Faster Multibeam Sonar Data Cleaning: Evaluation of Editing 3D Point Clouds using Immersive VR

Andrew H. Stevens and Thomas Butkiewicz
Center for Coastal and Ocean Mapping
University of New Hampshire
Durham, New Hampshire, USA

***Abstract*— Remote sensing technologies routinely generate point cloud datasets with billions of points. While automatic data cleaning algorithms exist, safety-critical applications (such as waterway surveys) still require that data be processed and verified by a human. This presents a significant bottleneck in the pipeline from surveys into navigational maps. The recent proliferation of low-cost, high-quality virtual reality systems presents an opportunity to explore how these technologies might be integrated into the point cloud data processing pipeline. Prior research has shown that stereoscopic viewing, head-tracked perspective, and bimanual interactions can lead to faster 3D task completion times and lower errors compared to traditional monoscopic, mouse-and-keyboard desktop systems. In this paper, we present a human factors study that compares 3D point cloud editing performance between a traditional interface and type types of immersive virtual reality interfaces. Our results showed that for complex datasets, the immersive interfaces generally led to faster task completion times than when using the desktop interface. Participants also reported a strong subjective preference for the immersive interface.**

Keywords— virtual reality; point clouds; editing; annotation

I. INTRODUCTION

Remote sensing technologies, such as multi-beam sonar and LIDAR, generate massive point cloud datasets with billions of individual points. Researchers have developed robust automatic data cleaning algorithms [11] that can identify and remove many of the unwanted data points that appear in these sets. However, these algorithms are not perfect, and in some fields, data quality is critical to safety of navigation. Thus, many point cloud datasets are still manually inspected and finalized by a human.

Cleaning point-cloud data is a notoriously tedious and time-consuming task. For example, a challenging hydrographic multi-beam sonar survey can take many times longer to clean than it does to collect. Any technologies that can reduce this workload are welcome.

The majority of industry-standard applications used for point cloud cleaning are designed for use with the ubiquitous 2D desktop mouse & keyboard interface. However, point cloud data, and the interactions required to work with it, are inherently 3-dimensional. Research has shown that for such 3D tasks, 3D interfaces are more effective solutions compared to collapsing the data/task to 2D [26], [50]. The increased effectiveness of 3D interfaces results from addressing both perceptual and interaction issues.

From a perceptual standpoint, point cloud data is challenging because many of the visual cues we usually rely on, such as shading and texture, only apply to surfaces, not individual points. Stereopsis (binocular disparity) can provide useful depth cues for point data, but requires 3D monitors or head-mounted displays (HMDs) capable of providing different images to each eye.

Motion parallax is perhaps the strongest visual depth cue available for point data in 2D or 3D environments. While some experimental interfaces provide continuous motion parallax cues by keeping the data constantly in motion [3], [39], [51], it is usually the result of interactions. For desktop interfaces this would be the user translating or rotating the camera with the mouse. Head-tracked interfaces, however, update the rendering based on the location of the viewer. This means even subtle head movements generate natural motion parallax cues.

Virtual reality (VR) HMDs provide both stereoscopic 3D and head tracking, often over wide-areas (allowing natural locomotion in addition to local head movements). Because they are superior at providing the two most significant depth cues involved with perception of 3D point data, they are an attractive option for working with point clouds. An example of such a system is depicted in Figure 1.

In this paper, we investigate the potential advantages of using an interface combining an immersive VR HMD display and bimanual 6-degree-of-freedom (6DOF) controls for point cloud cleaning. A human factors experiment was conducted to compare performance between our immersive VR interface and a generic desktop monitor/mouse/keyboard-based interface representative of traditional software packages. The results of this point-flagging experiment showed a clear advantage when using either of our VR interfaces with regard to completion time, while errors were generally equivalent between the interfaces. Users overwhelmingly preferred the VR interfaces, demonstrating clear support that this technology is mature enough to be integrated into current point cloud and sonar data editing software.

Finally, as some users are reluctant to adopt head mounted displays in their workplaces, and there can be issues with motion sickness, we also experimented with a hybrid interface, which uses the same 6DOF controllers and interaction methods, but replaces the HMD with a stereoscopic monitor.

This project is supported under NOAA grant #NA15NOS4000200.

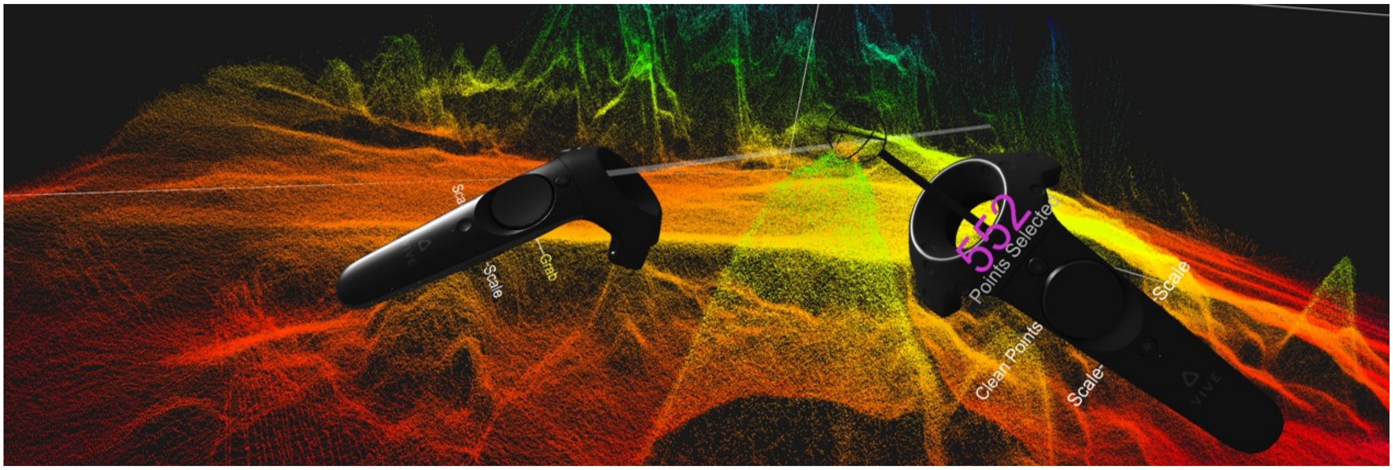

Fig. 1. Screenshot from our immersive VR data-cleaning interface, showing a user examining the point cloud from a swath of multi-beam sonar data

II. Related Work

Classic research in human perceptual theory has shown stereopsis, motion parallax, and kinetic depth (also referred to as external motion parallax) provide strong perceptual cues for understanding 3D distances and forms. Using random computer-generated dot patterns, Julesz was able to isolate stereopsis as a depth cue and demonstrate its power in conveying depth [22]. Rogers and Graham found that motion parallax on its own is sufficient for determining the 3D shape and depth of a surface using random dot patterns [34]. Kinetic depth can also convey an object's 3D form in the absence of other depth cues [46]. Additionally, 6DOF controllers allow for one-to-one direct interactions between the hand and virtual objects, yielding cues for its 3D structure through kinetic depth and kinesthetic correspondence [50].

A. 3D Interface Devices for 3D Tasks

Much research has shown strong empirical evidence of the benefits of matching the dimensionality of an interface to the data being manipulated. Early work combined one-way mirrors and stereoscopic displays with shutter glasses to provide a 3D virtual environment and 6DOF magnetic wand as input [36]. Positioning devices like the Bat [50] and the Lincoln Wand [33] use electromagnetic fields or ultrasonic delays, respectively, to achieve 6DOF capabilities. Grossman and Balakrishnan [19] present and evaluate a few indirect selection techniques using a 6DOF pointing device to cast a ray through an inaccessible 3D volumetric display.

Some evaluations have found free-hand 6DOF devices to be both less efficient and less accurate than a traditional 2DOF mouse for 3DOF object placement [7], though they note the presence of so-called parasitic motion when the free-hand device's trigger was clicked which may have significantly contributed to the relatively high errors that were measured. A later evaluation confirmed the dominance of a traditional mouse over a 6DOF 3D mouse (SpaceNavigator) in a similar 3DOF placement task, but found that the 3D mouse outperformed the traditional mouse in a more complex 6DOF docking task [26]. An evaluation of the virtual hand metaphor between an immersive projection system (e.g., a CAVE [13]) and an HMD found no difference in performance for a task manipulating virtual objects [40]. An experiment comparing multi-touch input to a mouse and 6DOF handheld controller for 3D selection and positioning found the 6DOF controller and mouse combination to be superior, performing equivalently in terms of accuracy, but faster for the 6DOF device when stereoscopic viewing was enabled [9].

Optical tracking methods using devices such as the Leap Motion Controller or the Microsoft Kinect can map a user's hands or gestures to a 3D interface as well. As with free-hand 6DOF devices, hand steadiness and fatigue are major factors affecting their performance; indeed, evaluations have found a traditional mouse to outperform the Leap Motion Controller in terms of accuracy and efficiency for 2D [5] and 3D [12] Fitts' Law pointing and selection tasks. Sensor inaccuracies are the primary issues preventing these devices from being adopted in interfaces for detailed 3D tasks, such as point cloud editing.

B. Head Tracking

One of the first head-tracked stereoscopic displays was created in the late 1960's at Harvard University by Sutherland [42], but its mechanical linkages limited the range of possible viewing positions and orientations.

Deering developed a head-tracked stereoscopic viewing system using shutter glasses and ultrasonic techniques on a CRT monitor [15]. While the system was described as "high resolution" for its time, a comparison shows its resolution to be approximately 25% of the resolution offered by current consumer-grade VR headsets, though it should be noted that the display in Deering's system is stationary and not head-mounted. Pausch et al. compared static displays with HMDs and found them to decrease a generic search task's completion time by 42%, with further reductions possible given additional practice with the HMD [29].

Fish tank VR [48] is a similar concept, and evaluations have shown the addition of head-tracking results in measurable improvements in performance, though generally the largest gains are made from the inclusion of stereoscopy [1], [2], [49]. A more recent study [23] found that the motion parallax from head-tracked viewing decreases fatigue and increases presence.

C. Bimanual 6DOF Controls

Yves Guiard developed the kinematic chaining model of bimanual actions [20] in which the left and right hands form a kinematic chain and the non-dominant hand forms a frame of reference which the dominant hand can use for its movements. An example of this type of asymmetric action would be painting a figurine by holding it in one hand and painting fine details with the other. An evaluation by Ullrich et al. found this mode of interaction to be superior over unimanual ones for a variety of tasks in virtual environments [43].

Wang et al. tracked users' hands with cameras and mapped unimanual interactions to translations and bimanual interactions to rotations in a CAD application [47]. Interestingly, a bimanual translation gesture was mapped to control the application's camera view. They report up to a 40% increase in efficiency for expert CAD users. An evaluation of another two-handed technique [35] found it suitable for precise positioning tasks, but for coarser tasks a one-handed method was superior.

A number of techniques can be found in [55] that explore the use of two independently-controlled cursors in 2D desktop interfaces. The Responsive Workbench [14] provides bimanual interactions via gloves, and presents an intuitive method for scaling an object by bringing the two gloved hands together or apart; we have adopted this "symmetric scaling" metaphor in our 3D interface.

D. Point Cloud Editing

The problem of trying to efficiently process large amounts of multi-beam sonar data to create seafloor maps is not new. Ware et al. presented a system designed to improve the workflow of sonar data processing [52]. The two most important components implemented in this design are the three sub-windows displaying the data being edited at different scales, and the use of statistical algorithms to identify and visualize outlier points which the hydrographer may want to flag for removal in the dataset. A large portion of the existing research on point clouds focuses on feature extraction [8] and segmentation/annotation of disparate objects captured by a depth camera [27], neither of which are activities associated with multi-beam sonar point cloud processing.

Similarly, most of the other relevant literature also focuses on the presentation of point clouds, and tools to find salient structures or features; whereas the research in this paper evaluates an entirely different interface for interacting with point clouds, both in terms of the display (head mounted stereo 3D) and the input devices (bimanual 6DOF). Most other prior research on point cloud interaction consists of approaches to simplify the process of manipulating and viewing point cloud data, usually by adding more devices to existing systems.

Sereno et al. [37] use a consumer-grade tablet's 6DOF to control a virtual camera along with its touchscreen input to delineate a 2D selection shape which can be moved around to create a 3D selection volume. In a slightly different approach, point clouds containing extractable structures can employ techniques like TeddyAware and CloudLasso [53] or CAST [54], which provide 2DOF input techniques for selecting density-based structures in the point data.

Dubois and Hamelin's "Worm Selector" [17] for selection within 3D point clouds on monoscopic desktop systems has users define successive planes throughout the 3D space, upon which contours are drawn to constrain the selection volume between linked contours on consecutive planes. They evaluated their novel method with a selection envelope built from spheres and cuboids, and found user preference for Worm Selection, as well as better precision and task completion times. It should be pointed out that the envelope was built up from statically-placed spheres and cuboids; an interactive, dynamic sphere or cuboid cursor may significantly speed up selection and improve accuracy through visual feedback of cursor motion.

A 3DUI conference contest in 2014 generated a number of interesting techniques to interact with point clouds [24], [44]. The Slice-n-Swipe technique [6], for example, is a progressive-refinement selection strategy controlled by single-finger free-air gestures through a Leap Motion device. The most relevant interface uses an Oculus Rift HMD (providing the same stereoscopic head-tracked viewing as the HMD used in the present study), hand and finger tracking within a small volume on a desk via a depth camera, and optional voice commands to annotate points selected by the virtual hand models [25]. Manipulation of the point clouds is accomplished through widgets, as opposed to the direct manipulations used in this research, and the solution is not completely immersive, in that the user must accomplish their work at a desk.

Of note is the system described by Cabral et al. which employs a Leap Motion Controller on a desktop computer to track both hands for 6DOF position and orientation control for annotating point clouds using a spherical cursor [10]. Unfortunately, only an informal evaluation of this system was performed, though it was noted that users needed to take frequent breaks due to the physical fatigue resulting from hovering one's hands over the Leap Motion Controller.

There are few user studies involving point cloud interaction and viewing techniques. A recent evaluation conducted by Bach et al. [4] empirically compared the efficiency of desktop-, tablet-, and HoloLens-based point cloud visualization environments, and found that while each had strengths in certain tasks, the desktop PC interface was generally superior due to the population's familiarity with its modality. Though the augmented reality environment displayed by the HoloLens was described as "immersive", its effective field of view was too limited to offer an experience comparable to VR HMDs.

An immersive visualization environment utilizing the Leap Motion Controller or Microsoft Kinect (or 3D space mouse) for interaction is described by Donalek et al [16]. They visualize point clouds from large digital sky surveys in a multiuser capable system, but offer no comparisons of visualizations or interaction techniques between the supported input devices.

Etemadpour et al. demonstrated the advantages of immersive VR over 2D desktop displays in perceiving localized distances between points [18]. This supports the use of immersive VR for analyzing sonar datasets, since the decision to flag points for removal is typically based upon the sounding's position and proximity to other points around it. VirtualDesk [45] was designed to take advantage of the strong visual depth cues provided by motion parallax and stereoscopy in VR, along with proprioceptive cues from directly using the hands as input devices in the virtual environment, but seated at a desk so as to alleviate the fatigue and discomfort arising from extended sessions where the body is unsupported (e.g., standing with nothing to rest the arms or hands upon. The VirtualDesk interface was tested against a comparable desktop interface for a variety of tasks involving 3D scatter plots (which can be considered a type of point cloud), and VirtualDesk was found to yield faster task completion times (except where the task required heavy interaction on the desktop controls) and lower error rates in tasks involving spatial analysis and comparisons..

III. Experiment

This experiment was designed to evaluate the relative advantages of annotating point clouds in an immersive room-scale virtual space, a more restricted seated virtual space, and a traditional desktop environment. The three interaction schemes in this experiment shared the same virtual environment (see Figure 2): a neutral background and a cube containing the point cloud dataset, which was centered in the view or space.

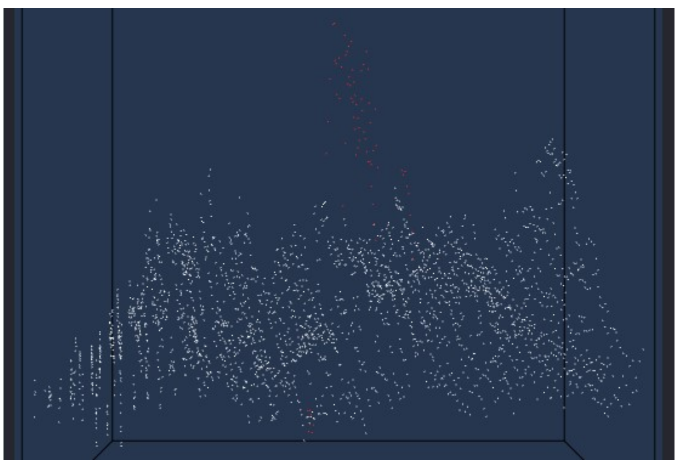

Fig. 2. An example of the virtual environment used in this study. Target points are marked in red and all others are colored white.

Our goal is to evaluate the potential for an immersive environment with 6DOF controllers to better interact with a 3D point cloud when compared to the current desktop-based systems. We also examine the relative benefits of doing work in an immersive virtual environment where the user is standing in an open physical space without any restrictions on movement, and doing the same work while seated with the support of armrests while interacting, inspired by the success reported in the VirtualDesk system [45]. Much of a hydrographer's time checking and validating a sonar dataset is spent exploring the data by moving the virtual camera to different viewpoints and levels of zoom. This represents a huge amount of effort and time spent on deliberate interaction, when the head-coupled perspective provided by an HMD turns this into a natural action that the user takes subconsciously, thereby eliminating the mental context switch between readjusting the viewpoint and making edits to the data.

A. *System Design*

A high-end gaming PC with an Nvidia GTX 980 graphics card ran the same custom software supporting both desktop and virtual reality displays, which was written in C++ using SDL2, OpenGL, and the OpenVR library. Both implementations displayed points as true points, which maintained a constant 3px square screen space size; participants were seated an appropriate distance away from the desktop monitor to maintain visual parity between the resolution of the points displayed on the desktop monitor and in the HMD to avoid visually biasing one system over another in this regard.

1) *Desktop Interface*

The desktop interface (shown in Figure 3) used a 24" Dell U2412M 1920x1200 px monitor, a standard keyboard, and a 3-button optical mouse with scroll wheel. The learning curves for commercial sonar editing software (e.g. CARIS, Qimera, MBEdit, etc.) were deemed too challenging to be practical for this study, so a custom interface, designed to be a distillation of these commercial applications, was implemented as a proxy.

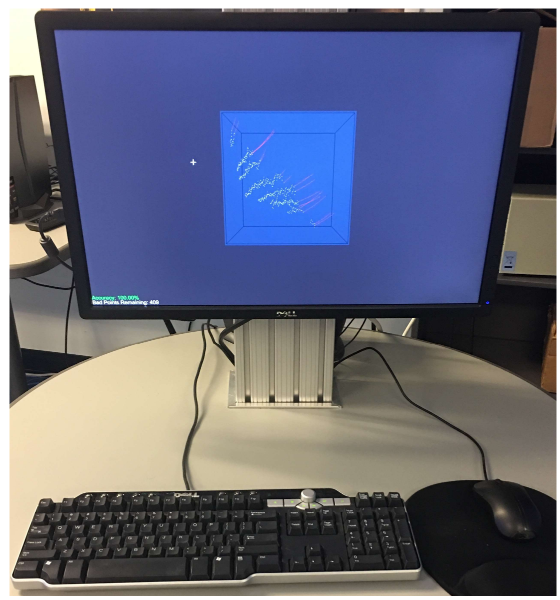

Fig. 3. The desktop interface

Rotation was accomplished through a standard arcball [38] and the left mouse button. Scrolling the wheel zoomed the view and clicking the scroll wheel would recenter the view via a planar translation. A keyboard key can be used to reset the position and orientation of the data volume.

The right mouse button is used to draw a lasso selection on the screen. While lassoing, a line connects the start of the lasso with the cursor to maintain a closed contour shape, visually indicating the area that will be selected when deactivated. Releasing the mouse button completes the lasso: connecting the beginning and end of the contour to create a closed selection area. To reset a selection, the user simply begins drawing a new contour or changes the camera view. When the user is satisfied with a selection, the spacebar on the keyboard is pressed to confirm, and any points inside the screen-space area of the contour are flagged and visually removed. This is generally how multi-point selection is implemented in each of the aforementioned commercial sonar data cleaning applications, and is one of the primary interactions we feel can be improved upon with handheld 3D devices.

The control-display (CD) ratios of rotations, zooming, and translations were made to match those of the commercial software and fine-tuned through pilot testing. The choice of interaction mappings were made to strike a balance between emulating the mappings of the commercial software and avoiding mappings that would lead to additional cognitive load, such as a multi-button combinations.

2) Virtual Reality Interface:

The immersive VR environment (shown in Figure 4) uses an HTC Vive system, which includes an HMD and two 6DOF tracked controllers. The HMD has a resolution of 1080x1200 px per eye, and refreshes 90 frames per second. The HMD was additionally equipped with a TPCast wireless module to untether it from the computer, allowing for unhindered movements and interactions within the 2.5x3m tracked space.

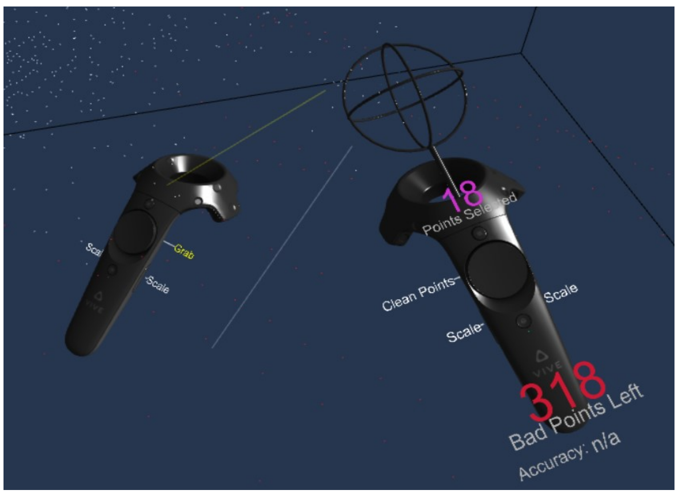

Fig. 4. An example of the VR interface showing the non-dominant left-hand controller manipulating the data volume while the dominant right-hand controller probes the points within its editing sphere

The controller in the non-dominant hand is used to directly reposition/reorient the data volume by engaging and holding the trigger. This mirrors the “one-handed grab” interaction in the Responsive Workbench [14] and was chosen to “preserve the mobile, dynamic role of the non-dominant hand as a base frame of reference” [21]. Since there is only a single object in the environment to interact with and interactions would be occurring locally (as opposed to at a distance, e.g. [30]), a simple direct manipulation metaphor was chosen [31]. When the two controllers are both triggered and brought together or apart, the data volume is scaled in a “symmetric scale” fashion [14], placing the origin at the centroid of the bounding volume, acting as an analog for desktop zooming. Integrated bimanual interactions are also supported, e.g. a flagging action can be performed in tandem with repositioning the data volume.

A spherical selection tool (volumetric cursor) is attached to the dominant controller, and flags points within it upon pulling the trigger. When a point is flagged using the tool, it is removed from the data volume and a small amount of vibration in the controller provides haptic feedback, as this has been shown to improve completion times for similar 3D tasks [1], [28], [32]. Three thin rings spin around the sphere to add more visual depth cues to the cursor to aid perception. The thumb on the primary controller can be swiped left on the controller’s touch-sensitive thumbpad to shrink the selection sphere radius, or right to increase the radius between 0.5mm and 0.5m, and is set to a default value of 5cm when new datasets are loaded.

Ray casting techniques were considered, but deemed impractical because selection of individual points was very imprecise due to small motions of the hand/wrist having large effects on pointer angle, especially in the case of unwanted parasitic motion when engaging the trigger to select. We found that keeping a desired point inside the sphere is much easier than trying to maintain a ray that constantly intersects a point. More advanced selection techniques such as Worm Selector [17] may be more analogous to the desktop lasso selection, but we argue that hydrographers are more interested in localized selection/flagging as opposed to setting up large and complex selections, the construction of which incurs a significant cognitive cost required by having a good spatial understanding of the selection planes and their individual frames of reference. We therefore chose a spherical cursor for its intuitiveness and low cognitive cost to learn and use it, though we recognize that it may not be the optimal design for such a tool.

In the standing scenario, the user has free reign within a 4 square meter area to walk around, into, or through the dataset. Our wireless HMD allows for leaning into or crouching down in front of the dataset, actions we found made us acutely aware of the tether that was present in our wired pilot-test prototype.

The seated scenario keeps the participant seated in the middle of the tracked space, in a chair that was allowed to swivel but not move along the floor. Two adjustable armrests were situated at a comfortable height for each participant, so that they may be used to rest and steady the arms while completing the experimental tasks. There were no additional obstructions in the physical tracked space or the virtual tracked space. Aside from this, the interactions and virtual environment are identical between the standing and seated scenarios.

B. Participants

The experiment was completed by 11 participants from the university student body who were recruited through email advertisements and received $20 compensation. Two of the participants were female, five were graduate students, one was left-handed, two had corrected vision, and they ranged in age from 18 to 42. Only one participant had not experienced any form of VR previously; six had tried either smartphone-based VR or dedicated VR; and four were acquainted with both forms of VR. Four of the participants were hydrographers or ocean mappers with extensive prior experience working with sonar data, and were considered expert users in the analysis.

C. Task

Participants engaged in a mock sonar data flagging task which was designed to eliminate the subjective aspects of identifying sonar system noise. The sonar datasets used in this study have already been processed and finalized by expert hydrographers (none of whom participated in this study), and were presented to the participants with the points that needed to be removed colored red, and all other points colored white to maintain a strong contrast between the two types of points, and prevent the task from devolving into a visual search.

Three noise types representing common noise patterns were identified for further examination, and are shown in Figure 5. The first type, called fliers, are groups of points set apart from the true seafloor surface, and are typically the easiest sensor noise to detect and flag. The next (and more challenging to flag) type of noise is ends, which occurs at the edge of a multibeam sonar swath and extends into the points comprising the seabed surface. The final type is embedded noise, which occurs in and near the seabed surface in poor surveying conditions; these are the most challenging class of error to correct because of their proximity to data that should be included in the final product.

For each of these three types of noise, five representative datasets were selected from actual multi-beam sonar survey data. These 15 datasets were presented one at a time in randomized order, using a simple color mapping of white for good data points and red for bad data points that needed to be flagged. Participants were asked to use each interface (desktop, seated VR, room-scale VR) to flag the red target points in all 15 datasets (for a total of 45 total tasks). Each interface provided participants with real-time feedback regarding their accuracy and the number of target points remaining. The datasets' total point counts ranged from 1030 to 8235 with a mean of 2655 points. The percentage of target points in each dataset ranged from 0.7% to 39.8% and averaged 16.7%.

D. Procedure

Participants began by reviewing and signing a consent form to take part in the study. An informal Snellen visual acuity test was administered to establish a baseline of at least 20/20 vision for the participant pool. Following this, the participant donned the HMD and adjusted the inter-pupillary distance (IPD) of the device until the clearest image possible was achieved. Another vision test was then conducted in the virtual environment presented by the HMD, using an optotype font and a series of six Sloan letters, where each set of six letters gradually subtended smaller visual angles until the letters could not be reliably read. All participants were able to distinguish the letters at a subtended visual angle of at least 6 seconds of arc.

After participants made any appropriate comfort adjustments, a training session interactively guided the user through the VR interface, environment, and study task. When participants felt adequately comfortable using the VR system, the HMD was removed and another training session started for the traditional desktop interface. Participants were instructed to favor accuracy over efficiency, reflecting the real-world importance of sonar point cloud accuracy when making nautical charts used for marine navigation.

Before the experiment began, the order of the three different interfaces was randomly determined for counterbalancing. They then completed 15 randomized trials using each interface, in the previously-decided order, for a total of 45 trials. The study concluded with a questionnaire to gather feedback. Total participation time per person was approximately two hours.

E. Design

The experiment was a 3x3 within-subjects design: three interfaces (desktop, seated VR, room-scale VR) and three noise types (fliers, ends, embedded).

Average time spent flagging is the time between when the first point was flagged and when 80% of the targets in the trial had been flagged; it was measured to help understand user efficiency. An 80% target threshold was chosen to compensate for some participants experiencing difficulty in locating targets in both the desktop display and HMD when only a few target data points remained. This is an effect of our using simple point-based rendering, and we note that the small number of participants who experienced these issues had the same difficulties in both display types, so we do not believe the results were biased towards one display over another.

Time to start flagging is the time from the start of the trial to the first point being flagged. This metric can help to reveal how long it takes to understand and set up the data for editing.

Finally, the percent error was measured as the number of non-target points flagged before the 80% target threshold divided by the number of targets. This shows how accurate participants were while flagging the data in a given condition.

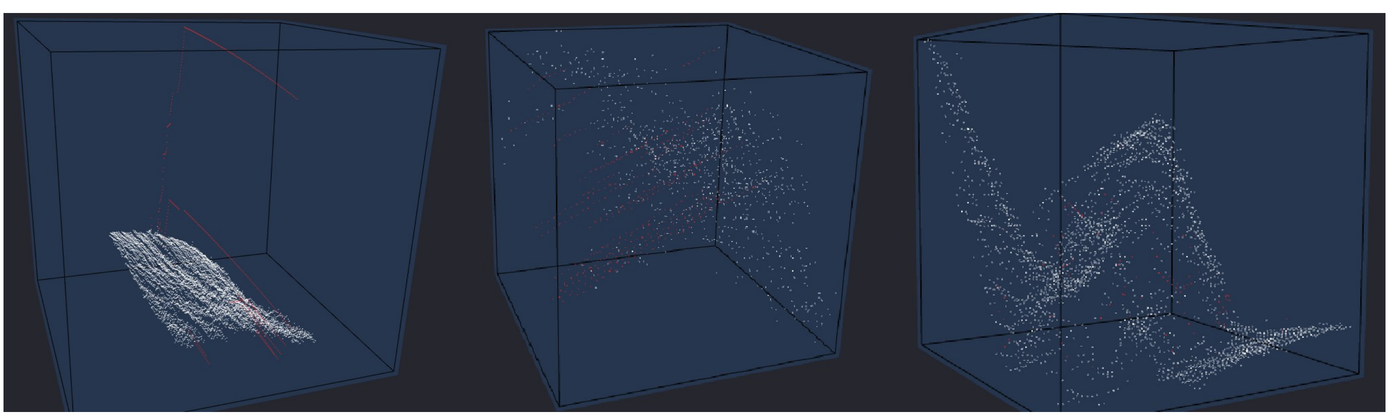

Fig. 5. Examples of the three most common types of noise patterns (red points) in 3D point cloud data collected by multi-beam echosounder sonar systems. From left to right: 'fliers', which extend above or below the actual seafloor; 'ends', a type of noise associated with the peripheral edges of wide swath coverage; and 'embedded' noise which occurs most often around ledges on the seafloor, and is regarded by hydrographers as the most challenging type of noise to flag

IV. Results

The results were analyzed with a linear mixed-effects model. Noise type and interface were included as fixed effects. Participant and dataset were designated as random effects.

An ANOVA found a significant effect of both noise type ($F(2, 12) = 11.8$, $p < 0.01$) and interface ($F(2, 464) = 18.6$, $p < 0.0001$) on the average time spent flagging data. The interaction between the noise type and interface was also a highly significant effect ($F(4, 464) = 7.76$, $p < 0.001$), suggesting that noise type and interface are not additive effects when it comes to flagging time. A Tukey's HSD test to group significantly different conditions at a 95% confidence level found four separate groups, as seen in Figure 6.

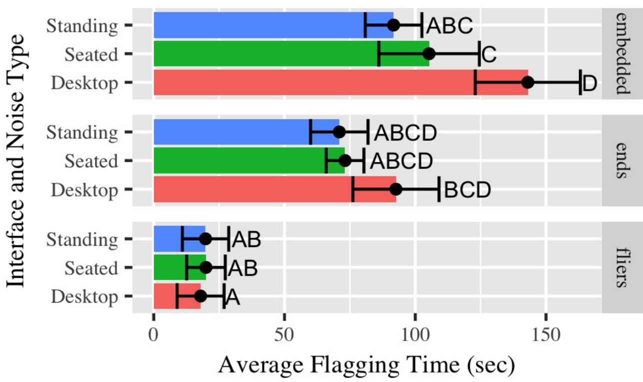

Fig. 6. Average time spent flagging, grouped by interface and noise type with 95% confidence intervals. Letters indicate Tukey's HSD test group membership for each condition at a 95% confidence level.

For average time to begin flagging data, both noise type ($F(2, 12) = 8.7$, $p < 0.01$) and interface ($F(2, 464) = 12.4$, $p < 0.0001$) were again significant effects, though the interaction between the two variables failed to reach significance. The results can be seen in Figure 7 along with the three groups found by Tukey's HSD test at a 95% confidence level.

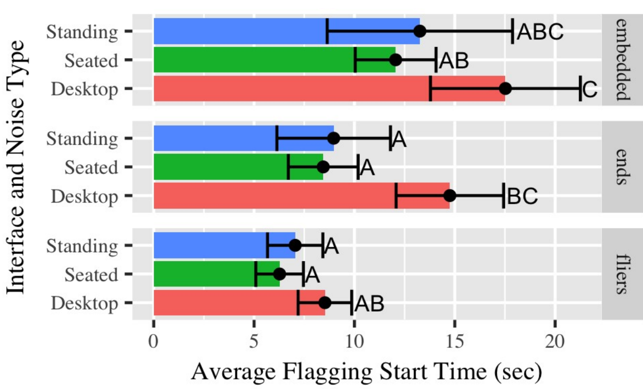

Fig. 7. Average time to start flagging data, grouped by interface and noise type with 95% confidence intervals and Tukey HSD group labels.

For average percent error, the noise type was a significant main effect ($F(2, 12) = 8.06$, $p < 0.01$), but the interface and the interaction between noise type and interface were not significant in explaining the variation in the response. Figure 8 shows these results and Tukey HSD groups at a 95% CI.

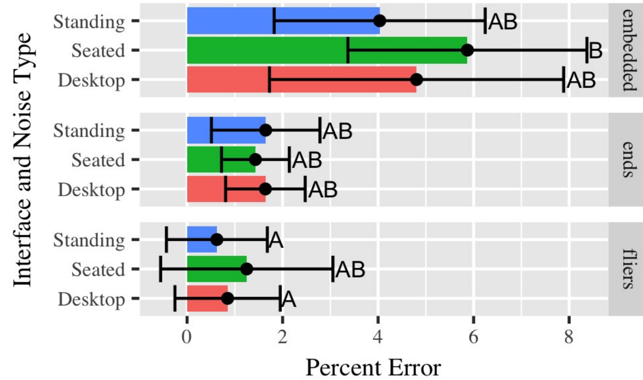

Fig. 8. Percent error, grouped by interface and noise type with 95% confidence intervals and Tukey HSD group labels.

The analysis found no significant difference in performance (task completion time nor percent error) for those participants who flagged more target points using bimanual interactions (e.g., manipulating the data volume while simultaneously flagging points).

A. Expert User Comparison

We split the participant pool by reported sonar data cleaning experience level into an expert group (n = 4) and an inexperienced group (n = 7) to better understand the effect of a priori task knowledge. In other words, if the expert users do a significantly better job than the inexperienced cleaners, the likely explanation is transfer effect from past experience.

Our analysis did not find any evidence that an expert designation accounted for a significant amount of the variation in the flagging time, percent error, or time to start flagging at a $p < 0.05$ level. This is important because the task was designed to be abstracted in such a way that obviates the need for subjective decisions about point removal, and there is no advantage to having any prior experience cleaning sonar data, which would add another confounding variable to the results.

B. Subjective Assessment

Participants completed a questionnaire after the experiment to gather subjective input about the interface designs and functionalities. Participants were asked to rate how much they agreed with statements regarding the three interfaces using a 5-point Likert scale. The results can be seen in Figure 9. On average, the room-scale VR interface was rated easiest to understand (4.64), easiest to use for task completion (4.91), most immersive (4.91), and least uncomfortable (2.09). The seated VR interface was rated similarly to the room-scale version, though participants found it to be the least tiring/fatiguing (2.73). The desktop interface was reported to interfere the most with efficiently completing the task (3.55).

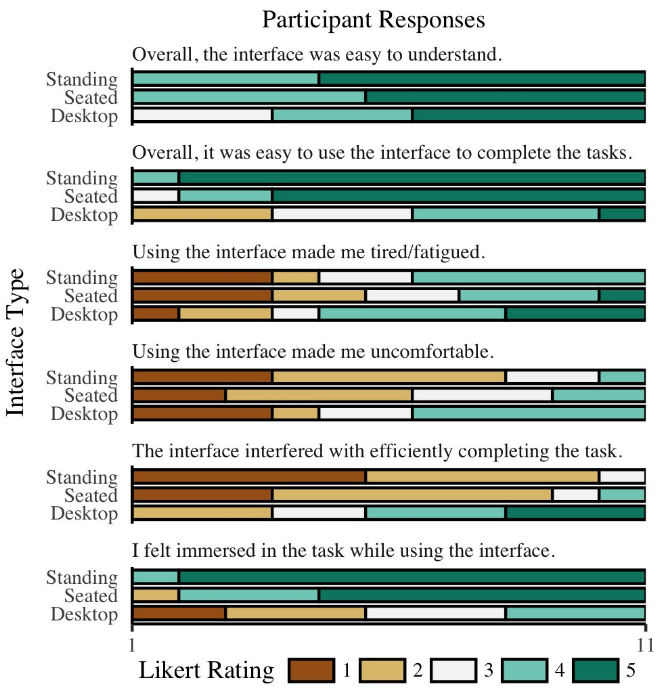

Fig. 9. A stacked chart showing the participant responses to a post-study questionnaire using the 5-point Likert rating scale. Participants were asked to what degree they agree with each of the six questions in the context of each experimental interface.

In the free-form feedback, an expert participant explicitly praised the integral action of changing the point-of-view on the data while also interacting with the data elements as a transformative contribution to the usual process:

> *The major boost in effectiveness and efficiency compared to the desktop interface, in my opinion, is the simultaneous manipulation and editing. This is typically a back-and-forth process in 2D, which now feels acutely inefficient!*

Another expert participant shared their appreciation for the improved viewing conditions, especially when dealing with more challenging noise patterns, a sentiment reflected in a majority of all participants' responses:

> *Full immersion within the environment; improved ability to clean isolated points; outstanding vision.*

V. Discussion

The results of the experiment demonstrate a clear speed advantage when using an immersive virtual environment and bimanual 6DOF controllers to flag 3D point cloud data. The average time spent flagging points was significantly accelerated for the most challenging noise patterns. This can be attributed to the addition of head-tracked viewing providing better depth cues and the interaction techniques allowing for a much more rapid editing process. A higher level of focus due to the increased immersion in the task (as reported in the questionnaire) may have also been a significant contribution. About half of the participants mentioned verbally or in their questionnaire that the thumbpads on the 6DOF controllers were too sensitive, and difficult to precisely adjust the length/radius of the editing sphere in the VR environment. This likely added a considerable amount of extra time to complete the task, and could be mitigated in future implementations by making further adjustments to the control-display (CD) ratio of the thumbpad to give more precise control to the user. This technique can also be used to help dampen the parasitic motions from clicking the trigger or other buttons.

An interesting result is the significant difference between the seated VR interface and the desktop interface in time to begin flagging the point clouds. The fact that the standing, room-scale configuration did not perform as quickly is because users would spend additional time physically moving into and around the dataset, whereas the seated configuration necessitated the majority of the data view manipulation to be carried out via the controllers, which accounts for the worse rating for the seated mode over standing mode in the context of interfering with the given task. Though the feedback does not provide a strong indication that one VR configuration is less fatiguing or more uncomfortable than the other, it is possible that the heavier reliance upon controller use in seated configurations would tire the user more quickly. If that is the case, other techniques can be employed to cut down on the amount of manual interaction required (e.g., by automatically oscillating the dataset around an axis [3], [39], [51]).

The percent errors were largely consistent between the different interfaces, demonstrating that our choices of interaction methods for the VR and desktop interfaces were comparable in efficacy and not biased to one implementation.

We also found that all but one of the participants took advantage of the bimanual interactions available in the VR interfaces. On average, participants cleaned 28% of the target points in the datasets while simultaneously manipulating the data volume containing them, and a majority of participants specifically called out those interactions as having the biggest impact on their performance. The experimental setup did not force participants to use bimanual interactions; a design requiring this type of interaction would better represent the performance of that technique both between and within participants, and allow for stronger comparisons between separated and integral actions for this type of task.

There are a number of directions this research could take going forward. The first and most obvious is to improve our point cloud visualization environment for better depth perception for both the desktop and VR interfaces. Points could be represented as spheres (or some other 3D geometry) with shading and possibly texturing, which would provide better shape-from-shading and occlusion depth cues.

While this study enabled the use of bimanual interactions, it did not require its use for completing the task. It would be interesting to further explore the relative performance differences of this and other asymmetric bimanual interactions, e.g., grab-and-scale [14].

VI. Followup Investigation

In our experience with actual domain users, we have noted some reluctance to adopting head mounted displays. This is understandable, as HMDs can decouple users from their normal workflows/spaces, and can induce motion sickness, especially as sonar data is sometimes cleaned on moving vessels while surveys are underway. While we have developed techniques for reducing such sickness [41], it likely cannot be entirely avoided. Thus, we have been experimenting with a Fishtank VR [48] hybrid solution, which couples the 6DOF handheld interaction with a stereoscopic desktop monitor instead of an HMD.

We suspected that much of the performance gain observed in our study may have been primarily due to the 6DOF hand-held interaction, as opposed to the head-coupled display. To investigate this, we conducted a limited follow-up evaluation of our hybrid Fishtank VR interface.

The interaction volume of the Fishtank VR interface is much smaller than the HMD VR interfaces. For our standing-VR condition, the egocentric interaction volume is a sphere with a radius of an arm's reach. For seated VR, one's lap and chair reduce this roughly to a hemisphere. The Fishtank VR interface has physical constraints from both the tabletop and the monitor, as well as significantly reduced field of regard (FoR): data must be kept within the viewing frustum of the monitor in order to get visual feedback on interactions, whereas the full FoR of an HMD interface allows users to place and view data anywhere around themselves. The usable interaction volume of our Fishtank VR interface is roughly an order of magnitude smaller than the unconstrained HMD interface. Thus, we expected users would need to perform a significantly higher number of manipulations in the Fishtank VR interface compared to the HMD interface.

The visual presentation of points was improved since our previous study, now being rendered as solid discs to make them more noticeable and increase occlusion cues, and being scaled by distance for better depth judgement. The Fishtank VR interface/controls mirrored the HMD-based interface, except controller models were not rendered (the real controllers were visible), and the cursor was offset ~50cm from the controller towards the monitor, directly mapping the interaction volume in front of the monitor to the virtual space behind it. This helps interactions stay at neutral or positive parallaxes that reduce eye strain, visual fatigue, and difficulty fusing stereo images with extreme negative parallax.

Four participants (including the two authors) evaluated the three interfaces with improved point cloud visualizations: desktop (see Section 3.A.1.), seated HMD-based VR (see Section 3.A.2), and hybrid Fishtank VR. The same task (see Section 3.C.) and a reduced set of three point clouds from each noise type category were used (nine total).

The results showed that, in general, error rates did not differ significantly, but cleaning the more-challenging datasets took longer in Fishtank. In these cases, there were significantly more manipulation actions (grabbing/scaling) performed on average per trial in the Fishtank VR condition than in the seated HMD-based condition; for embedded-type noise, there were twice as many manipulation actions (34.4 and 12.2, respectively), and for ends-type noise, there were almost three times as many (27.5 vs. 9.5, respectively). This is likely a result of only being able to edit smaller portions of the data at a time due to constrained interaction volume, and needing to move the dataset to see off-screen portions, whereas in an HMD, users can simply turn their heads.

VII. Conclusion

Point cloud data is a commonly encountered, but tedious to work with form of data in remote sensing and the sciences. This research explored the potential for consumer grade virtual reality interfaces, specifically the combination of a head-mounted stereo display and bimanual 6DOF controllers, for improving the speed and efficiency of processing point cloud data. A controlled experiment provided both empirical and subjective evidence supporting the theory that bimanual interactions, natural movements replacing camera navigation commands, and improved perceptual cues can lead to faster and less taxing (both physically and mentally) point cloud processing sessions.

Further investigation suggests that for users who are reluctant to adopt head mounted displays, using the 6DOF handheld controllers with a stereoscopic monitor instead of an HMD may provide many of the same benefits without decoupling them from their traditional workspace.

Additional research should be conducted to further explore and develop symmetric and asymmetric bimanual interaction techniques, as well as developing VR-display-specific improvements to point cloud visualization techniques to provide more and better cues for understanding depth and structure within point cloud datasets.

These results indicate that now-affordable VR technology may be practical for adoption into actual real-world point cloud data cleaning applications, as we have shown it has potential to provide significant workflow benefits to a task that will likely continue to require human input for the foreseeable future.

Acknowledgments

Research supported by NOAA Grant #NA15NOS4000200. Sonar data provided by the Ocean Exploration Trust.

References

[1] R. Arsenault, C. Ware, "Eye-hand Co-ordination with Force Feedback", *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, pp. 408–414, ACM, 2000.

[2] R. Arsenault, C. Ware, "The Importance of Stereo and Eyecoupled Perspective for Eye-hand Coordination in Fish Tank VR", *Presence: Teleoperators & Virtual Environments*, 13(5):549–559, 2004.

[3] E. Aygar, C. Ware, D. Rogers, "The Contribution of Stereoscopic and Motion Depth Cues to the Perception of Structures in 3D Point Clouds", *ACM Trans. on Applied Perception (TAP)*, 15(2):9, 2018.

[4] B. Bach, R. Sicat, J. Beyer, M. Cordeil, H. Pfister, "The Hologram in my Hand: How Effective is Interactive Exploration of 3D Visualizations in Immersive Tangible Augmented Reality?", *IEEE Trans. on Visualization and Computer Graphics*, 24(1):457–467, 2018.

[5] D. Bachmann, F. Weichert, G. Rinkenauer, "Evaluation of the Leap Motion Controller as a New Contact-free Pointing Device", *Sensors*, 15(1):214–233, 2015.

[6] F. Bacim, M. Nabiyouni, D.A. Bowman, "Slice-n-swipe: A Free-hand Gesture User Interface for 3D Point Cloud Annotation", *IEEE 3DUI*, pp. 185–186, 2014.

[7] F. Berard, J. Ip, M. Benovoy, D. El-Shimy, J.R. Blum, J.R. Cooperstock, "Did 'Minority Report' Get it Wrong? Superiority of the Mouse over 3D Input Devices in a 3D Placement Task", *IFIP Conf. on Human-*

Computer Interaction, pp. 400–414, Springer, 2009.

[8] M. Berger, A. Tagliasacchi, L.M. Seversky, P. Alliez, G. Guennebaud, J.A. Levine, A. Sharf, C.T. Silva, "A Survey of Surface Reconstruction from Point Clouds", *Computer Graphics Forum*, vol. 36, pp. 301–329, Wiley, 2017.

[9] T. Butkiewicz, A.H. Stevens, C. Ware, "Multi-touch 3D Positioning with the Pantograph Technique", *ACM SIGGRAPH I3D*, 13:1–9, 2019.

[10] M. Cabral, A. Montes, O. Belloc, R. Ferraz, F. Teubl, F. Doreto, R. Lopes, M. Zuffo, "Bi-manual Gesture Interaction for 3D Cloud Point Selection and Annotation Using COTS", *3DUI*, pp. 187–188, 2014.

[11] B.R. Calder, L.A. Mayer, "Automatic Processing of High-rate, High-density Multibeam Echosounder Data", *Geochemistry, Geophysics, Geosystems*, 4(6), 2003.

[12] J.C. Coelho, F.J. Verbeek, "Pointing Task Evaluation of Leap Motion Controller in 3D Virtual Environment", *Creating the Difference*, 78:78–85, 2014.

[13] C. Cruz-Neira, D.J. Sandin, T.A. DeFanti, "Surround-screen Projection-based Virtual Reality: The Design and Implementation of the Cave", *Proc. Comp. Graphics and Interactive Techniques*, pp. 135–142, 1993.

[14] L.D. Cutler, B. Frohlich, P. Hanrahan, "Two-handed Direct Manipulation on the Responsive Workbench", *Proc. Symp. on Interactive 3D Graphics*, 1997.

[15] M. Deering, "High Resolution Virtual Reality", *ACM SIGGRAPH Computer Graphics*, 26(2):195–202, 1992.

[16] C. Donalek, S.G. Djorgovski, S. Davidoff, A. Cioc, A. Wang, G. Longo, J.S. Norris, J. Zhang, E. Lawler, S. Yeh, A. Mahabal, M. Graham, A. Drake, "Immersive and Collaborative Data Visualization using Virtual Reality Platforms", *IEEE Int. Conf. on Big Data*, pp. 609-614, 2014.

[17] E. Dubois, A. Hamelin, "Worm Selector: Volume Selection in a 3D Point Cloud through Adaptive Modelling", *Int. Journal of Virtual Reality*, 17(1):1–20, 2017.

[18] R. Etemadpour, E. Monson, L. Linsen, "The Effect of Stereoscopic Immersive Environments on Projection-based Multi-dimensional Data Visualization", *IEEE Information Visualisation*, pp. 389–397, 2013.

[19] T. Grossman, R. Balakrishnan, "The Design and Evaluation of Selection Techniques for 3D Volumetric Displays", *ACM UIST*, pp. 3–12, 2006.

[20] Y. Guiard, "Asymmetric Division of Labor in Human Skilled Bimanual Action: The Kinematic Chain as a Model", *Journal of Motor Behavior*, 19(4):486–517, 1987.

[21] K. Hinckley, R. Pausch, D. Proffitt, N.F. Kassell, "Two-handed Virtual Manipulation", *ACM TOCHI*, 5(3):260–302, 1998.

[22] B. Julesz, "Binocular Depth Perception of Computer-Generated Patterns", *Bell System Technical Journal*, 39(5):1125–1162, 1960.

[23] S. Kongsilp, M.N. Dailey, "Motion Parallax from Head Movement Enhances Stereoscopic Displays by Improving Presence and Decreasing Visual Fatigue", *Displays*, 49:72–79, 2017.

[24] H. Krammes, M.M. Silva, T. Mota, M.T. Tura, A. Maciel, L. Nedel, "The Point Walker Multi-Label Approach", *3DUI*, pp. 189–190, 2014.

[25] P. Lubos, R. Beimler, M. Lammers, F. Steinicke, ¨Touching the Cloud: Bimanual Annotation of Immersive Point Clouds", *IEEE 3DUI*, pp. 191–192, 2014.

[26] E. Mattheiss, J. Schrammel, M. Tscheligi, "Navigating the Space: Evaluating a 3D-Input Device in Placement and Docking Tasks", *Proc. BCS Conf. On Human-Computer Interaction*, pp. 309–314, 2011.

[27] R. Monica, J. Aleotti, M. Zillich, M. Vincze, "Multi-Label Point Cloud Annotation by Selection of Sparse Control Points", *Int. Conf. on 3D Vision*, pp. 301–308, IEEE, 2017.

[28] J. Murayama, L. Bougrila, Y. Luo, K. Akahane, S. Hasegawa, B. Hirsbrunner, M. Sato, "SPIDAR G&G: A Two-Handed Haptic Interface for Bimanual VR Interaction", *Proc. of Eurohaptics*, pp. 138–146, 2004.

[29] R. Pausch, M.A. Shackelford, D. Proffitt, "A User Study Comparing Head-Mounted and Stationary Displays", *Proc. 1993 IEEE Research Properties In Virtual Reality Symp.*, pp. 41–45, 1993.

[30] I. Poupyrev, M. Billinghurst, S. Weghorst, T. Ichikawa, "The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR", *ACM UIST*, pp. 79–80, 1996.

[31] I. Poupyrev, T. Ichikawa, S. Weghorst, M. Billinghurst, "Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques", *Computer Graphics Forum*, 17:41–52, 1998.

[32] A. Prouzeau, M. Cordeil, C. Robin, B. Ens, B. Thomas, T. Dwyer, "Scaptics And Highlight-Planes: Immersive Interaction Techniques for Finding Occluded Features in 3D Scatterplots", *Proc. CHI Conf. on Human Factors in Computing Systems*, 2019.

[33] L.G. Roberts, "The Lincoln Wand", *Proc. 1966 Fall Joint Computer Conf.*, pp. 223–227, ACM, 1966.

[34] B. Rogers, M. Graham, "Motion Parallax as an Independent Cue for Depth Perception", *Perception*, 8(2):125–134, 1979.

[35] M. Schlattmann, R. Klein, "Efficient Bimanual Symmetric 3D Manipulation for Markerless Hand-Tracking", *Virtual Reality Int. Conf. (VRIC)*, vol. 2, pp. 1–10, 2009.

[36] C. Schmandt, "Spatial Input/Display Correspondence in a Stereoscopic Computer Graphic Work Station", *ACM SIGGRAPH Computer Graphics*, vol. 17, pp. 253–261, 1983.

[37] M. Sereno, M. Ammi, T. Isenberg, L. Besançon, "Tangible Brush: Performing 3D Selection with Portable and Position-aware Devices", *IEEE VIS2016*, 2016.

[38] K. Shoemake, "Arcball: A User Interface for Specifying Three-dimensional Orientation Using a Mouse", *Graphics Interface*, vol. 92, pp. 151–156, 1992.

[39] R.L. Sollenberger, P. Milgram, "Effects of Stereoscopic and Rotational Displays in a Three-dimensional Path-tracing Task", *Human Factors*, 35(3):483–499, 1993.

[40] A. Steed, C. Parker, "Evaluating Effectiveness of Interaction Techniques Across Immersive Virtual Environmental Systems", *Presence*, 14(5):511–527, 2005.

[41] A.H. Stevens, T. Butkiewicz, "Reducing Seasickness in Onboard Marine VR Use through Visual Compensation of Vessel Motion", *IEEE VR*, pp. 1872–1873, 2019.

[42] I.E. Sutherland, "A Head-Mounted Three Dimensional Display", *Proc. 1968, Fall Joint Computer Conf., Part I*, pp. 757–764, ACM, 1968.

[43] S. Ullrich, T. Knott, Y.C. Law, O. Grottke, T. Kuhlen, "Influence of the Bimanual Frame of Reference with Haptics for Unimanual Interaction Tasks in Virtual Environments", *IEEE 3DUI*, pp. 39–46, 2011.

[44] M. Veit, A. Capobianco, "Go'Then'Tag: A 3-D Point Cloud Annotation Technique", *IEEE 3DUI*, pp. 193–194, 2014.

[45] J.A. Wagner Filho, C.M.D.S. Freitas, L. Nedel, "VirtualDesk: A Comfortable and Efficient Immersive Information Visualization Approach", *Computer Graphics Forum*, 37:415–426, Wiley, 2018.

[46] H. Wallach, D.N. O'Connell, "The Kinetic Depth Effect", *Jour. of Experimental Psychology*, 45(4):205, 1953.

[47] R. Wang, S. Paris, J. Popovic, "6D Hands: Markerless Hand-Tracking for Computer Aided Design", *ACM UIST*, pp. 549–558, 2011.

[48] C. Ware, K. Arthur, K.S. Booth, "Fish Tank Virtual Reality", *Proc. INTERACT'93 and CHI'93*, pp. 37–42, ACM, 1993.

[49] C. Ware, G. Franck, "Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions", *ACM Trans. on Graphics*, 15(2):121–140, 1996.

[50] C. Ware, D.R. Jessome, "Using the Bat: A Six-Dimensional Mouse for Object Placement", *IEEE CG&A*, 8(6):65–70, 1988.

[51] C. Ware, P. Mitchell, "Reevaluating Stereo and Motion Cues for Visualizing Graphs in Three Dimensions", *Proc. Applied Perception in Graphics and Visualization*, pp. 51–58, ACM, 2005.

[52] C. Ware, L. Slipp, K.W. Wong, B. Nickerson, D. Wells, Y.C. Lee, D. Dodd, G. Costello, "A System for Cleaning High Volume Bathymetry", *The Int. Hydrographic Review*, 69(2), 1992.

[53] L. Yu, K. Efstathiou, P. Isenberg, T. Isenberg, "Efficient Structure-Aware Selection Techniques for 3D Point Cloud Visualizations with 2DOF Input", *IEEE TVCG*, 18(12):2245–2254, 2012.

[54] L. Yu, K. Efstathiou, P. Isenberg, T. Isenberg, "Cast: Effective and Efficient User Interaction for Context-Aware Selection in 3D Particle Clouds", *IEEE TVCG*, 22(1):886–895, 2016.

[55] R.C. Zeleznik, A.S. Forsberg, P.S. Strauss, "Two Pointer Input for 3D Interaction", *ACM Symp. Interactive 3D Graphics*, pp. 115–120, 1997.