

A Tool-free Neuronavigation Method based on Single-view Hand Tracking

Fryderyk Victor Kögl^{a,b,*}, Étienne Léger^{a,*}, Nazim Haouchine^a, Erickson Torio^a, Parikshit Juvekar^a, Nassir Navab^{b,c}, Tina Kapur^a, Steve Pieper^{a,d}, Alexandra Golby^a and Sarah Frisken^a

^aHarvard Medical School, Brigham and Women’s Hospital, Boston, MA, USA;

^bComputer Aided Medical Procedures, Technische Universität München, Munich, Germany;

^cWhiting School of Engineering, Johns Hopkins University, Baltimore, USA;

^dIsomics, Inc., Cambridge, MA, USA;

ARTICLE HISTORY

Compiled September 12, 2022

ABSTRACT

This work presents a novel tool-free neuronavigation method that can be used with a single RGB commodity camera. Compared with freehand craniotomy placement methods, the proposed system is more intuitive and less error prone. The proposed method also has several advantages over standard neuronavigation platforms. First, it has a much lower cost, since it doesn’t require the use of an optical tracking camera or electromagnetic field generator, which are typically the most expensive parts of a neuronavigation system, making it much more accessible. Second, it requires minimal setup, meaning that it can be performed at the bedside and in circumstances where using a standard neuronavigation system is impractical. Our system relies on machine-learning-based hand pose estimation that acts as a proxy for optical tool tracking, enabling a 3D-3D pre-operative to intra-operative registration. Qualitative assessment from clinical users showed that the concept is clinically relevant. Quantitative assessment showed that on average a target registration error (TRE) of 1.3cm can be achieved. Furthermore, the system is framework-agnostic, meaning that future improvements to hand-tracking frameworks would directly translate to a higher accuracy.

KEYWORDS

Image-guided Neurosurgery - Neuronavigation - Computer-aided Interventions - Visualization - Augmented Reality

1. Introduction

Neuronavigation systems are a crucial aid during neurosurgery Marcus et al. (2015). They make it possible to spatially align preoperative scans (such as MRI or CT scans) with the intraoperative surgical field providing surgeons with orientation and guidance to treat targeted brain lesions. Neuronavigation was shown to reduce hospital stays, severe complication rates, procedure costs, patient discomfort, and recovery time Grunert et al. (2003). However, existing commercial systems suffer from being

* These authors contributed equally

CONTACT Étienne Léger Email: etienne.leger@mail.concordia.ca

expensive and cumbersome to use and setup Léger et al. (2022), making them unsuitable for bedside neurosurgical interventions (such as external ventricular drain) Robertson et al. (2021) and out of reach for most centers in low income regions of the world where there is an enormous and growing need for neurosurgery Dewan et al. (2019).

Neuronavigation systems are used primarily for burr hole placement. Several authors claim that it is its central role Wagner et al. (2000); Spivak and Pirouzmand (2005), due to problems such as mis-registration and brain shift that make their use in later parts of the surgery more difficult. Burr hole placement does not require sub-millimeter accuracy to be useful and to bring tangible clinical benefits relative to freehand methods, that are used in low- and middle-income countries (LMIC) or for bedside procedures. Despite this lower requirement and the important need for navigation, neuronavigation for bedside procedures and in LMIC remains lacking.

The method presented here aims to fulfill this unmet need using low-cost hardware without the need for proprietary tracking tools by proposing the first camera-based tool-free neuronavigation method that can be used with very low hardware requirements. The central idea behind our method is to replace optically-tracked instruments by camera-tracked surgeon’s fingers enabling an intuitive pre-operative to intra-operative data registration. Our system has five main advantages:

- It is inexpensive, since it only requires a commodity camera and a laptop computer, hardware that can be sourced for under a few hundred USD and are likely already available at most centers, thereby potentially incurring no cost.
- It is fast, with a complete registration process that can be achieved under 2 minutes.
- It is built entirely using open-source components and is framework-agnostic.
- It achieves low registration error with an average target registration error (TRE) of 1.3 cm, which is clinically relevant according to Rai et al. (2019).
- It is less error prone than manual methods, rendering wrong side surgeries less likely to occur, for instance.

We strongly believe that this novel method is a step forward in providing neurosurgical guidance for the wider populations of patients that may not have access to advanced surgical facilities or where neuronavigation systems are considered too clinically disruptive and cumbersome. Additionally, considering its minimal setup time and negligible cost, our method can be used in conjunction with manual methods to verify their correctness, thereby reducing errors.

2. Related Works

Augmented reality (AR) has been proposed as a more intuitive and less cumbersome alternative to standard neuronavigation. AR systems for neurosurgery can be divided into three categories based on their technical implementation, according to Chidambaram et al. (2021). The first category uses head-mounted displays (HMDs), to overlay objects, such as segmentations of the tumour or vessels, directly onto the user’s view (Low et al. (2010),Incekara et al. (2018)). Their main selling point is the fact that they can be used hands-free, allowing the surgeon to utilize both hands for surgical tasks and secondarily, maintaining sterile operative conditions. Approaches from this category are particularly advantageous during the early stages of surgery, such as during craniotomy and trajectory planning. They cannot be used during more

critical stages such as tumor resection due to their inherent accuracy limitations and the inability to account for brain shift (Incekara et al. (2018)). The second category consists of approaches that use a tablet or smartphone to overlay such information onto the video feed of its camera (Watanabe et al. (2016)Léger et al. (2020)). The main advantage of these approaches is their simplicity, and in contrast to the HMDs, they are more comfortable. The last category involves overlaying information onto the video feed from a surgical microscope (Haouchine et al. (2021b), Alfonso-Garcia et al. (2020), Sun et al. (2016)). Methods from these last two categories are accurate enough to be used intraoperatively (Chidambaram et al. (2021)). Using AR in such a way, facilitates maximal safe resection of gliomas, which is a strong predictor of patient survival (Lacroix et al. (2001), McGirt et al. (2008), Sanai et al. (2011), Smith et al. (2008)).

An application of AR in neuronavigation is the placement of the burr hole (small hole that a neurosurgeon makes in the skull), which is then used to establish a craniotomy or place a shunt. It has been shown that AR can play an important role in optimizing the size and shape of the craniotomy (Cho et al. (2020), Kersten-Oertel et al. (2016)). Kersten-Oertel et al. used a custom built neuronavigation workstation, a camera and a tracking system to overlay structures such as tumors and vessels on the video feed of the camera. Watanabe et al. (2016) proposed a method where a tablet with a camera is tracked in the room with a system of six external infrared cameras. Then, structures extracted from the pre-operative images are superimposed on the video feed of the tablet. This system was evaluated on 6 patients and was successfully used for planning the skin incision as well as the localization of the craniotomy.

All methods mentioned so far do not reduce the complexity of setup of a standard neuronavigation system, on the contrary, most of them have even greater requirements. However, another approach, which requires significantly less equipment, was developed by Hou et al. (2016). In this method, a sagittal slice of an MRI together with a 2D in-plane tumor segmentation was superimposed over the video feed from an iPhone. The task of the user was to manually align the camera field of view (FOV) with the MRI. Once this manual step was completed, the location of the tumor, i.e. the projection of the tumor, could be outlined on the patient’s skin. This was then used to centre the burr hole and create the craniotomy. This method, requiring manual alignment of images by the user, is thus intrinsically less accurate. Additionally, it is limited to displaying an AR overlay from only one specific view-point (sagittal).

Another aspect of burr hole placement that can be considered is the optimization of its size and location, accounting for brain shift. This helps ensure that the relevant brain structures can still be accessed from the planned access point. Optimizing the craniotomy has been addressed in retrospective studies for evaluating the insertion trajectory accuracy (Chen and Nakaji (2012)) and burr hole placement (Rai et al. (2019)). Both studies focused on burr holes, however, with more than 50 patients in each of them, they clearly show the significance of determining the optimal location and size of a craniotomy. Recently, an image-based method to optimize the craniotomy opening was proposed (Haouchine et al. (2021a)). This method combines physics-based simulation and neural image analogy to predict both the geometry and the appearance of the brain surface at chosen locations before opening the skull.

Tool-free methods for burr hole and craniotomy placement are methods that do not require a significant amount of equipment, contrary to commercial products or most of the previously mentioned methods. These methods will expectedly have lower precision, but they will be easier to use and set-up and can be made available in low income settings. The previously mentioned example developed by Hou et al. (2016)

relies solely on a mobile phone. While its alignment procedure is quite rudimentary and can be prone to error, it still likely offers better precision than using no guidance.

Our Contributions: In this paper, we present a tool-free, neuronavigation framework for burr hole placement. On the one hand, it is *tool-free*, in that it does not require any specialised hardware (e.g. multiple cameras, depth cameras, optical or magnetic trackers or HMDs). It only requires a monocular RGB camera (webcam, phone camera), that is used to track the user’s hand, and a laptop to perform the required processing steps and display the guidance. On the other hand, it follows *well-established principles* of commercial products, by using the already established workflow of choosing a set of points on the patient to determine the patient-image registration and offers real-time positioning relative to the patient after registration. As illustrated in Figure 1, this system tracks the user’s hand (which functions as a proxy to a standard tracked pointer) to collect points for registration and to visualise its position in relation to the pre-operative scans for the purpose of defining the location of the burr hole. To the best of our knowledge, no similar approaches have been proposed in the literature.

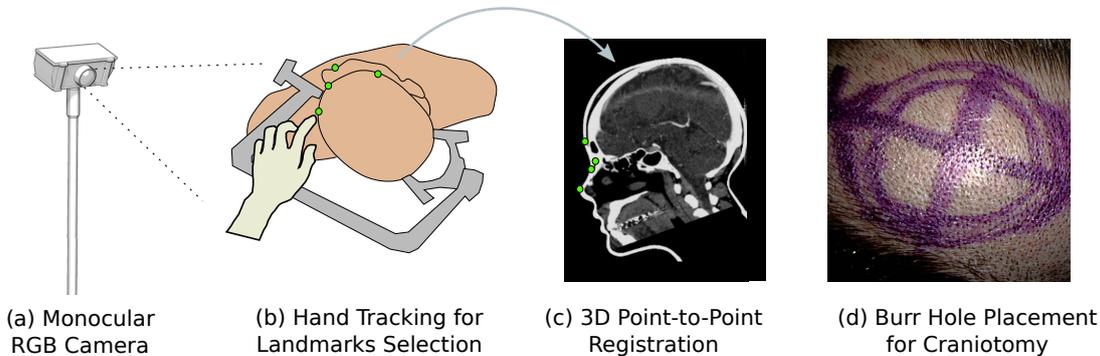


Figure 1.: Approach overview: using solely a single-view RGB camera in (a) our approach estimates the 3D pose of the surgeon’s hand and tracks it over time to select a set of facial landmarks on the patient’s face in (b); these landmarks are registered with a set of pre-defined anatomical landmarks in the MRI/CT scan (c) to finally allow the surgeon to optimally perform the craniotomy in (d).

3. Methods

3.1. Three-dimensional Finger Tracking

Tracking the fingers allows surgeons to interact in a naturally intuitive way with the system to identify the anatomical landmarks used for the registration. Hand tracking has been an active research topic that attained a high level of accuracy and robustness thanks to the successful integration of machine learning techniques. However, most of the existing methods provide only a 2.5D pose estimation using a monocular RGB camera, i.e. depth is estimated relative to the center of the hand, by opposition to a real 3D pose, where the depth would be defined in absolute terms, relative to the scene or the camera. In order to obtain a true 3D pose estimation, w.r.t the camera position, specialized hardware, e.g. depth sensors are required, which is not available

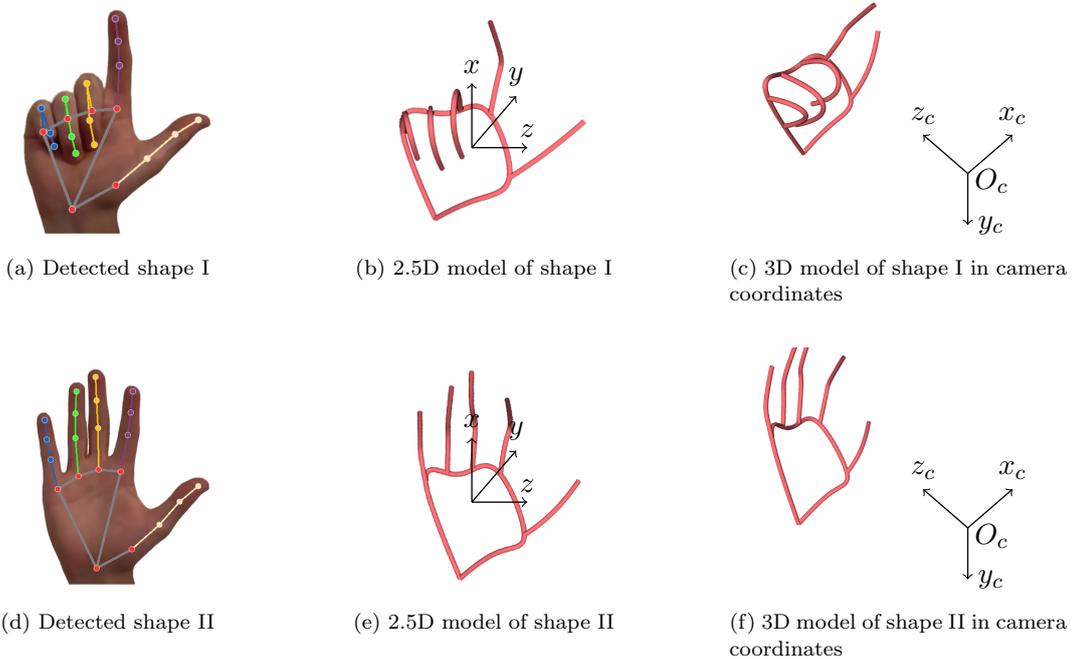


Figure 2.: 3D Hand Pose Estimation: our method uses learning models to detect and track hand landmarks in 2D and 2.5D and solves a PnP problem to estimate the full 3D hand. The top and bottom row depict two shapes as detected in a frame of the video feed and the corresponding reconstructed 2.5D (where the centre of the coordinate system is in the approximate geometric centre of the hand) and 3D models.

in our configuration. In the following, we describe our method to obtain a full 3D hand pose using solely a commodity monocular RGB camera.

In order to track the hand in the images and predict its 2.5D pose we rely on a machine learning method described by Zhang et al. (2020). This method uses a combination of two deep neural networks to output the position of multiple hand landmarks. This model uses the hand landmarks topology proposed by Simon et al. (2017) wherein 21 landmarks are defined to model hand articulations. The training relies on both real-world images and synthetic images. The real world images are used to learn the 2D coordinates, while the synthetic ones are used to learn a relative depth. This depth is calculated w.r.t. the hand center, thus limiting the pose to a 2.5D prediction only. This is not sufficient to select 3D points of interests on the patient’s face, since the center moves over time.

In order to obtain the 3D coordinates of the hand landmarks in a fixed coordinate frame (the camera space) we need to estimate the camera pose relative to the hand. This can be done by solving a classic Perspective- n -Point (PnP) problem Terzakis and Lourakis (2020). The aim of the PnP problem is to determine the position and orientation of a camera given its intrinsic parameters and a set of correspondences between an object’s 3D points and their 2D projections. With the estimated camera pose, we finally can estimate the full 3D hand pose illustrated in Figure 2.

3.2. Anatomical-based 3D-3D Registration

Preoperatively, a set of salient landmarks are selected on the patient scan. For this purpose, as detailed by Gerard et al. (2015), chosen landmarks need to be points that can be easily recognized and picked both on the preoperative scan and on the patient during surgery. However, in our case, since the pointing device, a finger tip, is blunter than a typical neurosurgical pointer, some commonly used landmarks, such as both *medial canthi*, aren't suitable. Then, intraoperatively, the surgeon uses the tip of one of their fingers to point out the preoperatively selected anatomical landmarks physically on the patient's head. To make this procedure more reproducible, the user is tasked to choose a specific point on their finger tip (e.g. center point of edge of nail) and reuse the same point for every acquisition. To adjust for the high level of noise present in the 3D hand landmark positions, two filtering steps are performed: reconstructed 3D positions temporal consistency is enforced using a low-pass filter with a two-frame sliding window (chosen empirically) and mean-based outlier removal is used for recording the registration points. The 3D position is recorded continuously for a duration of 100 frames. Of the resulting set of 100 points, those further than two standard deviations from the mean are removed (empirically around 40%) and the remaining are averaged.

The resulting two sets of points (i.e. recorded points and points defined on the scans) are brought into alignment using the fiducial registration method from Horn (1987). This method produces the optimal rigid transformation between the two sets of landmarks. Considering the amount of uncertainty in the finger tip position, two additional checks are done at this stage to ensure that all picked points are reliable.

First, if the standard deviation in the positions of the 100 individual points used in the temporal averaging is too high, the point is repicked. After having registered the two point sets, individual fiducial registration error (FRE) for each corresponding pair of points is computed. A larger FRE for one of the point pairs is likely to indicate that the pose estimation was inaccurate for that specific point acquisition. Points with a large FRE relative to the others are repicked until the FREs on all points are similar.

3.3. Craniotomy Placement Planning and Visualization

Using the registration described above, we can display the navigational information in two ways: 1) a standard neuronavigation view where the real-time position of the hand can be shown in relation to the patient's head in the 3D virtual view and individual MRI slices, and 2) Augmented Reality view where the preoperative scan and planing (trajectory or craniotomy placement) is superimposed on the camera image. Both of these visualizations are equally straightforward to produce since, with our system, the camera is used as the reference frame and the camera's intrinsic parameters are already determined. This means that going to and from camera space is only a matter of inverting the transformation matrix.

After registration, a deformable hand model is displayed at its correct position and with the correct pose relative to the patient in the virtual 3D view, as well as in the three orthogonal cardinal planes. This allows the surgeon to use their finger tip as a pointer to navigate and plan the craniotomy.

3.4. Implementation Details

As an initial preparation step, the camera’s intrinsic parameters and lens distortion need to be obtained. This step only needs to be done once, as this matrix is constant over time. It was derived using Bouguet’s Matlab implementation of the standard method from Zhang (2000), with the lens distortion estimation from Heikkila and Silven (1997).

Real-time image capture processing and display is done with the OpenCV toolbox Bradski (2000). The 2.5D hand pose is recovered using the MediaPipe machine learning framework Zhang et al. (2020). Ad-hoc corrections were made to adapt the MediaPipe detection framework to our need and improve the final 3D pose accuracy. First, due to the data it was trained with, MediaPipe returns positions on the front face of the hand, relative to the camera. In order to obtain the hand skeleton, all points were back-projected along the camera axis by half of the hand thickness. This thickness was empirically determined for every user and adjusted as a parameter. Second, MediaPipe also returns finger tip positions that are consistently too short. To alleviate this problem, again, a compensation parameter was determined for every user and adjusted. To solve the PnP problem, the method from Terzakis and Lourakis (2020) is used. At every frame, the recovered 3D positions of all 21 hand landmarks are streamed to 3D Slicer Fedorov et al. (2012) using the OpenIGTLink Tokuda et al. (2009) protocol.

Registration points are recorded in 3D Slicer using a custom-made module. Once all necessary points have been recorded, the registration transform is computed in 3D Slicer. From that point onwards, the computed transform is applied in 3D Slicer to the received streamed points, meaning that the hand is displayed in real-time relative to the preoperative scan.

The entire code-base is publicly available and shared with an open-source license². It includes both the code for the custom 3D Slicer extension and for running the hand detection on a video feed.

4. Results

To assess system accuracy, a full system performance assessment on a head phantom was performed in a laboratory environment (See Figure 3). Furthermore, baseline accuracy, using manual methods, was established. Additional tests were done specifically on the hand pose estimation to assess its accuracy and potential failure points. These experiments aimed at uncovering potential limitations in the currently used framework and inform potential improvements that can be made to it in the future to increase overall system performance.

4.1. Baseline

In order to establish a baseline and to be able to compare our concept with manual methods, we asked a neurosurgeon, experienced in neurosurgery without navigation, to place six points on a phantom using two analog methods. The first method did not use any tools - it involved carefully studying the axial, coronal and sagittal slices of a pre-operative scan without three-dimensional reconstruction and comparing features

²<https://github.com/koegl/HandNav>

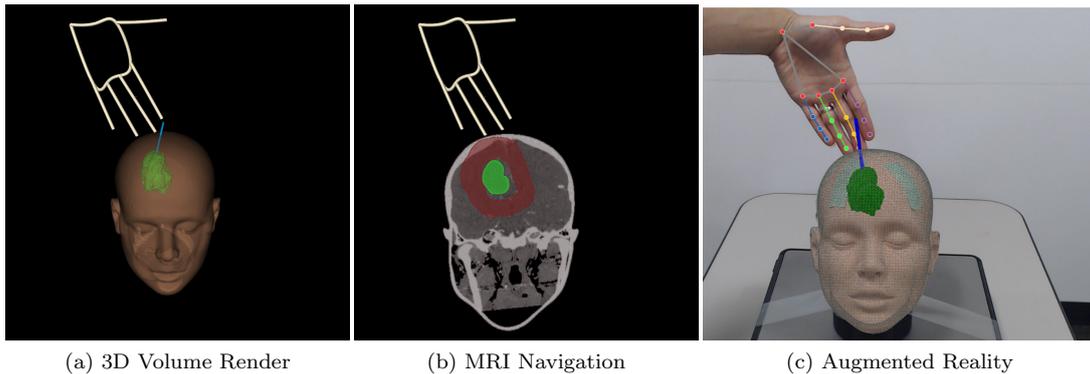


Figure 3.: (a): Implemented registered hand displayed overlaid in real-time on the segmented skin model, tumour and trajectory planning; (b): Depiction of the tracked hand over a selected MRI slice; (c): Depiction of the achievable AR view where a pre-segmented 3D virtual model is re-projected onto the live camera feed.

found on the patient and in the scan (e.g. midline, canthi, nasion, parietal boss,inion, tragi, among others) in order to find the correct surface point. The second method involved using a measuring tape and ruler. First, landmarks and distances were measured in the pre-operative scan (axial, coronal, and sagittal without three-dimensional reconstruction). Second, the same landmarks and distances were found on the phantom using the aforementioned ruler and tape. The points selected were compared to the points established before the experiment by another operator with the Brainlab Curve Brainlab (2022), which acted as the ground truth. The mean distances and standard deviations for the six points found with each method are shown in Table 1.

Table 1.: Baseline accuracy of two manual methods

Method type	Mean distance (mm)	STD (mm)
Tool-free estimation	11.4	1.6
Measurement-based estimation	5.8	4.6

4.2. Full System Evaluation

In this experiment, a phantom head was computer tomographed, the skin surface was segmented and six external landmarks were selected (both lateral canthi, both tragi and oral commissures) on that segmentation. Six target points were marked on the phantom as craniotomy locations (one frontal and one occipital along the midline, and a parietal and temporal on each side). All of the skin landmarks, as well as six targets points, were captured using an optical tracker (Optitrack V120:Duo, NaturalPoint, Inc., Corvallis, OR, USA).

The full pipeline, as described in section 3, was run to obtain the patient-to-image registration transform with our system, using the corresponding six skin landmarks. The distance between the target points picked with our system were then compared with those acquired with the optical tracking system that we used as our ground truth.

We obtained a mean landmarks registration error (FRE) of 7.97 mm \pm 3.73 mm (standard deviation), with a range of [2.48-12.46] mm and a mean target registration

error (TRE) of 13.31 mm \pm 3.36 mm, with a range of [10.30-19.63] mm.

In addition to the accuracy assessment, three users were asked to qualitatively assess the potential clinical usability of the devised system on a Likert scale (one neurosurgeon, one post-doctoral fellow and one medical student). There were six statements in the evaluation:

- S1: The system is clinically relevant.
- S2: You would prefer to use this system to not rely solely on manual methods.
- S3: You would feel comfortable using this system if no other neuronavigation system is available.
- S4: The system workflow is intuitive.
- S5: The system is easy to set up.
- S6: The system can help to verify a manual craniotomy placement.

and five responses to each statement: 1-Strongly disagree, 2-Disagree, 3-Neither agree or disagree, 4-Agree, 5-Strongly agree. The results from this evaluation are shown in Table 2.

Table 2.: Clinical evaluation of the system

	S1	S2	S3	S4	S5	S6
Neurosurgeon	5	4	4	4	4	5
Research fellow	5	4	4	3	4	5
Medical student	4	4	3	5	5	4

4.3. Effectiveness assessment of different system components

Multiple hand pose detection models were tested in designing the system and while the one used was the most robust, it still showed limitations. Some experiments were therefore conducted to quantify these limitations and failure modes, in order to better inform useful improvements that could be made to it to increase system performance.

4.3.1. Consistency of point picking

A simple test was devised to test the consistency of re-picking the same point in 3D space. To do this, a user was asked to pick the same fixed point 16 times. Between each acquisition the hand was moved out of the frame. The mean distance of all 16 points to their geometric center was 1.95 mm. The standard deviation of the points in x, y and z was 0.322 mm, 0.237 mm and 2.210 mm, respectively. The greatest 3D distance between any two points was 6.60 mm.

4.3.2. Hand orientation and pose variation

The first test in this experiment consisted in leaving the finger tip in a fixed location, but varying the orientation of the hand relative to the camera. Results from this experiment are displayed in Figure 4. In (a) we can see the spread of the points in the camera plane. In (b) we can see the spread of the points perpendicular to the camera axis. It is clear, that the variability is greater along z (the camera axis), meaning that the depth estimation has a greater error than the in-plane estimation.

The second test consisted in leaving the finger tip in a fixed location, but varying the hand pose, such as bending the thumb or spreading the fingers. Results from this experiment are displayed in Figure 5. Similarly to the previous experiment, the

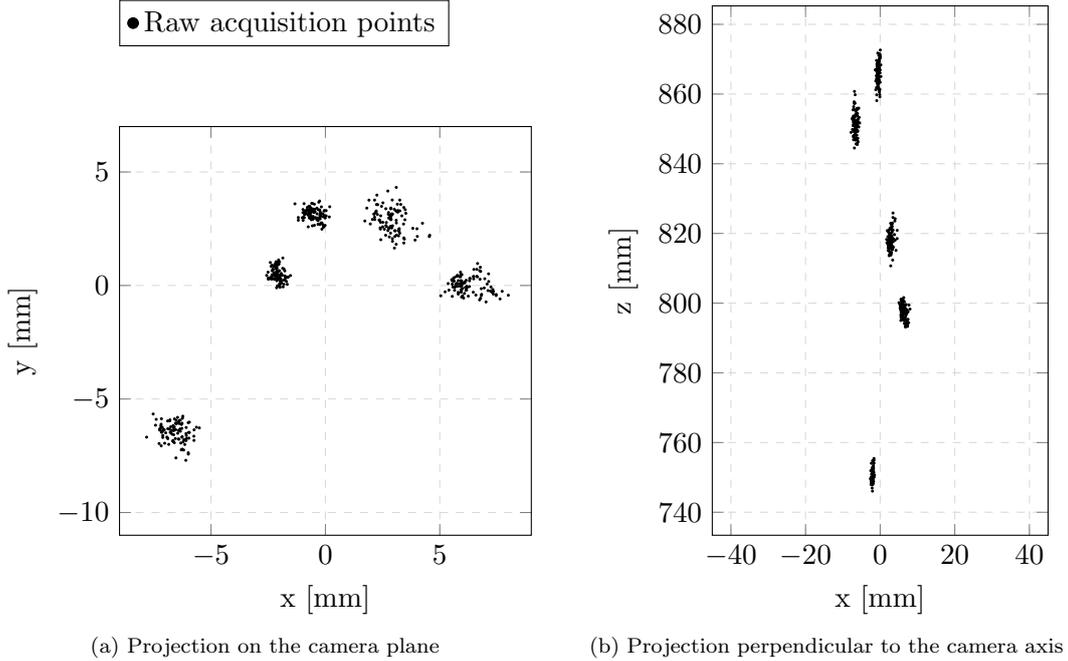


Figure 4.: Each of the five point clouds represents an acquisition of the same point at a different orientation of the hand. (a) Projection of the point clouds perpendicular to the camera axis. (b) Projection of the point clouds along the camera axis. Notice the difference in scale between (a) and (b).

variability in z is much greater than in-plane.

It can be clearly seen from this analysis that acquisitions are spread out in space, especially in the axis pointing to the camera. This would be a characteristic of all tracking methods trying to predict 3D coordinates from a 2D image, as the in-plane coordinates can be determined accurately directly from the pixel values. Furthermore, even a slight change in the orientation or pose of the hand has a big influence on the predicted positions. From our experience, uncommon hand-poses, that were probably not represented in the training set, were not accurately detected. Considering those limitations, special care was being placed on keeping a steady hand pose and orientation during acquisitions.

4.3.3. Ablation study

In addition to the framework assessment, we tested the effect of some components of our design by performing an ablation study. Two complete system runs were performed to compare against the previously reported full system results: one where no outlier removal was done and one where neither the outlier removal nor temporal filtering were done. Results from this experiment are reported in Table 3. This clearly shows that each post-processing step reduces the error of the system.

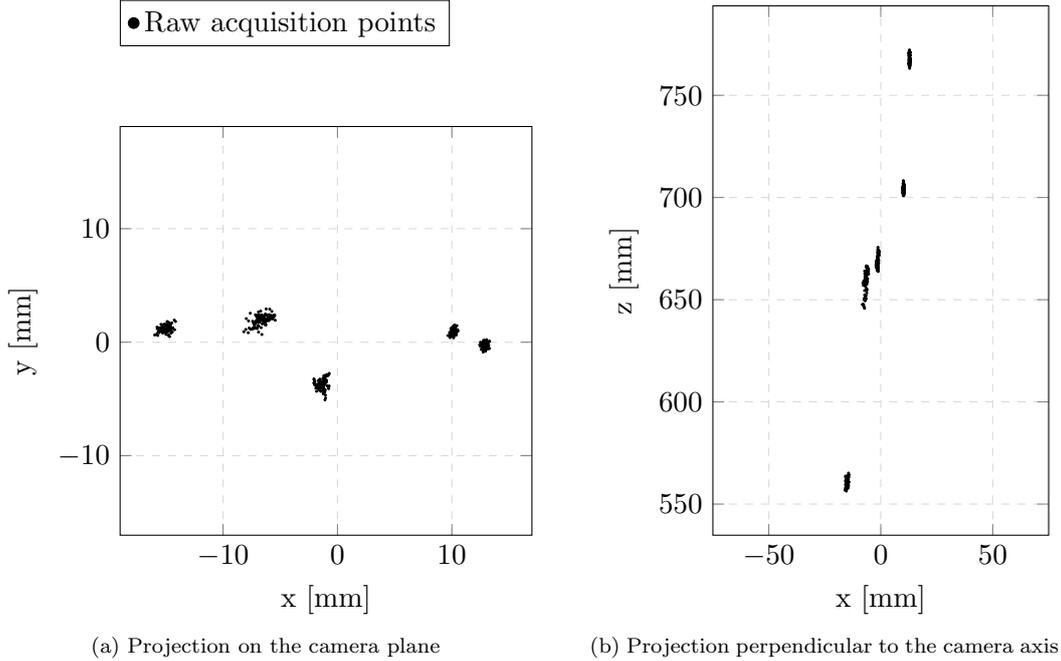


Figure 5.: Each of the five point clouds represents an acquisition of the same point at a different pose (shape) of the hand. (a) Projection of the point clouds perpendicular to the camera axis. (b) Projection of the point clouds along the camera axis. Notice the difference in scale between (a) and (b).

Table 3.: Results from the ablation study

Condition	Mean FRE (mm)	Mean TRE (mm)
Full system	7.97 ± 3.73	13.31 ± 3.36
Without outlier removal	13.77 ± 4.35	19.9 ± 10.44
Without outlier removal and temporal filtering	14.06 ± 4.73	22.79 ± 6.60

5. Discussion

The accuracy of our system detailed above in subsection 4.2 is currently on par with one of our analog baselines from subsection 4.1. Furthermore, it is likely only a lower bound as to what can be achieved with our proposed technique because predictive accuracy of machine learning tools for both 2D positions of hand landmarks as well as pose estimation will keep improving. 2.5D hand-tracking from single monocular camera is a currently very active area of research with a rapidly evolving state-of-the-art. Additionally, our method isn't dependent on a specific hand-tracking framework. That used in the presented prototype is the one that gave us the best results, but it could easily be swapped without affecting the rest of our pipeline. This means that our method will directly benefit from future developments of new frameworks and availability of larger and more complete training datasets.

According to the small system evaluation performed, as summarised in Table 2, the system was deemed of having potential clinical significance. While the small sample size prohibits from drawing detailed conclusions, it shows that this concept is promising for the intended application.

A second point highlighted by the data shown in subsection 4.3 is how currently available training sets have limitations that are hard to reconcile with the need for accurate positioning demanded by this application. The results from the ablation study also confirm that both noise reduction methods employed in our pipeline do achieve a significant increase in overall accuracy.

The hand-tracking framework we used was designed with the intent of providing hand pose for casual AR applications, typically for entertainment purposes or gesture detection. These applications have much lower accuracy requirements and can cope with using only a relative pose, unlike our application.

This is the reason why with the currently used framework, to achieve good accuracy, the user needs to maintain a constant pose and hand orientation relative to the camera. These limitations aren't inherent to our approach, but are rather specific to the machine learning framework and can easily be addressed in the future by training a tailored model.

Such a model, in order to achieve optimal patient registration accuracy should have high predictive accuracy of the finger tips in particular. It should also be trained on the most complete set of poses and hand orientations possible, spanning the entire 3D volume covered by the camera view. Indeed, contrary to casual AR applications, in our use case the hand will not necessarily be centered in the image and may present any orientation. Finally, it should cover all hand sizes, shapes and skin tones to ensure that the registration accuracy isn't influenced by inter-user variability of hand anatomy.

While it was not possible with our current setup because of significantly high signal to noise ratio, which required temporal smoothing, a future revision of the system could record the finger tip position over time, as it is slid across the patient's head to acquire a trace. This would enable our method to perform surface-to-surface registration in addition to landmark registration. This may perhaps help in improving accuracy further.

A significant advantage of our method is that it is compatible with other existing methods, facilitating integration into existing low-cost platforms like NousNav Léger et al. (2022) with minimal alterations, effectively replacing more costly tracking solutions and expanding the capabilities of such a low-cost full-suite neuronavigation solution.

In addition to identifying potential weaknesses of the current models, the results of subsection 4.3 can also inform camera position choices relative to the patient to optimize system accuracy. Each individual point-cloud has a much larger spread in the direction perpendicular to the image plane, about 3-7 times higher than in plane.

6. Conclusion

Our system shows that a burr hole placement system that doesn't require any specialized tools besides a commodity camera and provides 3D localization information in real-time can be built. This system, being tool-free, requires minimal setup, which makes it suitable for all instances where standard neuronavigation systems can't be used due to the logistical constraints of bedside procedures and emergency cases. Being low-cost and light-weight, the system is also well suited for lower resource settings and remote areas, where standard neuronavigation systems are not available. The devised system is also intuitive to use, making it less prone to errors than previously proposed tool-free systems. Furthermore, clinical users deemed the system and its accuracy to be clinically relevant, mainly in the context of error reduction and as a second pair

of eyes. Finally, it can be expected that the accuracy would significantly improve in the future, either if existing frameworks improve or if a purpose-made hand detection framework was built.

Funding

This work was supported by the following grants: NIH P41EB028741 and NIH R03EB032050 from the National Institute of Health, Jennifer Oppenheimer Cancer Research Initiative and Fonds de recherche du Québec - Nature et technologies.

References

- Alfonso-Garcia A, Bec J, Sridharan Weaver S, Hartl B, Unger J, Bobinski M, Lechpammer M, Girgis F, Boggan J, Marcu L. 2020. Real-time augmented reality for delineation of surgical margins during neurosurgery using autofluorescence lifetime contrast. *Journal of biophotonics*. 13(1):e201900108.
- Bradski G. 2000. The OpenCV Library. *Dr Dobb's Journal of Software Tools*.
- Brainlab. 2022. Brainlab curve. Available from: <https://www.brainlab.com/surgery-products/overview-platform-products/curve-image-guided-surgery/>.
- Chen F, Nakaji P. 2012. Optimal entry point and trajectory for endoscopic third ventriculostomy: evaluation of 53 patients with volumetric imaging guidance. *Journal of neurosurgery*. 116(5):1153–1157.
- Chidambaram S, Stifano V, Demetres M, Teyssandier M, Palumbo MC, Redaelli A, Olivi A, Apuzzo ML, Pannullo SC. 2021. Applications of augmented reality in the neurosurgical operating room: A systematic review of the literature. *Journal of Clinical Neuroscience*. 91:43–61.
- Cho J, Rahimpour S, Cutler A, Goodwin CR, Lad SP, Codd P. 2020. Enhancing reality: a systematic review of augmented reality in neuronavigation and education. *World Neurosurgery*. 139:186–195.
- Dewan MC, Rattani A, Fiegeen G, Arraez MA, Servadei F, Boop FA, Johnson WD, Warf BC, Park KB. 2019. Global neurosurgery: The current capacity and deficit in the provision of essential neurosurgical care. executive summary of the global neurosurgery initiative at the program in global surgery and social change. *Journal of Neurosurgery*. 130:1055–1064.
- Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, Bauer C, Jennings D, Fennessy F, Sonka M, et al. 2012. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*. 30:1323–1341.
- Gerard IJ, Hall JA, Mok K, Collins DL. 2015. New protocol for skin landmark registration in image-guided neurosurgery: Technical note. *Clinical Neurosurgery*. 11(3):376–381.
- Grunert P, Darabi K, Espinosa J, Filippi R. 2003. Computer-aided navigation in neurosurgery. *Neurosurgical review*. 26(2):73–99.
- Haouchine N, Juvekar P, Golby A, Frisken S. 2021a. Predicted microscopic cortical brain images for optimal craniotomy positioning and visualisation. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*. 9(4):407–413.
- Haouchine N, Juvekar P, Necessian M, Wells III WM, Golby A, Frisken S. 2021b. Pose estimation and non-rigid registration for augmented reality during neurosurgery. *IEEE Transactions on Biomedical Engineering*. 69(4):1310–1317.
- Heikkila J, Silven O. 1997. A four-step camera calibration procedure with implicit image correction. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; San Juan, Puerto Rico. IEEE Comput. Soc. p. 1106–1112.
- Horn BKP. 1987. Closed-form solution of absolute orientation using unit quaternions. *J Opt Soc Am A*. 4(4):629–642.

- Hou Y, Ma L, Zhu R, Chen X, Zhang J. 2016. A low-cost iphone-assisted augmented reality solution for the localization of intracranial lesions. *PLoS one*. 11(7):e0159185.
- Incekara F, Smits M, Dirven C, Vincent A. 2018. Clinical feasibility of a wearable mixed-reality device in neurosurgery. *World neurosurgery*. 118:e422–e427.
- Kersten-Oertel M, Gerard IJ, Drouin S, Petrecca K, Hall JA, Louis Collins D. 2016. Towards augmented reality guided craniotomy planning in tumour resections. In: *International Conference on Medical Imaging and Augmented Reality*. Springer. p. 163–174.
- Lacroix M, Abi-Said D, Fourney DR, Gokaslan ZL, Shi W, DeMonte F, Lang FF, McCutcheon IE, Hassenbusch SJ, Holland E, et al. 2001. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *Journal of neurosurgery*. 95(2):190–198.
- Léger É, Horvath S, Fillion-Robin JC, Allemang D, Gerber S, Juvekar P, Torio E, Kapur T, Pieper S, Pujol S, et al. 2022. Nounav: A low-cost neuronavigation system for deployment in lower-resource settings. *International Journal of Computer Assisted Radiology and Surgery*.
- Léger É, Reyes J, Drouin S, Popa T, Hall JA, Collins DL, Kersten-Oertel M. 2020. MARIN: an open-source mobile augmented reality interactive neuronavigation system. *International Journal of Computer Assisted Radiology and Surgery*. 15(6):1013–1021.
- Low D, Lee CK, Dip LLT, Ng WH, Ang BT, Ng I. 2010. Augmented reality neurosurgical planning and navigation for surgical excision of parasagittal, falx and convexity meningiomas. *British journal of neurosurgery*. 24(1):69–74.
- Marcus HJ, Pratt P, Hughes-Hallett A, Cundy TP, Marcus AP, Yang GZ, Darzi A, Nandi D. 2015. Comparative effectiveness and safety of image guidance systems in neurosurgery: a preclinical randomized study. *Journal of neurosurgery*. 123(2):307–313.
- McGirt MJ, Chaichana KL, Attenello FJ, Weingart JD, Than K, Burger PC, Olivi A, Brem H, Quinoñes-Hinojosa A. 2008. Extent of surgical resection is independently associated with survival in patients with hemispheric infiltrating low-grade gliomas. *Neurosurgery*. 63(4):700–708.
- Rai SKR, Dandpat SK, Jadhav D, Ranjan S, Shah A, Goel AH. 2019. Optimizing burr hole placement for craniotomy: A technical note. *Journal of neurosciences in rural practice*. 10(03):413–416.
- Robertson FC, Raahil MS, Amich JM, Essayed WI, Lal A, Lee BH, Prieto PC, Tokuda J, Weaver JC, Kirillos RW, et al. 2021. Frameless neuronavigation with computer vision and real-time tracking for bedside external ventricular drain placement: a cadaveric study. *Journal of Neurosurgery*. 136(5):1475–1484.
- Sanai N, Polley MY, McDermott MW, Parsa AT, Berger MS. 2011. An extent of resection threshold for newly diagnosed glioblastomas. *Journal of neurosurgery*. 115(1):3–8.
- Simon T, Joo H, Matthews I, Sheikh Y. 2017. Hand keypoint detection in single images using multiview bootstrapping. In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. p. 1145–1153.
- Smith JS, Chang EF, Lamborn KR, Chang SM, Prados MD, Cha S, Tihan T, VandenBerg S, McDermott MW, Berger MS. 2008. Role of extent of resection in the long-term outcome of low-grade hemispheric gliomas. *Journal of Clinical Oncology*. 26(8):1338–1345.
- Spivak CJ, Pirouzmand F. 2005. Comparison of the reliability of brain lesion localization when using traditional and stereotactic image-guided techniques: a prospective study. *Journal of neurosurgery*. 103(3):424–427.
- Sun Gc, Wang F, Chen Xl, Yu Xg, Ma Xd, Zhou Db, Zhu Ry, Xu Bn. 2016. Impact of virtual and augmented reality based on intraoperative magnetic resonance imaging and functional neuronavigation in glioma surgery involving eloquent areas. *World Neurosurgery*. 96:375–382.
- Terzakis G, Lourakis M. 2020. A consistently fast and globally optimal solution to the perspective-n-point problem. In: *European Conference on Computer Vision*. Springer. p. 478–494.
- Tokuda J, Fischer GS, Papademetris X, Yaniv Z, Ibanez L, Cheng P, Liu H, Blevins J, Arata J, Golby AJ, et al. 2009. Openlink: an open network protocol for image-guided therapy en-

- vironment. *The International Journal of Medical Robotics and Computer Assisted Surgery*. 5(4):423–434.
- Wagner W, Gaab M, Schroeder H, Tschiltschke W. 2000. Cranial neuronavigation in neurosurgery: assessment of usefulness in relation to type and site of pathology in 284 patients. *min-Minimally Invasive Neurosurgery*. 43(03):124–131.
- Watanabe E, Satoh M, Konno T, Hirai M, Yamaguchi T. 2016. The trans-visible navigator: a see-through neuronavigation system using augmented reality. *World neurosurgery*. 87:399–405.
- Zhang F, Bazarevsky V, Vakunov A, Sung G, Chang CL, Grundmann M, Tkachenka A. 2020. Mediapipe hands: On-device real-time hand tracking. In: *CVPR Workshop on Computer Vision for Augmented and Virtual Reality*.
- Zhang Z. 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 22(11):1330–1334.