


Multimedia Tools and Applications (2021) 80:31039–31058
<https://doi.org/10.1007/s11042-021-10520-z>

1173: INTERACTION IN IMMERSIVE EXPERIENCES



Exploring gestural input for engineering surveys of real-life structures in virtual reality using photogrammetric 3D models

Ślawomir Konrad Tadeja¹  · Yupu Lu¹ · Maciej Rydlewicz² · Wojciech Rydlewicz² · Tomasz Bubas² · Per Ola Kristensson¹

Received: 2 April 2020 / Revised: 28 October 2020 / Accepted: 5 January 2021 /
Published online: 30 January 2021

© The Author(s) 2021

Abstract

Photogrammetry is a promising set of methods for generating photorealistic 3D models of physical objects and structures. Such methods may rely solely on camera-captured photographs or include additional sensor data. Digital twins are digital replicas of physical objects and structures. Photogrammetry is an opportune approach for generating 3D models for the purpose of preparing digital twins. At a sufficiently high level of quality, digital twins provide effective archival representations of physical objects and structures and become effective substitutes for engineering inspections and surveying. While photogrammetric techniques are well-established, insights about effective methods for interacting with such models in virtual reality remain underexplored. We report the results of a qualitative engineering case study in which we asked six domain experts to carry out engineering measurement tasks in an immersive environment using bimanual gestural input coupled with gaze-tracking. The qualitative case study revealed that gaze-supported bimanual interaction of photogrammetric 3D models is a promising modality for domain experts. It allows the experts to efficiently manipulate and measure elements of the 3D model. To better allow designers to support this modality, we report design implications distilled from the feedback from the domain experts.

Keywords Virtual reality · Photogrammetry · Industrial visual analytics · Immersive analytics · Digital twinning · Virtual reality content

1 Introduction

A major obstacle to the mass adaptation of Virtual Reality (VR) is a lack of high-quality content. Even though computer-aided design (CAD) models can be visualized in VR, in the

✉ Ślawomir Konrad Tadeja
skt40@eng.cam.ac.uk

majority of cases such models will still have to be manually prepared, which is a resource-intensive task. Further, in order to generate increasingly realistic models that resemble real-life objects and structures, there is a need to apply high-quality textures to the models. This is particularly important if there is a desire to enhance the immersiveness experienced by the users [29].

One of the most promising approaches to semi-automatic generation of such high-quality 3D models is *photogrammetry* [27]. This approach uses 2D information, such as object's photographs, that can be combined with additional data gathered by a wide array of various sensors, such as GPS location data, to reconstruct digital photo-realistic 3D models of real-life objects [2] (Fig. 1).

However, photogrammetric models may suffer from certain deficiencies due to insufficient or corrupted input data. The resulting models may incur “sharp edges”, “blank spaces,” or break into multiple disconnected parts. This may affect the way in which the user would like to interact with the data.

Nevertheless, photo-realistic models generated from real-world content can be used to gather additional information about the as-it-is current condition of the modeled object. For example, depending on the model and its purpose, it may be feasible to carry out various measurements of such models without the need to actually measure the existing object in its physical space [1, 25], which is often expensive and cumbersome in practice.

Moreover, the capability of capturing the real-world objects and structures in the as-they-are state shows the great promise of the photogrammetric methods for the purpose of digital twinning. If such 3D models are of sufficient quality, digital twins can be used as an effective archival representation and may offer a substitute for site visits and inspections of heritage [3, 20], engineering, manufacturing or industrial sites [8] as well as in asset management [15]. Saddik et al. remarks that in order to unlock the full potential of the digital twins, a number of key technologies have to be utilized, listing among others, the Virtual and Augmented Realities [8].

In recent years, the VR community has successfully used photogrammetry to augment and enhance the fidelity of the user's experience while using an immersive interface, see

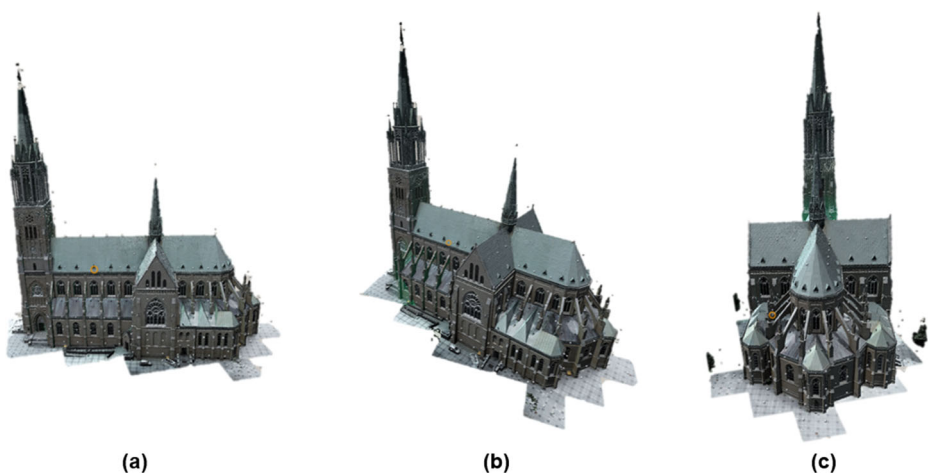


Fig. 1 The 3D photogrammetric model of the Archcathedral Basilica of St. Stanislaus Kostka in Łódź, Poland: (a) side view; (b) three-quarter view; (c) rear view

for example Vajak et al. [29]. Moreover, Antlej et al. [3] noted that when acquiring data for their system, models prepared using photogrammetry offered satisfactory results for VR applications. However, there is little prior work focusing on interacting with the generated 3D models in VR using virtual hands and unencumbered fully tracked gestural input driven by hand tracking. Prior work has investigated non-photogrammetric object manipulation in VR using mid-air interaction [7, 16] and bimanual input techniques [24]. However, despite this area being under-explored, the literature does indicate potential for unencumbered hand tracking-driven interaction of photogrammetric 3D models. For example, a usability study conducted by See et al. [20] concluded that using even low-fidelity virtual hands facilitated by hand-held controllers were better than using virtual device avatars to represent controllers.

In this paper we investigate gaze-supported bimanual interaction to allow domain experts to manipulate a complex photogrammetric 3D structure model and take distance, surface and volume engineering measurements on it. Various measurements of the as-they-are objects are common and important in many application areas, such as surveying, computer-aided design (CAD), structural engineering and architecture. Gaze-tracking and head-tracking are used as means to select a given interactive item. In this paper we estimate gaze-tracking using head-tracking. It is not as precise as eye-tracking as the directions of the user's head and eyes are not necessarily always aligned with each other [21, 22]. The bimanual interaction technique has been an item of interest and studies in the human-computer interaction for many years, for instance, see Buxton et al. [6], Guiard [9] or Balakrishnan et al. [5]. In the latter paper, the authors reported that in their experiments involving carrying out tasks in a 3D virtual scene, the participants had strongly favored the bimanual technique over the unimanual interaction [5].

We report on an engineering case study with six domain experts to better understand how novel gaze-supported bimanual interaction techniques would satisfy domain experts' needs and wants in the task of surveying complex 3D structures, such as buildings. The immersive VR environment used in this study is discussed in more detail in Tadeja et al. [25].

In engineering contexts, CAD models are usually representing idealized and generic items that are typically used for design and evaluation. Environments to support these models must allow not only repositioning but also modification, for example, scaling, separation, and combination of individual parts. Together, these interactions may support more advanced and more complex tasks such as computer simulations.

However, 3D photogrammetric models are often representative of specific real-world objects or structures, such as buildings, captured in their as-they-are states. Such 3D models can be applied to surveying, preservation and maintenance. As such, effective manipulation and interaction techniques for 3D photogrammetric models give rise to different constraints than the ordinary CAD models.

For example, where parts of the CAD models may be freely resized, in surveying with 3D photogrammetric models, the elements of the model must not only reference the actual physical characteristics of the original object but must also be scaled together when resized. It is essential, for example, that relative dimensions are not modified as this could invalidate the survey's results. Instead, interactions of this type are limited to the viewport. This type of specialist evaluation is not concerned with the simulated properties of the model, but the recorded physical characteristics that might vary considerably from the idealized reference. Measurement tools are therefore of crucial importance for evaluating distortions in the surveyed object. Unlike simulations that are designed to support experimentation, real-world

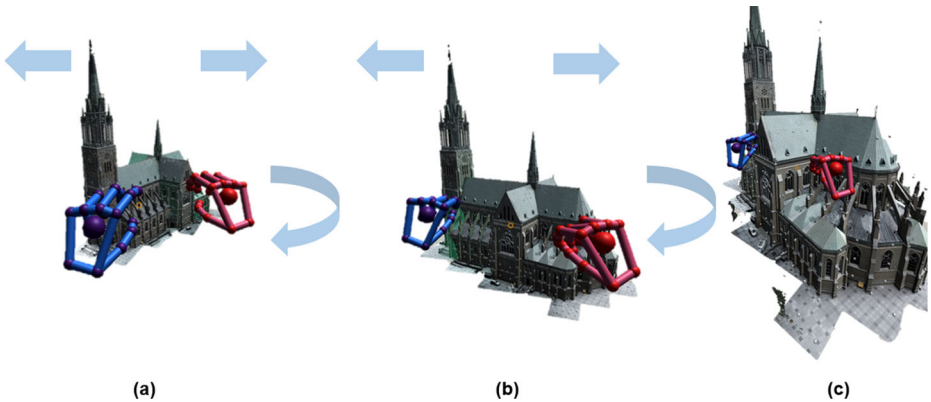


Fig. 2 Manipulation of an object: **a–c** the user rotates the model by making a pinch gesture with both hands and moving one hand (in red) in an arc whilst the other hand serves as the rotation axis (in blue). Simultaneously, by moving the hands apart, the user enlarges the model

properties of photogrammetric models might highlight critical safety concerns, such as sagging beams, subsidence, or dangerous corrosion. Such information, native to photo-realistic 3D models, would have to be additionally provided to ensure the same value for the user when working with CAD models.

2 Minimal design requirements

In this section, we describe the requirements associated with the two main components of the VR system designed to conduct an engineering survey of a real-life structure. These two interlinked functions are (1) model manipulation (Fig. 2); and (2) model measurement (Fig. 3). The ability to manipulate the model is a significant, and potentially limiting, factor in the measurements the user may make. The initial functionality was based on the authors' experience as active interaction design researchers and on prior observation of professionally active engineers. Guided by observations of participants' behavior and their suggestions, we have then extended our system with new capabilities that allow users to take

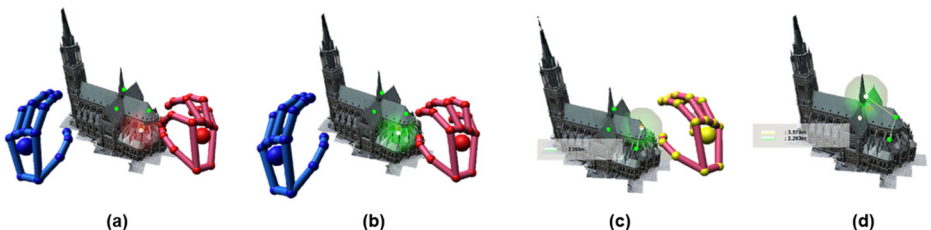


Fig. 3 The user's field of view **a–c** as the user is taking measurements of the model carried out with help of our system: **a** placing the ruler's markers; **b** selecting the marker to be connected with a another marker; **c** selecting a previously connected marker to be a starting point of another ruler; **d** presents the measurements on the HUD (Heads-Up Display): the yellow line has a length of 3.579 [m], whereas the green line has a length of 2.263 [m]

surface and volume measurements of such models. This enhanced functionality necessitated the model measurement requirements to be split into three sub-components.

Here, for simplicity, we are focusing on the minimal viable cases that can be easily extended to cover more complex tasks as well. For instance, taking a distance measurement using the Cartesian coordinates, understood as Euclidean distance in 3D space, requires at least two points. The length of the line connecting these two points constitutes the distance measurement. Similarly, in the case of the surface and volume measurements, consistent with common mesh-based 3D environments, we require coordinates of at least three and four points to calculate the triangle's surface area (Fig. 6) and pyramid volume (Fig. 7) respectively. These calculations have to be adjusted to take into account that the triangles and pyramid surfaces can have various inclination angles in 3D space. Since 3D objects are essentially constructed out of triangle meshes, it is possible to use these basic measuring tools to measure dimensions of more complex structures.

Model manipulation (DR1) The system should provide an easy-to-learn and effective method for manipulating the 3D model so that the interaction technique does not negatively affect the user's main task of identifying and taking desired measurements of the 3D model (Figs. 2 and 4).

Model measurement (DR2) The system should allow the user to take essential engineering measurements of the model, including various distances (for example, height and width), surface and volume of the whole model or its part.

These measurement requirements can be further decomposed into the following three sub-components:

Distance measurement (DR2D) Taking the distance measurements should allow the user to effectively measure any distances (in chosen unit of length, such as meters [m]) between any two chosen points in the 3D space, be it on, within, or near the vicinity of the model (Figs. 5 and 3).

Surface measurement (DR2S) Taking the surface measurements should allow the user to effectively measure any surface area (in a chosen unit of surface area, such as square meters [m^2]) spanning over at least three points regardless of the surface's angle of inclination in the 3D space. Such surfaces could be localized on top, in the near vicinity, within, or crossing over the borders of the model (Fig. 6).

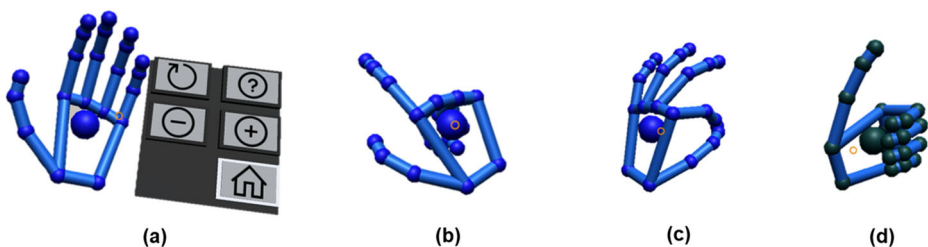


Fig. 4 The four gestures recognized by the hand-tracking: **a** the *left-hand palm up* calls the menu; **a** the *pointing finger* to press the menu buttons; **c** the *single-hand pinch* gesture; and **d** the *thumbs-up* gesture. Icons by *Icons8* (<https://icons8.com>)

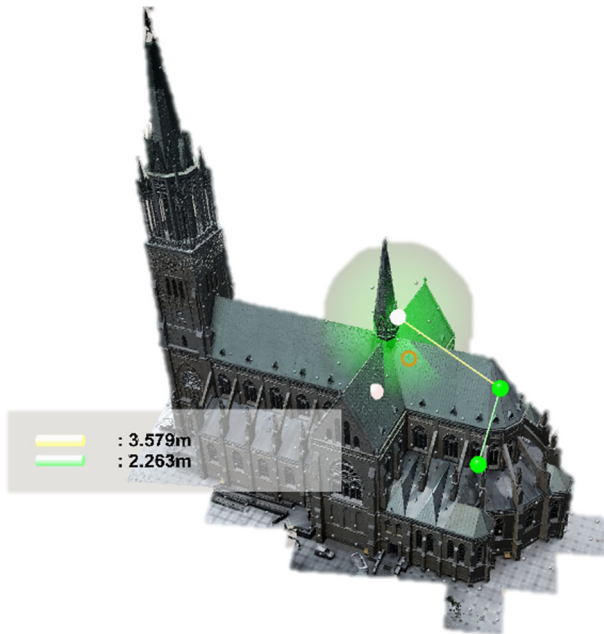


Fig. 5 Example of distance measurements of a photo-realistic recreation model of the Archcathedral Basilica of St. Kostka in Łódź, Poland. As can be seen on the HUD (Heads-Up Display), the yellow line has a length of 3.579 [m], whereas the green line has a length of 2.263 [m]

Volume measurement (DR2V) Taking the volume measurement should allow the user to effectively measure any volume (in chosen unit of volume, such as cubic meters [m^3]) given by at least four chosen points in the 3D space, be they on, within, or near the vicinity of the model (Fig. 7).

3 System description

3.1 Visualization framework

The hardware supporting our system consisted of an Intel i5-9400 CPU and NVidia GeForce GTX 1660 Ti GPU. The head-mounted display (HMD) was an Oculus Rift [17]. The hand-tracking was powered by the Leap Motion Sensor (LMS) [14]. The visualization was written using the Unity game engine and various supporting assets, and the LMS Software Development Kit (SDK) [14].

3.2 Photogrammetric models

Archcathedral model We used a 3D photorealistic model of the Archcathedral Basilica of St. Kostka in Łódź in Poland for capturing distance measurements in VR (see Fig. 1, and <http://softdesk.pl/katedra>). The photos were captured by an unmanned aerial vehicle (UAV). 1754 1754 photos were taken during six independent flights. Those were conducted on

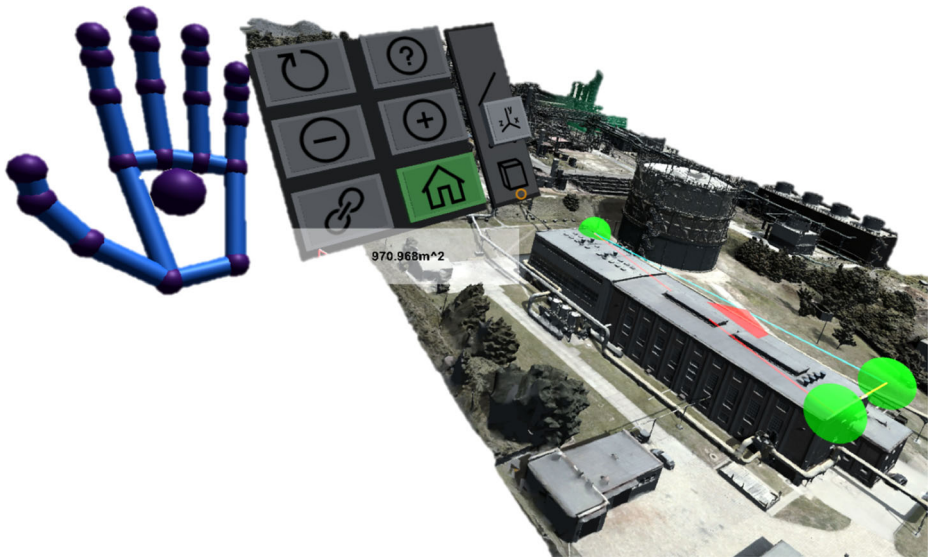


Fig. 6 The surface measurement of a building rooftop at the plant. The surface is approximated by the surface of a triangle whose vertices are given by the three rulers' markers (spheres) placed by the user in the rooftop corners. As can be seen on the HUD (Heads-Up Display), the generated triangle has a surface of 970.968 [m^2]

different dates, which influenced the accuracy of the image metadata as the altitude measurements depend highly on atmospheric pressure, which differed between the flights. As a result, the acquired data was not properly georeferenced, which caused loss of accuracy.

Coking plant model We used a 3D photorealistic model of a coking plant to test the features of capturing surface and volume measurements in VR (see Figs. 6 and 7 respectively, and <http://softdesk.pl/koksownia>). The imagery and sensor data was obtained during a few UAV flights and took around 2.5 hours. After aerotriangulation processed with a root mean square reprojection error of 0.49 pixels, a 3D model was reconstructed from 1491 photos. The sensor data was used to georeference and scale the final model.

3.3 Snapping grid

The 3D model is highly detailed and, as a result, it is impractical to expect a user to directly interact with individual points in the model. Instead, to facilitate an easy-to-understand measurement basis, we surrounded the 3D model by a snapping grid generated in the form of snapping points (represented as small spheres), positioned equidistantly around the model's mesh. Whenever the user was placing a ruler's marker in a vicinity of a snapping point, the marker would automatically snap to its position.

Further, it is possible to automatically place a snapping grid around the larger part of the model, such as its roof. It can be done relatively simply by approximation techniques using the mesh information alone. However, the ability to automatically place the snapping grid around smaller features of the model remains an open problem in the literature.

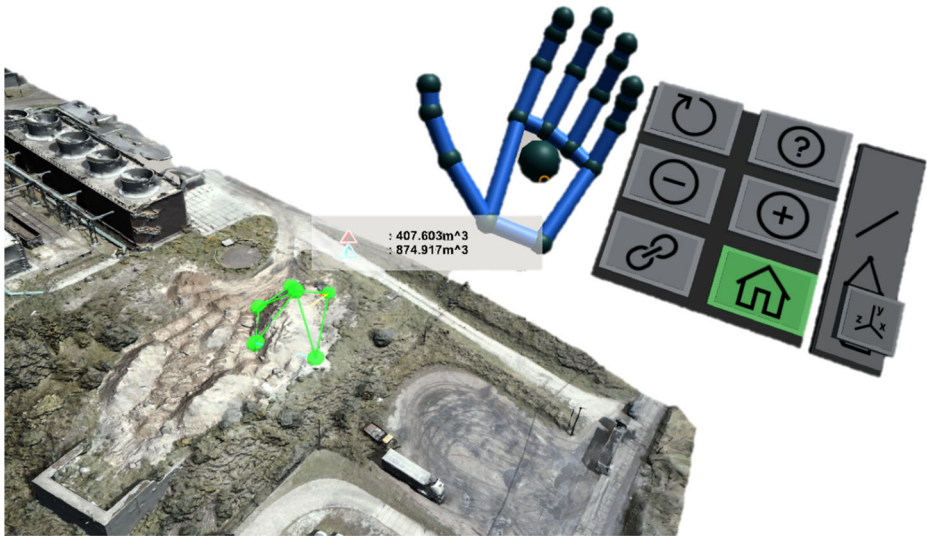


Fig. 7 The volume measurement of a pile of sand taken over a photogrammetric model of the plant. The user placed several ruler markers (spheres) and obtained volumetric measurements of two convex polyhedrons (tetrahedrons). As can be seen on the HUD (Heads-Up Display), the two tetrahedrons have a volume of $407.603 [m^3]$ and $874.917 [m^3]$ respectively

3.4 Interaction methods

The system used a mixture of interaction techniques. It combined the hand-tracking facilitated by the LMS mounted in front of the HMD with gesture-recognition capabilities afforded by the LMS SDK. The system also used gaze-tracking, which was built upon the *Unity VR Samples Pack* and used a cross-hair metaphor and ray-tracing to approximately estimate which object the user was gazing at.

As reflected by Slambekova et al. [23], combining gaze and gesture may positively influence the interaction results as more information about the user's intent is communicated to the system. Since the user can take as many measurements as desired, there may be a substantial number of interactive objects (for instance, the ruler's markers) present in the 3D scene, which were previously generated by the user. Without the use of gaze-tracking, we would have to apply a different, more complex mechanism to allow the user to swiftly select between these objects. For example, the user would have to use their hands to first grab an object. This action could cause an unintentional shift of the ruler's marker prior to taking a distance measurement by creating a spurious connection with another marker. To remove this risk, we would have to create another operation mode invoked either through a new hand gesture, inclusion of a new button, using the left-hand menu (Fig. 4), or a combination of both. Another option would be to implement raycasting selection that extend a ray pointing towards the selected interactive object. Again, in such a case we would have to couple the selection with an additional gesture made with the user's other hand. In both cases the user's gaze would most likely be directed towards the near vicinity of the object the user desires to interact with. Hence, using gaze-tracking to assist in the object's selection procedure is straightforward.

The system recognized four main gestures (illustrated in Fig. 4): (a) the *left-hand palms up* gesture which called up the menu; (b) the *pointing finger* gesture which was used to press

a button on the menu; (c) the *pinch* gesture which was used to drive all primary interaction; and (d) the *thumbs-up* gesture which was used to release the hold on an object. The size and placement of the menu loosely followed the design guidelines distilled by Azai et al. [4].

3.5 Object manipulation methods

The system provided the user with three main methods for manipulating the 3D model, illustrated in Fig. 2. These relied mainly on the double or single-handed pinch gesture (Fig. 4 and the LMS SDK documentation for more details) and allowed the user to: (1) move and reposition the model in the 3D space; (2) decrease or increase the model's size; and (3) rotate the model in the X - Z plane.

3.6 Measurement method

The user could acquire the model's dimension measurements by generating and placing rulers' markers around the object's mesh. Each pair of those markers could then be connected with a ruler, that is, a 3D vector where the magnitude was the measurement, revealed on the head-up display (HUD; Fig. 5).

4 Engineering surveys in real-world environment

Undertaking an engineering survey of existing, real-world asset e.g., a building or another structure, may be a time consuming and often cumbersome process. The goal of such a survey is to capture information about the physical characteristics of a given object or structure [1, 25]. This extracted data can be later on used to plan necessary maintenance and conservation works and to estimate the costs of such endeavors.

In some cases, such data cannot be retrieved directly from the construction documentation and design plans as these may be incomplete, contain errors, or not be even available. This is often the case when dealing with heritage sites where an engineer may be tasked with measuring the height or width of a given structure or dimensions of its individual elements (e.g. racks, towers, rooftop surface). Frequently, the ancient heritage monuments and landmarks, such as ancient places of worship (Fig. 1), does not have such data available on the spot or this data has not been documented thus far. Similar needs may arise when we are dealing with constructions or structures that occupy large areas, contain hazardous zones (Fig. 6) or are in remote locations.

In all of these cases, it may be more feasible and safer to first prepare a digital twin [8] of a given asset. Such digital representation, if of high fidelity, can be used to greatly simplify the task of conducting an engineering survey.

In this paper, the interaction context involves an engineer tasked with estimating the costs of rooftop replacements of a heritage building. As can be seen in Fig. 1, the building in question posses large, rather complex rooftop structure with lots of smaller features and details typical for heritage buildings from its era. Due to the age of the building itself i.e. it was completed in 1912, prior to two World Wars, the existing documentation may be incomplete, inaccurate, or no longer be available. In such a case, the normal order of business would be to (i) close areas near the building, (ii) put in place a scaffolding, and (iii) manually measure the required rooftop dimensions and assess its current condition. This process may not only be lengthy and costly but may also bring additional risks to the workers and pedestrians as well as being heavily depended on the weather conditions.

Instead, we have generated a photo-realistic digital representation of the Archcathedral using a photogrammetry technique. As it was prepared using drone-captured imagery data, there was no need to close nearby areas to set up the scaffolding, and the entire process was efficiently conducted in a short amount of time. In the case of the Archcathedral's building it took six independent drone flights with the total flight time estimated at about 3 hours. Including the test flights, equipment preparations and flight trajectory planning the data acquisition took less than a day. This 3D model was then plugged into the VR-based environment where it could be safely inspected and measured by domain-experts from the safety of their own offices.

5 Observational study

We carried out a small user study with six domain experts, all of whom had an engineering higher education background and years of experience in acquiring accurate measurements of physical buildings.

Given the importance of domain experts for this particular application, we opted for an observational study focusing on a qualitative approach with some ancillary quantitative data. Such evaluation methodology is recognized in, for example, Lam et al. [28], as one of the commonly adopted user research methods among visualization researchers.

5.1 Participants

The study involved six participants, denoted P1–P6. All participants reported little or no prior experience in either VR or hand tracking technology.

Participant 1 (P1) reported basic prior familiarity with VR and no previous exposure to any hand-tracking technology. She was 38 years old and for the past six years has been an office worker in an organization focused on technical inspections for the safe operation of technical equipment in Poland.¹ She was responsible for organizational knowledge-management, preparation of project documentation and managing employee training. She also reported suffering from astigmatism which had a detrimental effect on how she perceived the gaze-tracking cross-hair.

Participant 2 (P2) was 25 years old and reported no prior experience with either a VR or a hand-tracking interface. She held an engineering degree in power engineering and for the past three months had been an intern with the same technical organization as P1.

Participant 3 (P3) reported basic prior familiarity with VR and no previous experience with hand-tracking technology. He was 32 years old and held an engineering degree in power engineering.

Participant 4 (P4) reported basic previous experience with both VR and hand-tracking technologies. He held an engineering degree in mechanical engineering and was 44 years old. He reported having 18 years of experience with CAD systems.

¹Urząd Dozoru Technicznego *eng. Office of Technical Inspection* (UDT) is an EU Notified Body No. 1433 (<https://www.udt.gov.pl>)

Participant 5 (P5) was 47 years old and reported basic prior familiarity with both hand-tracking and VR technologies. He held an engineering degree in power machinery and had 21 years of related work experience.

Participant 6 (P6) reported neither using VR nor hand-tracking technology. He was 57 years old and held an engineering and a master degree in civil engineering. For the previous 10 years, he has been working as a project manager on construction sites. He reported sometimes perceiving a double cross-hair in the head-set and usually wore corrective glasses.

6 Results

6.1 Questionnaire survey results

The number of participants was small. Hence, a statistical analysis is inappropriate. Therefore, the descriptive quantitative statistics below only refer to the sample of participants for this study—the domain experts, and we do not attempt to generalize the findings to a wider population.

Notably, SSQ [13] results revealed only very slightly elevated levels of oculomotor strain among all participants, whereas we observed no effects of nausea. FSS [18] results revealed the lowest observed flow score was 58.57% (P6) with 92.8% (P2) being the highest which suggests a high level of engagement. An analysis of the NASA TLX [10] results revealed that P1 reported a 59/100 score whereas all the other participants (P2–P6) reported scores lower than 50/100 with P4 reporting the lowest score of 13/100. These results suggest relatively low levels of cognitive load across the participants. For the IPQ [11] analysis, as in Schwind et al. [19], we averaged the seven-point scale by the number of questions answered by the individual participants. The IPQ results revealed the majority of the scores approached 50%, with the lowest observed 2.79/7.0 (P6) and the highest 4.0/7.0 (P5). This

Table 1 The participants' results are presented next to the total possible score for each questionnaire

Questionnaire survey results							
Participant	Duration	IPQ	SSQ		FFS		NASA TLX
<i>No</i>	<i>Minutes</i>	<i>Average</i>	<i>Nausea</i>	<i>Oculo-motor</i>	<i>Flow</i>	<i>Anxiety</i>	<i>Total score</i>
P1	25 : 38	3.50/7	0/27	4/21	6.2/7	4.5/7	59.(3)/100
P2	34 : 08	3.71/7	0/27	2/21	6.5/7	5.5/7	30.(0)/100
P3	17 : 53	3.14/7	0/27	4/21	4.2/7	5.0/7	48.(6)/100
P4	16 : 38	3.64/7	0/27	2/21	5.2/7	2.0/7	13.(0)/100
P5	19 : 80	4.00/7	0/27	1/21	6.2/7	3.5/7	42.(0)/100
P6	22 : 37	2.79/7	0/27	1/21	4.1/7	2.5/7	31.(6)/100

The second column reports the approximate time spent in VR by each participant. For SSQ [13], we only report scores after the experiment was completed. Here, half of the participants (P2, P3, and P5) reported oculo-motor strain levels of 4.0 before the experimental phase commenced. The numbers in parentheses in the NASA-TLX [10] column are repeated decimals. In case of the IPQ [11] and FSS [18] the higher the score, the higher the “feel of presence” experience and engagement respectively, as opposed to NASA-TLX and SSQ where we want to observe the lowest cognitive workload and sickness symptoms

Table 2 Quantitative feedback: (r) the model's maximal rotation angle in degrees and (s) minimal and maximal building scale

Participant		Min	Max	Distance
P1	(r)	–	99.4°	21.3
	(s)	.698	38.26	
P2	(r)	–	66.7°	66.5
	(s)	.618	347.8	
P3	(r)	–	71.6°	21.1
	(s)	.282	4.475	
P4	(r)	–	66.0°	6.24
	(s)	.779	4.877	
P5	(r)	–	129°	32.9
	(s)	.090	1.359	
P6	(r)	–	109°	11.6
	(s)	.310	1.777	

The [distance] column shows the total displacement of the model by each participant

demonstrates a passable level for participants' perceived sensation of presence. All of the results from the questionnaires surveys can be found in Table 1.

While the participants were in the experimental phase we gathered additional quantitative feedback (Table 2). This data has to be analyzed with caution as it was not acquired in a timed, fully controlled experimental setting, and was gathered from a small number of domain experts. It was gathered to gain insights about to what extent the participants used the available techniques to manipulate objects (Fig. 2). As such, we were interested in the total angle by which they rotated the object, as a small rotation may be mistakenly applied while moving the model with both hands. We were also looking for the maximal and minimal scale of the model applied by the participants. In addition, we investigated the total displacement distance of the model, that is, the absolute Euclidean distance of the model's movement trajectory. This data indicated that P2's behavior slightly deviated from the other participants. Comparing to the other participants, she had, at some point, substantially enlarged the model zooming in on it (Table 2). She also moved the model around over twice as much as the next participant (Table 2). A possible factor may be her relatively young age and fewer years of professional experience.

6.2 Observed behavior

The tool was designed for the specific purpose of conducting inspection and surveying of engineering assets. As it is difficult to recruit representative volunteers with sufficient expertise, we opted to use a mixed-methods approach to gain an as deep as possible understanding of how the domain-experts would behave when using immersive interfaces to conduct an engineering survey of a 3D photogrammetric model. This approach included a think-aloud protocol as well as a semi-structured interview carried out immediately after the second task was completed. The captured quantitative data allowed us to confirm users' comments by inspecting logged data related to their execution of the tasks. For example, we could inspect by how much a participant had shrunk or increased the size of the model in order to place the ruler's marker.

We analyzed the video recordings that captured both the participants' gestures and the computer display streaming the participant's field of view. The video footage of P1 excluded nine minutes in the middle, which was lost due to a battery outage. We analyzed the videos to first identify a number of modalities along which we were focusing when re-examining the recorded data for the second time. As such, we were particularly interested in studying the following patterns:

Users' movement patterns . We were interested in how the users move (such as moving vs. standing in place) in the real-world which was instantaneously mapped onto their position in the VR environment. Half of the participants (P1, P2, P5) walked around, whereas the others (P3, P4, P6) preferred to remain in a fixed location while using the system. It might be related to the fact that these participants (P1, P2, P5) chose to operate on a largely magnified model in VR space (Table 2), which in turn also made them stretch their arms and hands to place the markers more frequently than the other participants.

The model's movement patterns After the initial moments of the experimental phase when the participants were told to familiarize themselves with the interaction techniques, we observed that all the participants have frequently used the available manipulation methods, such as (r) rotation, (d) movement and (s) scaling (Table 2). With the exception of a single participant (P2), the participants typically chose to manipulate the model by rotating it by small angles, moving it by small distances, or changing its size by a relatively small amount. The participants' manipulation outputs were probably constrained by the length (for movement and rotation) and degrees of freedom of a human arm (for movement). We also observed that P2, who, for a while, extensively increased the model's size as compared to the others (Table 2), seemed to generally prefer to work on a larger model. However, having increased the magnitude of the model too promptly caused her to startle, which most likely speaks of how realistic the model appeared to her. She (P2) also commented that *[it's] intuitive, it's a matter of familiarity with it, I think, because I had a problem with how to rotate it whilst it change its size simultaneously*. No other participant has had troubles with it. For example, P6 commented *I liked rotating, moving, zooming in and out*.

P3, on the other hand, only changed the model's size once when he selected an appropriate magnitude (for him). He also wished to rotate the model about more axes and not just in the $X - Z$ plane. A similar question about the rotation limits was raised by P5 as well: *Can it be rotated in any plane?*

P5 also tried to increase the size by spreading his right-hand fingers similarly to the gesture known from touchscreen interfaces. He suggested that in such case the user[...] *could enlarge with one hand and, for example, reduce [the model's size] with the other [hand]*.

Measurement patterns The participants (P1, P2, P3, P4, P5) chose to connect the rulers and take the measurements only after they felt that all the necessary markers were placed. Only in a single case (P6) a participant decided to take the measurements in two smaller chunks. However, the measurement patterns were still similar to the other participants.

Some users (P1, P2) preferred to look for the previously generated and not-used markers instead of generating new ones. After being suggested that it may be easier to generate a new marker they chose to do that the next time. Moreover, P2 chose to generate a number of markers first to place them one by one later on, whilst the other participants preferred to generate no more than a few markers at once.

Gesture usage In the beginning the participants had to be frequently reminded about the *thumbs-up* gesture (Fig. 4d) used to release the handle over a selected ruler's marker. P5 suggested that when one hand is placing the marker the other hand can be used to make the *thumbs-up* gesture.

All participants kept pointing with their fingers or entire arms towards the model to explain what they did or what they wanted to do to further illustrate what they were communicating. This behavior confirms the high level of engagement indicated in the questionnaires when immersed in the system.

In a single case (P3) a participant kept his left hand alongside his body during the majority of the experiment. P3 also suggested that he would like to see *The Projections of the point's coordinates on the coordinate system's axes*.

We observed that when the participants placed the marker on the Archcathedral Tower (Figure 1), all participants extended their arm to place the marker on top of it. This happened despite all participants being familiar with the manipulation methods and knowing that they could easily decrease the size of the model or lower it to prevent the need to extend their arm.

Gaze-tracking usage As we were using head-tracking as an approximation for the gaze-tracking, there were some issues related to this approach as remarked by the participants. For instance, P1 remarked: *The worst were the two [cross-hairs] that I see all the time [...] they annoy me because I feel confused*. However, she also commented that *[...] I have been diagnosed with astigmatism, maybe that's why? I never did anything with it [correcting the vision]*. P2 remarked that when grabbing the markers he *Forget to aim with this circle [the cross-hair]*. P6 commented that he did not intuitively recognized the connection between the glowing effect and the ability to manipulate the glowing marker. He said, *At first glance, I did not see this relationship*.

Pop-up menu usage At the beginning of the experiment, P1 kept the pop-up menu continuously visible (Fig. 4). When involuntary or voluntary change in the hand gesture occurred, she closed and opened her fist to recall the menu. During the experimental phase, P1 frequently chose to keep the pop-up menu visible while simultaneously making gestures with the other hand, such as removing or placing the rulers' markers around the model. Such behavior was not observed among the other participants with the exception of P6 who for a prolonged amount of time kept his left-hand in the upper hand position.

The participants frequently forgot that the markers can be connected only in the default mode, that is, no buttons should be selected on the left-hand menu.

Further, P5 suggested a vertical heads-up menu instead of the pop-up menu: *Can't such a menu be pinned permanently at the side-corner of the view so it would be visible [...] [constantly]*.

Desired precision of measurements and movements We observed that the main issue with the precision of movements and measurements had to do with the grid overlying the 3D model and a too large radius of the snapping distance. However, despite these inconveniences, the participants were able to carry out the necessary measurements. For instance, P1 remarked that *[...] the precision of the movements, I can see that with each such procedure it is getting better and better. Therefore, it seems to me that [...] it is a matter of training. The more a person interacts with this program, just like with a computer game, [that person] becomes more efficient and proficient.*, but she has also commented on the other occasion that *[...] the manipulation is not precise enough for me*. P3 commented that

The general idea is good only with precision [...] If it was more accurate, it would be better to move these points [rulers' markers] to the places I want, it's a very nice thing for sure.

7 System extensions: surface and volume measurements

The natural extension of the dimension measurement capabilities that were tested during the observational study are the abilities to capture surface and volume measurements as well. To carry out a surface or volume measurement, the user first has to use the additional slider menu (Figs. 6 and 7) attached to the left-hand pop-up menu to select the desired measurement type: (i) surface; or (ii) volume. Once selected, the user can place the ruler's markers around the model. A minimum of three (for surface) or four (for volume) markers have to be placed and connected to carry out the measurement.

In both cases, the measurements themselves are taken and presented on the HUD in chunks (Figs. 6 and 7). The surface measurements are reported as a triangulation of a surface. Volume data is captured as a series of volumes of tetrahedra. This allows the user to measure higher-complexity objects constructed from triangular meshes.

8 Discussion

Thanks to the immersiveness bestowed by VR, there is a substantial difference between how the user perceives, operates, and behaves within an immersive environment as opposed to a traditional workstation setup. There are a number of factors that contribute to this differentiation. For instance, various workstation-based specialized engineering software (such as CAD packages) have been in constant development for decades, which has resulted in well-established and mature technology. Engineers are also regularly taught how to use these programs and use them continuously throughout their professional careers. This prior experience in established practice and sunk cost in learning to master it has most likely biased domain-expert users towards familiar 2D WIMP (windows, icons, menus, pointer) environments. This bias, and the fact that VR technology is still not fully mature, means it is not meaningful to directly compare and contrast these two interfaces.

The main goal of the study was to consider the feasibility of using a gesture-controlled environment coupled with photo-realistic 3D digital twins of real-life structures to conduct a relevant task that domain-experts would be interested in when working with such software. 3D photogrammetric models allow a very close and detailed representation of real-life objects. They have the ability to capture physical information about them, such as their dimensions or their surface structure drawn from imagery information, including, for instance, cracks and discoloration of a rooftop. Such information can be used by experts to reason about the physical state of a given structure. This is not possible when working with computer-aided design (CAD) models, which we have investigated in prior studies [26]. Even carefully crafted CAD models do not represent the current visible state of real-life structures, as they are not generated based on recently captured data. Hence, they cannot be used to conduct an engineering inspection of an existing asset. In contrast, 3D photogrammetric models can be useful in effectively extracting such information from models instead of conducting a costly and time-consuming outdoor inspection. The goal of this study was to explore and trim the vast design space of immersive interfaces with such a purpose in mind by observing and reporting on how domain-experts would behave when using VR environments to take measurements of photogrammetric models.

Some of the features of the system, such as the snapping-grid, were not favored by the participants. This sentiment appear to be exacerbated by their professional and educational background in engineering where the ability to take very detail measurements is crucial (P3: *In general, the idea is good, only the [issues] with precision [of measurements]*). However, as commented by some participants, the snapping grid can be useful in other ways. For example, P1 had an idea of using the grid points as control points for a checklist used by an installation inspector while executing an actual task. The snapping grid could also be useful to fix the marker's position so that it would temporarily not follow the user's hand. This eliminates the need for using the *thumbs-up* gesture. Such a grid could be recalled and hidden on demand and thereby allow more flexibility with marker placement when needed. This suggests that the snapping grid idea, if tweaked to take into account the domain experts' feedback, can serve as a useful tool. A snapping grid may also positively impact effectiveness and ease of use when measuring surfaces and volumes. When the user has to place and connect a large number of markers it is feasible for the system to provide the user with an initial estimation of where the next points should be placed, as well as automatic connection suggestions, with nearby points generated as a group rather than individually. We conjecture that such a feature design would increase performance when taking measurements of the model.

Even though each measurement is automatically saved to a file and the three latest changes are visible on the HUD, P4 wanted to see more of them at once and have some constant fixed measurement as a reference frame. P3 suggested that the markers have their coordinates projected as vectors in the model's local coordinate system. To some extent, these suggestions can be adopted by allowing the user to recall a wire-frame of the rulers connecting a set of essential mesh vertices on demand. These, in turn, can be determined, for example, by using a convex hull algorithm or any other form of more advanced mesh analysis using computer vision.

Based on our observations and the comments by P1 and P2, it appears feasible to explore a bimanual manipulation technique where a gesture carries different semantics depending on the articulating hand. This allows the user to use one hand to rotate the model and the other to move it (P1); or use one hand to increase the model's size and the other to decrease it (P2).

8.1 Design implications

Model manipulation (DS1) Observations and qualitative feedback highlight the strengths of bimanual manipulation of a 3D photogrammetric model. These manipulation methods include: (1) moving and repositioning the model in 3D space; (2) increasing or decreasing the model's size; and (3) rotating the model in the X - Z plane. The participants quickly became fluent in these actions and, after a very short period, did not require any additional guidance from the researchers in order to effectively execute these techniques.

As such, the system should assist the user in effective manipulation of the 3D model by allowing the following three main operations on the model: (1) moving and repositioning the model in 3D space; (2) decreasing or increasing the model's size; and (3) rotating the model by any chosen virtual axis. The latter action removes the X - Z rotation constraint in our study. Participants were able to execute these actions using bimanual manipulation with minimal training. Further, these actions were easily understood, accepted by the participants and lead to successful outcomes.

Model measurement (DS2) Observations and answers to questions during the tasks' execution in the study suggest that the system needs a clear distinction for the measurement mode, as this was the most important task alongside model manipulation. These two tasks are interconnected and the domain experts executed them in a different order depending on their needs, such as a need to readjust their view of the model in order to decide on the next measurement, or to make the next measurement easier. Therefore the user should be able to quickly and reliably switch between these two modes. This requirement could be addressed by, for example, displaying an additional button on the menu with the default mode set to the model's manipulation mode.

The system should support three basic types of direct or indirect measurements: **(DS2D)** distance or length of any of the parts of the model; **(DS2S)** surface area (2); and (3) **(DS2V)** volume. Since these measurement types can be considered higher-order extensions of each other, that is, **DS2D**→**DS2S**→**DS2V**, the interface should provide an efficient and easy-to-learn method for switching between these measurement types. When starting from the measurement type requiring fewer points (distance) and switching to more complex measure type (surface and volume), each measure incurs only a small additional cost (more markers), or no extra cost whatsoever when going the other way, as these markers would already be placed in 3D space. Engineers may be interested not only in the overall surface area or volume but also in details of the measurements on which these calculations are based upon, such as the lengths of a triangle's or a pyramid' sides. In the VR environment this is easily achieved by extending the left-hand pop-up menu with a slider (Figs. 6 and 7) that provides an easy-to-learn and efficient method for switching between the measurement types without compromising previously taken results. Additionally, the model structure itself, that is, the triangular mesh, can be leveraged further by providing automatic suggestions on placing the markers in pre-selected spots around the model.

8.2 Future work

There are several fruitful avenues for future work. It is possible to revise the snapping grid functionality by, for example, keeping the snapping range constant and fairly small, even when the user increases or decreases the size of the model. It is also possible to refine the algorithm for snapping grid generation to be more efficient and effective in terms of placing the snapping points by automatically choosing points of interest, such as mesh vertices. These feature enhancements should be evaluated in a series of controlled experiments to assess if, and by how much, the improved snapping grid is useful and effective.

It is also possible to extend the system with additional capability for handling multiple models at once. In such a scenario, the models could be manipulated by the user as either a single object or as a group. This will most likely require the system to be able to recognize additional, more complex, gestures. Additionally, some studies suggest that using a *grasp* gesture instead of a *pinch* gesture may be more favorable among users [12].

9 Conclusions

This paper has studied bimanual interaction techniques for supporting engineers conducting engineering surveys of photo-realistic digital twins of 3D structures in a VR environment. The efficiency and ease of use of the techniques were studied in a case study with six domain experts.

Based on our observations of user behavior and the participants' own comments we conclude that enabling domain experts in surveying to use their own hands to directly manipulate—grabbing, moving, rotating and scaling—a photogrammetric 3D model in VR is perceived very favorably by such users. All of the participants were promptly able to use their own hands to directly manipulate and take engineering measurements of a complex real-world model in the VR environment. The domain experts also noted that their performance would likely improve with practice. However, the design of the left-hand menu and the overall measurement toolkit should be refined and extended to fulfill the domain experts' needs and wants.

The case study explored interaction techniques that allowed a domain expert to capture not only dimensions but also, as a by-product, calculate the surface and volume measurements of the desired model or its individual parts. Subsequently, after the analysis of the study results, we have extended the system's toolbox with new functionality that allows users to take surface and volume measurements (Figs. 6 and 7) without the need of calculating these values themselves. The way in which these features were implemented flows naturally from the fact that the models are constructed from triangular meshes, and from the initial method of placing and connecting two points in 3D space with a vector. The user does not have to spend time learning intricate novel interaction techniques and there is no need for the system to support elaborate gesture recognition of complex gestures.

The use of photogrammetric models that digitally recreate real-life structures in fine detail has great potential as a means of obtaining high-quality VR content. In addition, such models can help bring to life objects and structures that are well-known to users from everyday life, thus enhancing the VR experience. The case study with domain experts indicates that it is viable to support engineering measurements of complex photogrammetric 3D structures using easy-to-use and efficient interaction techniques. In turn, these observations and results further support the suggestion that photogrammetry has the potential to become a very feasible option for generating digital twins of existing real-life objects, structures or even larger areas.

Acknowledgements This work was supported by an EPSRC-1788814 studentship and the Tsinghua Academic Fund for Undergraduate Overseas Studies and the Fund from Tsien Excellence in Engineering Program. We would also like to thank Krzysztof Kutt for verifying our translations of the questionnaires, to Wioleta Cabala, for advice on the SSQ translation, and to Jason Jacques and Daria Domagała for proofreading the manuscript. We are also thankful to *Bentley Systems Incorporated* for the access to the *ContextCapture* package used to generate the 3D models. We would also like to thank Jerzy Sokołowski and Roman Górecki from the *Office of Technical Inspection (UDT)* in Łódź for their support.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References


1. Aber JS., Marzolf I, Ries JB. (2010) Chapter 3 - Photogrammetry. *Small-Format Aerial Photography*, Elsevier, Amsterdam, pp 23–39. <https://doi.org/B9780444532602100031>

2. American Society for Photogrammetry and Remote Sensing, What is ASPRS?, <https://www.asprs.org/organization/what-is-asprs.html>. Accessed: Apr 2019
3. Antle J K., Bykersma M., Mortimer M., Vickers-Rich P., Rich T., Horan B. (2018) Real-world Data for Virtual Reality Experiences: Interpreting Excavations. 2018 3rd Digital Heritage International Congress held jointly with 2018 24th International Conference on Virtual Systems Multimedia, pp 1–8
4. Azai T, Otsuki M, Shibata F, Kimura A (2018) Open Palm Menu: A Virtual Menu Placed in Front of the Palm. ACM Proceedings of the 9th Augmented Human International Conference, pp 17:1–17:5
5. Balakrishnan R, Kurtenbach G (1999) Exploring bimanual camera control and object manipulation in 3D graphics interfaces. In: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, Pittsburgh, pp 56–62. <https://doi.org/10.1145/302979.302991>
6. Buxton W, Myers B. A (1986) A study in two-handed input. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, pp 321–326. <https://doi.org/10.1145/22627.22390>
7. Caputo FM, Giachetti A (2015) Evaluation of Basic Object Manipulation Modes for Low-cost Immersive Virtual Reality. Proceedings of the 11th Biannual Conference on Italian SIGCHI Chapter, pp 74–77
8. El Saddik A. (2018) Digital Twins: The Convergence of Multimedia Technologies. IEEE MultiMed 25(2):87–92
9. Guiard Y Asymmetric Division of Labor in Human Skilled Bimanual Action. J Motor Behav Routledge 19(4):486–517, 1987 <https://doi.org/10.1080/00222895.1987.10735426>
10. Hart SandraG, Staveland Lowelle (1988) Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Adv Psychol, Elsevier 52:139–183
11. Igroup. Survey on experiences in virtual worlds. Available: <http://www.igroup.org>
12. Jude A, Poor G. M, Guinness D (2016) Grasp, Grab or Pinch? Identifying User Preference for In-Air Gestural Manipulation. ACM Proceedings of the 2016 Symposium on Spatial User Interaction, pp 219–219
13. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG (1993) Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. Int J Aviat Psychol, pp 203–220
14. Leap Motion, <https://www.leapmotion.com/>. [Online]. Last accessed: March 2019
15. Macchi M, Roda I, Negri E, Fumagalli L (2018) Exploring the role of Digital Twin for Asset Lifecycle Management. 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018 51(11):790–795. <https://doi.org/10.1016/j.ifacol.2018.08.415>
16. Manders C, Farbiz F, Yin TK, Miaolong Y, Chong B, Guan CG (2008) Interacting with 3D Objects in a Virtual Environment Using an Intuitive Gesture System. Proceedings of The 7th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, pp 5:1–5:5
17. Oculus Rift, <https://www.oculus.com>. [Online]. Last accessed: Nov 2018
18. Rhei nberg F, Vollmeyer R, Engeser S (2003) Assessment of flow experiences. Diagnosis of Motivation and Self-Concept (Tests and Trends N.F.2), eds. F. Rheinberg and J. Stiensmeier-Pelster (Göttingen: Hogrefe), pp 261–279
19. Schwind V, Knierim P, Haas N, Henze N (2019) Using Presence Questionnaires in Virtual Reality. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp 360:1–360:12
20. See ZS, Santano D, Sansom M, Fong CH, Thwaites H (2018) Tomb of a Sultan: A VR Digital Heritage Approach, 2018 3rd Digital Heritage International Congress held jointly with 2018 24th International Conference on Virtual Systems Multimedia, pp 1–4
21. Sidenmark L, Gellersen. H (2019) Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection. Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19). ACM, New York, pp 1161–1174. <https://doi.org/10.1145/3332165.3347921>
22. Sidenmark L, Gellersen H (2019) Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality. ACM Trans Comput-Hum Interact 27(1), Article 4 (January 2020), 40. <https://doi.org/10.1145/3361218>
23. Slambeckova D, Bailey R, Geigel J (2012) Gaze and Gesture Based Object Manipulation in Virtual Worlds. Proceedings of the 18th ACM Symposium on Virtual Reality Software and Technology, pp 203–204
24. Song W, Goh B, Hutama W, Fu C-W, Liu X (2012) A handle bar metaphor for virtual object manipulation with mid-air interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12). ACM, New York, pp 1297–1306. <https://doi.org/10.1145/2207676.2208585>
25. Tadeja SK., Lu Y, Rydlewicz W, Bubas T, Rydlewicz M, Kristensson PO (2019) PhotoTwinVR: An Immersive System for Manipulation, Inspection and Dimension Measurements of the 3D Photogrammetric Models of Real-Life Structures in VR. <https://arxiv.org/abs/1911.09958>
26. Tadeja SK, Lu Y, Seshadri P, Kristensson PO (2021) Digital Twin Assessments in Virtual Reality: An Explorational Study with Aeroengines. In: Proceedings of the 2020 IEEE Aerospace Conference, Big Sky. <https://doi.org/10.17863/CAM.47663>

27. Thompson MM (1966) Manual of photogrammetry. Geoscience Abstracts, American Society for Photogrammetry
28. Vajak D, Livada Č (2012) Empirical Studies in Information Visualization: Seven Scenarios. IEEE Trans Vis Comput Graph 18:1520–1536
29. Vajak D, Livada Č (2017) Combining photogrammetry, 3D modeling and real time information gathering for highly immersive VR experience. 2017 Zooming Innovation in Consumer Electronics International Conference, pp 82–85

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Sławomir Konrad Tadeja¹  · Yupu Lu¹ · Maciej Rydlewicz² · Wojciech Rydlewicz² · Tomasz Bubas² · Per Ola Kristensson¹

Yupu Lu
yl737@cam.ac.uk

Maciej Rydlewicz
mry@softdesk.pl

Wojciech Rydlewicz
wry@softdesk.pl

Tomasz Bubas
tbu@softdesk.pl

Per Ola Kristensson
pok21@cam.ac.uk

¹ Department of Engineering, University of Cambridge, Cambridge, UK

² Centrum Systemów Softdesk, Łódź, Poland