Point Cloud Alignment through Mid-Air Gestures on a Stereoscopic Display

Katja Krug*†
Katja.Krug@tu-dresden.de
Interactive Media Lab Dresden,
Technische Universität Dresden
Dresden, Germany

Marc Satkowski*‡
msatkowski@acm.org
Interactive Media Lab Dresden,
Technische Universität Dresden
Dresden, Germany

Reuben Docea
reuben.docea@nct-dresden.de
Translational Surgical Oncology,
National Center for Tumor Diseases
Dresden, Germany

Tzu-Yu Ku
tzu-yu.ku@mailbox.tu-dresden.de
Interactive Media Lab Dresden,
Technische Universität Dresden
Dresden, Germany

Raimund Dachselt†‡§
dachselt@acm.org
Interactive Media Lab Dresden,
Technische Universität Dresden
Dresden, Germany

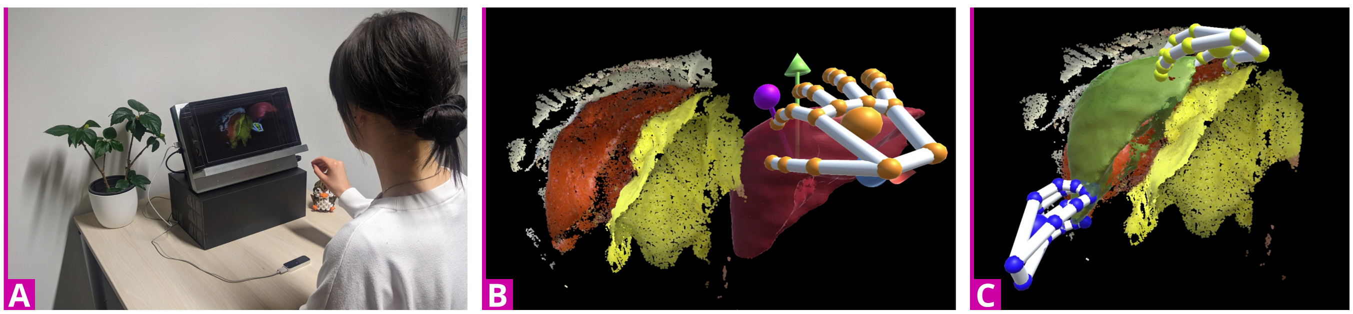

Figure 1: Our proposed and implemented system for the manual point cloud alignment. (A) shows a user interacting with the system consisting of a stereoscopic display and hand tracking device. (B) displays a point cloud (left) and a solid model (right). The model can be interacted with directly or via attached handlers. In (C) the same model is scaled by the use of both hands.

ABSTRACT

Manual point cloud registration is often a crucial step during the mapping of 3D point clouds and usually performed on a conventional desktop setup with mouse interaction. Since 3D point clouds are inherently spatial, these 2D applications suffer from impaired depth perception and inconvenient interaction. Nonetheless, there are few efforts to improve the usability of these applications. To address this, we propose an alternative setup, consisting of a stereoscopic display and an external hand tracker, allowing for enhanced depth perception and natural interaction without the need for body-worn devices or handheld controllers. We developed interaction techniques for point cloud alignment in 3D space, including visual feedback during alignment, and implemented a proof-of-concept prototype in the context of a surgical use case. We describe the use case, design and implementation of our concepts and outline future work. Herewith we provide a user-centered alternative to desktop applications for manual point cloud registration.

CCS CONCEPTS

• **Human-centered computing** → **Gestural input**; • **Hardware** → *Emerging interfaces*; • **Applied computing** → Health care information systems.

KEYWORDS

stereoscopic display, mid-air interaction, point cloud alignment, point cloud registration, image guidance navigation systems

ACM Reference Format:
Katja Krug, Marc Satkowski, Reuben Docea, Tzu-Yu Ku, and Raimund Dachselt. 2023. Point Cloud Alignment through Mid-Air Gestures on a Stereoscopic Display. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3544549.3585862

*These authors contributed equally to this research.
†Also with Centre for Tactile Internet with Human-in-the-Loop (CeTI), Technische Universität Dresden.
‡Also with Centre for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig, Germany
§Also with Cluster of Excellence Physics of Life (PoL), Technische Universität Dresden.

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3585862

1 INTRODUCTION & BACKGROUND

Point clouds are a collection of positions in 3D space, created from images or scans of physical objects or environments. They play an integral role in areas such as photogrammetry and computer vision. Point cloud models encompassing an object or scene are often built by aligning individual point clouds of the same subject to one another. These subsets are either captured from different perspectives or positions or contain new measurements. These registrations can be performed automatically and manually, using techniques such as the Iterative Closest Point (ICP) algorithm [1, 4], which is widely utilized to minimize the distance between two point clouds by iteratively refining an initial estimated transformation between them. In static environments, this initial transformation can be automatically approximated by processing contextual information, e.g., by capturing individual environment scans in direct succession, as seen in ICP variations implemented for self-driving vehicles [19]. However, automatic alignment processes can be error-prone and fail occasionally, a possibility which needs to be avoided in application areas which strongly depend on reliability and accuracy, such as during surgery. These cases call for manual registration which relies on human perception, e.g., by identifying and manually selecting corresponding regions in different 3D scans of the internal human body [8]. The application areas for manual point cloud registration are generally quite diverse. Sometimes, real-world objects just need to be reconstructed into CAD models [11]. Other datasets have evolved between captures, such as the shape and roughness of an aircraft wing through the accretion of ice [15] or the the dendrometic characteristics of trees in different seasons [2].

However, the platforms used to perform the manual alignment are rarely optimized to compensate for common difficulties regarding depth perception and interaction with 3D data on a 2D setup. All of the aforementioned works employ a conventional desktop setup with mouse interaction to manually register 3D point clouds. Projecting them into two dimensions causes information loss and can hinder proper spatial perception. Based on our experience with region-based registration methods [5], we assume that these interfaces need a certain amount of training to be operated efficiently. To address these issues, we propose an alternative to a conventional desktop setup by supporting proper depth perception and offering natural, intuitive interaction techniques with user-centered feedback for 3D point cloud alignment. Since 3D data representation is clearly beneficial for shape understanding [20], and stereoscopic perception has been proven to be especially beneficial when performing depth related tasks [16, 17], we design our concepts to be visualized on a stereoscopic display. While previous work on 3D point cloud manipulation in a stereoscopic environment has been largely focused on Mixed Reality (MR) applications [3, 14, 28, 31], we want to circumvent the need for additional handheld or body-worn devices, by incorporating an external hand tracking device.

Within this work, we explore a touchless 3D alternative to conventional desktop applications used to manually register two point clouds, that aims to achieve adequate alignment through natural, intuitive interaction (see Fig. 1). We developed user-centered interaction and feedback concepts specifically catered to the task of 3D point cloud alignment. We argue that this approach is specifically useful in a medical application context and illustrate our concepts against the background of point cloud registration during minimally invasive surgery. However, we are confident that our approach is broadly beneficial in areas that necessitate manual point cloud registration.

In summary, we contribute:

- The *concept* of a touchless alternative to manual desktop point cloud registration using mid-air gesture interaction in the context of a *medical use case*,
- *Interaction techniques* and *built-in feedback* that support gesture-based manual point cloud alignment in 3D space, including 3D point cloud placement techniques, and visual and functional alignment aids,
- *A proof-of-concept prototype* running on a autostereoscopic 3D display for improved depth perception (see Fig. 1).

2 MOTIVATION AND USE CASE

We aim to explore how the process of manual point cloud registration can be facilitated by tailoring the interaction with respective tools to this specific task and its application area. Despite a diverse spectrum of possible application areas, we recognize a pattern of frequent usage within minimally invasive surgery (MIS) applications, specifically in the context of liver surgery [8, 12, 24, 30]. We believe that designing a system with the constraints of the operating room in mind results in concepts that are broadly applicable, even in sensitive, restrictive environments. Therefore, we want to illustrate our concepts against the background of a use case inspired by the liver navigation pipeline by Docea et al. [8].

While MIS is known to improve patient safety and cost-effectiveness [9], it also increases the difficulty of the procedure due to the lack of depth perception and the inability to directly palpate organs [26]. Image Guidance Navigation Systems (IGS) can be employed to compensate for these drawbacks, operating on the basis of 3D maps of the inner body. The IGS proposed by Docea et al. specifically revolves around navigation in the context of liver resection, aiming to support surgeons by correctly overlaying important anatomical structures, such as vessels and tumors, over the endoscopic video stream. Here, these structures are acquired preoperatively via imaging techniques such as CT and MRI scans, and are subsequently manually registered to a point cloud, representing the inner body environment, which is intraoperatively built utilizing visual SLAM [7]. The manual rigid alignment is based on a region-based registration method[1] [5], within which the user defines multiple matching regions on two point clouds. Using these areas, a weighted ICP is subsequently performed to calculate a transform between the pre- and intraoperative liver surfaces, which is then handed to a fast non-rigid dynamic correction [23] for refinement and to adapt the preoperative model to the current shape of the organ.

In this system the manual registration is performed on a desktop application which is operated via mouse after successful intraoperative mapping of the patient's internal abdomen. Since this process takes place during surgery, directly handling a mouse would lead to the dominant hand becoming non-sterile, necessitating either another subsequent sterilization or to precautionarily wear a second layer of gloves during the interaction. Sadri et al. [25] address

[1]https://gitlab.com/arailis-public/mediassist3_region_register

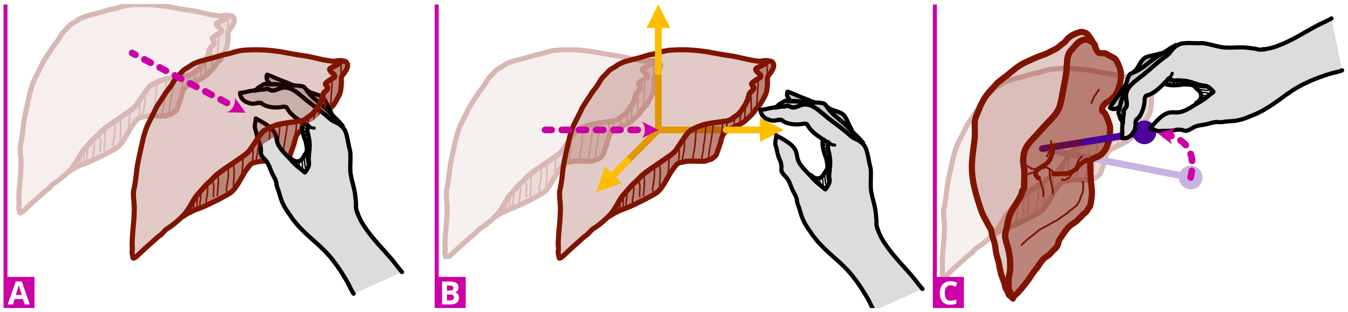

Figure 2: Different interaction techniques to alter the model's pose in varying levels of degrees of freedom (the semi transparent model indicates its previous position): (A) illustrates a 6 DoF translation and rotation by grasping and releasing. (B) shows a 1 DoF translation by grasping the translation handles and moving it along fixed axes. (C) displays a 3 DoF rotation by using the rotation lever to rotate the model around its pivot point.

this problem by introducing a hands-free patient model transformation approach via small head movements tracked by an MR HMD. Since the task of manual point cloud alignment is considerably more complex, we instead opt for mid air gesture interaction to circumvent this issue. We also aspire for the process to be performed on a stereoscopic display for improved depth perception, purposefully differentiating our concepts from MR applications. The work in the operating room requires clear communication between medical professionals and high attentiveness to the crowded, yet dynamic environment and its sensitive machinery. Beyond the discomfort, wearing a MR HMD could negatively influence the wearer's perception of their social and physical environment and hinder collaboration.

3 DESIGN

Our objective is to allow users to align two point clouds as closely as necessary, with the least amount of obstacles regarding perceptual difficulties and interaction complexity. Since many of the issues experienced during the manipulation of virtual objects in 3D space are perception and interaction related, we developed our concepts around the assumption that a stereoscopic display is used in combination with mid-air interaction techniques. Since manual point cloud registration usually merely provides an initial estimation for a subsequent refining algorithm, such as an ICP, the alignment does not need to be optimal. In our specific use case, the preoperative scan of the liver rarely matches the intraoperative scans, because of deformation due to breathing and how the patient is positioned. Perfect alignment is therefore mostly impossible. Therefore, we focus on interaction and guidance that facilitates intuitive and natural-feeling task completion with sufficiently accurate alignment. We illustrate these concepts in the context of organ alignment during MIS, specifically aligning two point clouds of a liver.

To reduce complexity, the transformations will only be performed on the point cloud of the preoperative organ model, in the following referred to as *the model*, while the intraoperative environment point cloud, referred to as *the point cloud*, has a fixed pose and dimension. To easily differentiate both while placing, a solid mesh covers the model while the point cloud consists of individual points. In order to keep our interaction metaphors consistent, we map specific interaction targets to individual hands. Interactions targeting the model are performed with the dominant hand (in our figures: the right hand), and those targeting the scene camera with the non-dominant hand. As an exception, scaling the model and zooming into the scene are performed with both hands.

In the following, we present techniques and design choices regarding 3D model transformation (Sec. 3.1), alignment guidance (Sec. 3.2), and 3D navigation (Sec. 3.3).

3.1 3D Model Transformation

We propose three major transformation strategies in varying degrees of freedom (DoF), as depicted in Fig. 2: a 6 DoF 3D placement, a 1 DoF translation, a 3 DoF rotation. A possible workflow is to first ensure a rough alignment by picking the model up and placing it within close vicinity of the point cloud and subsequently refining the alignment via 1 DoF translation and 3 DoF rotation. If necessary, the model can also be scaled dynamically. These strategies are in accordance with the guidelines for mid-air object manipulation established by Mendes et al. [18], who describe 6 DoF manipulations as suitable for coarse transformations and single DoF separation as *"very desirable for precise transformations, typically fine-grain adjustments"*.

Inserting a virtual hand into the model beyond its surface bounds allows one to grasp and release the model intuitively in 3D space (see Fig. 2A). By default, the model is displayed with X, Y, and Z axes protruding from it, always facing the virtual camera, with the origin of these axes positioned at the model's pivot point, which is initially placed at its centroid. The user can grab onto the handles of the individual axes and refine the model's position by pulling or pushing it in the desired direction in 1 DoF (see Fig. 2B). An additional fourth bar, the rotation lever, also protrudes from the pivot point of the model and dynamically aligns itself with the fingers of the user's hand, always pointing in its direction. Grabbing this lever results in the model being firmly attached to it. It is now possible to rotate the model in 3 DoF around its pivot point (see Fig. 2C). A similar manipulation approach was introduced by Gloumeau et al. [10] who position pins on object surfaces to define pivot points for the constrained rotation of virtual objects in immersive environments. Additionally, the user can also resize the model by grasping it with both hands on opposite sides and moving them towards or away from each other (see Fig. 1C).

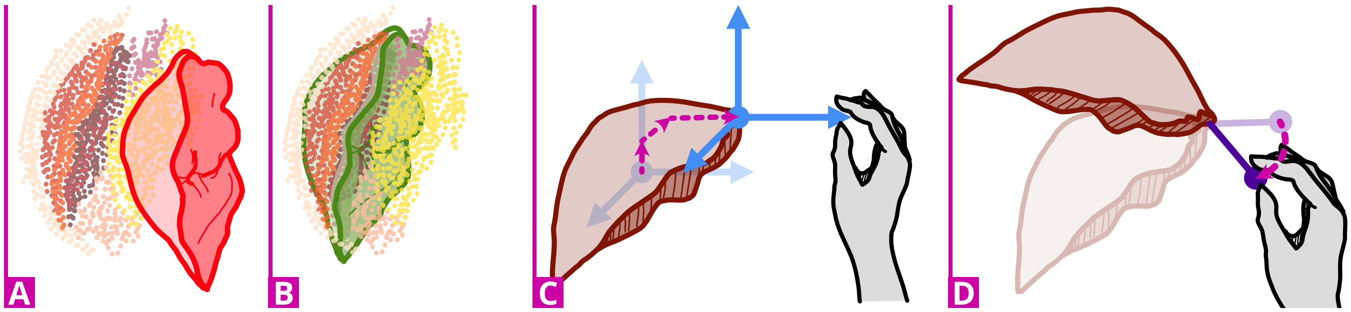

Figure 3: Alignment guidance functionalities of our design: (A+B) show color-coded visual feedback affecting the model, to clearly communicate the current alignment accuracy with red (A) for a low and green (B) for a high rating. (C+D) demonstrates the translation of the pivot point of the model (C) to the change of the center of the rotation (the semi transparent model indicates its previous position) (D).

To avoid ambiguous interaction, whenever an interaction is performed on either the model, the handles, or the lever, all unused transformation modalities are hidden. The rotation lever also becomes invisible when the hand is within a certain distance threshold to one of the translation axes. Any transformation modalities only appear when the dominant hand is recognized in the scene. This also helps reducing the visual clutter in the scene and therefore the cognitive load of the user.

3.2 Alignment Guidance

While stereoscopic displays allow for easier perception of the placement of objects in 3D space, it is still challenging to judge whether two objects, such as the model and point cloud, are sufficiently aligned. Therefore we propose to communicate an indication of the alignment quality during the model placement. This way the user can ponder next steps to correct insufficient alignment, or make an informed decision on when to commit to the result. We suggest an intuitive color gradient ranging from dark red to bright green, for visual feedback that can be perceived at a glance. The color is applied to the surface of the model (see Fig. 3A+B), so as to not detract attention from the point cloud, while also communicating the feedback easily visible in the viewer's focus. The quality of the alignment needs to be calculated during runtime for immediate visual feedback. To achieve this, we adapted an existing approach, a point-to-plane metric with least squares used in ICP, to the real time demands of our system.

On the other hand, users might find it easier to identify distinct features within the point cloud and the model, and prioritize to overlay them first and then proceed from there. For this approach, we propose the determination of a customizable pivot point for subsequent rotation. After having placed the model in a way so that previously identified distinct regions are superimposed, the user can switch into the pivot mode by flipping the dominant hand, while extending all fingers, over so that the palm faces upwards. Now, the current pivot point is represented by a sphere, which sits in the origin of 1 DoF translation handlers and a 3 DoF rotation lever, analogous to the transformation modalities described in Fig. 3.1. The pivot point can now be repositioned to the aligned distinct region (see Fig. 3C) and subsequently rotated by grabbing the rotation handler (see Fig. 3D). To leave this mode again, the previous mode switch hand gesture is repeated.

3.3 3D Navigation

During the alignment the user might need to get an overview or different perspective of the 3D scene by adjusting the position of the virtual camera. The camera is centered onto the centroid of the 3D point cloud. Using the non-dominant hand, the camera can be rotated around the centroid of the point cloud using a 3 DoF rotation lever with protrudes from it, analogous to the model rotation lever depicted in Fig. 2C. Here, instead of merely facing the direction of the fingers, the lever is constantly attached to the left hand and can be grasped at any distance to rotate the scene in the direction of the hand movement. It appears as soon as the left hand is detected in the interactive volume. To zoom into the scene for closer inspection, we propose to utilize a common zooming metaphor, where both hands are either moved towards each other to zoom out, or stray apart to zoom in. This is analogous to the gesture used to scale the model, only applied to the scene and performed with extended fingers, instead of a grasping hand posture.

4 PROTOTYPE & IMPLEMENTATION

To illustrate our design and test the feasibility of our concepts, we implemented a small-scale proof-of-concept prototype, which we will describe in the following sections.

4.1 Apparatus & System

We aim to combine a stereoscopic and holographic display with mid-air interaction (see Fig. 4A and Fig. 1A). For the output, we used a Looking Glass[2] Pro with a 16 inch display providing a viewing angle of 50°. As the device renders over 45 distinct views of a 3D scene simultaneously, it allows for the presentation of a stereoscopic video to multiple users simultaneously, without the need for head-mounted displays. The mid-air gesture input is captured using an Ultraleap Leap Motion Controller[3]. It has a field of view of 140° × 120° and a tracking range of 10cm up to 80cm (optimal up to 60cm). The leap motion controller is positioned on a surface in front of

[2]https://lookingglassfactory.com/
[3]https://www.ultraleap.com/product/leap-motion-controller/

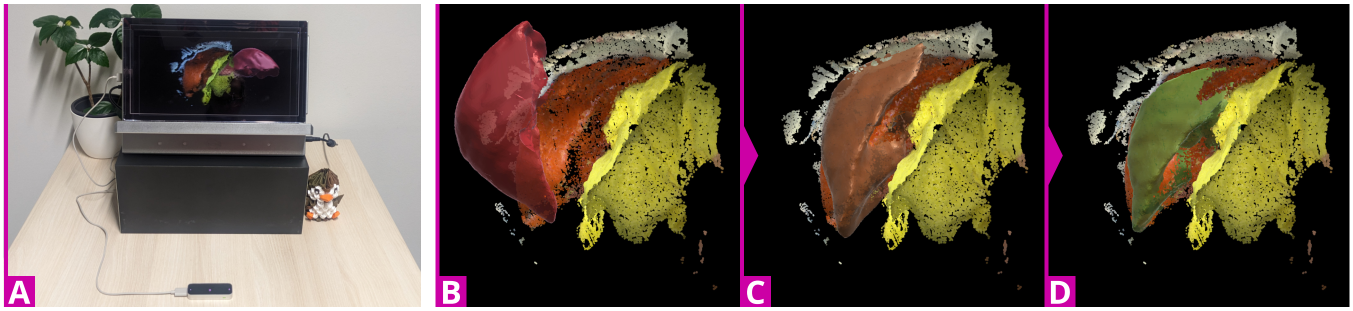

Figure 4: Images from our prototype system. (A) shows the setup of our system, consisting of a LookingGlass Pro as a stereoscopic output device and a LeapMotion Controller to track the hands for the purpose of input. (B–D) display our color-based alignment guidance feature, which changes the color of the model over a gradient between red (B) and green (D).

the stereoscopic display. For best ease of use, the display is slightly elevated in comparison to the controller (see Fig. 4A).

4.2 Software Prototype

The prototype was implemented in the Unity 3D engine which is supported by both the LookingGlass and the Leap Motion. The used preoperative liver model is a 3D model from the OpenHELP Heidelberg laparoscopy phantom[4]. It is an STL file with 30,726 individual triangles, which was converted into an OBJ file and then placed into the Unity scene. Other file types might necessitate pre-processing to create a solid mesh spanning the point cloud coordinates. Using the SLAM module ORB-SLAM2 [21, 22] together with a custom point cloud fusion module [8], we produced two reference point clouds[5] by scanning a human body phantom with an Aesculap EinsteinVision 3.0 FullHD stereo laparoscope: One of the inner body environment surrounding the liver, and one which specifically segments the liver, utilizing a segmentation algorithm [8]. The full point cloud contains 57,169 individual points, and the segmented point cloud 17,968. We used the segmented point cloud for color-based alignment guidance while we present the non-segmented point cloud to the users. For the latter, we assume that the increased number of visual markers can lead to a facilitated orientation within the 3D scene during interaction.

We implemented all the features conceptualized in Sec. 3 and tested them in the context of our proposed use case (e.g., Fig. 1B). For the color-based alignment guidance (see Fig. 4B-D), we implemented a simulated point-to-plane evaluation ICP [13] to judge the alignment during runtime by summing the closest distance of every point of the segmented point cloud to the models surface. The average distance is divided by the number of points in the segmented point cloud and subsequently mapped to a color gradient. Through the optimizations provided by Unity, i.e., the convex hull for meshes which reduces the number of planes, and the Unity methods of *Collider.ClosestPoint* and *Physics.ComputePenetration* - this calculation can be performed with over 60 fps.

5 FUTURE WORK & METHODOLOGY

With our design and prototype, we took the first steps into an exploration of alternative interfaces for natural and intuitive manual point cloud alignment. However, as this design is still in its early iterations, there is room for improvements and extensions. In the following, we will discuss our ideas for improving current and designing future features (Sec. 5.1) and describe plans on how to investigate the benefits of our design through user studies (Sec. 5.2).

5.1 Design & Feature Improvements

As our system is still work in progress, we can envision multiple further features and improvement of current features. In Sec. 3.2 we presented two strategies to support the alignment of point clouds in 3D space. We adapt the color of the liver model by judging the accuracy of alignment through a simulated point-to-plane evaluation as used in ICP [13], though there are several different possible approaches to do that, e.g., by comparing it to ground truth measurements. Further alignment aids could be targeted at the environment point cloud instead of the model, e.g., by changing the color or size of the points in aligned areas. Interaction directions could be given by highlighting translation axes, where interaction with whom would result in a closer alignment, similar to the AR widgets designed by Dastan et al. [6], or the AR orientation-guidance indicators presented by Sukan et al. [29]

We can further evaluate the current gesture set regarding efficiency and intuitiveness and enhance it by incorporating user-driven design techniques. Apart from the final gestures described in Sec. 3, we have designed and implemented two alternatives. In one alternative most of the interactions are performed with the right hand, except for two handed scaling. Zooming could be performed by grabbing the center of the point cloud and pulling or pushing. However, the ambiguity of these interactions caused difficulties. The other alternative always incorprates both hands, with the left continuously indicating one of three modes by facing the palm in different directions: camera transformation, model transformation, and pivot point placement. Meanwhile, the interactions are performed using the right hand.

The current design could also be applied to other use cases and evaluated regarding its applicability and to inspire new functionality. Additionally, since our implementation is easily deployable to AR or VR setups due to its Unity code base, we see possibilities of comparison or adaptation for use cases that are not as restrictive as the operation room.

[4]https://opencas.webarchiv.kit.edu/?q=node/27
[5]Provided in the supplemental material.

5.2 Planned Studies

With our presented design we explored a new alignment approach for 3D models. We suspect that our concepts not only surpass current desktop applications in terms of usability, but that they can additionally match them in alignment accuracy, if not outperform them. To investigate this claim, we aim to perform two studies which we will briefly describe the methodology for in the following.

5.2.1 Comparison Study. First, it is important to understand if our system helps users to perform better in a 3D model alignment task. This can be done by comparing an already existing desktop based system, such as the one used by Docea et al. [8], to our own system.

The task presented in this within-subjects study is the same for both conditions: aligning a preoperative 3D liver point cloud to an intraoperative point cloud of the internal abdomen. For that, we will record several abdomen point clouds to use in both conditions, where we alter the initial rotation and/or position of either the environment point cloud or the liver model. This in combination with an alternating order (between participants) of the conditions allows us to reduce carry-over and anchoring effects. After solving all provided tasks in one condition, a post-condition questionnaire has to be filled out. It should contain questions about the general user experience of the systems, using, e.g., scales (7-point Likert scales with 4 questions each) from the User Experience Questionnaire Plus (UEQ+)[6] [27], like Efficiency, Controllability (or Dependability), Usefulness, and Fun-To-Use (or Stimulation). Additionally, for each task, the task completion time and a measurement representing the quality of the alignment (similar to the color-based alignment guidance value) will be collected. Lastly, in a post-study questionnaire, comparison questions like "Which of the two systems do you like more?" or "Comparing the following properties, which was better?" can be asked.

5.2.2 Component Analysis Study. With the results of the first study, we expect to find a better user experience and a similar or improved performance with our proposed design. However, it is important to understand which of the major components that differentiate our design from conventional applications, especially the mid-air interaction and stereoscopic display, results in the biggest increase of either dependent variable. For that, four conditions should be part of this study: the desktop reference system, a system using a common desktop monitor in combination with mid-air interaction, a system using a stereoscopic display in combination with mouse and keyboard interaction, and lastly our system. The rest of this study would be comparable to the comparison study previously described in section Sec. 5.2.1.

6 CONCLUSION

Manual point cloud registration is an essential step during the mapping of 3D point clouds in many application areas. However, conventional desktop tools rarely compensate for common difficulties regarding depth perception or interaction with 3D data on a 2D setup. To address this, we began to explore how manual point cloud registration can be supported by tailoring respective tools to the task and application area. This was done by not only proposing a specific setup, consisting of stereoscopic display and mid-air gesture interaction, but also by designing and implementing interaction techniques and task-related visual feedback for this setup. This design allows for enhanced depth perception and natural interaction without the need for body-worn devices or handheld controllers. We suggest that our approach is broadly applicable to many application areas. However, since manual point cloud registration is specifically prominent in surgical navigation, we focused on a use case revolving around minimally invasive surgery. Here, our design has the additional benefits of providing sterile, touch-less interaction and allowing for unobstructed environment perception and facilitated collaboration in the operating room, compared to an HMD-based approach. We designed corresponding interaction and feedback concepts to be as intuitive and natural as possible. They include functionality for 3D point cloud transformation, alignment guidance and 3D navigation. Since our system is work in progress, we aim to use our working prototype as a basis for in-depth evaluation regarding usability and efficiency. For this, we described two specific methodologies for planned user studies. We believe that our concepts showcase the potential of enhancing the usability of manual point cloud registration applications and that our chosen use case emphasizes the necessity of adapting conventional tools for this task to their application areas.

[6]English version: http://ueqplus.ueq-research.org/

ACKNOWLEDGMENTS

This work was supported by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) under Germany's Excellence Strategy – EXC-2068 – 390729961 – Cluster of Excellence "Physics of Life" and EXC 2050/1 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of TU Dresden, and by the German Federal Ministry of Education and Research (BMBF, SCADS22B) and the Saxon State Ministry for Science, Culture and Tourism (SMWK) by funding the competence center for Big Data and AI "ScaDS.AI Dresden/Leipzig". We also acknowledge the funding by Sächsische AufbauBank (SAB), project ID 100400076 "Augmented Reality and Artificial Intelligence supported Laparoscopic Imagery in Surgery" (ARAILIS) as TG 70, and by the European Union's Horizon Europe research and innovation programme under grant agreement No. 101092646.

REFERENCES

[1] P.J. Besl and Neil D. McKay. 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14, 2 (1992), 239–256. https://doi.org/10.1109/34.121791
[2] Alexander Bucksch and Kourosh Khoshelham. 2012. Localized Registration of Point Clouds of Botanic Trees. *IEEE GEOSCIENCE AND REMOTE SENSING LETTERS* 10 (01 2012), 1–5. https://doi.org/10.1109/LGRS.2012.2216251
[3] Simon Burkard and Frank Fuchs-Kittowski. 2020. User-Aided Global Registration Method using Geospatial 3D Data for Large-Scale Mobile Outdoor Augmented Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 104–109. https://doi.org/10.1109/ISMAR-Adjunct51615.2020.00041
[4] Yang Chen and Gérard Medioni. 1992. Object modelling by registration of multiple range images. *Image and vision computing* 10, 3 (1992), 145–155.
[5] Logan W. Clements, William C. Chapman, Benoit M. Dawant, Robert L. Galloway Jr., and Michael I. Miga. 2008. Robust surface registration using salient anatomical features for image-guided liver surgery: Algorithm and validation. *Medical Physics* 35, 6Part1 (2008), 2528–2540. https://doi.org/10.1118/1.2911920
[6] Mine Dastan, Antonio Emmanuele Uva, and Michele Fiorentino. 2022. Gestalt Driven Augmented Collimator Widget for Precise 5 DOF Dental Drill Tool Positioning in 3D Space. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 187–195. https://doi.org/10.1109/ISMAR55827.2022.00033

[7] Reuben Docea, Micha Pfeiffer, Sebastian Bodenstedt, Fiona R. Kolbinger, Lukas Höller, Ines Wittig, Ralf-Thorsten Hoffmann, Esther G. C. Troost, Carina Riediger, Jürgen Weitz, and Stefanie Speidel. 2021. Simultaneous localisation and mapping for laparoscopic liver navigation : a comparative evaluation study. In *Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling*, Cristian A. Linte and Jeffrey H. Siewerdsen (Eds.), Vol. 11598. International Society for Optics and Photonics, SPIE, 115980B. https://doi.org/10.1117/12.2582121

[8] Reuben Docea, Micha Pfeiffer, Jan Müller, Katja Krug, Matthias Hardner, Paul Riedel, Martin Menzel, Fiona R. Kolbinger, Laura Frohneberg, Jürgen Weitz, and Stefanie Speidel. 2022. A Laparoscopic Liver Navigation Pipeline with Minimal Setup Requirements. In *2022 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. 578–582. https://doi.org/10.1109/BioCAS54905.2022.9948587

[9] Åsmund Avdem Fretland, Davit Aghayan, and Bjørn and Edwin. 2019. Long-term survival after laparoscopic versus open resection for colorectal liver metastases. *Journal of Clinical Oncology* 37, 18_suppl (2019), LBA3516–LBA3516. https://doi.org/10.1200/JCO.2019.37.18_suppl.LBA3516

[10] P Christopher Gloumeau, Wolfgang Stuerzlinger, and JungHyun Han. 2020. Pinnpivot: Object manipulation using pins in immersive virtual environments. *IEEE transactions on visualization and computer graphics* 27, 4 (2020), 2488–2494.

[11] Ammar Hattab and Gabriel Taubin. 2018. 3D Rigid Registration of Cad Point-Clouds. In *2018 International Conference on Computing Sciences and Engineering (ICCSE)*. 1–6. https://doi.org/10.1109/ICCSE1.2018.8373991

[12] Jon S. Heiselman, Logan W. Clements, Jarrod A. Collins, Jared A. Weis, Amber L. Simpson, Sunil K. Geevarghese M.D., T. Peter Kingham M.D., William R. Jarnagin M.D., and Michael I. Miga. 2017. Characterization and correction of intraoperative soft tissue deformation in image-guided laparoscopic liver surgery. *Journal of Medical Imaging* 5, 2 (2017), 021203. https://doi.org/10.1117/1.JMI.5.2.021203

[13] Kok-Lim Low. 2004. Linear Least-Squares Optimization for Point-to-Plane Icp Surface Registration. *Chapel Hill, University of North Carolina* 4, 10 (2004), 1–3.

[14] Paul Lubos, Rüdiger Beimler, Markus Lammers, and Frank Steinicke. 2014. Touching the Cloud: Bimanual annotation of immersive point clouds. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 191–192.

[15] Stephen McClain. 2016. Manual Point Cloud Registration for Combined Ice Roughness and Ice Thickness Measurements. https://doi.org/10.2514/6.2016-3590

[16] John P. McIntire, Paul R. Havig, and Eric E. Geiselman. 2012. What is 3D good for? A review of human performance on stereoscopic 3D displays. In *Head- and Helmet-Mounted Displays XVII; and Display Technologies and Applications for Defense, Security, and Avionics VI*, Peter L. Marasco, Paul R. Havig II, Daniel D. Desjardins, and Kalluri R. Sarma (Eds.), Vol. 8383. International Society for Optics and Photonics, SPIE, 83830X. https://doi.org/10.1117/12.920017

[17] John P. McIntire, Paul R. Havig, and Eric E. Geiselman. 2014. Stereoscopic 3D displays and human performance: A comprehensive review. *Displays* 35, 1 (2014), 18–26. https://doi.org/10.1016/j.displa.2013.10.004

[18] Daniel Mendes, Filipe Relvas, Alfredo Ferreira, and Joaquim Jorge. 2016. The Benefits of DOF Separation in Mid-Air 3D Object Manipulation *(VRST '16)*. Association for Computing Machinery, New York, NY, USA, 261–268. https://doi.org/10.1145/2993369.2993396

[19] Ellon Mendes, Pierrick Koch, and Simon Lacroix. 2016. ICP-based pose-graph SLAM. In *2016 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*. IEEE, 195–200.

[20] Tamara Munzner. 2014. *Visualization analysis and design*. CRC press.

[21] Raul Mur-Artal, J. Montiel, and Juan Tardos. 2015. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics* 31 (10 2015), 1147 – 1163. https://doi.org/10.1109/TRO.2015.2463671

[22] Raul Mur-Artal and Juan Tardos. 2016. ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras. *IEEE Transactions on Robotics* PP (10 2016). https://doi.org/10.1109/TRO.2017.2705103

[23] Micha Pfeiffer, Carina Riediger, Stefan Leger, Jens-Peter Kühn, Danilo Seppelt, Ralf-Thorsten Hoffmann, Jürgen Weitz, and Stefanie Speidel. 2020. Non-Rigid Volume to Surface Registration Using a Data-Driven Biomechanical Model. (09 2020). https://doi.org/10.1007/978-3-030-59719-1_70

[24] Gian Andrea Prevost, Benjamin Eigl, Iwan Paolucci, Tobias Rudolph, Matthias Peterhans, Stefan Weber, Guido Beldi, Daniel Candinas, and Anja Lachenmayer. 2019. Efficiency, Accuracy and Clinical Applicability of a New Image-Guided Surgery System in 3D Laparoscopic Liver Surgery. *Journal of Gastrointestinal Surgery* 24 (10 2019). https://doi.org/10.1007/s11605-019-04395-7

[25] Shirin Sadri, Shalva A. Kohen, Carmine Elvezio, Shawn H. Sun, Alon Grinshpoon, Gabrielle J. Loeb, Naomi Basu, and Steven K. Feiner. 2019. Manipulating 3D Anatomic Models in Augmented Reality: Comparing a Hands-Free Approach and a Manual Approach. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 93–102. https://doi.org/10.1109/ISMAR.2019.00-21

[26] Crispin Schneider, Moustafa Allam, Danail Stoyanov, D.J. Hawkes, K. Gurusamy, and Brian Davidson. 2021. Performance of image guided navigation in laparoscopic liver surgery – A systematic review. *Surgical Oncology* 38 (07 2021), 101637. https://doi.org/10.1016/j.suronc.2021.101637

[27] Martin Schrepp and Jörg Tomaschewski. 2019. Eine Modulare Erweiterung Des User Experience Questionnaire. *Mensch und Computer – Usability Professionals* (2019), 148–156. https://doi.org/10.18420/muc2019-up-0108

[28] Jonathan Dyssel Stets, Yongbin Sun, Wiley Corning, and Scott W. Greenwald. 2017. Visualization and Labeling of Point Clouds in Virtual Reality. In *SIGGRAPH Asia 2017 Posters* (Bangkok, Thailand) *(SA '17)*. Association for Computing Machinery, New York, NY, USA, Article 31, 2 pages. https://doi.org/10.1145/3145690.3145729

[29] Mengu Sukan, Carmine Elvezio, Steven Feiner, and Barbara Tversky. 2016. Providing Assistance for Orienting 3D Objects Using Monocular Eyewear. In *Proceedings of the 2016 Symposium on Spatial User Interaction* (Tokyo, Japan) *(SUI '16)*. Association for Computing Machinery, New York, NY, USA, 89–98. https://doi.org/10.1145/2983310.2985764

[30] Stephen Thompson, Johannes Totz, Yi Song, Stian Johnsen, Danail Stoyanov, Sébastien Ourselin, Kurinchi Gurusamy, Crispin Schneider, Brian Davidson, David Hawkes, and Matthew J. Clarkson. 2015. Accuracy validation of an image guided laparoscopy system for liver resection. In *Medical Imaging 2015: Image-Guided Procedures, Robotic Interventions, and Modeling*, Robert J. Webster III and Ziv R. Yaniv (Eds.), Vol. 9415. International Society for Optics and Photonics, SPIE, 941509. https://doi.org/10.1117/12.2080974

[31] Florian Wirth, Jannik Quehl, Jeffrey Ota, and Christoph Stiller. 2019. PointAtMe: Efficient 3D Point Cloud Labeling in Virtual Reality. In *2019 IEEE Intelligent Vehicles Symposium (IV)*. 1693–1698. https://doi.org/10.1109/IVS.2019.8814115