which could result in increased FLE as the imaging depth increases. Additionally, 15 calibration fiducials were used to obtain the calibrations regardless of imaging depth. The Z-axis error could be improved by collecting more calibration fiducials and collecting data with more variation in the needle angles for larger imaging depths.

It should be noted that the point-reconstruction method used to analyze the accuracy has

many limitations as many sources of error could affect the accuracy of the analysis. The phantom calibration required to position the cross-wire target in tracker space has many sources of error, including i) the error associated with the pivot calibration required for the stylus used for the cross-wire phantom registration affected how accurately the divots on the phantom were digitized and resulted in FLE, ii) the error associated with the fiducials extracted from the CAD model of the cross-wire phantom also produced FLE, and iii) the fiducial-registration between the phantom divot coordinates and the digitized coordinates also has an associated FRE which is a product of the FLE resulting from the first two sources of error. Thus, there is an error associated with the 3D coordinate that represents the position of the cross-wire with respect to the tracker's coordinate system. The US coordinate that represents the centroid of the cross-wire reflection is also manually selected, which resulted in additional TLE. To minimize these sources of error every aspect of the accuracy analysis was performed carefully by an experienced user. Despite the various sources of error, both the point-to-line calibration method with both manual and automatic localization was found to be accurate and achieved a normalized absolute error less than 2 mm compared to the ground-truth phantom position. We expect this value may be slightly larger than the true error associated with the calibration, due to the accumulation of errors within the phantom calibration process required for the cross-wire validation. The methods used to compare the accuracy of the point-to-line calibration with automatic and manual localizations provides more insight on if the automatic localization introduces any bias into the calibrations compared to the manual equivalent, rather than the true accuracy of the calibration approach. As the only information that changed during the accuracy analysis was the calibrations obtained using the manual and automatic localizations, the errors introduced into the analysis are consistent across both methods allowing for them to be compared against one another. The reported accuracy values indicate the system accuracy rather than the calibration accuracy. Additionally, the cross-wire reflection was also only captured in three positions throughout the US image, which may not adequately capture the rotational errors associated with the calibration. To properly quantify the accuracy of the point-to-line calibration a more comprehensive point reconstruction analysis should be performed with samples evenly spaced out throughout the US image.

The point-to-line calibration with automatic fiducial localization has the ability to improve the usability and robustness of the calibration approach. Removing the required manual localization could improve the time associated with the calibration approach, especially if a fully automatic solution is developed. The automatic localizations can enable novice users to obtain calibrations with the same level of accuracy as an expert, and would likely improve the calibration reproducibility across a group of novice users by decreasing the FLE associated with manual localization. Additionally, a novice user no longer has to learn how to perform accurate

manual localization, which is a time-consuming learning process that requires some form of teaching phase. The increased usability of this calibration approach could have large implications on the translation of this method between research groups. The ability for a novice user to obtain an accurate calibration with limited manual intervention could also encourage clinical implementation of this method.

2.5 Conclusions

Overall, this work presented and validates a point-to-line US calibration approach with machine learning-enabled automatic fiducial localization. A CNN was trained to localized the needle reflections (fiducials) required for the point-to-line calibration and was capable of selecting the centroid pixel coordinate within less than 0.25 mm from the manually localized ground-truth. The trained CNN was integrated into the calibration algorithm to replace the manual localization required for the original implementation, which is the last manual portion of the calibration process. The point-to-line calibrations produced average normalized absolute distance errors of less than $1.5 \text{ mm} \pm 1 \text{ mm}$ from the ground-truth point target for calibrations obtained using both the automatic and manual fiducial localization methods. The inclusion of the automatic localization did not have any effect on the trueness or precision of the calibrations compared to the manual equivalent. Both the manual and automatic fiducial localization approaches are capable of achieving accurate calibrations. The integration of the automatic fiducial localization into this point-to-line calibration improves the usability and ease of translation of this approach.

Chapter 3

Accuracy Assessment for the Co-registration Between Optical and VIVE Head Mounted Display Tracking

This chapter is adapted from the following manuscript:

- Groves, L. A., Carnahan, P., Allen, D. R., Rankin, Adam., Peters, T. M., & Chen, E. C. S. (2019). Accuracy assessment for the co-registration between optical and VIVE head-mounted display tracking. *International journal of computer assisted radiology and surgery*, 14(7), 1207-1215.

3.1 Introduction

Computer-assisted surgical navigation systems aim to improve the visual information provided to the surgeon through the inclusion of computer-generated graphics [36, 30]. Early-stage augmented- and mixed-reality (MR) surgical navigation systems typically employed a monitor to visualize the surgical information [88, 87], with these screen-based systems showing success in areas such as percutaneous interventions [100], forensic medicine [71], tumour resection [18], and a variety of other surgical interventions. While using a monitor to present advanced surgical navigation information improves guidance compared to standard image-guidance methods, these systems divide the clinician's attention between the patient and the screen that displays the information. Advances in wearable technology such as head-mounted displays (HMDs) have permitted augmented-reality (AR) environments to render surgical information directly in the user's visual field. While the concept of using an HMD for surgical navigation is not novel, the technological advancements and increased accessibility to

consumer-grade HMD systems have caused a resurgence of research in this area. The three most common types of HMDs include virtual-reality head-mounted displays (VR-HMD) providing entirely computer-generated scenes; video pass-through head-mounted displays (VPT-HMD) that can render entirely virtual scenes with or without superimposing information from one or more camera(s) fixed to the HMD, and optical see-through head-mounted display (OST-HMD) where the screen itself is transparent allowing for computer graphics to be superimposed onto the real world. Many systems have been proposed that integrate medical images into optical OST-HMDs, which can be registered with and potentially anchored to the real patient anatomy [101, 68, 12, 43]. This technology allows the user to see the real world through a translucent display, while virtual computer-generated graphics are superimposed on the scene [101].

With the majority of HMD research focused on OST-HMDs there is a gap in the literature surrounding surgical navigation systems that utilize VPT- and VR-HMDs. Although there are many clear benefits of OST-HMD AR environments, VPT- and VR-HMDs offer different benefits. Rendering either an entirely virtual scene or a scene consisting of mostly virtual components with some real input (augmented-virtuality or AV) provides more control over the input presented to the user. Furthermore, VPT-HMD or VR-HMDs have fewer perceptual issues, can render black (as distinct from an OST system where “black” implied a lack of illumination of a surface, rendering it as transparent), and have larger fields of view [130]. Additionally, the intrinsic tracking of OST-HMDs and VPT- or VR-HMD systems use very different mechanisms. OST-HMDs typically do not require extrinsic devices to track their position as they employ simultaneous localization and mapping (SLAM) to orient the HMD in space [130]. As these systems have no external tracking components the HMD coordinate system is defined with respect to the HMD itself. VPT- and VR-HMDs often use a combination of external tracking components and SLAM to localize the HMD and commonly two associated hand-held controllers in space [130]. The use of extrinsic tracking components allows for improved tracking accuracy which is an additional benefit to VPT- and VR-HMDs [130]. For these systems, the pose of the HMD is known with respect to the coordinate system defined by the external tracking components. Regardless of the method to track the HMD, each HMD system defines a coordinate system that is specific to the HMD.

Chen *et al.* [23] developed a method to register an external optical-tracking system (OTS) and the tracking system associated with an OST-HMD to develop an AR surgical navigation system. This system was successfully used to insert percutaneous sacroiliac screws [155]. The integration of an external tracking system into an HMD requires calibration to solve for the registration between the external tracking system and the tracking system associated with the HMD, allowing for models of externally tracked surgical instruments to be accurately posi-

tioned and visualized in the HMD. The research conducted using OST-HMD's and OTS justifies developing surgical navigation systems for VPT- or VR-HMDs. However, due to the inherent differences in technology between these systems, many of the methods developed for OST-HMDs cannot be applied to VR or VPT-HMDs. In this work, a method is proposed to co-register the Polaris Spectra OTS (NDI, Canada) and the HTC VIVE Pro (HTC VIVE Pro, HTC, Taiwan, ROC) VPT-HMD, such that optically tracked spatial information can be integrated into the HMD. The system combines industry-standard tracking technology with relatively inexpensive AV or VR equipment and forms the foundations required to develop MR VPT- and VR-HMD surgical training or planning systems that require tracked surgical instruments.

3.2 Materials and Methods

3.2.1 Equipment

The HTC VIVE Pro VPT-HMD was selected to be used within this work as it has a high-resolution display, two stereo cameras, 6 DoF HMD and controller tracking, a wide FoV of 110° , the ability to be un-tethered from the computer, and developers can access the tracking information for the HMD and controllers¹. The VIVE and VIVE Pro are tracked using external tracking components called lighthouse base stations. These base stations emit infrared (IR) light, which is detected by the IR sensors that cover the surface of the HMD and controllers. The difference in detection time between the sensors is used to localize the HMD and controllers within the tracking area. The external tracking is supplemented with inertial measurement units (IMU). Niehorster *et al.* assessed the original VIVE's tracking technology for accuracy and precision, and the authors found that the tracking technology is unsuitable for scientific research [107]. However, the VIVE Pro HMD uses the VIVE base station 2.0 and due to the unreported differences between the first- and second-generation lighthouse technology the tracking accuracy associated with the VIVE Pro must be re-validated.

A computer equipped with an Intel 7820x processor, 32GB of RAM, and an Nvidia GTX1070Ti graphics card was used for this research, as it can support the real-time computation and rendering requirements of the hybrid system. The VIVE Pro HMD, the associated controllers, and the VIVE tracking system were used within this research. The spatial tracking system used for this research is the Polaris Spectra (NDI, Canada) OTS with passive markers. This tracking system was selected as it is the most common system used for commercial surgical navigation [41]. The Polaris Spectra and VIVE tracking systems were arranged such that their FoV's

¹<https://www.vive.com/ca/product/vive-pro/>

were optimally overlapped during the calibration and evaluation processes (Fig. 3.1). The Plus Toolkit [77] was used to obtain tracking information from the Spectra OTS and SlicerVR ² was used to obtain tracking information from the HTC VIVE Pro through SteamVR (Valve Corporation, USA).

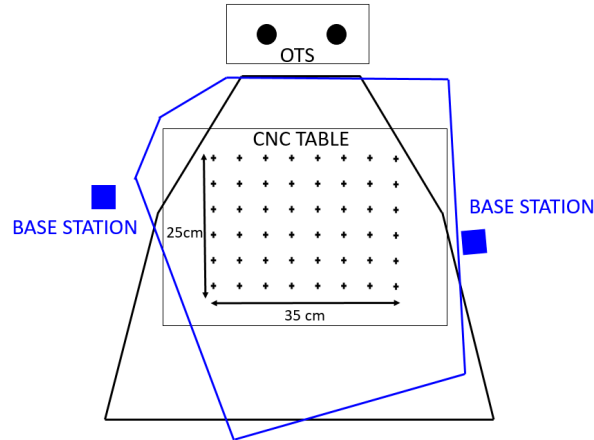


Figure 3.1: Schematic of the experimental setup. The black and blue region indicates the optimal tracking volume of the optical/VIVE Pro's tracking systems, respectively. A grid of locations was programmed in CNC coordinate system and is located in the overlapping, optimal, tracking volume of both systems.

A 3D printed apparatus was developed to co-register the VIVE and OTS as seen in Fig. 3.2. This apparatus was designed using SpaceClaim (SpaceClaim Corporation, USA) and constructed using the Ultimaker 3D printer (Ultimaker, NL) with 0.35 mm resolution. A reference optical dynamic reference frame (DRF) is rigidly fixed to the apparatus by a support post and the pose of the DRF is known with respect to the OTS. The VIVE controller has a contact fit with the apparatus through a negative imprint of the controller. Six hemispherical divots in known locations with respect to the origin of the VIVE controller were embedded into the surface of the apparatus. The divots were carefully milled to have a 6.35 mm diameter. An optically tracked pointer tool with a 6.35 mm diameter ball tip was used to digitize these divots within the co-calibration process. The optically tracked pointer tool was “pivot and spin” calibrated to provide the relationship between the ball tip and the optical DRF fixed to the tool, such that the pose of the stylus tip was tracked with respect to the optical tracker’s coordinate system. Pivot calibration is the process of pivoting the tool around its tip such that the motion of the tracking sensor forms the surface of a sphere about the stylus tip [25]. Sphere fitting was used to solve for the relationship between the collected data points and the centre of the sphere

²<https://github.com/KitwareMedical/SlicerVirtualReality>

using a least-squares approach [163]. Following pivot calibration spin calibration is required to provide the correct axis of rotation [152]. This pointer tool calibration resulted in a reported root-mean-square-error of 0.24 mm, an error of approximately 0.25 mm can be consistently achieved using this approach.

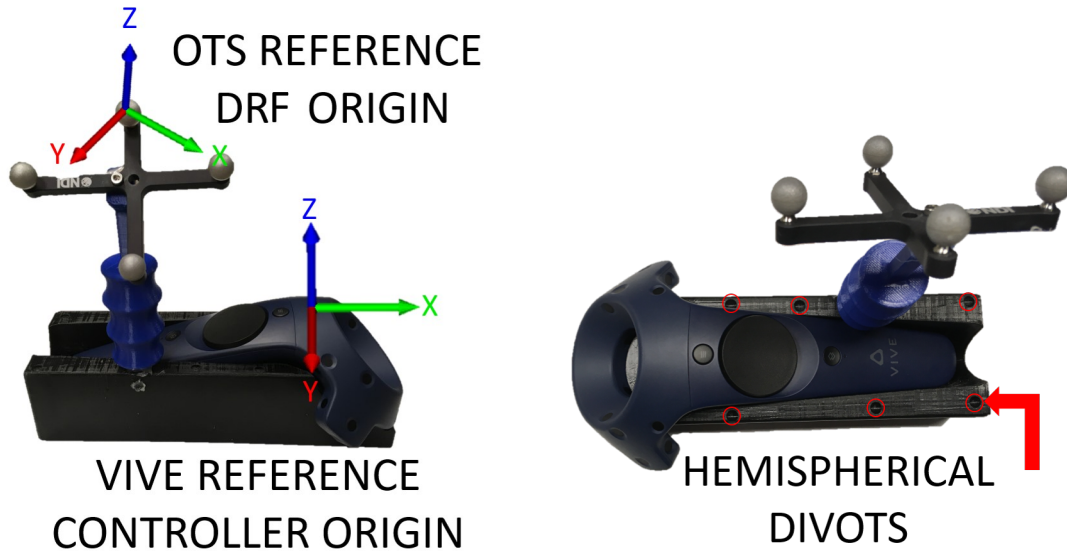


Figure 3.2: calibration device with optical DRF and hemispherical divots.

3.2.2 Co-calibration Process

The calibrated stylus was used to digitize each of the hemispherical divots on the co-calibration apparatus to obtain the 3D coordinate associated with each divot with respect to the reference optical DRF on the co-calibration apparatus, as seen in Fig. 3.3 a). The divot locations were collected in the VIVE tracker's coordinate system through extracting their locations with respect to the VIVE controller's origin from the CAD files of the co-calibration apparatus and the VIVE controller. This produced two corresponding sets of point locations: one in the optical tracker's coordinate system and one in the VIVE's coordinate system, that were registered to each other using closed-form fiducial-registration with singular value decomposition least-squares regression approach [5]. The registration process resulted in a 4×4 rigid transformation matrix that provided the fixed relationship between the reference optical DRF origin and the VIVE controller origin with a root mean squared error of 0.41 mm as seen in Fig. 3.3a). A typical error achieved with this co-calibration process is less than 0.50 mm.

Although this registration does not directly align the optical and VIVE coordinate systems, the hybrid system is defined by the concatenation of three matrices representing the pose of

the optically tracked tool, the registration between the optical DRF and the controller's origin, and the pose of the VIVE's controller with respect to the VIVE's coordinate system. The pose of the optically tracked tool was provided with respect to the optical DRF on the apparatus. Therefore, the aforementioned transformation chain can be applied such that the spatial tracking information can be used to visualize tracked surgical tools within the HMD. A visual depiction of this process can be seen in Fig. 3.3b).

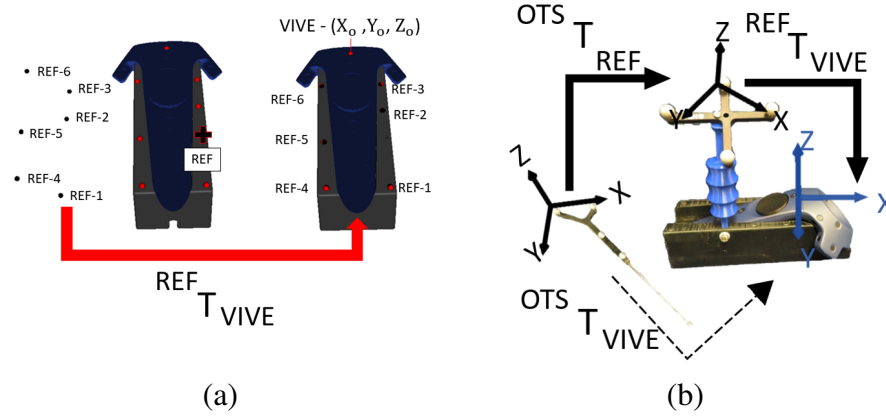


Figure 3.3: (a) Visual depiction of the point-based registration, (b) Visual depiction of the transformation chain used for the hybrid system.

3.2.3 Accuracy Analysis

To evaluate the accuracy of the hybrid system, the OTS and the VIVE's tracking system, a validation tool was 3D printed with a negative imprint of the VIVE controller and an opening for an optical DRF. The front of this tool has a cylindrical opening in which a 9.53 ± 0.01 mm diameter carbon fibre stylus with a 6.35 mm diameter ball tip was inserted and fixed to create a dual tracked stylus. The validation tool can be seen in Fig. 3.4. The validation tool was pivot [25] and spin [152] calibrated twice to solve for the pose of the stylus tip with respect to both the optical DRF origin and the VIVE controller origin fixed to this tool. The associated calibration errors were 0.29 mm and 0.71 mm for the optical and VIVE tracking systems, respectively. On average these processes can consistently achieve errors less than 1 mm. This validation tool was used within the position and rotation accuracy assessments.

3.2.4 Positional accuracy assessment

A Computer Numerical Control (CNC) machine outfitted with the ACU-RITE SENC 150 linear encoder (ACU-RITE, US) can position the centre of a vice used to hold the apparatus

within $2.54\text{ }\mu\text{m}$. This system was employed to serve as the ground truth for the evaluation of the positional accuracy of the OTS, the hybrid system, and the VIVE Pro tracking system. The validation tool was mounted in the vice attached to the CNC machine at the centre of the combined tracking FoV. The reference apparatus was positioned such that it was centred in the area that data was collected highlighted by the red arrows in Fig. 3.4.

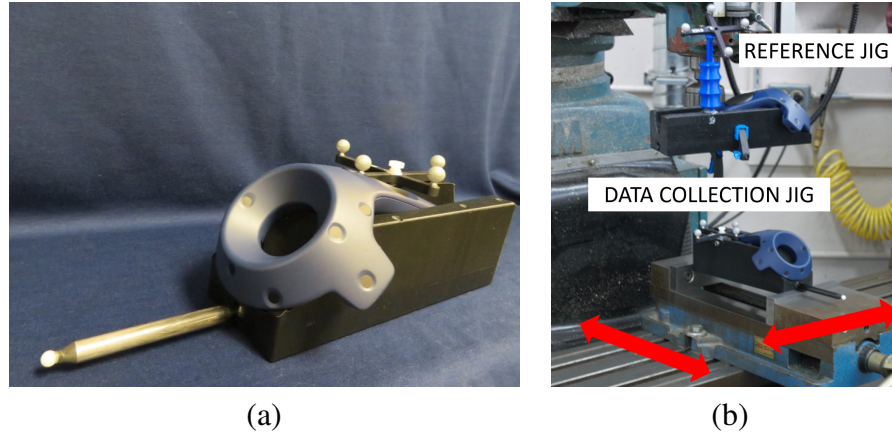


Figure 3.4: (a) Dual tracked validation tool outfitted with a stylus, (b) Position accuracy validation set-up with arrows to highlight the data collection path.

The CNC machine was programmed to collect a grid of 48 points spanning 350 mm by 250 mm with 50 mm spacing in both directions. The CNC machine mechanically moved the vice to each location, where 25 samples were collected for the spatially tracked positions reported by the optical DRF, the calibrated optical DRF, and the VIVE controller. The reported CNC location was recorded to be used as a ground truth.

Prior to analysis, the singular value decomposition least-squares regression approach was used to register all point grids to a common coordinate system for comparison [5]. The collected data resulted in 48 point clouds with 25 samples each for each point in the grid (1200 samples in total per system). For all three tracking systems the same approach was used to evaluate the positional accuracy. The absolute distance from each collected sample to its corresponding CNC ground truth position was calculated, resulting in 1200 distance values. The average absolute distance is indicative of the trueness of the tracking system, while the standard deviation between distance values represents the precision. These values along with the 95 % confidence interval (CI) are summarized in Table 3.1.

Rotational accuracy assessment

To assess the rotational accuracy of the tracking systems a hemispherical phantom with 13 cylindrical openings at a variety of angles was manufactured using the Ultimaker 3 3D printer.

To eliminate issues associated with the rough finish of 3D printed components and to ensure the stylus on the validation tool had a snug fit, tubing with an outer diameter of 12.70 ± 0.25 mm and inner diameter of 9.53 ± 0.25 mm was selected to fit precisely with both the cylindrical openings in the hemispherical phantom and the stylus. The final device is shown in Fig. 3.5 (a). A CT scan of the hemispherical phantom was taken using a GE Lightspeed VCT (General Electric, USA), and the cylindrical openings were segmented to extract the ground truth rotations. The hemispherical phantom was secured by clamping the rectangular base affixed to the phantom using two mechanical arms fixed to a stationary table. The Polaris Spectra tracking FoV was situated such that it maximized the ability to capture the optical DRF while maintaining maximum overlap with the VIVE tracker's FoV.

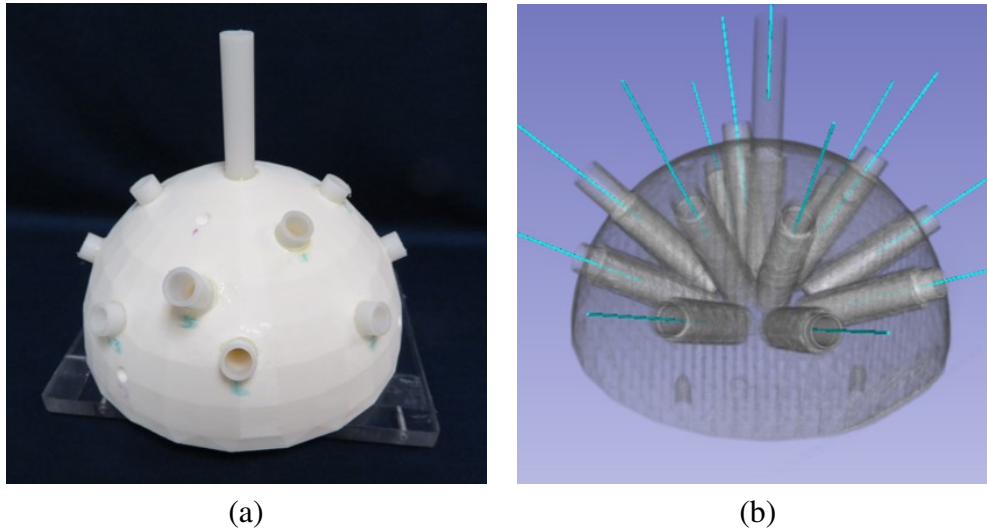


Figure 3.5: Hemisphere phantom for rotation validation. Image a) depicts the 3D printed hemisphere phantom with cylindrical openings and Image b) shows the CT scan of the hemisphere phantom and the rotation vectors for each cylindrical opening.

The calibrated stylus on the validation tool was inserted into each of the 13 cylindrical openings in the hemisphere. For each insertion, tracking information consisting of 25 rotation values stored as quaternions were collected for the OTS, the VIVE, and the hybrid system, respectively. To produce gold standard rotation values, the cylindrical openings were manually segmented from the CT volume of the hemisphere. For each opening, a cylinder was fitted to the manual segmentation using an M-estimator SAMple Consensus algorithm [147]. The result of this process can be seen in Fig. 3.5 b).

The orientations collected with respect to each tracking system were rotated to the coordinate system of the gold standard values measured from the CT scan using a least-squares absolute orientation procedure [62]. The absolute angular difference between the gold standard orientation and the angle reported by each tracking system was calculated for each sample. The

average, standard deviation, and 95 % CI for the absolute angular error of each tracking system is presented in Table 3.2.

3.3 Results

The positional and rotational accuracy presented as mean, standard deviation and 95 % CI for the OTS, VIVE tracking system, and hybrid tracking system are presented in Tabs. 3.1 and 3.2, respectively. These results are also summarized as box plots in Fig. 3.6. The reported values are compared to ground-truth positions and rotations, and therefore the smaller the value the more accurate the system is.

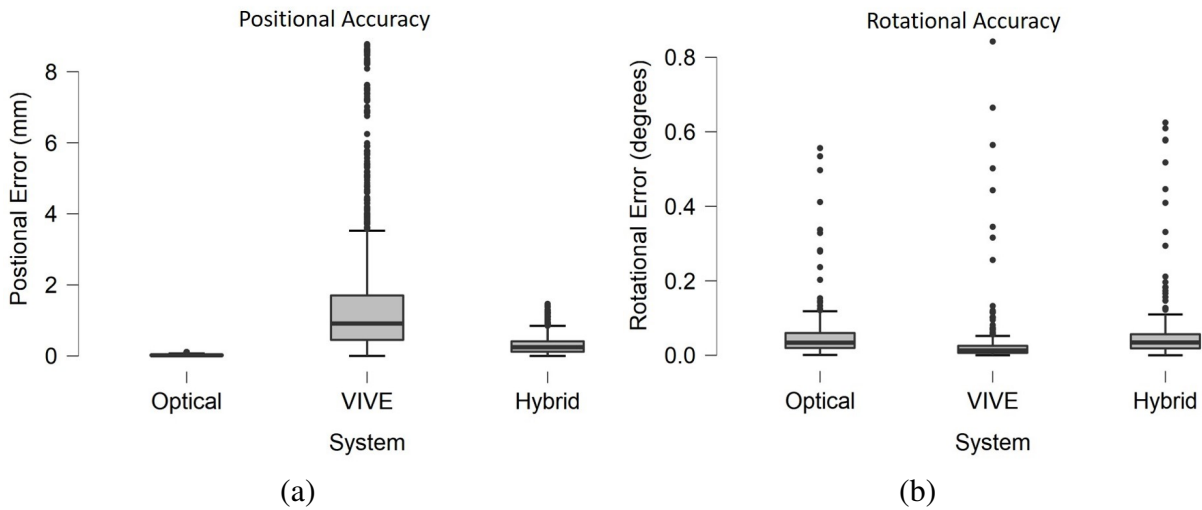


Figure 3.6: Positional (a) and rotational (b) accuracy summarized in box plots

Table 3.1: The average, standard deviation, and 95 % CI calculated using the absolute distances between each collected sample and its corresponding CNC location.

Positional Accuracy		
Tracking Method	Mean \pm std	95% CI
Optical	0.026 ± 0.02 mm	[0.03 mm, 0.03 mm]
VIVE	1.52 ± 1.76 mm	[1.42 mm, 1.62 mm]
Hybrid	0.30 ± 0.25 mm	[0.29 mm, 0.32 mm]

Table 3.2: The average, standard deviation, and 95 % CI calculated using absolute angular differences between each collected sample and its corresponding ground truth angle extracted from the cylindrical phantom CT scan.

Rotational Accuracy		
Tracking Method	Mean \pm std	95% CI
Optical	$0.057 \pm 0.11^\circ$	$[0.04^\circ, 0.07^\circ]$
VIVE	$0.035 \pm 0.09^\circ$	$[0.02^\circ, 0.05^\circ]$
Hybrid	$0.060 \pm 0.11^\circ$	$[0.05^\circ, 0.07^\circ]$

3.4 Discussion

This work has established the positional and rotational accuracy for the Polaris OTS, the VIVE PRO tracking system, and the proposed hybrid tracking system that registers optically tracked tools to the VIVE tracker's coordinate system. The resultant positional accuracy was sub-millimetre for the OTS and the hybrid system, and on the order of a millimetre for the HTC VIVE tracking system. The rotational accuracy values were less than one degree for all three systems. The hybrid system has better positional tracking accuracy than the HTC VIVE system despite the fact that the controller's pose is used within the calibration process. The results suggest that the HTC tracking system has a global trueness bias which is likely due to the IMUs used to track the VIVE controller. The proposed method of co-registration is unaffected by the lower positional accuracy of the VIVE system since the VIVE controller is stationary within the co-calibration apparatus, which results in a larger reliance on the precision of the system. However, the hybrid system is highly dependent on the tracking accuracy of the OTS, as any error associated with the tracking of an optical DRF would propagate and affect the ability to render the accurate pose of the virtual representation of the surgical tool.

As the accuracy required for surgical navigation is highly dependent on the application, there is no definite threshold to judge accuracy values. However, since the Polaris Spectra is a commonly used commercial tracking system for medical imaging and surgical navigation research [136], the accuracy values achieved by this system can provide a benchmark to compare the hybrid tracking system's accuracy against. The hybrid system achieved positional and rotational accuracy of $0.30 \text{ mm} \pm 0.25 \text{ mm}$ and $0.06^\circ \pm 0.11^\circ$, respectively. The 95% confidence interval indicates that there is a high probability that the positional and rotational error values will be consistently less than 0.50 mm and 0.50° . The accuracy of the hybrid system is slightly inferior to the OTS due to the propagation of error during the calibration process.

However, the differences in trueness values between the two systems were only 0.28 mm and 0.00°, for the position and rotation respectively. The positional and rotational accuracy promote the continued use of the hybrid tracking approach to integrate tracked tools into the VIVE Pro.

The work published by Niehorster *et al.* [107] established the accuracy and precision of the original HTC VIVE system. While the results presented in that paper cannot directly be compared to our results due to differences in the reported metrics, broader comparisons can be drawn. They concluded that there was a systemic drift in the direction perpendicular to the direction in which they collected their grid of samples [107]. Their conclusion was not consistent with our work, as the displacement of samples was only in the direction parallel to the data collection. Updates to the software, or changes between the original HTC VIVE and the HTC VIVE Pro have potentially mitigated these issues. Furthermore, Niehorster *et al.* [107] concluded that the tracking of the HTC VIVE is not suitable for the medical field. However, the results obtained from our work counters this as we achieved millimetre and sub-degree accuracy for the HTC VIVE Pro, which is further improved with the hybrid system.

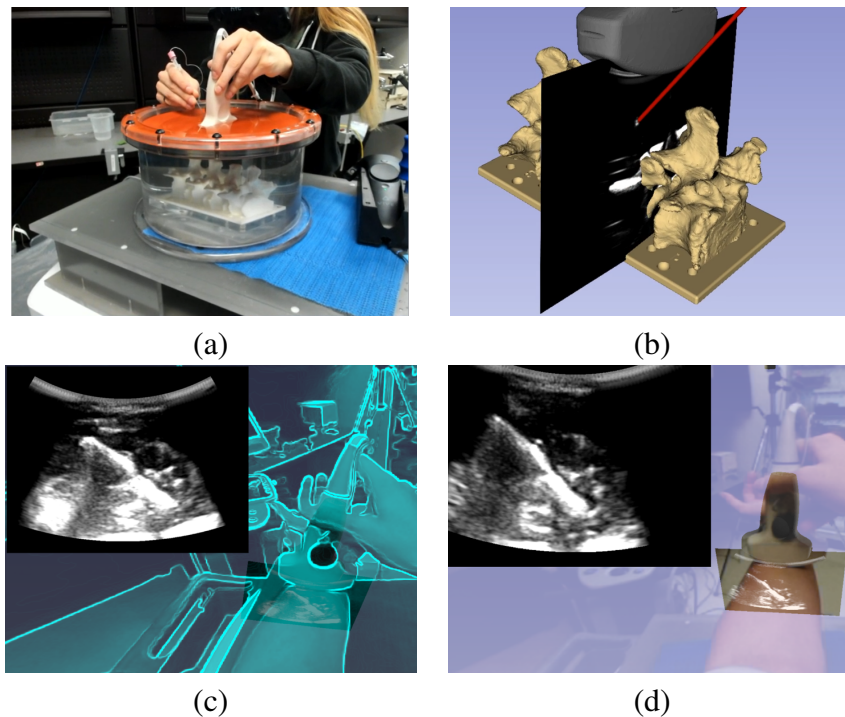


Figure 3.7: Examples of first-person Point-of-View MR visualization for surgery: (a) a VR system for spine needle intervention where streaming ultrasound and real-time needle advancements are visualized using HMD, b) VR representation of this procedure as viewed within the HMD, c) AV scene with streaming ultrasound and the edge mapping feature for the stereo camera video input, d) AV scene with streaming ultrasound and the translucent feature for the stereo camera video input

The ability to integrate spatial tracking information into the HMD allows for VR training or surgical practice systems to be trivially developed. One application of the system as it currently stands is training or practicing US-guided needle navigation. A training system for US-guided needle insertions can be simply implemented in this framework through the inclusion of a tracked and calibrated surgical needle, ultrasound image and virtual anatomical model as seen in Fig.3.7. Further, any surgical navigation system that presents spatial tracking information and image data in a 2D VR environment can use the hybrid tracking system to display the information directly into the user's line of sight to a substantially improved accuracy compared to the VIVE's tracking capabilities.

As the mathematical foundation of the hybrid tracking system is based on the concatenation of transforms, each component of this transformation chain introduces sources of error. The error associated with the co-registration was a product of cumulative errors throughout the calibration process. The requirement of pivot calibrating the stylus prior to collecting the landmark locations resulted in a registration error of 0.24 mm propagating into the point-based registration. This error could be improved through performing pivot calibration with a three-point ball contact instead of a single conforming surface for a more accurate implementation [85]. Furthermore, the locations of the divots within the VIVE controller coordinate system were extracted from the CAD model. The 3D printing manufacturing process may introduce errors in this fiducial localization. However, extracting fiducial locations from a CT scan of the co-calibration apparatus is not feasible as there is no means to relate the CT coordinate system to the VIVE's controller coordinate system directly. We suggest that this co-calibration apparatus be manufactured using an alternative, more precise method, such as a better 3D printer or CNC milling. In this study, the fiducial localization error was minimized through precisely milling each divot on the calibration apparatus, and the cumulative effects of these sources of error resulted in a fiducial registration error of 0.41 mm. The position accuracy could be further influenced by tracking errors, despite the fact that the validation set-up was optimized to ensure that the majority of data were collected in the centre of the overlapping tracking FoV. The locations at the extremes of the data collection grid may have been too close or too far from the base stations resulting in a less accurate controller pose decreasing the VIVE positional accuracy. The rotational accuracy results were affected by the propagation of error from pivot and spin calibrating the stylus fixed to the validation tool.

Looking forward, continued development on the usability and logistics of the system is required to promote more clinical usability. The function of the hardware can be improved through the inclusion of the wireless adapter for the HTC HMD and the additional base stations to increase the size of the tracking area and reduce line-of-sight issues. This would promote a more clinically applicable system, as the user would be less restricted through the un-tethered

HMD and an increased tracking FoV to room-scale tracking. The usability of the system can be expanded through the use of the varying views provided by the HMD's stereo cameras, allowing for various implementations that span across the MR spectrum as seen in Fig. 3.7. In the future, the effects of including opaque video, colour video, and edge mapping need to be studied to gain an understanding of what is most useful in a surgical context. Alternatively, accessing the video feed from the stereo cameras such that the developer has more control over how to integrate the real world video input into the system could allow for interesting AV environments to be developed. The inclusion of real world information would allow the system to be expanded from surgical training and planning to clinical surgical navigation applications.

Our approach could be trivially extended to a hybrid tracking system between any spatial tracking system and other similar HMD systems. For example, to address line of sight issues associated with optical tracking, a version of the system could be developed through co-registration of a magnetic tracking system (e.g. the NDI Aurora) and the VIVE's tracking system. A three-way hybrid tracking system the registers optical, magnetic, and VIVE coordinate systems could also be trivially developed. Furthermore, these methods could be replicated for other similar VR- or VPT-HMD systems that utilize controllers or some form of extrinsic tracking components. As the co-calibration apparatus is tracked simultaneously by multiple tracking systems, a data-fusion scheme (such as the Unscented Kalman Filter [154]) could be implemented to further improve tracking accuracy.

3.5 Conclusions

This work developed and evaluated a co-registration technique that accurately integrates optical tracking information into a commercially available cost-effective VPT-HMD. The hybrid tracking system uses a calibration apparatus that is simultaneously tracked by the OTS and the VIVE tracking system, such that spatial tracking information can be used within the HMD with surgically acceptable accuracy. The positional and rotation accuracy of the hybrid tracking system were $0.30 \text{ mm} \pm 0.25 \text{ mm}$ and $0.06^\circ \pm 0.11^\circ$, respectively. Thus, this hybrid tracking system has the potential to be used in the clinical environment as it achieves submillimetre and subdegree accuracy tracking surgical tools within a consumer-grade HMD system. The use of our co-calibration apparatus allows for a single calibration to be used repeatedly within any overlapping tracking region between the OTS and the VIVE. This work validated the use of a co-calibration apparatus to integrate spatially tracked instruments into a VPT-HMD and sets the foundations for the development of an AV or VR immersive first-person point of view tracked surgical navigation system.

Chapter 4

Towards a mixed-reality first-person point of view needle navigation system for central line insertions

This chapter is adapted from the following manuscript:

- Groves, L., Li, N., Peters, T. M., & Chen, E. C. (2019, October). Towards a mixed-reality first-person point of view needle navigation system. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 245-253). Springer, Cham.

4.1 Introduction

Central venous catheterizations (CVCs) are performed over 6 million times annually between the United States and Europe [126, 90, 13], with the internal jugular vein (IJV) being the most utilized insertion site [17]. Ultrasound-(US)-guided CVC is becoming the preferred technique as it has the potential to reduce complications including accidental punctures to structures such as the carotid artery (CA) [134, 141]. The US-guided approach relies on real-time US video, depicting cross-sections of the anatomy on a 2D monitor, to guide the needle insertion (Fig. 4.1b). Despite US-guidance improving complication rates, clinical studies have found rates of CA puncture to be approximately 7 % for US-guided cases [93]. The high rate of arterial puncture under US-guidance may be a product of the inability to visualize all relevant anatomy and the needle tip within the limited 2D US FoV. US-guidance also requires the user to focus on the surgical navigation information being visualized on a monitor, while performing the insertion on the patient resulting in an uncoupling of the user's visual and motor fields.

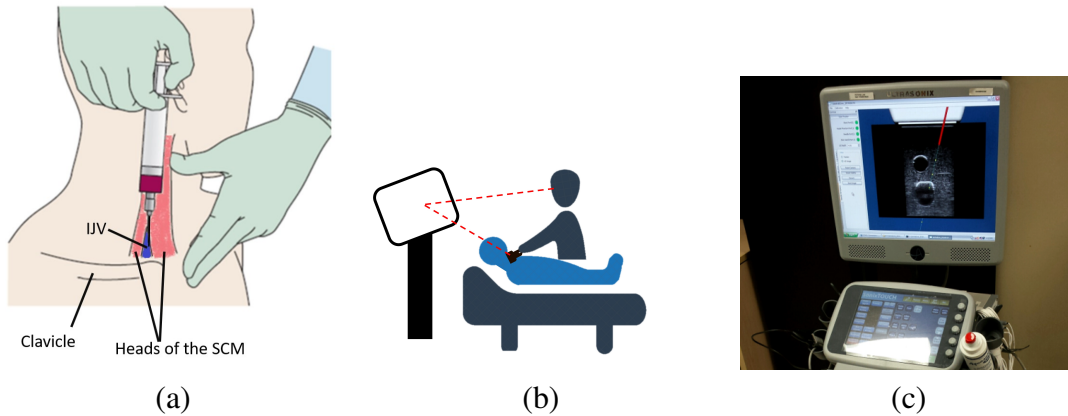


Figure 4.1: Comparison of guidance techniques for CVC a) landmark guidance b) US-only guidance and c) AV guidance.

To address the aforementioned limitations of US-guided CVC, Ameriet *al.* [2] developed an augmented-virtuality (AV) monitor-based surgical navigation system for needle guidance. This system employed magnetic tracking to render tracked virtual models of the US probe, needle and needle trajectory, overlaid on a front-facing US image on a traditional 2D monitor (Fig. 4.1c), such that the user can simultaneously visualize the CA, IJV and needle tip. The tracking information was used to update the position of the needle and trajectory throughout the insertion. The pose of the US image and probe were held constant throughout the insertions, maintaining the front-facing visualization that is typical of US-guidance as seen in Fig. 4.1c). This system supplemented the US-guidance visualization with the inclusion of the tracked needle and needle trajectory but did not demonstrate significant improvement in the complications associated with the needle insertion compared to the US-only technique for expert users [2]. The success of this system may have been influenced by the fixed front-facing US visualization, as the user was still reliant on a 2D environment to perform a 3D procedure, which may have resulted in difficulties interpreting the 3D relationship between the tools and the anatomy. Additionally, the outcomes of this experiment may have been affected by the maintained visual and motor disparity produced by using a monitor exterior to the visual field of the phantom where they are performing the insertion. Despite the inconclusive results, this promising work has motivated the development of advanced surgical navigation environments to reduce complications associated with CVC.

Here we investigate the efficacy of a first-person immersive mixed-reality (MR) system for CVC needle navigation. Toward the long-term goal of clinical deployment, we first aim to understand how the visualization paradigm associated with surgical information affects the rate of complication during US-guided CVC. We developed a surgical navigation system that comprises tracked models of the US probe, needle, and needle trajectory in a common coor-

dinate system with the tracked free-hand streaming US. Our implementation did not constrain the US image to a single view but allowed for the movement of the US probe and image to be updated in real-time. For this work, we compare US-only guidance to our MR guidance system displayed on a 2D monitor, and within an HMD. We hypothesize that the use of the HMD for visualization will improve the success of needle insertions compared to those that employ US-only guidance, or display the US image and needle trajectory on a 2D monitor. This work aims to highlight the importance of coherent visual and motor fields for surgical applications through a comprehensive user study involving 33 clinicians and 20 medical students.

4.2 Materials and Methods

A phantom comprising two hollow (wall-less) vascular structures representative of the CA and IJV, embedded in polymerized Poly (vinyl alcohol) cryogel (PVAc) [145], was constructed (Fig. 6.1a). This phantom was constructed using the same methods as those used by Ameri *et al.* for their central line surgical navigation study [2]. The positive models of both the CA and IJV were manufactured using 3D printer based on the manual segmentation of a patient computed tomography (CT) provided from the work performed by Ameri *et al.* [2] and reused for this phantom development.

To create realistic vascular impressions in the PVAc, the phantom mould used CA and IJV 3D-printed models based on the manual segmentation of a patient computed-tomography (CT) neck vasculature scan (Fig. 4.2a) [2, 50]. The mould was filled with non-polymerized PVA gel and underwent two freeze-thaw cycles for the PVA gel to polymerize into PVAc [2, 50]. The 3D printed vessels were then removed from the phantom after polymerization to produce hollow vessels [2, 50]. The phantom was housed in a plastic container with 8 hemispherical fiducial markers and a 6 Dof NDI reference Disk sensor and scanned with a cone-beam CT (O-Arm, Medtronic, USA). The positions of these fiducials were digitized with respect to the reference sensor fixed to the phantom using a pivot calibrated ball tip stylus [163] with a conforming 6.35 mm diameter. These fiducials were also manually extracted from the CT scan, such that the fiducial pairs can be used to solve for the relationship between the phantom's and the tracker's coordinate systems using a least-squares solution [5]. This phantom serves as a surrogate for patient anatomy, producing US images where the CA and IJV can be trivially interpreted by the user. An US image of the phantom is depicted in Fig. 4.2b), which can be compared to those obtained from healthy volunteers in Fig. 4.2c) and d).

The surgical guidance system comprises a monitor or mixed-reality (MR) HMD (HTC VIVE Pro, HTC, Taiwan ROC), a magnetic tracking system (Aurora, NDI, Canada), a clinical US scanner (SonixTouch, BK Medical, USA), and a surgical hypodermic needle (7 cm with

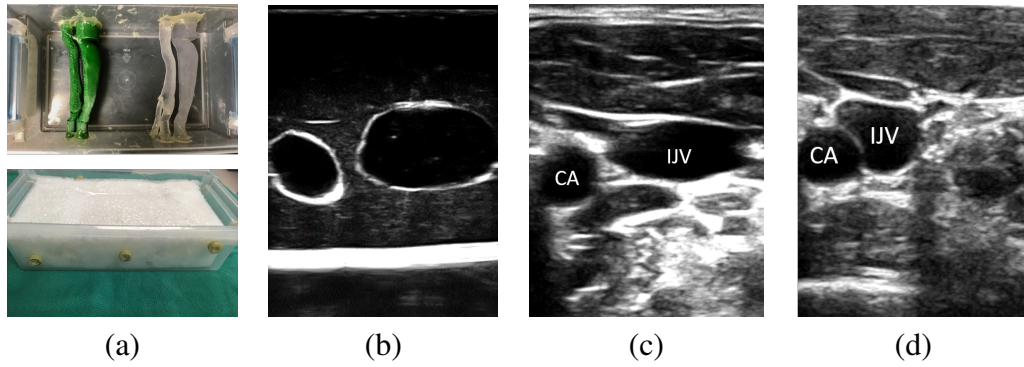


Figure 4.2: a) Phantom development, b) Phantom under US, c) and d) health human neck vasculature under US. Images c) and d) courtesy of healthy volunteers.

10 ml syringe, Fig. 4.3b). A linear transducer (L14-5, BK Medical, USA) was used to acquire real-time images of the phantom (depth of 6 cm with a gain of 42 %). The US transducer, neck phantom, and surgical needle were magnetically tracked, spatially calibrated, and registered into a common coordinate system. As the HTC VIVE Pro has its own tracking system, it is co-registered with the magnetic tracking through a simultaneously-tracked apparatus that registers the magnetically tracked tools into the HMD coordinate space (Fig. 4.3a)), which is a magnetically tracked variation of the system developed in Chapter 3 [52]. This co-calibration method was validated using a Computerized Numerical Control machine with reported trueness and precision values less than 1 mm and 1° , for position and rotation respectively [52]. The magnetically tracked US transducer was calibrated using a point-to-line calibration approach solved using an Anisotropic Orthogonal Procrustean Analysis [35] with manual localization as described in Chapter 2. To track a surgical hypodermic needle, a 6-DoF magnetic pose sensor was integrated into a 3-way Lure-lock connector, used to connect the metallic needle with the plastic syringe. A tracked “template” with the exact negative imprint of the needle assembly was used to calibrate the spatial pose of the needle assembly with respect to its magnetic pose sensor (Figure 4.3b). A visual assessment of the system accuracy, consisting of tracker co-calibration, tool calibration, and patient registration is shown in Fig. 4.3c).

Three modes of visualization were implemented and evaluated: 1) the traditional US-guidance technique with a 2D front-facing US image displayed on the US scanner, 2) an MR system displayed on a 2D monitor, 3) an MR system displayed in a first-person perspective in the HMD, as depicted in Fig. 4.4. The MR system comprised calibrated streaming US video, a virtual representation of the tracked probe and needle, and a needle trajectory represented as a 10 cm blue extension from the needle tip, all registered in a common coordinate system. This system was displayed in a first-person front-facing perspective in the HMD as shown in Fig. 4.4 c). The surgical navigation system displayed on the monitor provides the information

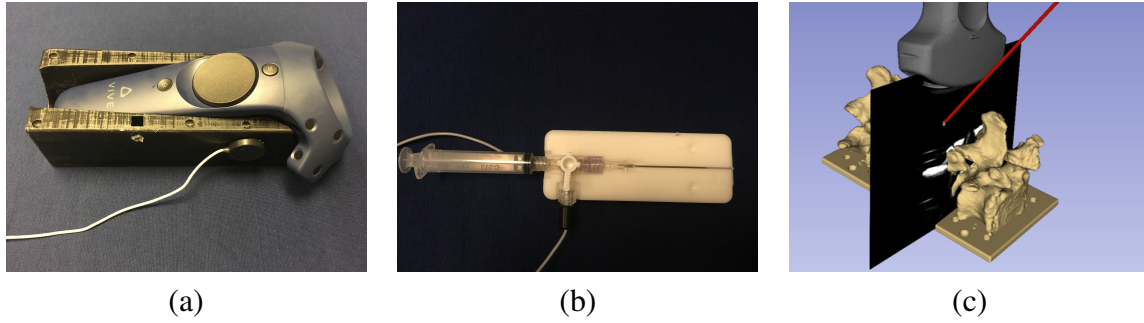


Figure 4.3: a) The co-calibration apparatus tracked by the VIVE controller and magnetic pose sensor b) the calibration apparatus for the syringe, and c) Visual representation of an example of tracked tools registered in the HMD's coordinate system, where the alignment between the models of the spine and needle and their reflections in the US image indicated the total system accuracy

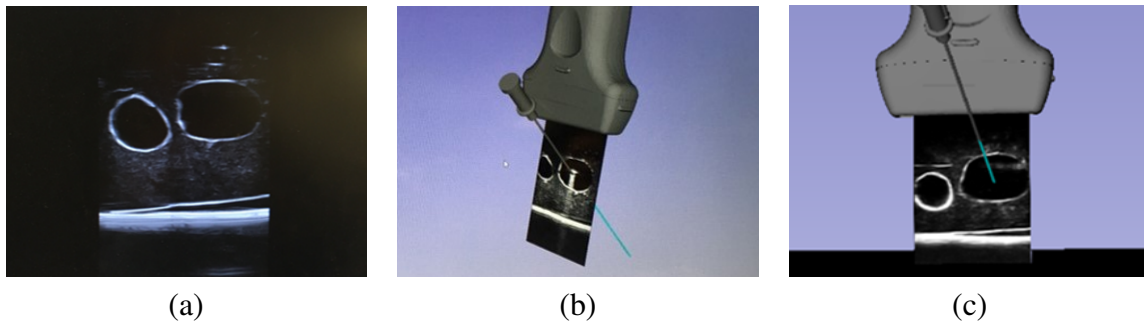


Figure 4.4: Visual representation of each mode of visualization where a) is the US-only system, b) is the monitor system, and c) is the HMD system. Images b) and c) comprise models of the US probe, needle, needle trajectory and the calibrated US image.

on an oblique angle such that the user has a 3D perspective, as depicted in Fig. 4.4 b).

Two groups were recruited for this study. The expert group comprised 31 clinicians, who have been trained in US-guided central line insertions and have performed over 15 clinical insertions as it has been shown that a user becomes proficient in US-guided CVC after 8 clinical insertions [105]. The expert cohort have performed 319 ± 359 clinical US-guided insertions with an average of 5.5 ± 3.3 years of experience. The second group consisted of 20 medical students, who did not have training or experience with US-guided CVC but are knowledgeable in neck vascular anatomy. All participants were recruited with consent according to the local REB regulation (Western University REB 107254). Each user was assigned a randomized order of the systems that was counter-balanced such that there was an equal number of users assigned to the same order. Prior to the experiment, each participant was briefed on and introduced to the needle insertion required for CVC using the neck phantom. The vasculature on the left side of the neck phantom was used to train the users on all of the systems. The participants were given time to perform needle insertions using the US-only, MR on a monitor, and MR in the HMD

system, in their assigned order, until they felt they were comfortable with all of the modes of visualization. The study was conducted using the vessels on the right-hand side of the phantom. The participant was required to perform one insertion into the vessel on the right-hand side of the phantom for each of the modes of visualization in their assigned order. Sufficient time was provided in between switching modes of visualization to allow the participant to rest and adapt to the new mode, such that the recorded data is not skewed by the clinician's memory of the spatial orientation of the phantom. The streaming US video, time, and tracked trajectories for the entire simulated procedure were recorded for post-analysis. After the experiment, the users filled out a questionnaire which is included in Appendix A Fig. A.1.

The recorded data was post-processed to extract procedure time and needle insertion accuracy to evaluate user performance. We defined procedure time as the time from the needle on the surface of the phantom until the user completed their insertion. Three separate metrics were used to analyze the needle insertion accuracy. Insertion success was defined as a binary metric where an insertion was denoted as successful if the final needle tip position was within the bounds of the target IJV mesh. Two continuous success metrics were obtained by measuring the distance from the final needle tip position to the central line of the vessel and the closest point on the vessel wall. The distance to the vessel was signed as it was denoted as negative if the final needle tip position was outside of the target IJV. The questionnaire responses were in the form of a continuous scale where the centre and two ends were anchored with written descriptions. If the user agreed with one of the given responses they could mark that part of the scale or anywhere along the scale. The questionnaire responses were converted into a numeric 10.0 scale, as the physical scales are 10 cm in length, and summarized in Tab. 4.5.

4.2.1 Statistical Analysis

The continuous results, such as time, distance to the central line and vessel wall, were confirmed to be normally distributed using the Anderson-Darling test. These continuous metrics were individually compared across the three conditions: US-only, MR system on a 2D monitor, and MR system in the HMD, using a repeated-measures ANOVA. The metrics that returned a p-value less than 0.05 from the ANOVA analysis underwent a least squared distance multi-comparison post-test with the Bonferroni correction to compare between each pair of conditions. This test returns the six possible comparisons between the metric being analyzed for the US-only, monitor, and HMD results with an associated p-value. For this test, a p-value less than 0.05 indicates that there is strong evidence that the means of the two groups are significantly different. Metrics that produced significantly different means are summarized in the graphs in Fig. 4.5. While, the time and number of CA punctures were not found to have any significant

difference between systems they are summarized in Tab. 4.4 as they are relevant. A chi-squared test was performed to compare the success rates between groups as the success rate is a discrete metric. For this test, a p-value less than 0.05 indicates that there is strong evidence that the frequency rates in the contingency table are significantly different across the groups. The success rates and p-values less than 0.05 are denoted in Tab. 4.3. The user questionnaire results were converted from the continuous scale to numerical values and summarized in Tab. 4.5.

4.3 Results

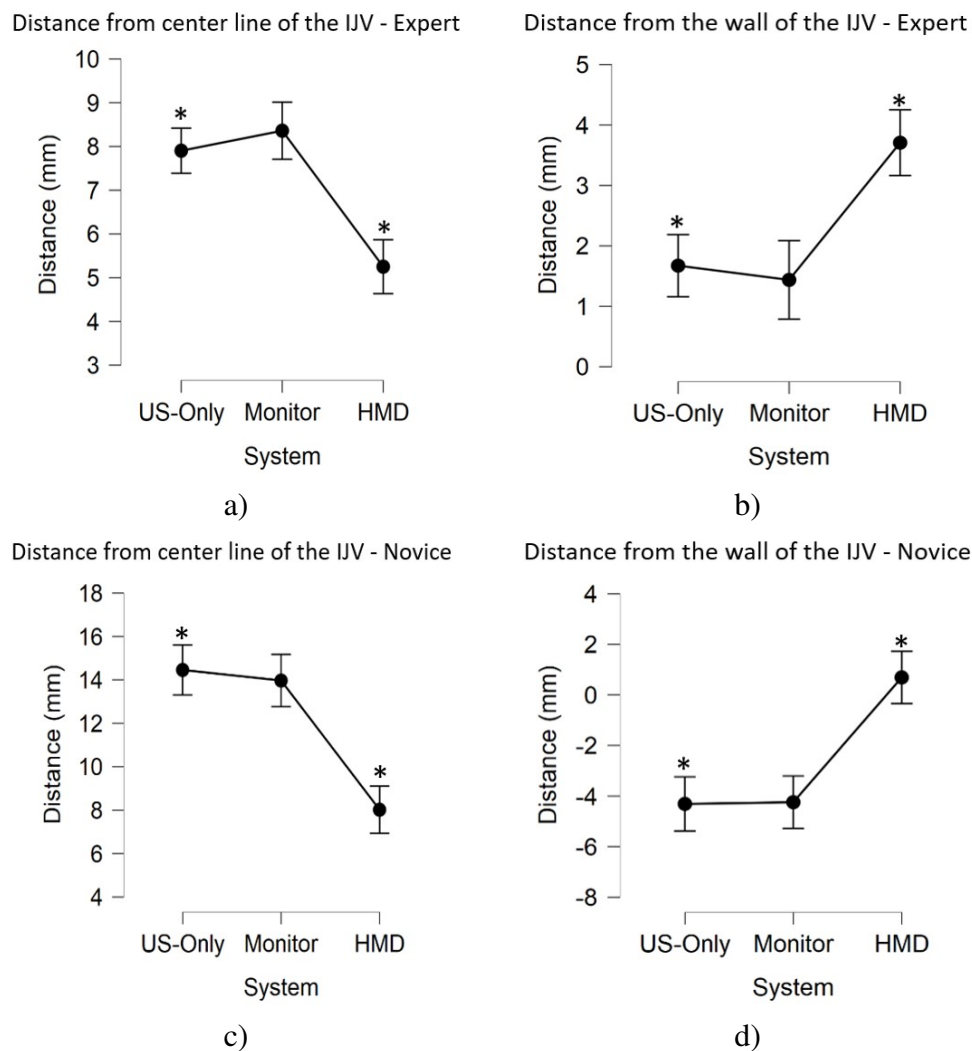


Figure 4.5: Distance metrics presented with respect to system where a) and c) summarize the distances from the final needle tip position to the closest point on the centre line of the vessel for experts and novices respectively and b) and d) summarize the distances from the final needle tip position to the closest point on the vessel wall for experts and novices respectively

Table 4.1: Post Hoc test results following repeated-measures ANOVA analysis for the distance from the final needle tip position to the closest point on the centre line of the vessel for both experts and novices.

Experts			95% CI for Mean Difference			
		Mean Difference	Lower	Upper	t	P _{bonf}
US Only	Monitor	-0.457	-2.538	1.623	-0.541	1.00
	HMD	2.652	0.571	4.732	3.136	0.008**
Monitor	HMD	3.109	1.028	5.189	3.677	0.001**
Novices			95% CI for Mean Difference			
		Mean Difference	Lower	Upper	t	P _{bonf}
US Only	Monitor	0.488	-3.593	4.569	0.300	1.00
	HMD	6.439	2.358	10.520	3.962	0.001**
Monitor	HMD	5.951	1.870	10.032	3.662	0.002**

** p < .01

Note. P-value and confidence intervals adjusted for comparing a family of 3 estimates (confidence intervals corrected using the bonferroni method).

Table 4.2: Post Hoc test results following repeated-measures ANOVA analysis for the distance from the final needle tip position to the closest point on the vessel wall for both experts and novices

		Mean Difference	95% CI for Mean Difference		t	p _{bonf}
			Lower	Upper		
US Only	Monitor	0.235	-1.756	2.227	0.291	1.00
	HMD	-2.035	-4.026	-0.043	-2.514	0.044*
Monitor	HMD	-2.270	-4.261	-0.279	-2.805	0.020*
		Mean Difference	95% CI for Mean Difference		t	p _{bonf}
			Lower	Upper		
US Only	Monitor	-0.067	-3.786	3.652	-0.045	1.00
	HMD	-5.001	-8.720	-1.282	-3.377	0.005**
Monitor	HMD	-4.934	-8.653	-1.215	-3.331	0.006**

* p < .05

Note. P-value and confidence intervals adjusted for comparing a family of 3 estimates (confidence intervals corrected using the bonferroni method).

Table 4.3: Summary of the success rates by system and the associated statistics

Success Rate (%)	US-only	Monitor	HMD	χ^2	P-value
Experts	67.7%*	64.5%	93.5%*	8.034	0.018*
Medical Students	25%*	25%	70%*	13.71	0.001*

Table 4.4: Summary of insertion time and number of users who punctured the CA under each mode of visualization.

Insertion Time	System		
	US-Only	Monitor	HMD
Experts	11.46 ± 6.42	9.94 ± 10.45	7.82 ± 4.41
Novices	11.48 ± 7.45	12.23 ± 7.26	11.49 ± 6.47
CA punctures	US-Only	Monitor	HMD
Experts	1	0	0
Novices	2	0	0

Table 4.5: User questionnaire results

Question	Average Score
How viable is the HMD to use in the OR?	4.35
If the system was clinically available how often would you use it?	4.75
How useful do you think the HMD system would be for training US-guided CVC?	7.04
How was the comfort associated with using the HMD?	7.64

4.4 Discussion

The experimental results demonstrate that the use of the HMD system to provide a first-person MR visualization significantly improved the accuracy of the IJV needle insertion required for CVC compared to the US-only system, for both expert users and medical students. Ninety-four percent (93.50 % or 29/31) of expert clinicians performed successful IJV insertions using the HMD system, an improvement from the US-guidance case where only 68 % (21/31) of expert clinicians performed successful insertions ($p=0.018$). Seventy percent (14/20) of medical students performed successful insertions when using the HMD system, compared to a 25 % (5/20) success rate for the same cohort using the US-only system (Tab. 4.3).

The HMD system also enabled these participants to perform more consistent targeting of the centre of the IJV compared to the US-only mode of visualization, for both expert users and

medical students. The study results showed that the average distance between the final needle position and the centre line of the IJV decreased from 7.9 ± 4.50 mm to 5.4 ± 2.30 mm ($p=0.008$) when the expert users performed the insertion with the HMD-system instead of US-guidance. The same trend was observed for novice medical students, where the average distance from the vessel wall decreased from 14.5 ± 7.10 mm under US-guidance to 8.0 ± 4.20 mm under HMD-guidance. Targeting the centre of the vessel is an ideal outcome for IJV cannulation as positioning the needle at the centre of the IJV decreases the risk of accidental punctures during the guidewire insertion [9].

Between the US-only and the MR system displayed on the monitor, our data showed that there was no significant improvement for either clinicians or novice medical students in terms of targeting accuracy. In our monitor-based MR system, the virtual representations of a tracked needle, needle trajectory, and US probe were visualized in a common coordinate system with the calibrated free-hand US image. The pose of the US video was un-constrained, and participants could adjust the vantage point freely, with most participants preferring an oblique viewing angle for an improved 3D perception. These findings are consistent with the work of Ameri *et al.* [2], who also found no significant improvements in the expert cohort when monitor-based MR guidance was used compared to the US-only mode of visualization. Their system maintained a stationary front-facing US image throughout the insertion to retain the familiar view of US images on an US console [2], whereas in our system the user can select an arbitrary 3D vantage point. In both systems, virtual representations of the tracked needle and its trajectory supplemented the 3D view of the surgical scene, with the intention to provide improved needle navigation. The lack of improvement in our monitor-based system and that proposed by Ameri *et al.* is likely due to the decoupling between the clinician's motor and visual field the US console or a monitor was used as visualization devices.

In contrast, the use of HMD, delivering an MR guidance system using a first-person immersion vantage point, significantly improved the outcome of CVC insertion. The CVC insertion success rate and accuracy (measured as the distance between the final tip location to the centre line of the vessel and from the vessel wall) were improved when compared to US-only guidance. These results emphasized the importance of using an HMD to ensure a coherent visual and motor field during needle guidance. In the HMD, this is achieved by bringing the needle guidance information directly into the line of sight of the clinician. While the monitor-based system would be more readily be integrated into a clinical workflow, our results suggested there are benefits for using an HMD for needle guidance, promoting the continued pursuit of research related to the use of HMDs in similar clinical scenarios.

In terms of complication rates, there was no significant difference in the rates of CA puncture between these 3 modes of visualization implemented for our experiment, as the overall

rates of CA puncture were low. Despite the lack of significant differences we reported the CA punctures for relevance. One expert and two novices punctured the CA and all punctures occurred under US-guidance. The observed low rates of CA puncture are likely due to the simplicity of our anthropomorphic neck phantom, as the IJV and CA had a simple orientation with limited overlap as it is laterally positioned to the IJV (Figure 6.1). An additional experiment designed to test the system on a more realistic and representative anthropomorphic neck phantom is planned for future work. As depicted in Figure 6.1, the appearance and configuration of the neck vasculature are variable. The utilization of US to guide needle insertion using has the inherent inability to track the position of the needle tip once it is traversed beyond the US image plane. In a realistic clinical scenario, the inadvertent posterior IJV wall punctures could result in damage to a critical anatomical structure adjacent to the IJV, including CA, but it is not observed often using our phantom. While our study did not show that the first-person immersive MR system had a significant effect on the rates of CA puncture, the HMD system resulted in a higher rate of successful insertions, as the participants positioned the needle closer to the centre of the vessel with less posterior wall punctures when compared to US-only guidance.

The feasibility of using an HMD in a clinical setting is a critical concern for using the first-person immersive MR for needle guidance. Results of the questionnaire responses suggested that the first-person immersive MR HMD system may be more impactful when used for training rather than for clinical deployment, as on average these clinicians ranked the clinical viability of the system a 4.35/10 compared to a 7.04/10 for usefulness for training. Most clinicians indicated that they would consider using advanced visual guidance on a case-by-case basis, as more complicated cases would benefit from a higher level of visual guidance. We are currently utilizing the HTC VIVE Pro as a virtual reality display device even though it has stereo cameras and can be used as an augmented-reality video-pass-through display. Incorporation of the stereo camera feeds for visual guidance may facilitate the clinical acceptability of this first-person immersive technology, as the clinical would be able to visualize the guidance information while maintaining direct views of the reality simultaneously. However, such an approach is not feasible with current devices, as the stereo camera image resolution and fidelity are low. Alternatively, hologram-based optical-see-through HMD devices such as the Microsoft HoloLens can be used, but extensive evaluation on accuracy tracking (camera hand-eye calibration) and careful integration with optical- or magnetic- tracking systems would be required.

Our results, in conjunction with those presented by Ameri *et al.*, suggest that visualizing the 3D spatial relationship of the surgical tools in a common coordinate system with a 2D US image does not provide sufficient information for a user to perceive the 3D context of the surgical scene, regardless of the viewing perspective of the US image. In contrast, screen-

based systems that use 3D US information, have been developed and successfully applied to focal liver tumour ablation [45], prostate brachytherapy [39], and breast biopsy [38] suggesting there is a potential pathway to implement a clinical-feasible monitor-based guidance system for CVC insertion. Developing and incorporating advanced 3D visualization technique to improve 3D perception using 2D monitor is one of the future projects. Alternatively, a tablet-based display that could be situated closer to the surgical site could facilitate a more coherent visual and motor field coordination.

4.5 Conclusion

This chapter describes the development of an advanced needle guidance system that renders tracked tools such as the US probe, needle, and needle trajectory as well as the calibrated US image on a 2D monitor or within an HMD. This research aimed to compare needle insertion performance using US-only guidance, the advanced guidance system on a monitor, and the advanced guidance system in the HMD. Thirty-one expert clinicians and 20 medical students were trained on all three systems and then used each system to perform a needle insertion on the phantom. The HMD system significantly improved the number of successful needle insertions, as 94% clinicians and 70% of medical students had a final needle tip position within the vessel's lumen compared to 68% and 25% for US-only guidance, for clinicians and medical student's respectively. The HMD system also significantly improved the distance from the final needle tip position to the central line of the vessel. Clinicians using the HMD system had an average distance 5.40 ± 2.29 mm compared to 7.90 ± 4.51 mm using the US-only system, meaning that they were consistently closer to the centre of the lumen of the vessel compared to the US-only approach. The monitor system did not show any significant improvements compared to the US-only system for either group, as it was limited by the 2D US information. The user's ability to interpret the relationship between the tools and anatomy is imperative to successful needle navigation and can be achieved by using an HMD to align the visual and motor fields to provide a 3D perspective.

Chapter 5

Automatic segmentation of the carotid artery and internal jugular vein from 2D ultrasound images for 3D vascular reconstruction

This chapter is adapted from the following manuscript:

- Groves, L. A., VanBerlo, B., Veinberg, N., Alboog, A., Peters, T. M., & Chen, E. C. S.(2020). Automatic segmentation of the carotid artery and internal jugular vein from 2D ultrasound images for 3D vascular reconstruction. *International Journal of Computer Assisted Radiology and Surgery*, 15(11), 1835-1846.

5.1 Introduction

Percutaneous internal jugular vein (IJV) needle insertions are used to access the central venous system [17], with carotid artery (CA) punctures being one of the most common and severe complications that can occur during IJV cannulation [17]. Ultrasound-(US)-guided needle insertions have the potential to reduce complications by providing clinicians with a real-time cross-sectional view of the neck anatomy to visualize the relationship between the IJV and CA in 2D [49, 134]. The high variability in human neck vascular morphology across the patient population [49, 148] has motivated research efforts in the development of advanced US-based surgical navigation systems [2, 50], along with the characterization of neck vasculature morphology to further assist and improve central venous catheterization (CVC) [49, 148].

US imaging has been employed to analyze the effect of the anatomical relationship between

the IJV and CA on CVC [49, 148] and the relationship between head rotation and diameter of the vessels [95, 156]. For these applications, anatomical structures must be segmented from the US images. Manual segmentation by an expert is the gold standard, but the process is labour-intensive and susceptible to human error [108]. Patient data derived from 2D US alone has limitations, as a single cross-sectional slice cannot adequately represent the entire structure. One example of a measurement that requires 3D information, is the assessment of the variability of the location of the CA bifurcation, which to date has been performed using excised vessels from cadavers [82, 160]. Vascular dissection is a time-consuming process that sacrifices the structural integrity and normal physiological properties found *in vivo*. The preferred situation would be the use of 3D automatic segmentation of the vessels from US to reflect patient positioning at the time of intervention.

The degree of manual analysis required to quantify trends in vascular anatomy has prompted work such as automatic segmentation of the media-adventitia and lumen-intima boundaries of the CA from 3D US images [164], the inner lumen of the CA in a longitudinal orientation [161], and CA plaques [150]. As far as we are aware, there is no method in the current literature to simultaneously and automatically delineate both the IJV and CA within a 2D transverse US image. Such a procedure would allow for the automatic analysis of the morphology and anatomical relationships of these vessels, and to enable accurate reconstruction of 3D volumes of the neck vasculature without exposing the patient to radiation, removing barriers for further research on the morphology of neck vasculature. Other applications of these vascular reconstructions include, but are not limited to, real-time intra-operative use or during preoperative planning to augment guidance for CVC. Therefore, the secondary motivation of this work is the development of 3D models of the vasculature, which could be used to develop a more clinically relevant navigation system, while maintaining 3D information.

The U-Net Convolutional Neural Network (CNN) architecture has been applied to automatically segment regions of interest associated with the CA [164, 150]. U-Net is a semantic segmentation architecture, meaning the goal is to label each pixel in the image with a corresponding class that was provided in training examples to obtain pixel-wise label maps [131]. Each pixel is classified as either belonging to the background or to one of the foreground classes that were provided during training [131]. For certain U-Net applications, false segmentations occur due to the inability of the network to differentiate between regions that contain pixels of a specific class and regions that contain pixels with similar features to the class of interest. Two methods to compensate for this issue of false segmentation include i) post-processing steps to retain the largest segmentation [161], or ii) cropping the input to a region of interest (RoI) that contains only the anatomy of interest [150]. Mask R-CNN is an alternative to U-net for automatic segmentation and was inspired by Faster R-CNN [58] for object detection [129]. The

Mask R-CNN algorithm is capable of producing highly accurate segmentations as the architecture includes a Region Proposal Network (RPN) that determines possible bounding boxes or regions of interest (RoI) and probabilities of the occurrence of each object class within the bounding box [58]. The remainder of the algorithm predicts pixel segmentations within each RoI [58]. Thus, the Mask R-CNN algorithm can automatically predict objects without pre-processing to crop the image to a RoI. Mask R-CNN has recently been applied to medical image processing tasks including the detection and segmentation of meniscus tears [31] and segmentation of the prostate gland and prostatic lesions in MRI images [32]. Other applications include a modified Mask R-CNN for breast tumour detection and segmentation in US images [81]. These successes have motivated the investigation of a Mask R-CNN deep learning solution to automatically segment the CA and IJV from tracked 2D US images and to reconstruct the 3D vessels' surfaces for guiding intra-operative interventions.

The objectives of this chapter are twofold. First, we aim to develop an automatic segmentation framework capable of delineating both the CA and IJV from transverse US images, with an accuracy comparable to that obtained by manual segmentation. We then aim to formulate a vessel reconstruction pipeline to utilize these automatic vascular segmentations and spatial tracking to reconstruct the 3D geometries of the CA and IJV, with an accuracy comparable to that provided by reconstructions from CT angiography. These capabilities have the potential to automate vascular measurements in 2D and 3D and to improve US-guided needle interventions by applying the algorithm during or after the scan.

5.2 Materials and Methods

5.2.1 Data Collection

All images were collected using an Ultrasonix US scanner (SonixTouch, BK Medical, USA) with the L-14-5 Linear US transducer. As vascular structures can be as deep as 5.50 cm [49], an imaging depth of 6 cm was used to acquire neck vascular US images to enable the inclusion of all human vascular configurations. This US probe was spatially calibrated [20] and tracked using a magnetic tracking system (Aurora Tabletop, NDI, Canada). The US calibration provides the spatial pose of the US image with respect to the magnetic tracker's coordinate system.

The scanning protocol begins with the transducer placed between the two heads of the sternocleidomastoid muscle just above the clavicle, and ends at the mandibular border, proceeding in an inferior-to-superior direction. Nine normal control US scans of healthy volunteers were performed by a medical student specifically trained in this procedure, with each subject being imaged in two positions employed in clinical practice: supine on a horizontal table. A third-

year anesthesia resident performed an additional 6 scans on patients in a local hospital, with patients laying horizontally in a standard hospital bed. All data were collected with a sampling rate of 60 frames per second. All participants were recruited with consent according to the local REB regulation (Western University REB 114604). The CA and IJV were manually segmented from these US images by a medical student with experience in US neck imaging using 3D Slicer, such that each image had a corresponding mask for the CA and IJV ¹.

The complete data set comprises 2439 US images from 15 subjects containing cross-sectional views of the neck vascular anatomy. The US images were stored as 8-bit bitmaps, having pixel intensities in the range of $[0, 255]$. To validate our machine learning models a 4-fold cross-validation approach was performed. First, this data set was partitioned into 4 unique training, test and validation sets. Each training set comprised a unique combination of scans and their masks from 11 subjects (70-78% of the data set). Each test and validation set consisted of unique combinations of images from both a normal control and a patient, as well as their respective labels. The test and validation sets comprise 15-23% and 5-7% of the data set, respectively. The number of images included in each data set is summarized in Table 5.1. The number of images within the sets differs for each fold as the number of images associated with each scan varies due to differences in subject anatomy, as the scanning protocol was defined based on anatomical landmarks rather than a specific scan length. The data sets were partitioned such that all images from the same subject are isolated to the same set resulting in a slightly different number of images in each set. All images included in either the test and or validation sets were excluded from the sets used to train the network, as they were employed solely for evaluation to avoid bias. Each of these training sets was augmented by randomly scaling by a factor in the range 0.80 to 1.20 and rotating by an angle in the range of -15° to 15° , to produce images that represent possible variation that may occur during scanning. These transformations were automatically applied during training. During this process, these test and validation sets were used to evaluate the Dice Score of the trained model to form a baseline accuracy across normal and patient data. However, for analysis, the images within the training and test sets for each fold were reorganized based on whether they had been derived from a normal control or patient subject. The images that comprise these control and patient data sets were not used to train the fold that they would be evaluating. These control and patient images were analyzed using the Dice score, recall, and precision. This control patient split was selected to provide a more in-depth analysis of the applications of these networks on control and patient data independently, as well as on the overall accuracy across a mixed cohort.

¹<https://github.com/VASST/AIVascularSegmentation>

Table 5.1: Summary of number of images allocated to each data set used to train and evaluate the networks. The percentage of normal control images in the training, test and validation sets are denoted in brackets. The number of normal and patient images for each fold are the images that were used to evaluate the networks.

Dataset	Fold A	Fold B	Fold C	Fold D
Training Images	1717 (81%)	1799 (81%)	1902 (87%)	1819 (81%)
Test Images	552 (95%)	475 (88%)	352 (80%)	481 (83%)
Validation Images	170 (53%)	165 (47%)	185 (42%)	139 (53%)
Control Images	623	498	360	474
Patient Images	99	142	177	146

5.2.2 Deep Learning Segmentation

Computational hardware used for training the networks included an Intel® Xeon® E5-2683 v4 CPU workstation running at 2.10 GHz and 2 NVIDIA® Tesla® P100 GPUs, each with 12 GB of memory. All code was written in Python and executed on SHARCNET (Compute Canada’s high-performance computing network). We trained two neural network models: one with the Mask R-CNN architecture, and one with U-Net CNN for automatic vessel segmentation [58]. Both networks were trained using identical data sets. Memory and computational requirements during training and inference were decreased by resampling the images from 589×374 to 256×256 pixels with bilinear interpolation.

The implemented U-Net architecture was motivated by the standard U-Net encoder-decoder architecture [131]. The encoder consisted of 3 blocks of 2 convolutions with a kernel size (k) of 3, followed by a max-pooling layer with $k = 2$. The bottleneck comprised 2 consecutive convolutions with $k = 3$, while the decoder used 3 blocks of up-convolutions and 2 subsequent convolutions with $k = 3$. The decoder’s blocks also received residual connections from the output of blocks in the encoder of the same shape. ReLU was used as the activation function for all intermediate layers. The output layer was a single convolution with $k = 1$ that employed the softmax activation function over the background and classes, producing an output with the same dimensions as the input image. The network was trained to minimize the categorical cross-entropy loss function. The learning rate (α) was set to 0.0001 at the start of training. During training, if the validation loss did not decrease after the most recent 3 epochs, α was multiplied by 0.50. To encourage regularization, early stopping was applied to halt training when the validation loss did not decrease during the 10 most recent epochs [122]. Each fold was trained over the following number of epochs: set A ran for 33 epochs, set B ran for 19 epochs, set C ran for 27 epochs, and set D ran for 19 epochs. As U-Net is susceptible to

false segmentations, a connected-component post-processing algorithm was applied to keep the largest connected segmentation for both the IJV and CA and remove all other segmentations, as described in work by Xie. *et al.* [161].

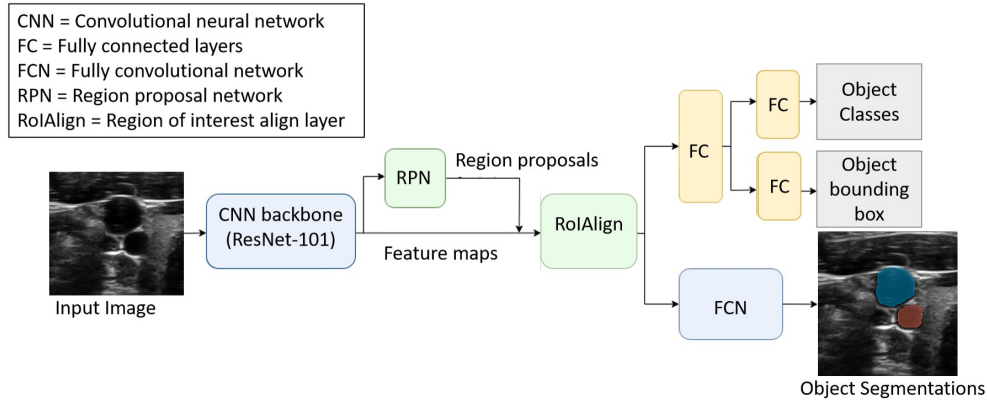


Figure 5.1: The Mask R-CNN architecture depicting the CNN backbone, the Region Proposal Network, and the RoIAlign Layer. The “box head” is a series of Fully Connected layers that outputs the predicted class and bounding box for each object. The “mask head” outputs binary segmentation masks for each object instance.

A Mask R-CNN model requires ground truth segmentation masks and bounding boxes for training. The bounding boxes were generated automatically by calculating the smallest rectangle that would enclose an individual vessel segmentation, defined by a 4-tuple consisting of two (x, y) coordinate pairs. The input to the Mask R-CNN model was the resized raw US image. The output of the model was a series of 256×256 masks, bounding boxes, and classes for each predicted vessel instance. In the rare case that there were more than two object masks predicted by the network, we considered only the two predicted with the highest confidence. The code to define and train the neural network model was adapted from Matterport’s Mask R-CNN implementation, which was built using the Keras library with the TensorFlow backend [59]. No changes were made to the core Mask R-CNN architecture. Our model segments objects of two classes: CA and IJV. Although the image background may be considered as a third class, no background segmentation masks are predicted by the network. Matterport’s implementation [59] offered the choice between ResNet-50 and ResNet-101 as the backbone of the network. ResNet-50 was chosen here because it contains significantly fewer parameters, lending itself to faster training and prediction time [57]. Multiple hyperparameters were tuned by performing several training experiments and adjusting the value of one while keeping others constant. The square anchor boxes used in the RPN had side lengths of 8, 16, 32, 64, and 128 pixels. Sixty-four regions of interest (RoIs) were fed to mask and classifier heads of the network for each image. The RPN non-max suppression threshold was set to 0.70, while

the learning rate (α) was set to 0.001 at the start of training. During training, if the validation loss did not decrease after the most recent 15 epochs, α was multiplied by 0.75. The batch size was 16 and was spread equally across 2 GPUs during training. The model was trained for 100 epochs to minimize the Mask R-CNN loss function $L = L_{cls} + L_{box} + L_{mask}$, where L_{cls} and L_{box} are defined as they were for Fast R-CNN [58], L_{cls} is the categorical cross-entropy loss for object classification, and L_{box} is the smooth L1 loss for bounding box localization [47]. Localization is defined as a 4-tuple consisting of an (x, y) coordinate, width, and height. L_{mask} is the mean per-pixel binary cross-entropy loss across segmentation masks for both classes [58]. The neural network was trained to minimize the loss function, with the object segmentation with the highest probability being selected for each class.

5.3 Vessel reconstruction

The automatically segmented label masks and the spatial tracking of the US probe were used to reconstruct the vessels in 3D. The probe tracking data was adjusted using the probe calibration provided the pose of each image with respect to the 3D spatial tracker's coordinate system, such that the automatic segmentations are positioned with respect to the field of view of the image from which it was captured. Three-dimensional binary morphological hole filling, with an annulus-shaped kernel of size $[30, 30, 30]$, was used to fill the gaps between the slices [140]. A 3D Gaussian blur filter with an α of 0.50 was applied to smooth the vessels, as visually depicted in Fig. 5.2. The four trained Mask R-CNN algorithms were used to obtain surface reconstructions on one of the patient left-side scans that was not used to train or evaluate any of the folds. As the US scanning protocol consistently collected scans beginning just superior to the clavicle, the CT segmentations started just superior to the clavicle and ended at approximately the same location as the most superior US image. The point data was extracted from the US and CT reconstructed vessels and was used to solve a rigid surface-based registration using an iterative closest point algorithm [8] to place the models in a common coordinate system for comparison. The reconstruction algorithm was evaluated through surface-to-surface distance comparisons between the US and CT reconstructed vessels after registration. The volume and surface area (SA) of the reconstructed vessels from US and CT were calculated and expressed as a ratio of the metric extracted from US to that extracted from CT, where the smaller value of the two was used as the numerator such that the ratios would be less than 1 to not bias the average.

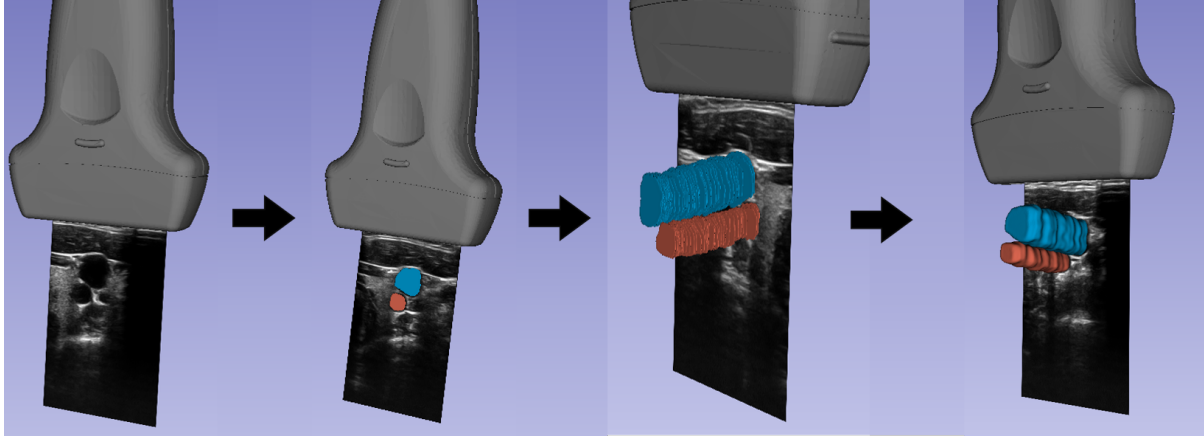


Figure 5.2: Visual depiction of the reconstruction process. The left-most figure shows a calibrated US image positioned and scaled to the true field-of-view of the US beam. The second graphic demonstrates a segmented and calibrated US image, where the CA and IJV have been delineated. The third panel depicts a vascular skeleton where each image in the tracked scan has been segmented and spatially positioned using the calibration to form a skeleton. The final figure depicts the closed surface reconstructed vessels after the application of binary morphological hole filling and Gaussian blur smoothing.

5.4 Results

Four-fold cross-validation was performed, whereby all 2439 collected and segmented images were allocated into training, test, and validation sets in four unique combinations. During training, the test and validation sets each comprised one patient and one normal control scan. The images that were excluded from training were reorganized into patient and control data sets for evaluation. The manual and automatic segmentations produced by the Mask R-CNN and U-Net algorithms were compared by calculating the Dice Score, recall, and precision across each class are summarized in Figures 5.3 and 5.4, for the CA and IJV respectively. The average Dice, recall and precision values across all folds and all evaluation images are summarized in Table. 5.2. Four sample images were selected to show the potential issues that occur with the U-Net segmentation and post-processing are depicted in Fig. 5.5.

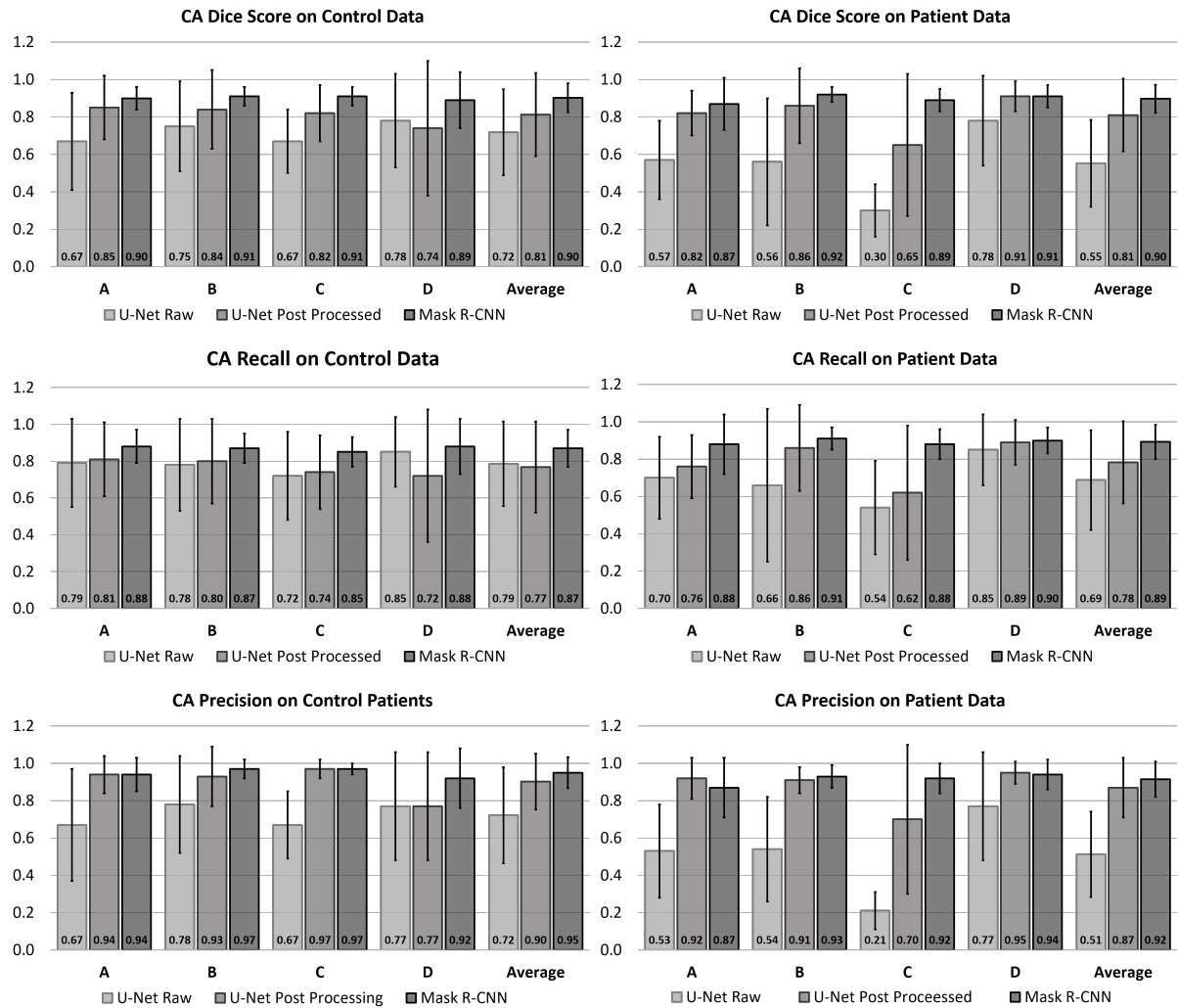


Figure 5.3: Average Dice, Recall, and Precision for the CA from each of the four-folds. These results are reported separately for normal and patient data.

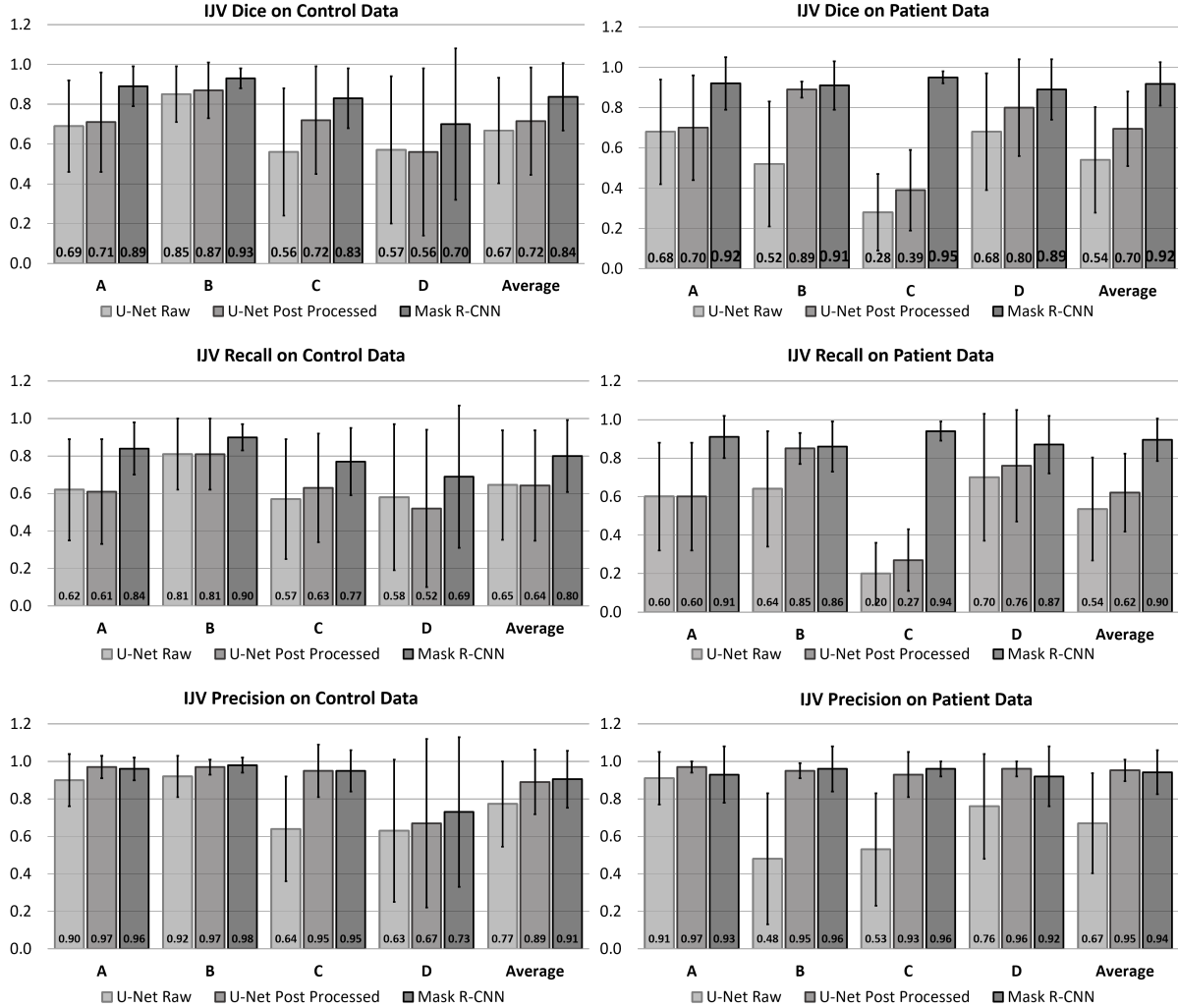


Figure 5.4: Average Dice, Recall, and Precision for the IJV from each of the four-folds. These results are reported separately for normal and patient data.

Table 5.2: Summary of the Dice, Recall, and Precision averaged across the patient and control data from all four-folds for the raw U-Net, post-processed U-Net, and Mask R-CNN. These data are presented for the IJV and CA separately.

Metric	U-Net Raw	U-Net Processed	Mask R-CNN
IJV			
Average Dice Score	0.60 ± 0.26	0.71 ± 0.23	0.88 ± 0.14
Average Precision	0.59 ± 0.28	0.63 ± 0.25	0.85 ± 0.15
Average Recall	0.72 ± 0.25	0.92 ± 0.12	0.92 ± 0.14
CA			
Average Dice Score	0.64 ± 0.23	0.81 ± 0.21	0.90 ± 0.08
Average Precision	0.74 ± 0.25	0.78 ± 0.23	0.88 ± 0.10
Average Recall	0.62 ± 0.24	0.89 ± 0.16	0.93 ± 0.09

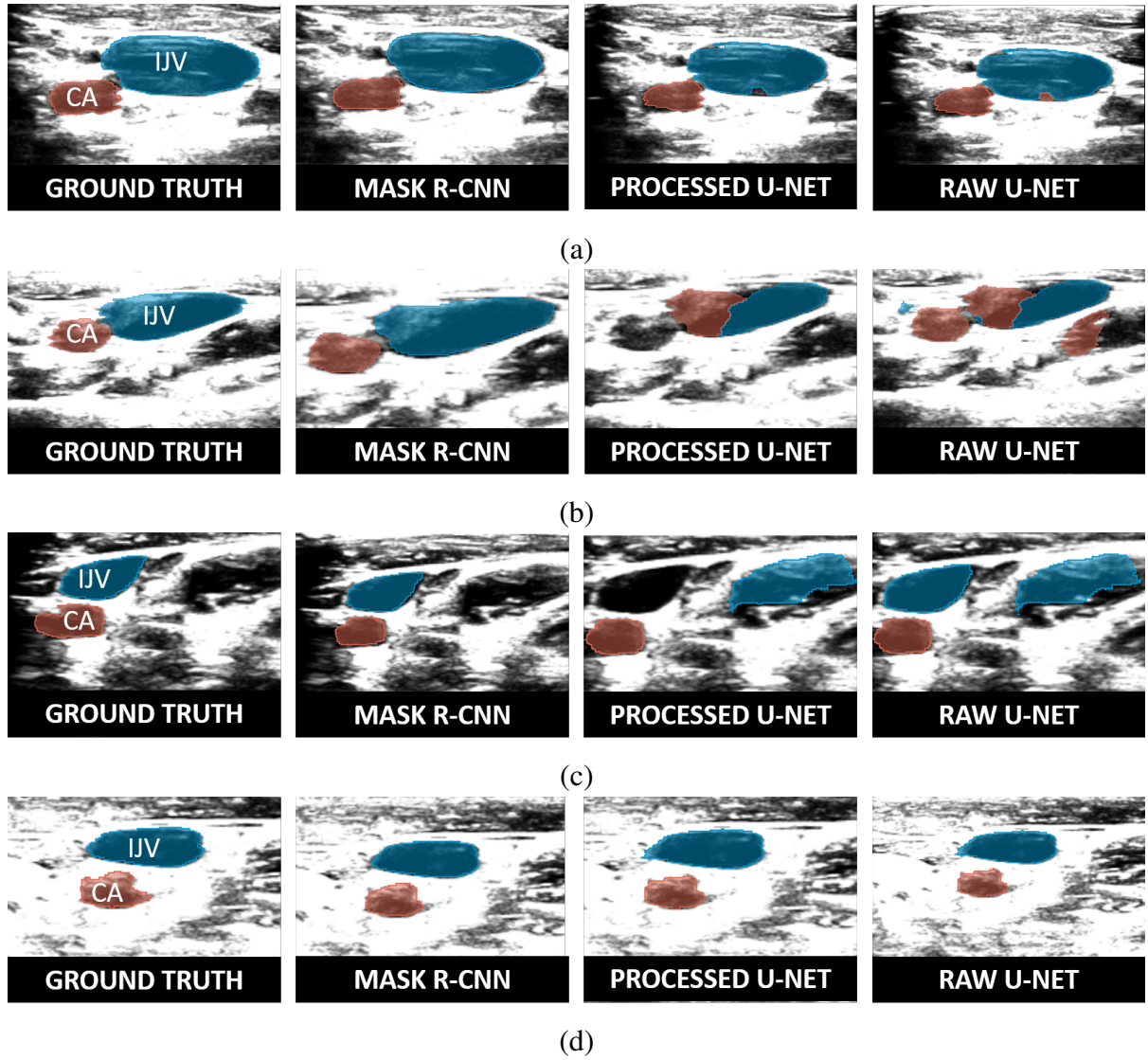


Figure 5.5: Four sample images with their respective outputs from the Mask R-CNN and U-Net with and without processing. Row a) shows an example of a small cluster of misclassified pixels from the U-Net. Row b) provides an example of a large group of pixels that have been misclassified as CA when they should be IJV in the U-Net output. Row c) depicts a vessel-like structure that has been misclassified as the IJV and the post-processing algorithm selecting this false segment as the IJV. Row d) depicts an image that has accurate outputs across all 3 algorithms. Despite the U-Net output from images a-c containing erroneous segmentation, the Mask R-CNN produced accurate segmentations across all sample images.

The surface-to-surface distance between the registered vessel models from all four-folds are depicted in Fig. 5.6 and 5.7, where colours progress from blue (cool) to red (hot) as the distance increases. The SA and volume ratios of the values extracted from the four Mask

R-CNN reconstructions and the CT vessels and the average are summarized in Table 5.3.

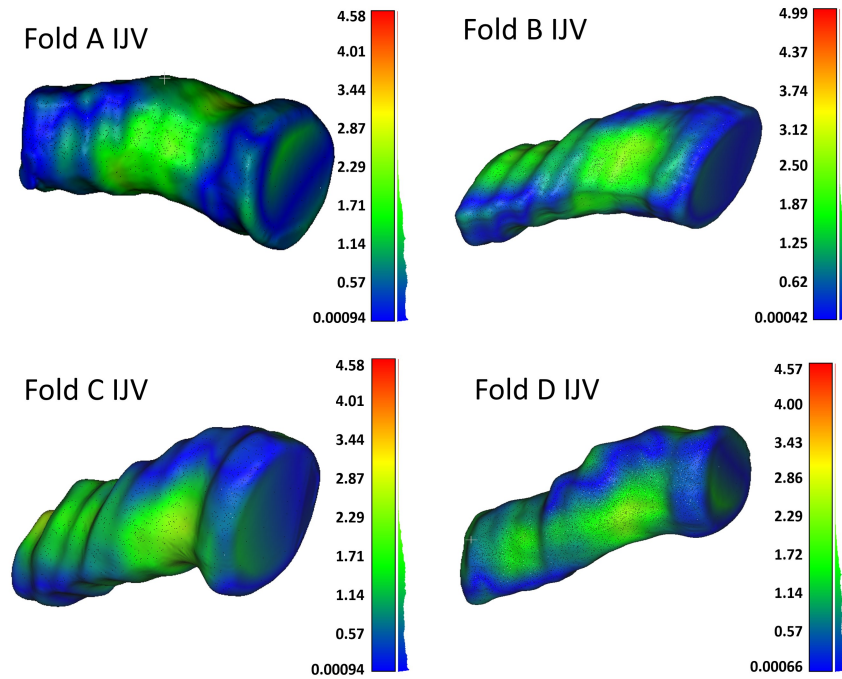


Figure 5.6: IJV surface-to-surface distances between the reconstructed US and the ground truth CT for all four-folds. The colour progresses to warm colours as distances increase.

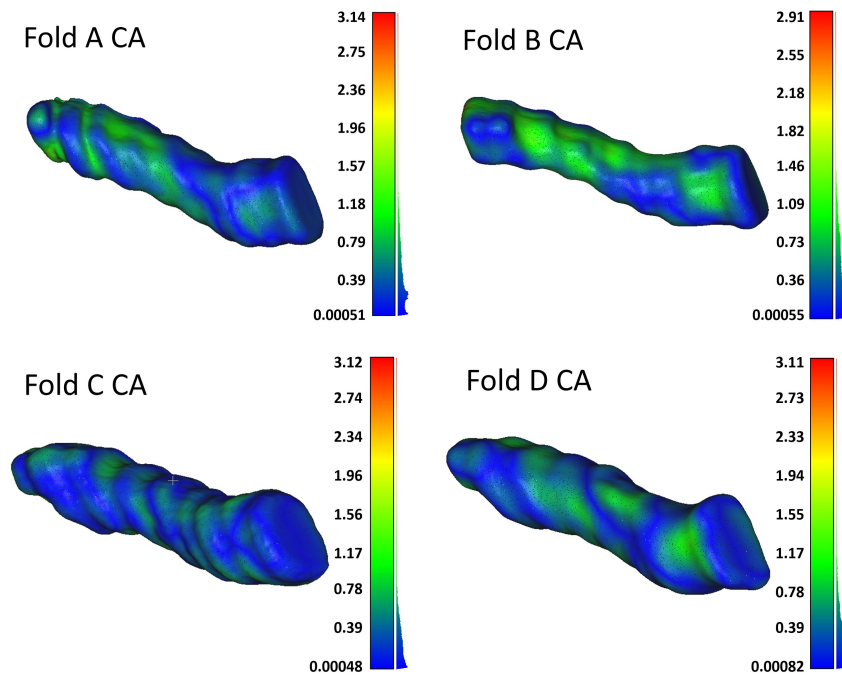


Figure 5.7: CA surface-to-surface distances between the reconstructed US and the ground truth CT. The colour progresses to warm colours as distances increase.

Table 5.3: Summary of the SA and volume ratio between the metrics produced from the US reconstructions from the four trained networks to the metrics extracted from the CT segmented vasculature.

Fold	Fold A	Fold B	Fold C	Fold D	Average
CA SA ratio	0.97	0.84	0.96	0.98	0.94
IJV SA ratio	0.94	0.85	0.81	0.93	0.88
CA volume ratio	0.87	0.70	0.94	0.93	0.86
IJV volume ratio	0.95	0.82	0.75	0.93	0.86

The four representative vasculature reconstructions are visualized with respect to the calibrated US image for reference in Fig. 5.8. These subjects did not have associated neck CT scans and therefore a more comprehensive analysis could not be performed.

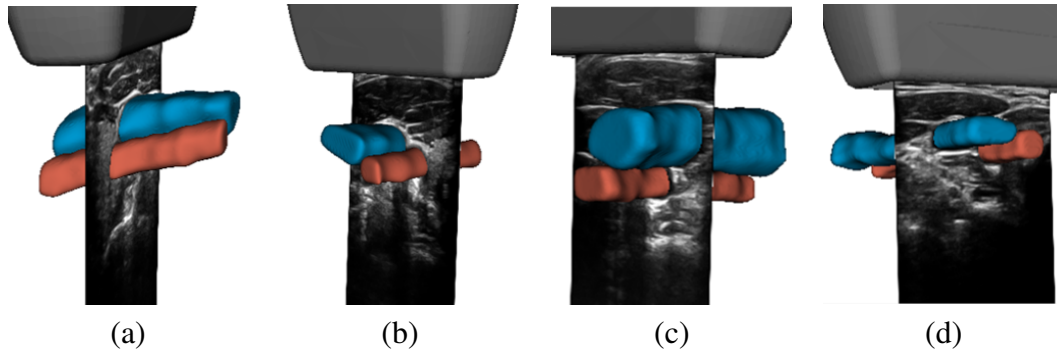


Figure 5.8: Each letter (a, b, c, d) represents a unique human subject who was not used to train the algorithm used to produce the segmentations.

5.5 Discussion

This work aimed to develop a pipeline that produces automatic vascular reconstructions for the CA and IJV from tracked transverse US scans, for the purpose of vascular morphology analysis or surgical navigation. In this work, we compare U-Net and Mask R-CNN algorithms, both capable of automatically segmenting the CA and IJV from transverse US images. The automatic segmentations produced eliminate the requirement for manual delineation of vessels to obtain vascular measurements. The Mask R-CNN segmentation algorithm was used within a vascular reconstruction pipeline that outputs closed 3D surface models of the vessels which can be used for 3D vascular morphology analysis or surgical navigation.

U-Net is a semantic segmentation algorithm where each pixel is assigned to a class. Our implementation produces a label map, where each pixel has been assigned to one of three

classes: background, CA, or IJV. The raw output of the U-Net may produce multiple clusters of pixels labelled as either the CA or the IJV, with some pixels being misclassified as seen in Fig. 5.5. These erroneous segmentations motivated the use of a post-processing step to identify one segmentation for each of the CA and IJV classes. A major factor that contributes to the high number of false segmentations is the non-unique appearance of the neck vascular structures under US. The CA and IJV are vascular trunks with several branching vessels that have similar features under US. The CA and IJV are the major vascular structures in the neck and should be the largest vascular structures in the US images acquired. For this reason, similar to the work of Xie. *et al.* [161], we applied a post-processing step that identifies the largest connected component for each of the CA and IJV classes. The average dice score for the CA and IJV for the post-processed U-Net are 0.71 ± 0.23 and 0.81 ± 0.21 , respectively. Applying this post-processing step to the U-Net segmentations improved the Dice Score by 0.11 and 0.17 compared to the raw U-Net output, for the IJV and CA respectively. This post-processing algorithm fails in cases where an erroneous segmentation has the largest number of connected components, and thus the post-processing selects the wrong cluster of pixels (Fig. 5.5c). Moreover, the U-Net output commonly misclassifies pixels between the CA and IJV (Fig. 5.5b), an issue that would persist regardless of the post-processing algorithm applied. Both of these issues contribute to the small change in Dice scores. As the accuracy of the post-processed U-Net was still lower than desired, for this application we investigated the use of Mask R-CNN.

The Mask R-CNN contains a regional proposal sub-network that identifies bounding boxes within the image where segmentations are most likely to occur. The algorithm then segments these structures within the bounding box and returns a probability that they belong to the class to which they have been assigned by the label. The output of our Mask R-CNN algorithm selects the segmentation with the highest probability of belonging to the CA or IJV class. Thus, our algorithm returns a single fully-connected segmentation for the CA and IJV based on a trained statistical probability with a reduced number of misclassified pixels. The average dice score for the IJV and CA for the Mask R-CNN are 0.88 ± 0.14 and 0.90 ± 0.08 , respectively. The Mask R-CNN improved the Dice Score by 0.17 and 0.09 compared to the post-processed U-Net, for the IJV and CA, respectively. Mask R-CNN provides an alternative to U-Net and is particularly beneficial for segmentation problems where there are many structures that contain similar features to the objects that aim to be segmented. Mask R-CNN does not require any pre- or post-processing steps, as the RPN selects the segmentation based on the statistical probability associated with the bounding box and thus the entire object segmentation, rather than a per-pixel basis. We note that while Mask R-CNN provides more accurate segmentations compared to U-Net, it does so at a higher cost of computational requirements. With advances in modern GPU hardware, the potential to execute Mask R-CNN for vascular reconstruction in

real-time is high. Overall, the Mask R-CNN achieved average Dice Scores, recall and precision of values above 0.85, which are sufficiently accurate to be used for vascular reconstruction and measurements pertaining to the relationship between vessels.

We used all four trained Mask R-CNN networks to obtain vascular US surface reconstructions of the CA and IJV on a patient scan that was not part of the training or evaluation data sets. Each reconstruction was compared to a manually segmented CT scan of the same patient's vasculature, using a surface-to-surface distance analysis (Fig 5.6 and 5.7). The reconstruction of the CA is slightly more accurate than the IJV, as the IJV is susceptible to deformation under the pressure of the US probe during scanning, and thus is more representative of the true accuracy of the reconstruction. We calculated the ratio of the SA and volume values extracted from the US to the values from the CT reconstructed vessels, as summarized in Table. 5.3. On average, the SA ratio was 0.94 and 0.88, for the CA and IJV respectively. The average volume ratio was 0.86 for both the CA and IJV. The errors present in the Mask R-CNN results are typically in the form of a loss of detail at the border of the vessel lumen, details that have minor effects on the ability to use these reconstructions for surgical navigation or vascular measurements. With the majority of points being within 2 mm of the CT reconstructed vessels and with sub-millimetre differences in resulting registration metrics, this algorithm is capable of producing accurate vascular reconstructions.

The US-based vascular reconstruction pipeline using Mask R-CNN could be useful for a variety of applications. We believe US-guided central line insertions may benefit from the integration of 3D reconstructed vasculature into advanced surgical navigation systems. This algorithm can also be used to quantify the relationship between the CA and IJV in 3D to expand upon the 2D vascular morphology research previously performed [49, 148]. Additionally, the multi-class segmentation using Mask R-CNN can trivially be extended to include additional pathologies or reconstruct various anatomical structures. One possible extension is the segmentation of calcified plaques which have a non-unique appearance in US images, and relying on a network such as U-Net or algorithms based on feature detection would likely result in many incorrect segmentations of plaques. As the size of plaques can vary drastically, a more rigorous post-processing selection algorithm would be required. The Mask R-CNN is more suitable than U-Net for this type of application, as it provides a statistical method for selecting the appropriate segmentation, which is important for multi-class segmentation problems where features are not inherently unique to the structure of interest.

5.6 Conclusions

In this work, we compared Mask R-CNN and U-Net algorithms developed to automatically segment the CA and IJV from transverse US images. The Mask R-CNN algorithm was more accurate than the U-Net alternative and achieved average Dice scores of 0.88 ± 0.14 and 0.90 ± 0.08 , for the IJV and CA respectively. The Mask R-CNN-based vascular reconstruction pipeline was accurate compared to the CT equivalent, with the majority of measured distances between the surfaces being less than 2 mm. These reconstructions were able to produce accurate metrics with the average ratio of the volume produced by the US, to the volume produced by the CT, being 0.86 for both the CA and IJV. This work can be used to analyze neck vasculature morphology in both 2D and 3D. Furthermore, the 3D models can be used for surgical planning or surgical navigation. Overall, we have developed and evaluated a highly accurate Mask R-CNN algorithm for instance segmentation of the CA and IJV in transverse US images, which can be used for vascular reconstruction.

Chapter 6

Improving central line needle insertions using in-situ vascular reconstructions

This chapter is adapted from the following manuscript:

- Groves, L. A., Li, N., VanBerlo, B., Veinberg, N., Peters, T. M., & Chen, E. C. (2020). Improving central line needle insertions using *in-situ* vascular reconstructions. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 1-7.

6.1 Introduction

Central line insertions at the site of the internal jugular vein (IJV) are prone to high-risk complications, such as carotid artery (CA) puncture, with clinical CA puncture rates around 7 % for US-guided cases [93]. To address these high complication rates in Chapter4, an advanced surgical navigation system was developed that visualized tracked models of the needle, needle trajectory, and US probe in a common frame of reference with the tracked and calibrated US image. This system was rendered in an HMD and on a screen, and a cohort of 31 clinicians and 20 medical students tested the systems to evaluate if they provide more accurate targeting compared to US-guidance for needle puncture into the IJV. The system displayed in the HMD demonstrated significant improvements in the success rates and targeting accuracy compared to US-only guidance [50]. However, when the system was displayed on the monitor in a 3D perspective, there was no significant difference in insertion accuracy compared to US-guidance. In general, the clinical practitioners that evaluated our system believed that the use of an HMD is not clinically viable, and has a preference for the screen-based system. The lack of significant improvements with regard to targeting accuracy for our screen-based system, as well as a similar screen-based system developed by Ameri *et al.* [2] compared to US-guidance, illustrated

that the inclusion of spatially tracked models of the needle in a common coordinated system with a tracked and calibrated 2D US image does not provide sufficient guidance for central line insertions. While these systems provide additional visual information, human's ability to perceive 3D visual information rendered on a 2D monitor is extremely limited [10]. The human visual perception system has to rely on monocular (2D) cues, such as occlusion, relative size, and shadows, to interpret the 3D relationship between objects when viewing information on a monitor [10]. When the system was rendered in the HMD the users were provided with a more intuitive 3D understanding of the relationship between the anatomical targets and the surgical instruments resulting in higher success rates. Thus, to develop an advanced monitor-based surgical navigation system for central line insertion the visualization must provide an intuitive relationship between the needle and vessels in 3D.

We hypothesize that visualizing 3D Z-buffered models of the anatomy, along with tracked models of surgical tools, will improve the targeting success of the needle insertion for CVC compared to the standard US-only approach. The contributions of this work are the development of an advanced needle guidance system for use in a clinical setting, which was evaluated for feasibility by 20 expert clinical practitioners. In this chapter, I build on the work described in Chapter 5 to develop an image-guidance platform that addressed the limitations outlined in Chapter 4. This system provides a method to reconstruct the neck vasculature in 3D without the use of ionizing radiation, which facilitates accurate needle guidance without causing harm to patients. This system provides 3D reconstructed anatomy visualized in such a manner that the position of the needle with respect to the vessel is obvious to the operator.

6.2 Materials and methods

6.2.1 System development

The equipment used to develop our 3D virtual-reality surgical guidance system comprises a magnetic tracking system (Aurora, NDI, Canada), a clinical US scanner (IE-33, Philips, USA), and a tracked surgical hypodermic 7 cm metallic needle with a 10 ml syringe. A linear transducer (L15-7io, Philips, USA) was used to acquire real-time images of the phantom at an imaging depth of 6 cm. The magnetically tracked US transducer was calibrated using the Procrustean point-to-line registration method [20] with manual localization described in Chapter 2, and the surgical needle was calibrated using a template-based approach [25]. A phantom comprising two hollow (wall-less) vascular structures representative of the CA and IJV, embedded in polymerized Poly (vinyl alcohol) cryogel (PVAc) [145], was constructed (Fig. 6.1a). The phantom construction process is outlined in the work performed by Ameri *et al.* [2], as

well as in Chapter 4. In short, 3D printed models of the CA and IJV were embedded into non-polymerized PVAc. These models were extracted from the phantom after it underwent two freeze-thaw cycles to polymerize the PVAc. This phantom serves as a surrogate for patient anatomy, producing US images where the CA and IJV can be interpreted by the user (Fig. 6.1a). The phantom was housed in a container with eight 6.35 mm diameter hemispherical divot and an NDI 6 Dof reference disk sensor fixed to the outside. This phantom was imaged with a CT scanner and the vascular structures were segmented from the resulting images to provide CT representations of the vessels, which were only used for part of the analysis of the targeting task. The vascular structures from the CT scan were not used for guidance and were not visible to the user at any point in the experiment. The positions of the hemispherical divots were manually extracted from the CT scan. These hemispherical fiducials were also digitized such that their position was known with respect to the reference sensor fixed to the phantom bin using a magnetically tracked and pivot calibrated ball tip stylus [163] with a conforming 6.35 mm diameter. A fiducial-registration was applied to the fiducial pairs to solve for the relationship between the phantom's and the tracker's coordinate systems using a least-squares solution [5]. The US transducer, neck phantom, and surgical needle were magnetically tracked, spatially calibrated, and registered into a common coordinate system as depicted in Fig. 6.4.

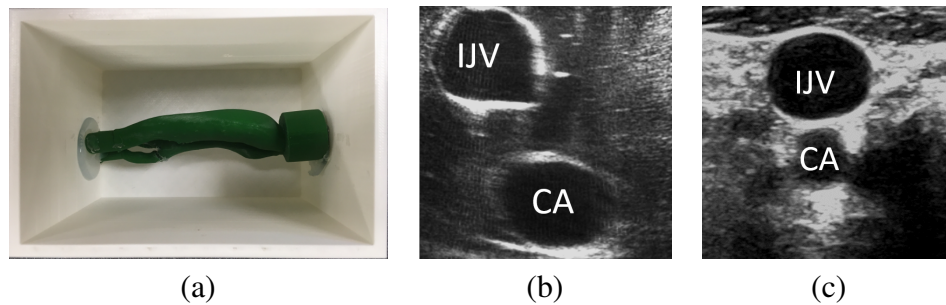


Figure 6.1: a) depicting the phantom mould, b) depicting an US scan of the phantom vasculature and b) depicting an US of human vasculature

As presented in Chapter 5, we developed a Mask R-CNN algorithm to automatically segment the CA and IJV from 2D transverse US images [53]. This algorithm achieved an average Dice score across a four-fold cross-validation on human data of 0.88 and 0.90, for the IJV and CA segmentations respectively [53]. These data sets used to evaluate the network were not used for training and contained images of healthy subjects and patients. The reconstruction pipeline is as follows: each image in the tracked scan is segmented by the Mask R-CNN algorithm; each segmentation is calibrated to its corresponding spatial location to form a skeleton of the vessel; a binary morphological closing filter with an annulus kernel shape of size [30,30,30] is used to fill the gaps between slices; a Gaussian blur with a 0.50 mm kernel is applied to smooth the models. Examples of the automatic segmentation and its corresponding reconstruction on

a healthy human subject are depicted in Fig. 6.2.

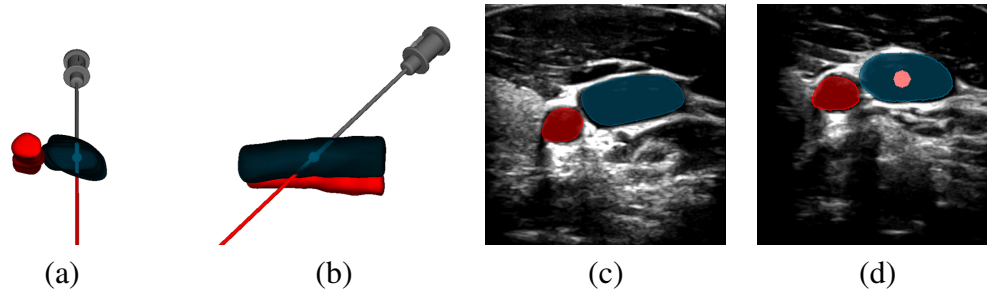


Figure 6.2: a) and b) depict vascular reconstructions from human US, with a needle superimposed on the scene to provide an example of the system for human vasculature. c) and d) provide example vascular segmentations, where image d) also contains a markup that is representative of the needle tip pose from images a) and b).

To integrate the US reconstructed vasculature into our system, the vessels in the phantom were reconstructed in 3D using our reconstruction pipeline developed in Chapter 5. The phantom that was developed is a simplified version of the neck anatomy containing only representative IJV and CA structures and none of the additional anatomy present in human scans. Visually the appearance of the phantom under US provides representations of the human neck vasculature that allow a user to perform US-guided insertions and this type of phantom has been used previously for central line insertion studies [2, 50]. However, on a pixel level, the features present in US scans of the phantom differ from human scans as PVAc produces slightly different pixel intensities compared to human tissue. Additionally, human neck US images contain other anatomies such as the neck muscles and other vessels that are not present in the phantom. Thus, the Mask R-CNN algorithm trained on human data did not achieve the same accuracy on the phantom. As this work aims to quantify the usability of 3D anatomy to guide US insertions, and not the accuracy of the phantom segmentations or reconstructions compared to human anatomy, the Mask R-CNN model was retrained on phantom images. A training set was developed by scanning a phantom with US and manually segmenting the vessels. This process resulted in 223 images and manually segmented labels that were used to train a phantom Mask R-CNN algorithm. A second phantom was developed using the same protocol described above. A scan of the new phantom produced a validation set of 215 images, which along with their manual segmentations, were used to evaluate the phantom Mask R-CNN algorithm. The Mask R-CNN achieved a Dice score of 0.99 for both the IJV and CA segmentations on the validation set. Since the methods used to develop the phantom were designed to be reproducible the two phantoms are nearly identical resulting in high accuracy. Again as the goal of this paper is not to analyze the accuracy of the Mask R-CNN phantom segmentation algorithm, as it is only to be used to obtain the vascular reconstruction for this study, a more robust accuracy

analysis was not performed. This network trained on phantom images was used in place of the network trained on human data within the pipeline to produce reconstructions of the CA and IJV. As these vascular models were produced using a calibrated US probe, they were obtained with respect to the tracker's coordinate system and thus are inherently registered to a common coordinate system with the tracked models of the surgical tools and calibrated US images.

While the clinical deployment of this system would require the clinician to perform the vessel reconstruction *in situ* just prior to the insertion of the needle, for this validation study, the vascular reconstruction was performed before the experiment, to keep the evaluation environment consistent between users. This is justified as the neck phantom is static and anatomical variations such as the effect of neck rotation were not simulated. The system comprising the 3D models of the IJV, CA, and the tracked tools were rendered on a 2D monitor. The CA was displayed as a fully opaque surface in red, and the IJV was displayed as a blue wire-frame model. Visualizing the IJV as a wire-frame model takes advantage of the standard computer graphics rendering pipeline relating to the computation of Z-buffer values. The Z-buffer is also known as the depth buffer as it improves depth perception of objects in simulated 3D environments by rendering overlapping virtual objects based on their depth (Z-value) with respect to the particular perspective of the virtual observer (virtual camera) such that the relative depth between objects can be interpreted. When objects are rendered in a simulated 3D environment on a 2D screen, multiple objects may have overlapping [X,Y] pixel coordinates, such that they occupy the same pixel location on the screen. However, as the scene is derived from 3D information these objects also have an associated Z or depth value. Z-buffering requires a depth test where if two objects occupy the same [X,Y] pixel location their Z-values are compared and the lowest Z-value is included in the Z-buffer. This process is repeated for all overlapping objects in the scene, such that the scene can be rendered with intuitive depth perception as closer objects occlude objects that are farther away from the virtual camera. The Z-buffer enhances depth perception as occlusion is one of the major monocular cues used by the human visual perception system to understand the relative depths of objects based on 2D information. Fig. 6.3, illustrates how rendering our 3D scene using a Z-buffer improves the 3D perception compared to rendering the scene with no depth culling. The appearance of the needle mesh changes naturally when it is situated between, in front or behind the wire-frame of the vasculature with respect to the viewing direction of the virtual camera, as seen in Fig. 6.4. This allows the user to have an intuitive 3D perspective based on the Z-buffer rendering and relative occlusion of the needle model, needle trajectory, and the IJV model. Rendering our VR scene using Z-culling allows the user to position their needle and trajectory in the centre of the vessel based on the relative colour change of the needle produced by the blending of the colours of the two overlapping virtual objects. If the needle is on the periphery of the vessel model, as depicted in

Fig. 6.4 a) and b), only a small portion of the needle appears blue. Performing the insertion at the position where the needle trajectory is maximally blue ensures the needle is centred within the vessel (Fig. 6.4 c). In addition to providing a more intuitive depth perception, Z-buffer rendering does not incur any additional computation cost during the rendering process.

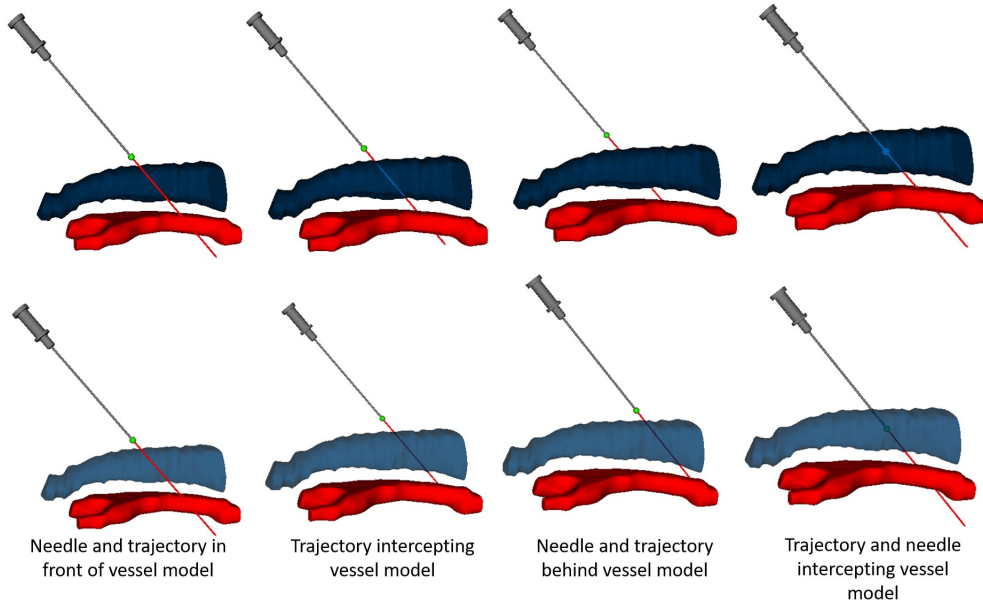


Figure 6.3: Visual comparison between using a wireframe and a solid but translucent model to visualize the IJV. This figure compares visualizations with: the needle and trajectory in front of the IJV, the trajectory intercepting the IJV, the needle and trajectory behind the IJV, and the needle and trajectory intercepting the IJV.

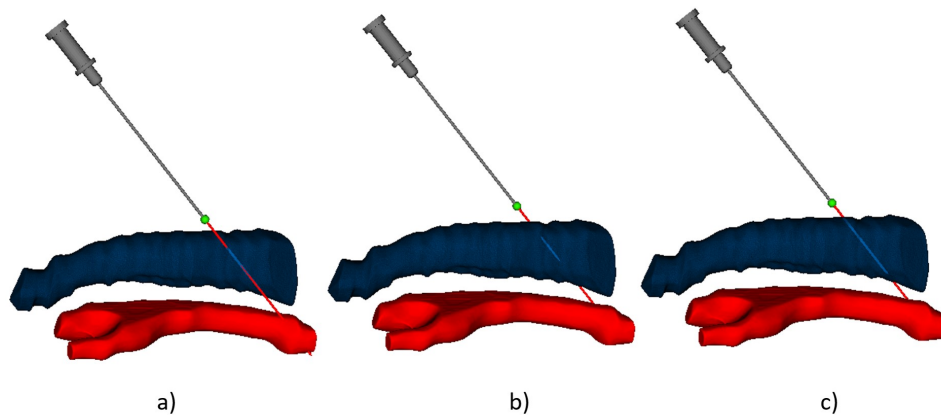


Figure 6.4: Visual depiction of how to position the needle at the centre of the vessel, where a) and b) depict the needle trajectory at the periphery and c) depicting the trajectory centred about the vessel through optimizing the amount of needle trajectory that is rendered blue.

During the experiment, the pose of the user's vantage point with respect to the vasculature remained consistent in a longitudinal orientation. This orientation replicated the vantage point of the clinicians when performing clinical central venous catheterization. The US and CT reconstructed surfaces were compared using a nearest neighbour Euclidean distance approach (Fig. 6.5 [46]). The discrepancies between the CT and US derived vascular models are mainly attributed to the pressure applied at the contact surface between the probe and phantom during US scanning. In our phantom design, the IJV is positioned superior to the CA within the phantom and is, therefore, more susceptible to deformation. The average diameter of the IJV reconstruction is 4.20 mm and 3.30 mm, for the US and CT respectively. As the IJV is also deformable in humans this system would require the patient to be imaged in the Trendelenburg position, which is commonly employed for IJV puncture during central venous catheterization [157]. Trendelenburg position requires the patient to be supine on the table with their head declined below their feet at an angle of roughly 16° , which allows the vessel to distend reducing the amount of deformation during scanning.

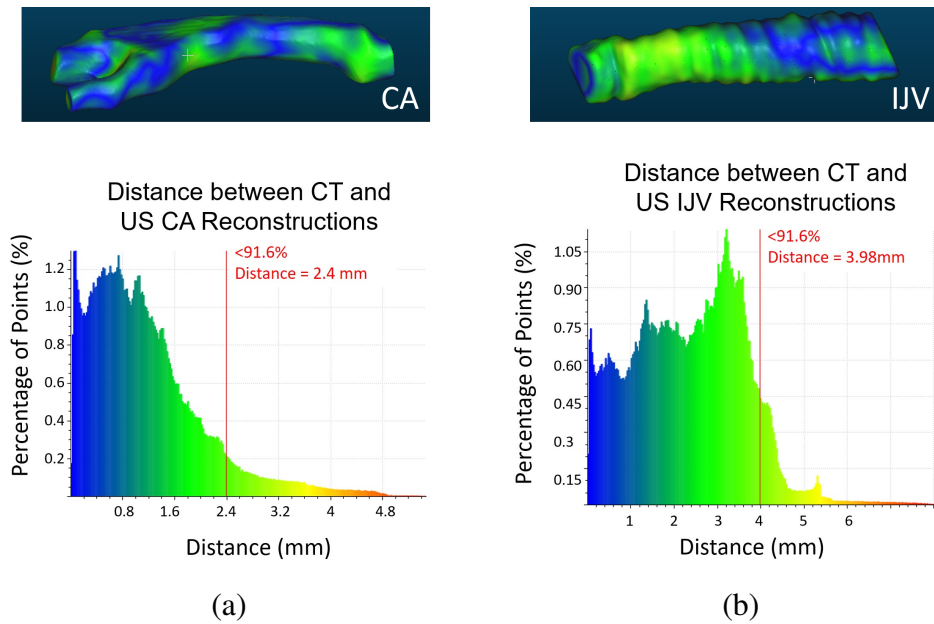


Figure 6.5: Histogram of the surface distance between the CT and US reconstructions of the phantom CA (a) and IJV (b)

6.2.2 Experimental Methods

Twenty clinicians with US-guided CVC training, who had performed over 15 clinical insertions, were recruited for this study. On average our participants had performed 211.44 ± 150.08 central line insertions with an average of 7.86 ± 4.93 years of experience. All participants were recruited with consent according to the local REB regulation (Western University REB

114734). Each participant was randomly assigned either the US-only or screen-based system as their first experienced system in a counter-balanced design. Each participant was given oral introductions relating to the system, and they were allowed a single training insertion using the system to which they were initially assigned. The user then performed an experimental insertion, in which the imaging and tracking data of the needle and probe were recorded. This process was then repeated for the second system. The experimental configuration is depicted in Fig. 6.6, where sub-figure a) is the tracked needle and its calibration block and sub-figure b) is the tracked and calibrated US probe. To keep the environment consistent between the two guidance methods being employed, a single monitor was used to display either the US image or the 3D system depending on which system was being employed. Following the experiment, each user completed a questionnaire pertaining to the 3D surgical navigation system, which is included in Appendix A Fig. A.2. As the phantom model is deformable all users were required to use the US probe to apply pressure to the phantom regardless of the navigation system being used. The requirement to maintain the use of the US probe during the insertions performed with the advanced navigation system allows for coordination between the tool and the probe to be a factor in both insertions. The use of the US probe during navigation would be required in a clinical scenario such that the portion of the vessel being imaged by the US probe would be continuously segmented and updated in real-time and allows for the calibrated US image to be integrated into the visualization.

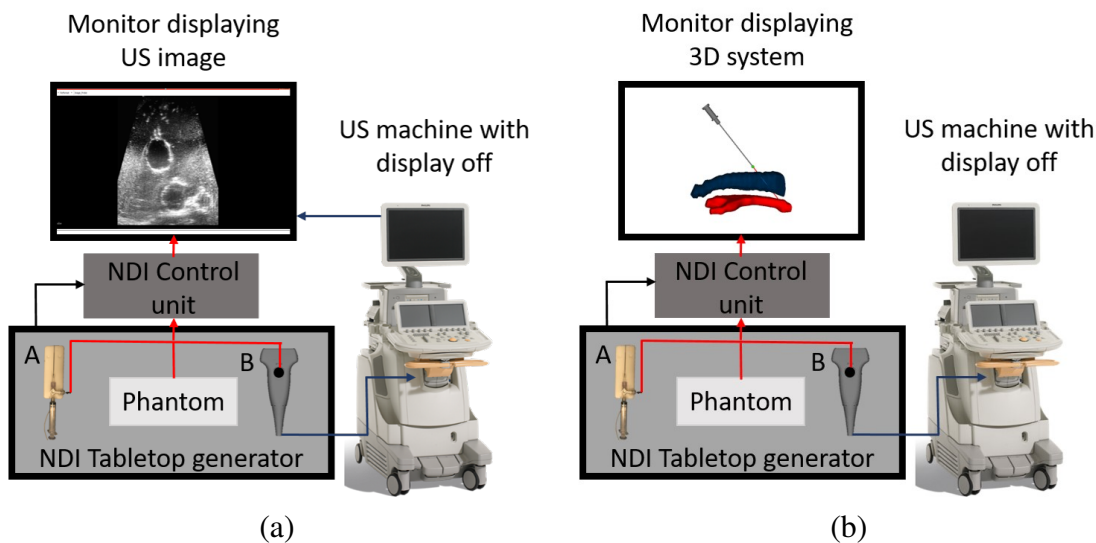


Figure 6.6: Experimental setup for a) the 3D system and b) US-only guidance. Where A is the tracked syringe and its tracked calibration block, and B is the tracked and calibrated US probe. This figure is drawn from the point of view of the user.

Several metrics were derived using the recorded tracking data for each insertion, these

metrics are consistent with those used in Chapter 4. The selected metrics include the time from the initial needle position on the surface of the phantom to their final needle placement; the insertion success defined as the final needle tip position within the vessel wall with respect to both the US and CT reconstructions; and the shortest distances from the final needle tip position to the IJV centre line [65] and the IJV wall with respect to both the CT and US reconstructions. The distance from the vessel wall was a signed result as was labelled as negative if the needle was outside the vessel lumen, thus a large positive result represents an insertion within the vessel, but far from the vessel wall. The results provided with respect to the US reconstruction are more telling than those provided with respect to the CT, as the clinicians were required to maintain pressure on the phantom with the US probe. The use of the CT vascular models is to provide an additional metric for accuracy, the discrepancies between the CT and US reconstructions should not have a major effect on our study as the CT information is never visualized to the user. All of the continuous metrics (time and distance) were deemed to be normally distributed following an Anderson-Darling test allowing statistical differences between means to be analyzed using a paired t-test. Since the insertion success is a discrete event, the McNemar Chi-squared test with Yate's correction was employed.

6.3 Results

The results, including time, the distance from the centre line (DFCL) of the US and CT reconstructions, the distance from the vessel wall (DFVW) of the US and CT reconstructions, and the number of successful insertions (SI), are summarized in Table. 6.1. The questionnaire results are summarized in Table. 6.2.

Table 6.1: Vessel targeting results for the US-only and 3D guidance systems across time, distance from the centre line (smaller is better), distance from vessel wall (larger is better), and number of successful insertions, all with respect to the US and CT reconstructions. Bold faced values are considered the better result.

	US-Only	3D System	P-value
Metric	<i>Mean ± Std</i> CI(95%)	<i>Mean ± Std</i> CI(95%)	
Time(sec)	20.5 ± 9.8 [15.9,25.0]	10.4 ± 5.5 [7.8,13.0]	<0.001***
US DFCL(mm)	4.2 ± 2.9 [2.9,5.6]	1.8 ± 0.9 [1.3,2.2]	0.002**
CT DFCL(mm)	3.6 ± 2.0 [2.7,4.5]	2.1 ± 1.3 [1.5,2.7]	0.02*
US DFVW(mm)	0.7 ± 4.3 [-1.4,2.7]	4.4 ± 1.6 [3.6,5.2]	<0.001***
CT DFVW(mm)	0.8 ± 4.5 [-1.2,2.9]	4.5 ± 1.5 [3.8,5.2]	0.001**
Metric	Total Number	Total Number of SI	
US SI	14	20	0.041*
CT SI	14	20	0.041*

Table 6.2: Summary of user questionnaire result (scored out of 10)

Question	Average Response
If the advanced guidance system was available how often would you use it?	6.4
Do you think the advanced system could be useful for training US guided CVC?	8.0
Do you think the advanced guidance system would be viable to use in the OR?	6.9
Did you find the relationship between the needle and vessels to be intuitive?	7.9

6.4 Discussion

A guidance system for CVC that visualizes 3D US surface reconstructions of the CA and IJV and tracked models of surgical tools was evaluated against traditional US-only guidance. As summarized in Table. 6.1, the 3D system provided superior needle guidance compared to the US-only method, with the use of the 3D system resulting in a successful insertion rate of 100 % compared to 70 % for US-only guidance ($p=0.04$) among 20 trained medical practitioners. The 3D system also significantly improved users' targeting accuracy, with distances from the centre line of the US reconstructed IJV being 1.8 ± 0.90 mm and 4.2 ± 2.90 mm ($p=0.00$), for the 3D and US-only systems respectively. The results indicate that the 3D system has distinct advantages for needle guidance during CVC.

The two previously described CVC navigation systems employed a screen-based 2D perspective [2] and an HMD system with 2D imaging in a 3D context [50] respectively. The HMD system (94 % success rate) and our 3D screen-system (100 % success rate) both significantly improved the insertion accuracy for expert users, as compared to the previously described screen-based system (64 % success rate) [2]. We believe that our system is beneficial for navigation, as it provides an intuitive 3D visualization of both the surgical tools and the anatomy, as depicted in Figures 6.3 and 6.4. Rendering the IJV using a Z-buffer enhances depth perception based on the colour changes and occlusions between models. If the needle or its trajectory is in front of the vasculature, they are fully visible but become occluded if they are behind it. If either is within the vasculature, the section intersecting the wire mesh turns blue (Fig. 6.3). Aligning the trajectory of the needle with the IJV wire mesh allows the user to fully comprehend the position and orientation of their needle with respect to the vasculature, based on the colour change (Fig. 6.4). If only a small portion of the trajectory is blue, it is close to the edge of the vessel, allowing users to intuitively position the needle until they maximize the blue component of trajectory, indicating they are aligned with the centre of the vessel

(Fig. 6.4). By following this path, they can navigate the needle to the target. We believe the success of the HMD and 3D system is a consequence of the user's ability to comprehend the spatial relationship between their tool and the vasculature in 3D, unlike the 2D screen-based system.

The high success rate of insertions performed using our 3D system emphasizes the importance of providing surgical navigation information in a 3D context. The rapidly expanding field of machine learning provides the potential for patient-specific 3D anatomical models to be obtained in real-time without exposing the patient to radiation through the use of tracked or 3D US and high-quality GPUs. This work also highlights the importance of the visualization properties used when designing the system. Small visualization decisions, such as using the wireframe for the IJV, allow for 2D monitors to be used successfully for 3D navigation. Developing 2D screen-based navigation systems decreases the barriers to clinical translation and decreases the system footprint in the operating room (OR). In Chapter 4, clinicians on average allocated a score of 4.35/10 in response to the clinical viability of the HMD system in the OR. In contrast, our 3D surgical navigation system achieved an average score of 6.9/10 for clinical viability. A screen-based system does not affect the line of sight with the patient and employs a display that is familiar to the users. Screen-based systems can also be integrated into clinical devices or systems, such as US machines or OR monitors, that are already present in the OR. Therefore, surgical navigation systems must simultaneously optimize both the clinical usefulness and feasibility by employing iterative design practices, based on expert user feedback, to increase the translation of successful advanced surgical navigation systems.

Despite our promising results, our study design has limitations and there are barriers to clinical translation. The results show that the time for needle insertion is significantly lower for the 3D system compared to the US-only approach. The reported time is associated with inserting the needle, which indicates the ease with which a user is able to navigate to the target. However, the process of obtaining patient-specific vascular reconstruction would increase the overall time required for the CVC insertion as compared to the US-only approach, as it would require additional time for the clinician to perform the surface scan and for the algorithm to output the vascular reconstructions prior to performing the insertion. To move this system towards clinical translation, the following aspects of the system must be evaluated. First, the reconstruction accuracy and precision, and the time required for human vascular reconstruction across a cohort of clinicians must be evaluated. This system should also be integrated into an US machine to maximize clinical usability. Lastly, a more robust user study should be performed to evaluate the system accuracy across repeated insertions on a variety of vascular configurations. Another limitation to this approach is that US scanning results in pressure applied to the contact surface which may distort the vasculature, as seen in Fig 6.5. For this

reason, we aim to evaluate the effects of continued use of US imaging, with or without the US image being visible in the scene when using the 3D system, to ensure that the pressure applied when performing the insertion is consistent with that during the scan used for surface reconstruction. Continued scanning would allow for the vessel models to be updated in real-time during needle insertion. We believe that the pipeline of using machine learning to develop patient-specific 3D models of anatomy for targeting could be useful for a number of other needle targeting applications, such as femoral artery line insertions, biopsy, and nerve blocks.

Overall, this work aims to provide further insight into human perception during targeting tasks, specifically emphasizing that when using a 2D monitor for targeting it is necessary to provide the information in a 3D context. With the expansion of machine learning, 3D models can be reconstructed and used for real-time navigation in 3D. This chapter forms a basis for the development of systems that employ machine learning to form real-time 3D patient-specific anatomical models that can be used for targeting tasks that are visualized via a monitor.

6.5 Conclusions

An advanced interventional navigation system comprising US surface reconstructions of the IJV and CA in a common coordinate system, with a model of a tracked needle and its trajectory on a 2D monitor for central line insertions on a phantom was developed and evaluated in this work. A Z-buffer visualization was used for the IJV to improve depth perception and provide an intuitive 3D relationship between virtual models projected onto a 2D surface, as past research illustrated limitations of 3D navigation on a 2D monitor. Twenty clinicians performed insertions into the phantom using both this system and the traditional US-only approach and demonstrated that the 3D guidance system provided superior guidance compared to the US-only system. The success rate of the 3D system was 100 % compared to 70 % for the US-only system. The 3D system provided more consistent targeting with the distance from the US reconstructed IJV's centre line being 1.8 ± 0.90 mm and 4.2 ± 2.90 mm for the 3D system and US-only guidance, respectively. Overall, this work highlights the importance of providing a 3D anatomical context for needle guidance. Providing users with sufficient 3D information by selecting visualization methods that encourage the 3D perception through the use of monocular cues can be feasible for clinical integration and provide superior navigation compared to standard image-guided alternatives while maintaining screen-based navigation.

Chapter 7

Conclusions and Future Directions

Ultrasound-(US)-guided insertions have become the routine method used for the needle insertion required during central venous catheterization (CVC). This thesis was motivated by the high prevalence of complication rates that result from 2D US-guided needle insertions for central venous access at the site of the internal jugular vein (IJV). This thesis aimed to develop and evaluate 3D US-based environments for needle guidance into the IJV, with the two main foci being 1) the development and evaluation of methods that form the basis of 3D US environments such as calibrations and machine learning algorithms and 2) the development and validation of 3D US-based needle navigation systems for central venous access at the IJV. The contributions of this thesis in these two areas are summarized in the following section.

7.1 Contributions of the Thesis

Ultrasound probe calibration is a fundamental building block used within most advanced US-guidance systems, as it provides the relationship between the US image and a spatial tracking sensor fixed to the US probe enabling the use of free-hand tracked US. A common US calibration technique is a point-to-line method that previously required manual fiducial localization and had only been theoretically validated. In Chapter 2, we developed a convolutional neural network (CNN) that is capable of localizing the fiducials required for this algorithm (needle reflections) within 0.25 mm of manually localized ground truth coordinates across 4-8 cm imaging depths from a single linear probe. This CNN was integrated into the point-to-line calibration approach to provide a formulation of the calibration approach with automatic fiducial localization. The calibration approach with both manual and automatic fiducial localizations was validated using point reconstruction accuracy, and found to have less than 1.50 mm of error for the normalized mean absolute distance from the ground-truth, with standard deviations less than 0.50 mm. Thus, this work validates the use of the point-to-line calibration approach

with manual and automatic fiducial localization for the development of advanced US-guided surgical navigation systems. These calibration approaches accurately position the US image in a common coordinate system with other tracked objects and enables free-hand tracked US, as well as 3D US reconstructions. The automatic fiducial localization improves the usability of this algorithm and allows novice users to obtain calibrations with a fiducial localization error similar to an expert user.

Optimizing methods to visualize advanced US-guided surgical navigation systems has been an ongoing research topic for years. Wearable technology or head-mounted displays (HMDs) place visual information directly into the user's field of view, and research on their use for surgical navigation dates back to the 1960s. Currently, there are many commercially available head-mounted displays on the market, and devices such as the HTC VIVE Pro HMD enable the development of surgical navigation systems rendered directly into the user's field of view. In order to use an HMD for surgical navigation systems that rely on spatial tracking and US calibration, the spatial tracking system must be calibrated to the tracking system associated with the HMD. In Chapter 3, we develop and evaluate a method to register an external optical tracking system to the HTC VIVE's tracking system, by means of a dual tracked co-calibration apparatus. This co-calibration approach achieved sub-millimetre and sub-degree accuracy compared to ground-truth values. Thus, it provides a method to accurately position tracked tools and calibrated US images within the coordinate system of an HMD, such that they can be visualized in a first-person perspective.

The point-to-line US calibration method and the co-calibration method used to register a spatial tracking system with the HTC VIVE Pro HMD, developed in Chapters 2 and 3, provide the backbone of a first-person needle navigation system rendered in the HTC VIVE Pro HMD that was developed in Chapter 4. The US calibration accurately positions the US image in the spatial tracker's coordinate system, and the co-calibration accurately positions the calibrated US image and the spatially tracked tools in the HMDs coordinate system, such that the interactions between the tools and the US image can be visualized in the HMD. The system we developed employing these principles renders the real-time tracked US in a common coordinate system with tracked models of the needle, needle trajectory, and US probe. We demonstrated the use of this advanced needle navigation system for the application of needle insertions at the IJV, using a phantom containing replicas of the IJV and CA vasculature. We compared our surgical navigation system, rendered on both a monitor and an HMD, to the US-only guidance method used clinically. The HMD system enabled more accurate needle insertions compared to both the monitor and US-only systems for the phantom experiment across 31 clinicians, illustrating the importance of a 3D perspective for needle navigation. The HMD system also significantly improved the needle insertion accuracy for 20 novice medical students, with no

central line insertion experience. This research highlights the benefits of HMDs for needle navigation research, and as technology advances and becomes more clinically usable, systems similar to that developed in this work should be used for surgical navigation tasks. Despite the success of the HMD system and the lack of significant improvements using the monitor-based system, our clinical practitioners preferred the advanced navigation system visualized on the monitor. Thus, the results from this chapter motivated investigating other advanced monitor-based systems that allow the user to visualize all relevant tools and anatomy in 3D to further investigate the ability to use a 2D display for surgical navigation.

In Chapter 3, the visual output of the advanced surgical navigation system displayed on the monitor was the clinician's preference but did not have any significant effect on the accuracy of the needle insertion compared to the US-only case. This system rendered the tracked and calibrated free-hand US image at an oblique angle, in a common coordinate system with a tracked model of the needle and needle trajectory. While this system provided a more advanced visualization for the user, there were limitations with the user's ability to accurately perceive the 3D information on a 2D monitor. Despite the fact that 2D tracked US provides a 3D perspective, in order to use a 2D display, the surgical navigation system must provide sufficient 3D context of the tools and anatomy for successful guidance. To improve the 3D information provided to the user, in Chapter 4 we developed a machine-learning-based vascular reconstruction pipeline that given a tracked US recording produced 3D models of the CA and IJV. The machine learning algorithm was capable of delineating the CA and IJV with a Dice score of 0.92 and 0.93, for the CA and IJV respectively. The US-derived vascular reconstructions were compared to manually segmented vasculature from a CT scan of the same patient. The average surface area ratio between the US and CT vascular models obtained was 0.94 and 0.88, for the CA and IJV respectively. The average volume ratio between the US and CT vascular models was 0.86 for both the CA and IJV. This vascular reconstruction pipeline is capable of accurately obtaining models of the CA and IJV given a tracked US recording of the neck.

The vascular reconstruction pipeline developed in Chapter 5 was used to reconstruct the vasculature in a phantom of the neck. These phantom reconstructions were used to develop another screen-based surgical navigation system where the reconstructed vessels were rendered in a common coordinate system with models of the tracked needle, needle tip, needle trajectory. To improve the users' ability to perceive the information in 3D, the IJV was rendered using a Z-buffer, which allows for the relationship between the needle models and the vasculature to be intuitively understood in 3D as the visualization of the needle changes based on its position with respect to the vessel model. This surgical navigation system was tested on a cohort of 20 clinicians trained in US-guided central line insertions. Each user performed experimental insertions using the advanced surgical navigation system and US-only guidance. The surgical

navigation system significantly improved the needle insertion accuracy compared to the US-only case and had an 100% success rate. Machine learning algorithms can be used to obtain 3D US reconstructions that have the capability to provide sufficient 3D information on a 2D monitor to guide needle insertions. Anatomical US reconstructions should be integrated into or supplement US-guided systems such that the user can understand the 3D context of their tools with respect to patient-specific anatomy.

7.2 Future Directions

The development of surgical navigation systems for needle interventions should be shifting towards the inclusion of 3D patient-specific environments with advanced visualizations. Future research should focus on the integration of machine learning algorithms into advanced US navigation and training systems and the continuous evaluation of novel display technologies.

Machine learning algorithms can be integrated into almost every component required for the development of advanced US-based needle navigation systems, from calibration to anatomical segmentation and reconstruction. machine learning methods applied to areas such as fiducial localization for US calibration may increase the potential for clinical translation of surgical navigation systems, as methods requiring manual intervention can be replaced with automatic methods. Research should focus on the development of systems that require minimal to no user intervention prior to the use of the system. Machine learning methods should be used to automate any steps that require user intervention or manual inputs to lower the barrier to translation and produce more robust and precise outputs. Machine learning algorithms can also be integrated into surgical navigation systems by automatically producing patient-specific anatomical reconstructions in a common coordinate system using tracked and calibrated US images. Through the development of fully automatic reconstruction pipelines, a quick pre-scan can provide 3D anatomical context for the clinician, which can be used to supplement the US information. 3D US environments should be developed and evaluated for needle navigation tasks with different targets such as biopsy, femoral vein central line insertions, or nerve blocks. Additionally, the vascular reconstruction pipeline developed in this work could be expanded to incorporate plaque segmentations to quantify and visualize plaque burden in real-time. In order to develop machine learning algorithms that are capable of accurately segmenting anatomy across a range of patients, US scanners, and sonographers, it is imperative that research groups form long-term and ongoing relationships with clinical practitioners who perform US-guided navigation. US-guidance information is typically not recorded or saved, meaning researchers are losing out on a plethora of clinical scans that can be used to improve the robustness of machine learning algorithms. These practices are highly encouraged for clinical applications

requiring the manual segmentation of anatomical structures in US images, as researchers benefit from high-quality gold-standard segmentations. The development of systems based upon machine learning algorithms developed using large clinical data sets is another way to lower the barrier of clinical translation of advanced surgical navigation systems. Machine learning offers the ability to reduce equipment requirements through the development of image-based tracked algorithms for needle tracking and sensor-less reconstruction pipelines. These methods do require a trade-off between aspects such as accurate real-time tracking of tools and US for a reduction in required equipment and calibrations. The development of accurate image-based tracking algorithms also requires large data set containing tracked patient scans. Overall, the use of machine learning within the development of advanced US environments is highly encouraged.

Another area that should be explored is the use of advanced surgical navigation systems, particularly those that provide 3D context of the anatomy, for US-guidance training. For central line insertions, the accuracy of the needle puncture is dependent on the experience of the operator. Systems that integrate US-derived anatomical reconstructions into a common coordinate system with the calibrated US image and the tracked tools, may offer the ability to improve the training of central line insertions. Novice users may benefit by gaining a more in-depth understanding of the 3D context of the US image with respect to the anatomy and surgical tools to advance their conceptual understanding of the procedure. Thus, these systems should be properly evaluated for their usefulness for training.

Research on the various types of displays used to visualize surgical navigation information has been ongoing for decades. As display technology and visualization approaches are continually advancing, this area of research requires ongoing investigation. As technology incorporated into commercially available HMDs advances, their role in the development of surgical navigation systems should be continuously re-evaluated. Specifically, video-pass through HMDs with advanced video integration and lightweight displays will likely be the most beneficial systems for surgical navigation as they offer the best trade-off between perceptual issues and real-world information. The tracking associated with these HMD systems will need to be evaluated as new HMDs or updated systems are released. The evaluation of various data fusion methods may be required to optimize the combination of real-world video streams, medical images, and surgical tool tracking within HMDs for surgical navigation. However, a realistic assessment of the current state of commercially available HMDs, suggests that their integration into surgical suites is a long-term goal. In the interim, screen-based surgical navigation systems that provide information in a 3D context should be developed and evaluated. Sufficient 3D information rendered with the use of monocular depth cues allows for accurate needle targeting on a monitor. The displays used for this type of navigation system should be large mobile mon-

itors that can display multiple visualizations and be placed in optimal viewing locations should first be implemented, as this most closely resembles practices clinicians are used to from CT and MRI. Clinicians will likely be more receptive to changes in visualizations over changes in displays, and research should focus on small steps that enable the clinical usability of advanced screen-based systems for surgical navigation, with the long-term goal of deployment in an HMD. The ultimate goal is to provide a visualization system that provides the most intuitive user interface between the clinician and the patient.

Overall, this work focused on the development of 3D US environments for needle navigation into the IJV. Machine learning can be integrated into methods required for the development of advanced surgical navigation systems such as US calibration and anatomical reconstruction. This thesis highlights the importance of providing users with systems that allow for the relationship between the needle and anatomy to be intuitively understood. The use of both an HMD to enable a first-person perspective, and a 3D reconstructed vasculature rendered with a Z-buffer, improved the needle insertion accuracy compared to US-guidance used clinically.

Bibliography

- [1] Fikri M Abu-Zidan, Ashraf F Hefny, and Peter Corr. Clinical ultrasound physics. *Journal of Emergencies, Trauma and Shock*, 4(4):501, 2011.
- [2] Golafsoun Ameri, John S. H. Baxter, Daniel Bainbridge, Terry M. Peters, and Elvis C. S. Chen. Mixed reality ultrasound guidance system: a case study in system development and a cautionary tale. *International Journal of Computer Assisted Radiology and Surgery*, 13(4):495–505, 2018.
- [3] M. E. Anderson, M. S. McKeag, and G. E. Trahey. The impact of sound speed errors on medical ultrasound imaging. *The Journal of the Acoustical Society of America*, 107(6):3540–3548, 2000.
- [4] Muhammad Arif, Adriaan Moelker, and Theo van Walsum. Automatic needle detection and real-time bi-planar needle visualization during 3D ultrasound scanning of the liver. *Medical image analysis*, 53:104–110, 2019.
- [5] K Somani Arun, Thomas S Huang, and Steven D Blostein. Least-squares fitting of two 3-D point sets. *IEEE Transactions on pattern analysis and machine intelligence*, PAMI-9(5):698–700, 1987.
- [6] Michael P. Bannon, Stephanie F. Heller, and Mariela Rivera. Anatomic considerations for central venous cannulation. *Risk Management and Healthcare Policy*, 4:27–39, 2011.
- [7] Parmida Beigi, Septimiu E Salcudean, Gary C Ng, and Robert Rohling. Enhancement of needle visualization and localization in ultrasound. *International Journal of Computer Assisted Radiology and Surgery*, pages 1–10, 2020.
- [8] Paul J Besl and Neil D McKay. Method for registration of 3-D shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pages 586–606. International Society for Optics and Photonics, 1992.

- [9] Michael Blaivas and Srikar Adhikari. An unseen danger: Frequency of posterior vessel wall penetration by needles during attempts to place internal jugular vein central catheters using ultrasound guidance. *Critical Care Medicine*, 37(8):2345–2349, 2009.
- [10] Rositsa Bogdanova, Pierre Boulanger, and Bin Zheng. Depth perception of surgeons in minimally invasive surgery. *Surgical innovation*, 23(5):515–524, 2016.
- [11] Matias Bruzoni, Bethany J. Slater, James Wall, Shawn D. St Peter, and Sanjeev Dutta. A Prospective Randomized Trial of Ultrasound- vs Landmark-Guided Central Venous Access in the Pediatric Population. *Journal of the American College of Surgeons*, 216(5):939–943, 2013.
- [12] Emanuele Maria Calabrò, Fabrizio Cutolo, Marina Carbone, and Vincenzo Ferrari. Wearable Augmented Reality Optical See Through Displays Based on Integral Imaging. In *Wireless Mobile Communication and Healthcare*, pages 345–356. Springer, Cham, 2017.
- [13] N. Calvert, D. Hind, R. G. McWilliams, S. M. Thomas, C. Beverley, and A. Davidson. The effectiveness and cost-effectiveness of ultrasound locating devices for central venous access: a systematic review and economic evaluation. *Health Technology Assessment*, 7(12):1–84, 2003.
- [14] Aladin Carovac, Fahrudin Smajlovic, and Dzelaludin Junuzovic. Application of ultrasound in medicine. *Acta Informatica Medica*, 19(3):168, 2011.
- [15] Jonathan Carr. *Surface reconstruction in 3D medical imaging*. PhD thesis, University of Canterbury, Christchurch, New Zealand, 1996.
- [16] L Cavanna, G Civardi, P Mordenti, D Vallisa, R Berte, and C Di Nunzio. Central venous catheter care for the patients with cancer: ultrasound-guided insertion should be strongly recommended for internal jugular vein catheterization. *Annals of oncology*, 24(11):2928–2929, 2013.
- [17] Anne Chao et al. Performance of central venous catheterization by medical students: a retrospective study of students’ logbooks. *BMC Medical Education*, 14(1):168, 2014.
- [18] Pauline Chauvet, Toby Collins, Clement Debize, Lorraine Novais-Gameiro, Bruno Pereira, Adrien Bartoli, Michel Canis, and Nicolas Bourdel. Augmented reality in a tumor resection model. *Surgical Endoscopy*, 32(3):1192–1201, 2018.

- [19] Chengqian Che, Tejas Sudharshan Mathai, and John Galeotti. Ultrasound registration: A review. *Methods*, 115:128–143, 2017.
- [20] Elvis C. S. Chen, Terry M. Peters, and Burton Ma. Guided ultrasound calibration: where, how, and how many calibration fiducials. *International Journal of Computer Assisted Radiology and Surgery*, 11(6):889–898, 2016.
- [21] Elvis C. S. Chen, Terry M. Peters, and Burton Ma. Which point-line registration? In *Proc. SPIE 10135, Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 1013509, 2017.
- [22] Elvis C. S. Chen, Terry M. Peters, and Burton Ma. Which point-line registration? In *Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 10135, page 1013509. International Society for Optics and Photonics, 2017.
- [23] Xiaojun Chen, Lu Xu, Yiping Wang, Huixiang Wang, Fang Wang, Xiangsen Zeng, Qiu-gen Wang, and Jan Egger. Development of a surgical navigation system based on augmented reality using an optical see-through head-mounted display. *Journal of Biomedical Informatics*, 55:124–131, 2015.
- [24] Xiaojun Chen, Lu Xu, Yiping Wang, Huixiang Wang, Fang Wang, Xiangsen Zeng, Qiu-gen Wang, and Jan Egger. Development of a surgical navigation system based on augmented reality using an optical see-through head-mounted display. *Journal of biomedical informatics*, 55:124–131, 2015.
- [25] Chen, Elvis C. S., Lasso, Andras, and Fichtinger, Gabor. External tracking devices and tracked tool calibration. In *Handbook of Medical Image Computing and Computer Assisted Intervention*, pages 777–794. Elsevier, 2020.
- [26] Carling L. Cheung, Chris Wedlake, John Moore, Stephen E. Pautler, and Terry M. Peters. Fused video and ultrasound images for minimally invasive partial nephrectomy: A phantom study. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, pages 408–415. Springer Berlin Heidelberg, 2010.
- [27] Carling L. Cheung, Christopher Wedlake, John Moore, Stephen E. Pautler, Anis Ahmad, and Terry M. Peters. Fusion of stereoscopic video and laparoscopic ultrasound for minimally invasive partial nephrectomy. In *Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling*, volume 7261, pages 96 – 105. International Society for Optics and Photonics, SPIE, 2009.

- [28] Sou Chen Chew, Zhi Yuen Beh, Vineya Rai Hakumat Rai, Mohamad Fadhil Jamaluddin, Ching Choe Ng, Karuthan Chinna, and M Shahnaz Hasan. Ultrasound-guided central venous vascular access—novel needle navigation technology compared with conventional method: A randomized study. *The journal of vascular access*, 21(1):26–32, 2020.
- [29] Jui-Ying Chiao, Kuan-Yung Chen, Ken Ying-Kai Liao, Po-Hsin Hsieh, Geoffrey Zhang, and Tzung-Chi Huang. Detection and classification the breast tumors using mask r-cnn on sonograms. *Medicine*, 98(19), 2019.
- [30] Kevin Cleary and Terry M. Peters. Image-guided interventions: technology review and clinical applications. *Annual Review of Biomedical Engineering*, 12:119–142, 2010.
- [31] V. Couteaux, S. Si-Mohamed, O. Nempont, T. Lefevre, A. Popoff, G. Pizaine, N. Villain, I. Bloch, A. Cotten, and L. Boussel. Automatic knee meniscus tear detection and orientation classification with Mask-RCNN. *Diagnostic and Interventional Imaging*, 100(4):235–242, 2019.
- [32] Zhenzhen Dai Dai, Eric Carver, Chang Liu, Joon Lee, Aharon Feldman, Weiwei Zong, Milan Pantelic, Mohamed Elshaikh, and Ning Wen. Segmentation of the prostatic gland and the intraprostatic lesions on multiparametric MRI using Mask-RCNN. *arXiv preprint arXiv:1904.02575*, 2019.
- [33] Ayon Dey. Machine learning algorithms: a review. *International Journal of Computer Science and Information Technologies*, 7(3):1174–1179, 2016.
- [34] Mingyue Ding and Aaron Fenster. Projection-based needle segmentation in 3D ultrasound images. *Computer Aided Surgery*, 9(5):193–201, 2004.
- [35] Burton Ma Elvis C. S. Chen, Terry M. Peters. Which point-line registration? In *Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 10135, page 1013509. International Society for Optics and Photonics, 2017.
- [36] R. Ewers, K. Schicho, G. Undt, F. Wanschitz, M. Truppe, R. Seemann, and A. Wagner. Basic research and 12 years of clinical experience in computer-assisted navigation technology: a review. *International Journal of Oral and Maxillofacial Surgery*, 34(1):1–8, 2005.
- [37] Caroline Ewertsen, Adrian Săftoiu, Lucian G Gruionu, Steen Karstrup, and Michael B Nielsen. Real-time image fusion involving diagnostic ultrasound. *American Journal of Roentgenology*, 200(3):W249–W255, 2013.

- [38] A Fenster, KJM Surry, GR Mills, and DB Downey. 3D ultrasound guided breast biopsy system. *Ultrasonics*, 42(1-9):769–774, 2004.
- [39] Aaron Fenster, Mingyue Ding, Ning Hu, Hanif M. Ladak, Guokuan Li, Neale Cardinal, and Dónal B. Downey. Visualization and Segmentation Techniques in 3D Ultrasound Images. In *Computer Vision Beyond the Visible Spectrum*, pages 241–269. Springer-Verlag, 2005.
- [40] Aaron Fenster and Donal B Downey. 3-D ultrasound imaging: A review. *IEEE Engineering in Medicine and Biology magazine*, 15(6):41–51, 1996.
- [41] D D. Frantz, Andrew D. Wiles, S E. Leis, and S R. Kirsch. Accuracy assessment protocols for electromagnetic tracking systems. *Physics in Medicine and Biology*, 48(14):2241–2251, 2003.
- [42] Alfred M. Franz, Tamas Haidegger, Wolfgang Birkfellner, Kevin Cleary, Terry M. Peters, and Lena Maier-Hein. Electromagnetic tracking in medicine -A review of technology, validation, and applications. *IEEE Transactions on Medical Imaging*, 33(8):1702–1725, 2014.
- [43] Jacob T Gibby, Samuel A Swenson, Steve Cvetko, · Raj Rao, Ramin Javan, and Raj Rao. Head-mounted display augmented reality to guide pedicle screw placement utilizing computed tomography. *International Journal of Computer Assisted Radiology and Surgery*, 2018.
- [44] Jacob T Gibby, Samuel A Swenson, Steve Cvetko, Raj Rao, and Ramin Javan. Head-mounted display augmented reality to guide pedicle screw placement utilizing computed tomography. *International journal of computer assisted radiology and surgery*, 14(3):525–535, 2019.
- [45] Derek J Gillies, Joseph Awad, Jessica R Rodgers, Chandima Edirisinghe, Derek W Cool, Nirmal Kakani, and Aaron Fenster. Three-dimensional therapy needle applicator segmentation for ultrasound-guided focal liver ablation. *Medical physics*, 46(6):2646–2658, 2019.
- [46] Girardeau-Montaut, Daniel. Cloudcompare-open source project. *OpenSource Project*, 2011.
- [47] Ross Girshick. Fast R-CNN. *Proceedings of the IEEE International Conference on Computer Vision*, 2015 International Conference on Computer Vision, ICCV 2015:1440–1448, 2015.

- [48] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the 13th International Conference On Artificial Intelligence and Statistics*, 9:249–256, 2010.
- [49] Andrew C. Gordon, Daniel Saliken, John C. and Johns, and Robin R. Owen, Richard and Gray. US-guided puncture of the internal jugular vein: Complications and anatomic considerations. *Journal of Vascular and Interventional Radiology*, 9(2):333–338, 1998.
- [50] Leah Groves, Natalie Li, Terry M. Peters, and Elvis C. S. Chen. Towards a mixed-reality first person point of view needle navigation system. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 245–253. Springer International Publishing, 2019.
- [51] Leah Groves, Adam Rankin, Terry M. Peters, and Elvis C. S. Chen. The effect of imaging and tracking parameters on ultrasound probe calibration robustness. In *Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 10951, page 109510X. International Society for Optics and Photonics, 2019.
- [52] Leah A. Groves et al. Accuracy assessment for the co-registration between optical and VIVE head-mounted display tracking. *International Journal of Computer Assisted Radiology and Surgery*, 14(7):1207–1215, 2019.
- [53] Groves, Leah A., VanBerlo, Blake, Veinberg, Natan, Alboog, Abdulrahman, Peters, Terry M., and Chen, Elvis C.S. Automatic segmentation of the carotid artery and internal jugular vein from 2d ultrasound images for 3d vascular reconstruction. *International journal of computer assisted radiology and surgery*, 2020.
- [54] Hengtao Guo, Sheng Xu, Bradford Wood, and Pingkun Yan. Sensorless Freehand 3D Ultrasound Reconstruction via Deep Contextual Learning. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 12263 LNCS, pages 463–472. Springer Science and Business Media Deutschland GmbH, 2020.
- [55] Ilker Hacihaliloglu, Parmida Beigi, Gary Ng, Robert N Rohling, Septimiu Salcudean, and Purang Abolmaesumi. Projection-based phase features for localization of a needle tip in 2D curvilinear ultrasound. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 347–354. Springer, 2015.

- [56] Charles R Hatt, Gary Ng, and Vijay Parthasarathy. Enhanced needle localization in ultrasound using beam steering and learning-based segmentation. *Computerized Medical Imaging and Graphics*, 41:46–54, 2015.
- [57] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [58] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017.
- [59] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):386–397, 2020.
- [60] Florian Heinrich, Luisa Schwenderling, Mathias Becker, Martin Skalej, and Christian Hansen. Holoinjection: augmented reality support for ct-guided spinal needle injections. *Healthcare Technology Letters*, 6(6):165–171, 2019.
- [61] Robert T Held and Tiffany T Hui. A guide to stereoscopic 3d displays in medicine. *Academic radiology*, 18(8):1035–1048, 2011.
- [62] Berthold K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, 4(4):629–642, 1987.
- [63] P. Hsu, R. W. Prager, A. H. Gee, and G. M. Treece. Freehand 3D Ultrasound Calibration: A Review. In *Advanced Imaging in Biology and Medicine*, pages 47–84. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [64] Victor F Humphrey. Ultrasound and matter—physical interactions. *Progress in biophysics and molecular biology*, 93(1-3):195–211, 2007.
- [65] Izzo, Richard, Steinman, David, Manini, Simone, and Antiga, Luca. The vascular modeling toolkit: a python library for the analysis of tubular structures in medical images. *Journal of Open Source Software*, 3(25):745, 2018.
- [66] Michele H Johnson. Head and neck vascular anatomy. *Neuroimaging Clinics of North America*, 8(1):119–141, 1998.
- [67] Naoki Kaneko, Makoto Sato, Taro Takeshima, Yoshihide Sehara, and Eiju Watanabe. Ultrasound-guided central venous catheterization using an optical see-through head-mounted display: A pilot study. *Journal of Clinical Ultrasound*, 44(8):487–491, 2016.

- [68] Naoki Kaneko, Mayumi Tsunoda, Masatsugu Mitsuhashi, Keisuke Okubo, Taro Takeshima, Yoshihide Sehara, Mutsumi Nagai, and Kensuke Kawai. Ultrasound-guided fine-needle aspiration in the neck region using an optical see-through head-mounted display: A randomized controlled trial. *Journal of Ultrasound in Medicine*, 36(10):2071–2077, 2017.
- [69] Dimitrios Karakitsos, Nicolaos Labropoulos, Eric De Groot, Alexandros P. Patrianakos, Gregorios Kouraklis, John Poularas, George Samonis, Dimosthenis A. Tsoutsos, Manousos M. Konstadoulakis, and Andreas Karabinis. Real-time ultrasound-guided catheterisation of the internal jugular vein: A prospective comparison with the landmark technique in critical care patients. *Critical Care*, 10(6):R162, 2006.
- [70] Ron Kikinis, Steve D Pieper, and Kirby G Vosburgh. 3d slicer: a platform for subject-specific image analysis, visualization, and clinical support. In *Intraoperative imaging and image-guided therapy*, pages 277–289. Springer, 2014.
- [71] Thomas Kilgus, Eric Heim, Sven Haase, Sabine Prüfer, Michael Müller, Alexander Seitel, Markus Fangerau, Tamara Wiebe, Justin Iszatt, Heinz-Peter Schlemmer, Joachim Hornegger, Kathrin Yen, and Lena Maier-Hein. Mobile markerless augmented reality and its application in forensic medicine. *International Journal of Computer Assisted Radiology and Surgery*, 10(5):573–586, 2015.
- [72] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [73] Roberta L Klatzky, Bing Wu, and George Stetten. Spatial Representations From Perception and Cognitive Mediation: The Case of Ultrasound. *Current directions in psychological science*, 17(6):359–364, 2008.
- [74] Gayatri Kompella, Maria Antico, Fumio Sasazawa, S Jeevakala, Keerthi Ram, Davide Fontanarosa, Ajay K Pandey, and Mohanasankar Sivaprakasam. Segmentation of femoral cartilage from knee ultrasound images using mask r-cnn. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 966–969. IEEE, 2019.
- [75] Jens Kowal, Christoph A Amstutz, Marco Caversaccio, and Lutz P Nolte. On the development and comparative evaluation of an ultrasound b-mode probe calibration method. *Computer Aided Surgery*, 8(3):107–119, 2003.

- [76] Daniela Kuhnt, Miriam HA Bauer, and Christopher Nimsy. Brain shift compensation and neurosurgical image fusion using intraoperative mri: current status and future challenges. *Critical Reviews™ in Biomedical Engineering*, 40(3), 2012.
- [77] Andras Lasso, Tamas Heffter, Adam Rankin, Csaba Pinter, Tamas Ungi, and Gabor Fichtinger. PLUS: Open-Source Toolkit for Ultrasound-Guided Intervention Systems. *IEEE Transactions on Biomedical Engineering*, 61(10):2527–2537, 2014.
- [78] Andras Lasso, Tamas Heffter, Adam Rankin, Csaba Pinter, Tamas Ungi, and Gabor Fichtinger. PLUS: Open-source toolkit for ultrasound-guided intervention systems. *IEEE Transactions on Biomedical Engineering*, 61(10):2527–2537, 2014.
- [79] Hillary Lia, Gregory Paulin, Caitlin T Yeo, Jessica Andrews, Nelson Yi, Hassan Haq, Steve Emmanuel, Kristian Ludig, Zsuzsanna Keri, Andras Lasso, et al. Hololens in suturing training. In *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 10576, page 1057628. International Society for Optics and Photonics, 2018.
- [80] Florentin Liebmann, Simon Roner, Marco von Atzigen, Davide Scaramuzza, Reto Sutter, Jess Snedeker, Mazda Farshad, and Philipp Fürnstahl. Pedicle screw navigation using surface digitization on the Microsoft HoloLens. *International Journal of Computer Assisted Radiology and Surgery*, 14(7):1157–1165, 2019.
- [81] Jun Liu and PengFei Li. A Mask R-CNN model with improved region proposal network for medical ultrasound image. In *Intelligent Computing Theories and Application*, pages 26–33. Springer International Publishing, 2018.
- [82] Albert Lo, Michael Oehley, Adam Bartlett, Dave Adams, Phil Blyth, and Saad Al-Ali. Anatomical variations of the common carotid artery bifurcation. *ANZ Journal of Surgery*, 76(11):970–972, 2006.
- [83] Jennifer K Logan, Soroush Rais-Bahrami, Baris Turkbey, Andrew Gomella, Hayet Amalou, Peter L Choyke, Bradford J Wood, and Peter A Pinto. Current status of mri and ultrasound fusion software platforms for guidance of prostate biopsies. *BJU international*, 114(5):641, 2014.
- [84] Shayne Longpre and Ajay Sohmshtetty. Facial keypoint detection. *Facial Detection Kaggle competition*, 2016.
- [85] Burton Ma, Niloofar Banihaveb, Joy Choi, Elvis C. S. Chen, and Amber L. Simpson. Is pose-based pivot calibration superior to sphere fitting? In *Medical Imaging 2017:*

- Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 10135, page 101351U. International Society for Optics and Photonics, 2017.
- [86] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3. Citeseer, 2013.
 - [87] Márcio Cerqueira de Farias Macedo, Antônio Lopes Apolinário Júnior, Antonio Carlos dos Santos Souza, and Gilson Antônio Giralddi. High-Quality On-Patient Medical Data Visualization in a Markerless Augmented Reality Environment. *SBC Journal on Interactive Systems*, 5(3):41–52, 2010.
 - [88] Nader Mahmoud, Óscar G. Grasa, Stéphane A. Nicolau, Christophe Doignon, Luc Soler, Jacques Marescaux, and J. M. M. Montiel. On-patient see-through augmented reality based on visual SLAM. *International Journal of Computer Assisted Radiology and Surgery*, 12(1):1–11, 2017.
 - [89] Christopher R. Mascott. Comparison of Magnetic Tracking and Optical Tracking by Simultaneous Use of Two Independent Frameless Stereotactic Systems. *Operative Neurosurgery*, 57(4):295–301, 2005.
 - [90] David C. McGee and Michael K. Gould. Preventing Complications of Central Venous Catheterization. *New England Journal of Medicine*, 348(12):1123–1133, 2003.
 - [91] John P McIntire, Paul R Havig, and Eric E Geiselman. Stereoscopic 3D displays and human performance: A comprehensive review. *Displays*, 35(1):18–26, 2014.
 - [92] Arian Mehrfard, Javad Fotouhi, Giacomo Taylor, Tess Forster, Nassir Navab, and Bernhard Fuerst. A comparative analysis of virtual reality head-mounted display systems. *arXiv preprint arXiv:1912.02913*, 2019.
 - [93] Ninfa Mehta, Walter Wallace Valesky, Allysia Guy, and Richard Sinert. Systematic review: is real-time ultrasonic-guided central line placement by ED physicians more successful than the traditional landmark approach? *Emergency Medicine Journal*, 30(5):355–359, 2013.
 - [94] Laurence Mercier, Thomas Langø, Frank Lindseth, and Louis D. Collins. A review of calibration techniques for freehand 3-D ultrasound systems. *Ultrasound in Medicine & Biology*, 31(2):143–165, 2005.
 - [95] Raymond L. Merritt, Michael E. Hachadorian, Kristof Michaels, Eric Zevallos, Kubwimana M. Mhayamaguru, Zuheily Closser, and Charlotte Derr. The effect of head

- rotation on the relative vascular anatomy of the neck: Implications for central venous access. *Journal of emergencies, trauma, and shock*, 11(3):193–196, 2018.
- [96] J. Michael Fitzpatrick, Jay B. West, and Calvin R. Maurer. Predicting error in rigid-body point-based registration. *IEEE Transactions on Medical Imaging*, 17(5):694–702, 1998.
- [97] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality: a class of displays on the reality-virtuality continuum. In *Telemanipulator and Telepresence Technologies*, volume 2351, pages 282 – 292. International Society for Optics and Photonics, SPIE, 1995.
- [98] A. H. Miller. Ultrasound Guidance versus the Landmark Technique for the Placement of Central Venous Catheters in the Emergency Department. *Academic Emergency Medicine*, 9(8):800–805, 2002.
- [99] G Edward Morgan, Maged S Mikhail, Michael J Murray, and C Philip Larson. *Clinical anesthesiology*, volume 361. Lange Medical Books/McGraw-Hill New York, 2006.
- [100] Michael Müller, Marie-Claire Rassweiler, Jan Klein, Alexander Seitel, Matthias Gondan, Matthias Baumhauer, Dogu Teber, Jens J. Rassweiler, Hans-Peter Meinzer, and Lena Maier-Hein. Mobile augmented reality for computer-assisted percutaneous nephrolithotomy. *International Journal of Computer Assisted Radiology and Surgery*, 8(4):663–675, 2013.
- [101] Nassir Navab, Joerg Traub, Tobias Sielhorst, Marco Feuerstein, and Christoph Bichlmeier. Action- and workflow-driven augmented reality for computer-aided medical procedures. *IEEE Computer Graphics and Applications*, 27(5):10–14, 2007.
- [102] NDI. NDI Aurora - User Manual.
- [103] NDI Inc. *IL 1070116 spectra User Guide Passive Polaris*, 2012.
- [104] Nigel Newbutt, Ryan Bradley, and Iian Conley. Using virtual reality head-mounted displays in schools with autistic children: views, experiences, and future directions. *Cyberpsychology, Behavior, and Social Networking*, 23(1):23–33, 2020.
- [105] Ba Vinh Nguyen, Gwenaél Prat, Jean Louis Vincent, Emmanuel Nowak, Nicolas Bizien, Jean Marie Tonnelier, Anne Renault, Mehdi Ould-Ahmed, Jean Michel Boles, and Erwan L’Her. Determination of the learning curve for ultrasound-guided jugular central venous catheter placement. *Intensive Care Medicine*, 40(1):66–73, 2014.

- [106] P.K.T. Nichols and E. Major. Central venous cannulation. *Current Anaesthesia & Critical Care*, 1(1):54–60, 1989.
- [107] Diederick C. Niehorster, Li Li, and Markus Lappe. The Accuracy and Precision of Position and Orientation Tracking in the HTC Vive Virtual Reality System for Scientific Research. *i-Perception*, 8(3):204166951770820, 2017.
- [108] Wiro J Niessen, Carolien J Bouma, Koen L Vincken, and Max A Viergever. Error metrics for quantitative evaluation of medical image segmentation. In *Performance Characterization in Computer Vision*, pages 275–284. Springer, 2000.
- [109] J Alison Noble and Djamal Boukerroui. Ultrasound image segmentation: a survey. *IEEE Transactions on medical imaging*, 25(8):987–1010, 2006.
- [110] American Institute of Ultrasound in Medicine et al. Aium practice parameter for the performance of selected ultrasound-guided procedures, 2016.
- [111] Francisco P.M. Oliveira and João Manuel R.S. Tavares. Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering*, 17(2):73–93, 2014.
- [112] Dinesh D Patil and Sonal G Deore. Medical image segmentation: a review. *International Journal of Computer Science and Mobile Computing*, 2(1):22–27, 2013.
- [113] Robert Patterson, Marc D. Winterbottom, and Byron J. Pierce. Perceptual issues in the use of head-mounted visual displays. *Human Factors*, 48(3):555–573, 2006.
- [114] Stephanie L Perkins, Michael A Lin, Subashini Srinivasan, Amanda J Wheeler, Brian A Hargreaves, and Bruce L Daniel. A mixed-reality system for breast surgical planning. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 269–274. IEEE, 2017.
- [115] Terry M. Peters, Cristian A. Linte, Ziv Yaniv, and Jacqueline Williams. *Mixed and augmented reality in medicine*. CRC Press, 2018.
- [116] Etta D Pisano, Henry Fuchs, Andrei State, Mark A Livingston, Gentaro Hirota, William F Garrett, and Mary C Whitton. Augmented Reality Applied to Ultrasound-Guided Breast Cyst Aspiration. *Breast Disease*, 103, 1998.

- [117] Arash Pourtaherian, Farhad Ghazvinian Zanjani, Svitlana Zinger, Nenad Mihajlovic, Gary Ng, Hendrikus Korsten, et al. Improving needle detection in 3d ultrasound using orthogonal-plane convolutional networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 610–618. Springer, 2017.
- [118] F Prada, M Del Bene, L Mattei, L Lodigiani, S DeBeni, V Kolev, I Vetrano, L Solbiati, G Sakas, and F DiMeco. Preoperative magnetic resonance and intraoperative ultrasound fusion imaging for real-time neuronavigation in brain tumor surgery. *ULTRASCHALL IN DER MEDIZIN*, 2015.
- [119] R. W. Prager, U. Z. Ijaz, A. H. Gee, and G. M. Treece. Three-dimensional ultrasound imaging. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 224(2):193–223, 2010.
- [120] R.W. Prager, R.N. Rohling, A.H. Gee, and L. Berman. Rapid calibration for 3-D free-hand ultrasound. *Ultrasound in Medicine & Biology*, 24(6):855–869, 1998.
- [121] Philip Pratt, Matthew Ives, Graham Lawton, Jonathan Simmons, Nasko Radev, Liana Spyropoulou, and Dimitri Amiras. Through the HoloLens™ looking glass: augmented reality for extremity reconstruction surgery using 3D vascular models with perforating vessels. *European Radiology Experimental*, 2(1):1–7, 2018.
- [122] Lutz Prechelt. Early Stopping — But When? In *Neural Networks: Tricks of the Trade: Second Edition*, pages 53–67. Springer, Berlin, Heidelberg, 2012.
- [123] Raphael Prevost, Mehrdad Salehi, Simon Jagoda, Navneet Kumar, Julian Sprung, Alexander Ladikos, Robert Bauer, Oliver Zettinig, and Wolfgang Wein. 3D freehand ultrasound without external tracking using deep learning. *Medical Image Analysis*, 48:187–202, 2018.
- [124] Raphael Prevost, Mehrdad Salehi, Simon Jagoda, Navneet Kumar, Julian Sprung, Alexander Ladikos, Robert Bauer, Oliver Zettinig, and Wolfgang Wein. 3d freehand ultrasound without external tracking using deep learning. *Medical Image Analysis*, 48:187 – 202, 2018.
- [125] Long Qian, Mathias Unberath, Kevin Yu, Nassir Navab, Bernhard Fuerst, Alex Johnson, and Greg Osgood. Technical Note: Towards Virtual Monitors for Image Guided Interventions Real-time Streaming to Optical See-Through Head-Mounted Displays. *arXiv preprint arXiv:1710.00808*, 2017.

- [126] I Raad. Intravascular-catheter-related infections. *Lancet (London, England)*, 351(9106):893–8, 1998.
- [127] Rafa Rahman, Matthew E Wood, Long Qian, Carrie L Price, Alex A Johnson, and Greg M Osgood. Head-mounted display use in surgery: a systematic review. *Surgical innovation*, 27(1):88–100, 2020.
- [128] Bernhard Reitinger, Alexander Bornik, Reinhard Beichel, Georg Werkgartner, and Erich Sorantin. Augmented reality based measurement tools for liver surgery planning. In *Bildverarbeitung für die Medizin 2004*, pages 274–278. Springer, 2004.
- [129] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 1–14, 2015.
- [130] Jannick P. Rolland and Henry Fuchs. Optical Versus Video See-Through Head-Mounted Displays in Medical Visualization. *Presence: Teleoperators and Virtual Environments*, 9(3):287–309, 2000.
- [131] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9351, pages 234–241. Springer Verlag, 2015.
- [132] Michael Rosenthal, Andrei State, Joohi Lee, Gentaro Hirota, Jeremy Ackerman, Kurtis Keller, Etta D Pisano, Michael Jiroutek, Keith Muller, and Henry Fuchs. Augmented reality guidance for needle biopsies: An initial randomized, controlled trial in phantoms. *Medical Image Analysis*, 6(3):313–320, 2002.
- [133] Roberto Sanz-Requena, David Moratal, Diego Ramón García-Sánchez, Vicente Bodí, José Joaquín Rieta, and Juan Manuel Sanchis. Automatic segmentation and 3D reconstruction of intravascular ultrasound images for a fast preliminar evaluation of vessel pathologies. *Computerized Medical Imaging and Graphics*, 31(2):71–80, 2007.
- [134] Bernd Saugel, Thomas W. L. Scheeren, and Jean-Louis Teboul. Ultrasound-guided central venous catheter placement: a structured review and recommendations for clinical practice. *Critical Care*, 21(1):225, 2017.
- [135] Gregory A. Schmidt, Julien Maizel, and Michel Slama. Ultrasound-guided central venous access: what’s new? *Intensive Care Medicine*, 41(4):705–707, 2015.

- [136] HJ Scholten, Arash Pourtaherian, N Mihajlovic, HHM Korsten, and R A. Bouwman. Improving needle tip identification during ultrasound-guided procedures in anaesthetic practice. *Anaesthesia*, 72(7):889–904, 2017.
- [137] Lei Shi, Tao Luo, Li Zhang, Zhongcheng Kang, Jie Chen, Feiyue Wu, and Jia Luo. Preliminary use of hololens glasses in surgery of liver cancer. *Journal of Central South University. Medical sciences*, 43(5):500–504, 2018.
- [138] Nahian Siddique, Sidike Paheding, Colin P Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access*, 2021.
- [139] Tobias Sielhorst, Marco Feuerstein, and Nassir Navab. Advanced medical displays: A literature review of augmented reality. *IEEE/OSA Journal of Display Technology*, 4(4):451–467, 2008.
- [140] Pierre Soille. *Morphological Image Analysis*. Springer Berlin Heidelberg, 2004.
- [141] Nilam J. Soni, Luis F. Reyes, Holly Keyt, Alejandro Arango, Jonathan A. Gelfond, Jay I. Peters, Stephanie M. Levine, Sandra G. Adams, and Marcos I. Restrepo. Use of ultrasound guidance for central venous catheterization: a national survey of intensivists and hospitalists. *Journal of Critical Care*, 36:277 – 283, 2016.
- [142] Geoffrey A Sonn, Edward Chang, Shyam Natarajan, Daniel J Margolis, Malu Macairan, Patricia Lieu, Jiaoti Huang, Frederick J Dorey, Robert E Reiter, and Leonard S Marks. Value of targeted prostate biopsy using magnetic resonance–ultrasound fusion in men with prior negative biopsy and elevated prostate-specific antigen. *European urology*, 65(4):809–815, 2014.
- [143] Angela Sorriento, Maria Bianca Porfido, Stefano Mazzoleni, Giuseppe Calvosa, Miria Tenucci, Gastone Ciuti, and Paolo Dario. Optical and electromagnetic tracking systems for biomedical applications: a critical review on potentialities and limitations. *IEEE reviews in biomedical engineering*, 13:212–232, 2019.
- [144] Andrei State, David T Chen, Chris Tector, Andrew Brandt, Hong Chen, Ryutarou Ohbuchi, Michael Bajura, and Henry Fuchs. Observing a volume rendered fetus within a pregnant patient. In *Proceedings Visualization’94*, pages 364–368. IEEE, 1994.
- [145] K. J. M. Surry, H. J. B. Austin, A. Fenster, and T. M. Peters. Poly(vinyl alcohol) cryogel phantoms for use in ultrasound and MR imaging. *Physics in Medicine and Biology*, 49(24):5529–5546, 2004.

- [146] Michele Tonutti, Daniel S Elson, Guang-Zhong Yang, Ara W Darzi, and Mikael H Sodergren. The role of technology in minimally invasive surgery: state of the art, recent developments and future directions. *Postgraduate medical journal*, 93(1097):159–167, 2017.
- [147] Philip HS Torr and Andrew Zisserman. Mlesac: A new robust estimator with application to estimating image geometry. *Computer vision and image understanding*, 78(1):138–156, 2000.
- [148] Ulku C. Turba, Renan Uflacker, Christopher Hannegan, and J. Bayne Selby. Anatomic relationship of the internaljugular vein and the common carotid artery applied to percutaneous transjugular procedures. *CardioVascular and Interventional Radiology*, 28(3):303–306, 2005.
- [149] Giuseppe Turini, Sara Condino, Paolo Domenico Parchi, Rosanna Maria Viglialoro, Nicola Piolanti, Marco Gesi, Mauro Ferrari, and Vincenzo Ferrari. A microsoft hololens mixed reality surgical simulator for patient-specific hip arthroplasty training. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*, pages 201–210. Springer, 2018.
- [150] E. Ukwatta, J. Awad, D. Buchanan, G. Parraga, and A. Fenster. Three-dimensional semi-automated segmentation of carotid atherosclerosis from three-dimensional ultrasound images. In *Medical Imaging 2012: Computer-Aided Diagnosis*, volume 8315, pages 206–211. International Society for Optics and Photonics, SPIE, 2012.
- [151] Tamas Ungi, Hastings Greer, Kyle R. Sunderland, Victoria Wu, Zachary M.C. Baum, Christopher Schlenger, Matthew Oetgen, Kevin Cleary, Stephen R. Aylward, and Gabor Fichtinger. Automatic Spine Ultrasound Segmentation for Scoliosis Visualization and Measurement. *IEEE Transactions on Biomedical Engineering*, 67(11):3234–3241, 2020.
- [152] Tamas Ungi, Andras Lasso, and Gabor Fichtinger. Open-source platforms for navigated image-guided interventions. *Medical Image Analysis*, 33:181–186, 2016.
- [153] Thomas Vaughan, Andras Lasso, Tamas Ungi, and Gabor Fichtinger. Hole filling with oriented sticks in ultrasound volume reconstruction. *Journal of Medical Imaging*, 2(3):034002, 2015.

- [154] E. A. Wan and R. Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373)*, pages 153–158, 2000.
- [155] Huixiang Wang, Fang Wang, Anthony Peng Yew Leong, Lu Xu, Xiaojun Chen, and Qiugen Wang. Precision insertion of percutaneous sacroiliac screws using a novel augmented reality-based navigation system: a pilot study. *International Orthopaedics*, 40(9):1941–1947, 2016.
- [156] Wanyu Wang, Xinyang Liao, Elvis C. S. Chen, John Moore, John S. H. Baxter, Terry M. Peters, and Daniel Bainbridge. The Effects of Positioning on the Volume/Location of the Internal Jugular Vein Using 2-Dimensional Tracked Ultrasound. *Journal of Cardiothoracic and Vascular Anesthesia*, 2019.
- [157] Menachem M Weiner, Paul Geldard, and Alexander JC Mitnacht. Ultrasound-guided vascular access: a comprehensive review. *Journal of cardiothoracic and vascular anesthesia*, 27(2):345–360, 2013.
- [158] Gavin Wheeler, Shujie Deng, Kuberan Pushparajah, Julia A. Schnabel, John M. Simpson, and Alberto Gomez. Virtual linear measurement system for accurate quantification of medical images. *Healthcare Technology Letters*, 6(6):220–225, 2019.
- [159] Andrew D. Wiles, David G. Thompson, and Donald D. Frantz. Accuracy assessment and interpretation for optical tracking systems. In *Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display*, volume 5367, page 421. SPIE, 2004.
- [160] Dawit Habte Woldeyes. Anatomical variations of the common carotid artery bifurcations in relation to the cervical vertebrae in Ethiopia. *Anatomy & Physiology: Current Research*, 4(3), 2014.
- [161] Meiyan Xie, Yunzhu Li, Yunzhe Xue, Randy Shafritz, Saum A. Rahimi, Justin W. Ady, and Usman W. Roshan. Vessel lumen segmentation in internal carotid artery ultrasounds with deep convolutional neural networks. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2393–2398. IEEE, 2019.
- [162] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9(4):611–629, 2018.
- [163] Ziv Yaniv. Which pivot calibration? In *Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 9415, page 941527. SPIE, 2015.

- [164] Ran Zhou, Aaron Fenster, Yujiao Xia, J. David Spence, and Mingyue Ding. Deep learning-based carotid media-adventitia and lumen-intima boundary segmentation from three-dimensional ultrasound images. *Medical Physics*, 46(7):mp.13581, 2019.

Appendix A

Questionnaires

Appendix A includes the questionnaires A.1 and A.2 that were completed by the user's of the study in Chapters 4 and 6, respectively.

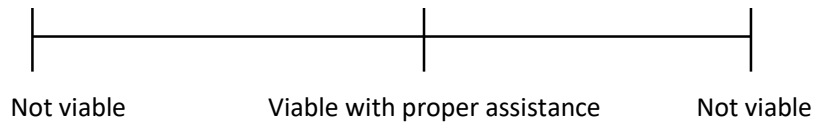
Figure A.1: User questionnaire used in Chapter 4, where each scale is 10 cm long and has three anchors with written descriptions at 0, 5 and 10 cm

Questionnaire

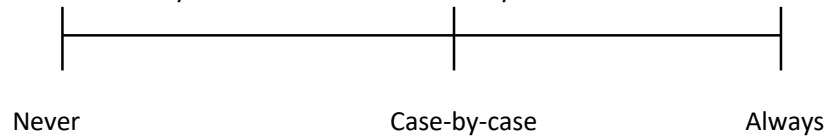
1. How many years of central line insertion experience do you have? Approximately how many central line insertions have you performed?

Years: _____ Number of insertions: _____

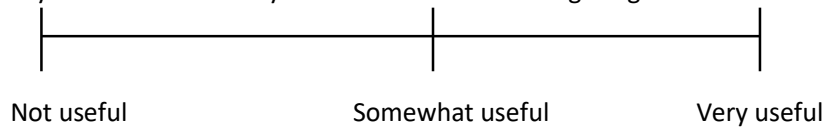
2. How viable is the HMD to use in the OR?



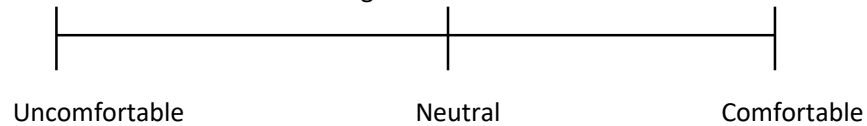
3. If the system was clinically available how often would you use it?



4. How useful do you think the HMD system would be for training US-guided CVC?



5. How was the comfort associated with using the HMD?



Other Comments: _____

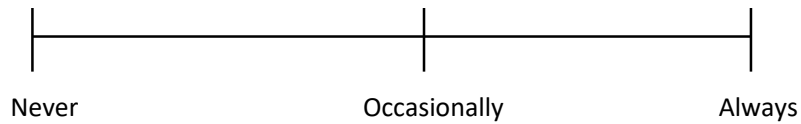
Figure A.2: User questionnaire used in Chapter 6, where each scale is 10 cm long and has three anchors with written descriptions at 0, 5 and 10 cm

Questionnaire

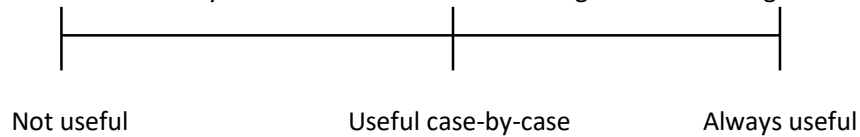
1. How many years of central line insertion experience do you have? Approximately how many central line insertions have you performed?

Years: _____ Number of insertions: _____

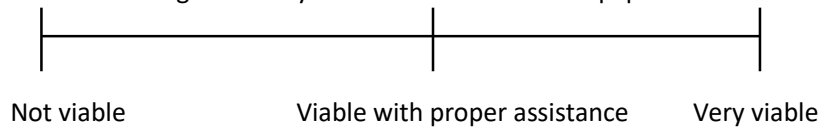
2. If the advanced guidance system was available how often would you use it?



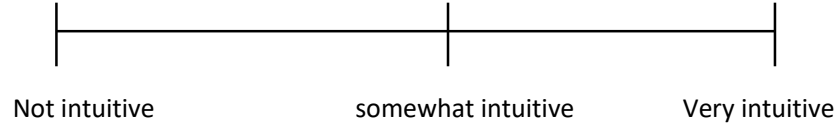
3. Do you think the advanced system could be useful for training for ultrasound guided CVC?



4. Do you think the advanced guidance system would be a viable equipment to use in the OR?



5. Did you find the relationship between the needle and vessels to be intuitive?



Other Comments:

Appendix B

Research Ethics Approval

Ethics approval was obtained from the **Health Sciences Research Ethics Board at Western University, London, Canada, and** for the studies presented in Chapters 4, 5 and 6.



Date: 20 December 2017

To: Elvis Chen

Project ID: 107254

Study Title: Validation of an augmented reality guidance system for needle placement in a phantom of the internal jugular vein

Application Type: Continuing Ethics Review (CER) Form

Review Type: Delegated

Full Board Reporting Date: January 9 2018

Date Approval Issued: 20/Dec/2017

REB Approval Expiry Date: 13/Jan/2019

Dear Elvis Chen ,

The Western University Research Ethics Board has reviewed the application. This study, including all currently approved documents, has been re-approved until the expiry date noted above.

REB members involved in the research project do not participate in the review, discussion or decision.

Western University REB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The REB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Kelly Patterson, Ethics Officer, on Behalf of Dr. Joseph Gilbert, HSREB Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).



Date: 8 January 2020

To: Dr. Daniel Bainbridge

Project ID: 114734

Study Title: Validation of a surgical navigation system for central line needle insertion in a phantom study

Application Type: HSREB Initial Application

Review Type: Delegated

Meeting Date / Full Board Reporting Date: 28/Jan/2020

Date Approval Issued: 08/Jan/2020

REB Approval Expiry Date: 08/Jan/2021

Dear Dr. Daniel Bainbridge

The Western University Health Science Research Ethics Board (HSREB) has reviewed and approved the above mentioned study as described in the WREM application form, as of the HSREB Initial Approval Date noted above. This research study is to be conducted by the investigator noted above. All other required institutional approvals must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
LOI	Written Consent/Assent	20/Dec/2019	3.0
Questionnaire VI	Paper Survey	03/Dec/2019	1.0
Research plan	Protocol	14/Nov/2019	2.0

Documents Acknowledged:

Document Name	Document Type	Document Date	Document Version
Flowchart study design	Flow Diagram	14/Nov/2019	1.0
References	References		

No deviations from, or changes to, the protocol or WREM application should be initiated without prior written approval of an appropriate amendment from Western HSREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University HSREB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Patricia Sargeant, Ethics Officer (ext. 85990) on behalf of Dr. Philip Jones, HSREB Vice-Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).



Date: 31 January 2020

To: Dr. Daniel Bainbridge

Project ID: 114604

Study Title: 3D Neck Vascular Structure Reconstruction using Freehand Ultrasound

Application Type: HSREB Initial Application

Review Type: Delegated

Meeting Date : 05/Nov/2019 13:00, 03/Dec/2019, 13:00 28/Jan/2020 13:00

Date Approval Issued: 31/Jan/2020 09:12

REB Approval Expiry Date: 31/Jan/2021

Dear Dr. Daniel Bainbridge

The Western University Health Science Research Ethics Board (HSREB) has reviewed and approved the above mentioned study as described in the WREM application form, as of the HSREB Initial Approval Date noted above. This research study is to be conducted by the investigator noted above. All other required institutional approvals must also be obtained prior to the conduct of the study.

Documents Approved:

Document Name	Document Type	Document Date	Document Version
Collection form	Other Data Collection Instruments	04/Dec/2019	2
Letter of Information and Consent 5.0	Written Consent/Assent	30/Jan/2020	5.0
Other_data_collection_instrument	Other Data Collection Instruments	03/Sep/2019	1.0
Research plan 4.0	Protocol	09/Jan/2020	4

No deviations from, or changes to, the protocol or WREM application should be initiated without prior written approval of an appropriate amendment from Western HSREB, except when necessary to eliminate immediate hazard(s) to study participants or when the change(s) involves only administrative or logistical aspects of the trial.

REB members involved in the research project do not participate in the review, discussion or decision.

The Western University HSREB operates in compliance with, and is constituted in accordance with, the requirements of the TriCouncil Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2); the International Conference on Harmonisation Good Clinical Practice Consolidated Guideline (ICH GCP); Part C, Division 5 of the Food and Drug Regulations; Part 4 of the Natural Health Products Regulations; Part 3 of the Medical Devices Regulations and the provisions of the Ontario Personal Health Information Protection Act (PHIPA 2004) and its applicable regulations. The HSREB is registered with the U.S. Department of Health & Human Services under the IRB registration number IRB 00000940.

Please do not hesitate to contact us if you have any questions.

Sincerely,

Nicola Geoghegan-Morphet, Ethics Officer on behalf of Dr. Joseph Gilbert, HSREB Chair

Note: This correspondence includes an electronic signature (validation and approval via an online system that is compliant with all regulations).

Appendix C

Reprint Permissions

For the chapters that were adapted from previously published articles, permission to reprint was obtained.

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Jun 28, 2021

This Agreement between Ms. Leah Groves ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5097711377153
License date	Jun 28, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	International Journal of Computer Assisted Radiology and Surgery
Licensed Content Title	Accuracy assessment for the co-registration between optical and VIVE head-mounted display tracking
Licensed Content Author	Leah A. Groves et al
Licensed Content Date	May 8, 2019
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	yes
Title	Accuracy assessment for the co-registration between optical and VIVE head-mounted display tracking
Institution name	Western University

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Jun 28, 2021

This Agreement between Ms. Leah Groves ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5097720533660
License date	Jun 28, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	Towards a Mixed-Reality First Person Point of View Needle Navigation System
Licensed Content Author	Leah Groves, Natalie Li, Terry M. Peters et al
Licensed Content Date	Jan 1, 2019
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	yes
Title	Accuracy assessment for the co-registration between optical and VIVE head-mounted display tracking
Institution name	Western University

SPRINGER NATURE LICENSE TERMS AND CONDITIONS

Jun 28, 2021

This Agreement between Ms. Leah Groves ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5097720324046
License date	Jun 28, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	International Journal of Computer Assisted Radiology and Surgery
Licensed Content Title	Automatic segmentation of the carotid artery and internal jugular vein from 2D ultrasound images for 3D vascular reconstruction
Licensed Content Author	Leah A. Groves et al
Licensed Content Date	Aug 24, 2020
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	yes
Title	Accuracy assessment for the co-registration between optical and VIVE head-mounted display tracking
Institution name	Western University



Improving central line needle insertions using in-situ vascular reconstructions

Author: Leah A. Groves, , Natalie U. et al

Publication: Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization

Publisher: Taylor & Francis

Date: jul 4, 2021

Rights managed by Taylor & Francis

Thesis/Dissertation Reuse Request

Taylor & Francis is pleased to offer reuses of its content for a thesis or dissertation free of charge contingent on resubmission of permission request if work is published.

[BACK](#)

[CLOSE](#)

Curriculum Vitae

Name: Leah Groves

Post-Secondary Education and Degrees: **Bachelors in Biomedical Engineering**
University of Guelph, Guelph Ontario
2013 - 2017

Honours and Awards: **Ontario Graduate Scholarship**
2019-2020

Related Work Experience: **Teaching Assistant**
The University of Western Ontario
2017 - 2021
Research Assistant
Mount Sinai Hospital, Toronto
2016-2018

Accepted Journal Manuscripts:

Groves, L. A., VanBerlo, B., Veinberg, N., Alboog, A., Peters, T. M., & Chen, E. C. (2020). Automatic segmentation of the carotid artery and internal jugular vein from 2D ultrasound images for 3D vascular reconstruction. *International Journal of Computer Assisted Radiology and Surgery*, 15(11), 1835-1846.

Groves, L. A., Li, N., VanBerlo, B., Veinberg, N., Peters, T. M., & Chen, E. C. (2020). Improving central line needle insertions using *in-situ* vascular reconstructions. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 1-7.

Groves, L. A., VanBerlo, B., Peters, T. M., & Chen, E. C. (2019). Deep learning approach for automatic out-of-plane needle localisation for semi-automatic ultrasound probe calibration. *Healthcare technology letters*, 6(6), 204-209.

Groves, L. A., Carnahan, P., Allen, D. R., Adam, R., Peters, T. M., & Chen, E. C. (2019). Accuracy assessment for the co-registration between optical and VIVE head-mounted display tracking. *International journal of computer assisted radiology and surgery*, 14(7), 1207-1215.

Conference Proceedings:

Groves, L., Li, N., Peters, T. M., & Chen, E. C. (2019, October). Towards a mixed-reality first person point of view needle navigation system. *In International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 245-253). Springer, Cham.

Groves, L., Rankin, A., Peters, T., & Chen, E. C. (2019, March). The effect of imaging and tracking parameters on ultrasound probe calibration robustness. *In Medical Imaging 2019: Image-Guided Procedures, Robotic Interventions, and Modeling* (Vol. 10951, p. 109510X). International Society for Optics and Photonics.

International Conference Presentations:

IPCAI 2020 Short Talk - Virtual

- Presented a 10 minute virtual talk titled Automatic 3D reconstruction of the carotid artery and jugular vein from tracked ultrasound

AECAI 2020 Short Talk - Virtual

- Presented a 5 minute virtual talk titled Improving central line needle insertions using in-situ vascular reconstructions

MICCAI 2019 Poster Presentation - Shen Zhen, China

- Presented a poster titled Towards a mixed-reality first person point of view needle navigation system

AECAI 2019 Short Talk - Shen Zhen, China

- Presented an 8 minute talk titled A deep learning approach for automatic out-of-plane needle localization for semi-automatic ultrasound probe calibration

IPCAI 2019 Short Talk - Rennes, France

- Presented an 8 minute talk titled Accuracy Assessment for the Co-registration Between Optical and VIVE Head Mounted Display Tracking

SPIE 2019 Long Talk- San Diego, USA

- Presented a 15 minute talk titled The effect of imaging and tracking parameters on ultrasound probe calibration robustness

Local Conference Presentations and Event Participation

Imaging Network Ontario 2021 - Virtual

- Pitch presentation titled Improving central line needle insertions using in situ vascular reconstructions

London Health Research Day 2020 - London, Canada

- Poster presentation titled Evaluation of a mixed-reality first person point of view needle navigation system for central line insertions

London Health Research Day 2019 - London, Canada

- Poster presentation titled The effect of imaging and tracking parameters on ultrasound probe calibration robustness

Imaging Network Ontario 2019 - London, Canada

- Oral presentation titled A mixed-reality first person point of view needle navigation system

MICCAI Educational Challenge 2018 - Granada, Spain

- Awarded third place in the MICCAI Educational Challenge 2018 for the development of an interactive blog on the topic of 3D Slicer and its use for image-guided therapy applications

London Imaging Discovery Day 2018 - London, Canada

- Short talk titled ultrasound Calibration: A Method for the Common User

London Health Research Day 2018 - London, Canada

- Poster presentation titled The effect of imaging and tracking parameters on ultrasound probe calibration robustness

Imaging Network Ontario 2018 - Toronto Canada

- Poster presentation titled The effect of imaging and tracking parameters on ultrasound probe calibration robustness

3 Minute Thesis Competition 2018 - London, Canada

- Western top 20 finalist: Bringing sight to surgery

Western University Health and Research Conference 2018 - London, Canada

- Keynote speaker: VAAST lab research overview

Supervisory Experience

ENGINEERING CAPSTONE TEAM 2020-2021

Assisted in the supervision and mentoring a capstone team on the development of a stand-alone ultrasound system capable of receiving, processing, and visualizing US images for image segmentation applications.

Blake VanBerlo 2019

Supervised Blake while he was a summer medical student in the VASST lab. I developed and managed the research project and provided Blake with the background, data and infrastructure necessary to complete a machine learning project over a 6 week placement. This resulted in a journal publication and continued collaboration over multiple journal papers.

Natalie Li 2018-2019

Supervised two of Natalie's undergraduate research projects resulting in a MICCAI conference proceeding and a Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization publication.

ENGINEERING CAPSTONE TEAM 2018-2019

Assisted and mentored students with their final capstone project. This involved teaching the students about 3D Slicer and a system I created using the platform, as well as advising them on their project.

Leadership and Volunteer experience

- Mental Health and Wellness Representative for the Biomedical Engineering Graduate Student Committee 2019-2021
- Wellness Ambassador for Biomedical Engineering 2018-2019
- Facilitated the Imaging Day for Western Summer Engineering Academy 2019
- Numerous demonstrations and presentations for Canadian Medical Hall of Fame Discovery Days 2017-2020
- Numerous demonstrations and presentations for Canadian Medical Hall of Fame School Museums 2017-2020