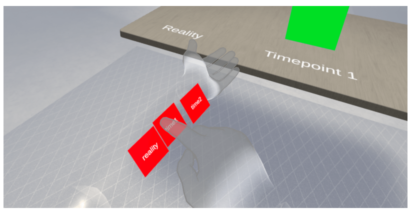
# Assessing Visualization and Interaction Techniques to Support Comparison Tasks in Virtual Reality

Henrike Weingärtner
LMU Munich
Munich, Germany
henrike.weingaertner@ifi.lmu.de

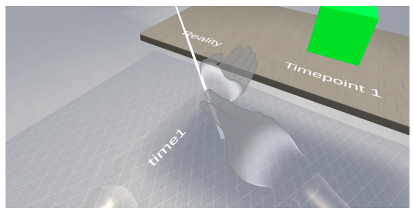
Julian Rasch
LMU Munich
Munich, Germany
julian.rasch@ifi.lmu.de

Nils Rothamel
LMU Munich
Munich, Germany
nils.rothamel@campus.lmu.de

Florian Müller
TU Darmstadt
Darmstadt, Germany
florian.mueller@tu-darmstadt.de

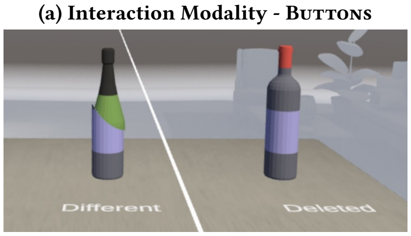



**(a) Interaction Modality - Buttons**

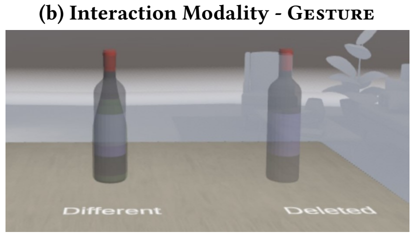



**(b) Interaction Modality - Gesture**

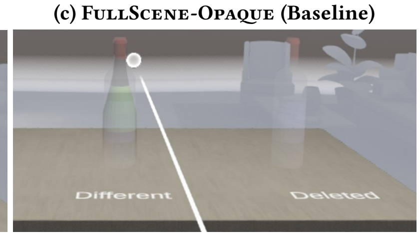
**(c) FullScene-Opaque (Baseline)**




**(d) Flashlight-Opaque**




**(e) FullScene-Transparent**




**(f) Flashlight-Transparent**

**Figure 1: We analyze different combinations of interaction (a, b) and visualization (c, d, e, f) techniques to support comparison of scenes in VR environments. (a, b) depict button and gesture-based interaction modalities. (c) shows the virtual study environment without any support visualizations, serving as a baseline. (d, e, f) present the combinations of the different visualization methods.**

## Abstract

Desktop screens are effective for supporting comparison tasks, but as the scale increases to room-sized or larger structures, context is lost. Users are forced to focus on isolated details through panning, zooming, and scrolling, making it difficult to maintain an overview while exploring finer details. Virtual Reality (VR) potentially offers a solution to this problem by immersing users in 3D spaces and enabling more intuitive comparisons. While related work has proposed many solutions for visualizing and interacting for comparison tasks in desktop environments, knowledge regarding the efficacy of supporting such tasks in VR environments is still lacking. We investigated varying visualization and interaction techniques in a controlled experiment with 24 participants. Our findings provide valuable insights for designing VR systems that improve usability, reduce workload, and enhance performance in comparison tasks.


ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3719737

## CCS Concepts

• **Human-centered computing** → **Virtual reality**; **Gestural input**.

## Keywords

Comparison Support, Visualization Techniques, Interaction Techniques, Virtual Reality

**ACM Reference Format:**
Henrike Weingärtner, Julian Rasch, Nils Rothamel, and Florian Müller. 2025. Assessing Visualization and Interaction Techniques to Support Comparison Tasks in Virtual Reality. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3706599.3719737

# 1 Introduction

Recognizing changes or subtle shifts is crucial for decision-making and comparison tasks across fields [11] like forensics [4], historical research [12], medical imaging [18], and automotive engineering [15]. We can easily perform such tasks on a desktop screen for small objects but as the scale increases, e.g., for comparing room-sized or larger structures, context is lost, and we are forced to focus only on isolated details while panning, zooming, and scrolling the environment. This hampers our ability to maintain a clear overview and examine finer details at the same time [9]. Further, the inherent 2D nature of desktop systems makes the task more difficult, as they lack the spatial depth and perspective needed to fully represent complex 3D environments.

While AI-driven solutions could theoretically encode and highlight differences explicitly, they would introduce new challenges, like distancing users from the process, leading to disengagement and reliance on automated judgments that may bypass human intuition. As another approach, Virtual Reality presents a promising alternative by immersing users directly into a 3D environment and, thus, potentially offers a powerful way to balance computational efficiency with active human involvement and understanding. However, while VR enables users to explore spatial relationships [20], it introduces a new challenge: How do we compare different objects or scenes in such an immersive setting? For desktops, research identified juxtaposition – where objects are separated in space or time – and superimposition – where an object overlays another in the same coordinate system – as typical comparison techniques. However, it is as of today unclear how these different visualization techniques perform in VR environments and how they need to be designed to ensure both efficiency and accuracy, but also to provide a pleasant user experience.

As a first step to achieve this, in this paper, we contribute a systematic investigation of common comparison techniques in a VR environment. We include both juxtaposition and superimposition as possible visualization techniques. As side-by-side comparisons are impractical in VR as they disrupt immersion, and we can only fully experience one 3D space at a time, we use temporal displacement (i.e., showing only one of the alternative versions of a scene at a time.) as a juxtaposition technique. As these approaches rely heavily on user interaction (i.e., selecting different scenes or visualization modes), this further raises questions about how the comparison visualizations should be applied — whether to entire scenes or specific areas — and how users can best control and engage with them.

In our study, we evaluated how different combinations of interaction techniques and visualization methods impact users' ability to identify differences between VR scenes. By examining their impact on usability, workload, and performance, we provide actionable insights for designing VR applications that effectively leverage human intuition for comparison tasks.

# 2 Related Work

Searching and comparing objects or scenes is a critical task across various domains [4, 11, 15]. While AI and computers excel at automating tasks and processing data efficiently, having a human in the loop is beneficial for several reasons [17]. Transparency and accountability are crucial for trust and accuracy, as seen in AI-driven medicine, where users need reliable explanations for AI decisions. The European Commission requires human oversight to ensure trustworthy AI [1]. Humans provide critical thinking and ethical judgment that machines cannot, ensuring decisions are aligned with nuanced contexts and values [10]. Their oversight also helps prevent errors and mitigate unintended consequences, particularly in complex or high-stakes scenarios.

Supporting users in effectively performing these tasks in VR requires thoughtful design of both visualization techniques and interaction methods. Object comparison techniques have been explored across various domains. Each method offers distinct advantages and challenges. Visual designs that aid comparison generally fall into three categories: juxtaposition (separating objects in space or time), superimposition (overlaying objects in the same coordinate system), and explicit representation of relationships (visually encoding connections, such as differences between objects) [3]. Gleicher et al. [3] highlight the significance of interaction in visual comparison. Another prominent solution for VR visualization are metaphor-based visualization approaches, such as using flashlights and other objects. These familiar objects help users intuitively navigate and interact with 3D visualizations [16]. This work also laid the groundwork for later research that utilized intuitive gestures and tools to navigate complex data in a more accessible and user-friendly way. On-Body-Interaction uses the human body as both an intuitive input and responsive output interface, providing a constantly available interaction platform [6]. Leveraging the body as an interactive surface has been shown to improve accuracy and reduce fatigue while preserving mobility and eliminating the need for extra devices. This creates a seamless, direct interaction space, minimizing reliance on external controllers and offering an always-accessible platform for engagement [2]. Body interaction can be enabled through various methods, like acoustic sensors [7] or wearables [13, 14]. Depth-sensing cameras can enable interaction with any surface [5].

# 3 Methodology

By combining these proven visualization and interaction techniques, our study aims to enhance user engagement and effectiveness in immersive comparison tasks. We conducted a controlled experiment to explore the impact of different interaction and visualization techniques on the accuracy, usability, and perceived workload of spotting differences between multiple versions of a VR scene. We tasked participants to discover as many differences as possible between three different versions of the same VR scene. In each condition, we presented our participants with one original scene and two varied scenes that contained in total 30 differences (15 in each varied scene) to the original scene. Depending on the condition, participants used different techniques to switch between scenes and experienced varying visualizations. We chose 30 differences through pre-testing the task, to suit the timeframe per condition and used two varied scenes to add a reasonable level of complexity without the risk of overwhelming the participants.. To minimize learning effects, each condition used a newly randomized original scene, and the two alternatives were derived from it.

To systematically assess the impact of different interaction and visualization techniques, we varied three independent variables:

**Interaction Modality** We varied the way participants activated and deactivated the display of a varied scene between BUTTON and GESTURE. Buttons were positioned between hand and elbow joints. Each button was labeled to indicate the corresponding scene. For the gesture, swiping forward or backward switched to the next or previous scene. A forearm display indicated the selected time point. The sequence was fixed (original scene - alternative scene 1 - alternative scene 2), with no cyclic navigation.

**Visualization Technique - Area of Effect** We varied the way the alternative scene was presented with two levels. In the FULLSCENE case, the alternative scene took over the participant's full field of view. In the FLASHLIGHT case, we used the metaphor of a flashlight held in the hand. The cone of light (15-degree angle) from the flashlight serves as a portal into the second scene, allowing users to selectively choose the area of interest.

**Opacity** As a final independent variable, we varied the blending in the area defined by the visualization technique between a semi-TRANSPARENT view, in which the original scene and the selected varied scene were visible at the same time, and an OPAQUE view, in which only the original or the alternative scene was visible.

We employed a repeated-measures design, resulting in a total of $2x2x2 = 8$ conditions (compare Figure 1), consisting of 4 visualizations and 2 interaction techniques. To avoid learning effects, we counterbalanced the order of conditions in a balanced Latin square design with 8 levels. For each condition, we measured the number of correctly and incorrectly identified differences as measures for the accuracy of our participants. Further, we assessed the usability of the system using the System Usability Scale (SUS), and the perceived workload using the NASA Task Load Index (NASA-TLX).

## 3.1 Apparatus

The study was developed using Unity, leveraging the Meta All-in-One SDK for hand tracking and interaction. The virtual environment was a realistic office setting, featuring 40 spawn points for item placement. The layout of the office—complete with desks and shelves remained constant across conditions, but specific objects like plants and decorations were altered between time points. This ensured logical and varied object placement. Spawn points were categorized (e.g., small items, large items, floor objects) to maintain realism, with items being randomly selected and placed within their category. Buttons and Gestures were implemented with inverse kinematics on the left forearm via the Unity humanoid model. Only the hands and arms were visible to participants. Visualizations were designed by employing various methods of superposition, where objects from multiple time points are displayed simultaneously.

## 3.2 Procedure

Participants signed a consent form and completed a demographic questionnaire. They were briefed on the study's objective to identify differences between time points in VR. This was followed by two tutorials on the interaction methods (button or gesture) and on the visualization techniques (Transparent, Flashlight, Flashlight Transparent). Participants then completed three 45-second runs per condition and interaction combination, identifying and correcting up to 30 possible differences per run. When participants identified a difference, they used a pinch gesture to interact with the element, causing the difference to disappear and be replaced by the corresponding object from the original scene. This prevented participants from selecting the same difference multiple times. The time limit prevented excessive searching and potential frustration over missing final differences. The number of differences exceeded what could realistically be found in the given time, ensuring engagement and a consistent challenge and preventing ceiling effects. After each trial, they were asked to complete the System Usability Scale (SUS) and NASA Task LoadIndex (NASA-TLX) assessments and answer a question about Immersion. Finally, we collected qualitative feedback in a semi-structured interview. Each participant's session lasted around 30 minutes.

## 3.3 Participants

A total of 24 participants took part in the study, with an average age of 26.33 years (13 male, 11 female). The sample consisted of individuals from diverse occupational and educational backgrounds. Half of the participants had prior VR experience, using it a few times a year, while the other half were novices. All participants provided the necessary consent before taking part in the study.

## 3.4 Analysis

We computed the RAW TLX score as proposed by Hart [8]. For analysis, we fitted linear mixed models (by REML) with our independent variables as fixed effects and the individual participant as a random effects. We assessed significance using Type III Wald chi-square tests. Where we found significant main or interaction effects we did Post-Hoc-Tests and applied the Bonferroni method for p-value adjustments. For non-parametric data (SUS), we first performed a aligned rank transformation according to Wobbrock [19], followed by the same linear regression models and significance tests. We fitted Poisson regression models and assessed significance using Type III Wald chi-square tests to analyze count data. Significant main or interaction effects were followed by post hoc tests with Bonferroni-adjusted p-values.

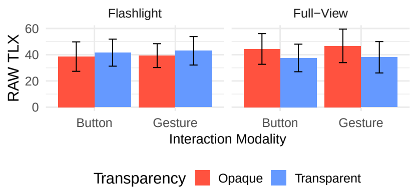
# 4 Findings

This section outlines the findings of our controlled experiment.

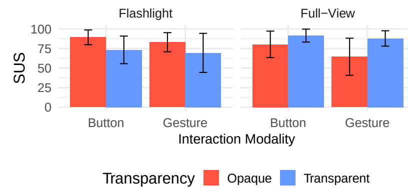
## 4.1 Raw TLX

Evaluating the raw TLX, we found values ranging from $M = 37.5, SD = 10.5$ (BUTTON-FULLSCENE-TRANSPARENT) to $M = 46.7, SD = 12.8$ (GESTURE-FULLSCENE-OPAQUE) (see Figure 2a). Our analysis on the TLX scores identified significant main effects of the AREA OF EFFECT ($\chi^2 = 7.15, p < 0.01$), but the post-hoc tests revealed no significant differences for the FLASHLIGHT and FULLSCENE conditions.

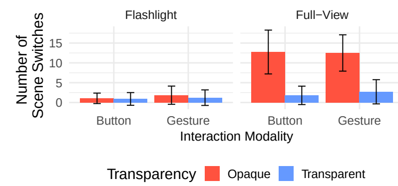
We found a significant effect for the interaction between the AREA OF EFFECT and OPACITY factors ($\chi^2 = 10.22, p < 0.01$). However, there were no significant main effects for INTERACTION



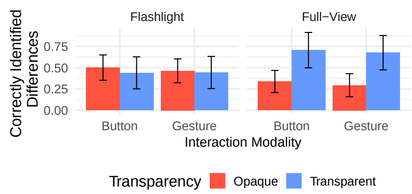

(a) RTLX



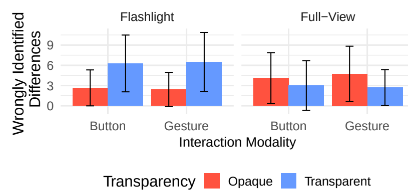

(b) SUS




(c) Number of Scene Switches




(d) Correctly Identified Differences in %




(e) # of Wrongly Identified Differences

**Figure 2: The mean results of our user study's log and count data. Error bars depict standard deviation.**

Modality or any other interaction effects. In the FullScene condition, FullScene-Opaque had a higher TLX score than FullScene-Transparent. However, no significant difference was found between Flashlight-Opaque and Flashlight-Transparent. While FullScene-Opaque had significantly higher scores compared to Flashlight-Opaque ($p < 0.001$), Flashlight-Transparent had significantly higher scores than FullScene-Transparent ($p < 0.001$).

## 4.2 SUS

Evaluating the SUS scores, we found values ranging from $M = 64.4, SD = 23.8$ (Gesture-FullScene-Opaque ) to $M = 91.5, SD = 8.27$ (Button-FullScene-Transparent) (see Figure 2b). Significance testing revealed a significant main effect of Input Modality ($F(1, 161) = 16.40, p < .001$) on SUS scores, while no significant main effects were found for the other factors. The post-hoc tests revealed significantly higher SUS scores for Button compared to Gesture ($p < 0.001$). We found a significant effect for the interaction between the Input Modality and Area of Effect factors ($F(1, 161) = 3.65, p < 0.05$). Button-Opaque had a significantly higher SUS score than Gesture-Opaque ($p < 0.001$), but no significant difference was found Button-Transparent and Gesture-Transparent. We also found a significant main effect of the interaction between the factors Area of Effect and Opacity ($F(1.161) = 74.10, p < 0.001$). Post-Hoc tests revealed, that when looking at the Flashlight factor, Flashlight-Opaque had significantly higher scores than Flashlight-Transparent ($p < 0.001$), while for FullScene, FullScene-Transparent had significantly higher scores than FullScene-Opaque ($p < 0.001$). Additionally, FullScene-Transparent scored significantly higher than Flashlight-Transparent ($p < 0.001$), whereas Flashlight-Opaque scored significantly higher than FullScene-Opaque ($p < 0.001$). Specifically, the Flashlight visualization worked better when paired with the Opaque condition, while the FullScene visualization performed better when paired with the Transparent condition.

## 4.3 Scene Changes

Evaluating the number of scene changes, we found values ranging from $M = 0.9, SD = 1.60$ (Button-Flashlight-Transparent) to $M = 12.7, SD = 5.53$ (Button-FullScene-Opaque) (see Figure 2c). We found significant differences for the interaction modality ($\chi^2 = 15.20, p < 0.001$) and Area of Effect ($\chi^2 = 436.95, p < 0.001$) factors. Post-hoc tests revealed significantly higher numbers of scene switches for Gesture than for Buttons ($p < 0.001$). Flashlight had a significantly lower numbers of scene switches than FullScene ($p < 0.001$). We found a significant main effect of the interaction between Modality and Area of Effect factors ($\chi^2 = 14.67, p < 0.001$). For both Flashlight-Gesture and FullScene-Gesture, there were significantly more switches compared to Flashlight-Button and FullScene-Button ($p < .001$). We also found a significant main effect of the interaction between Area of Effect and Opacity ($\chi^2 = 86.87, p < 0.001$). Post-Hoc tests revealed significant differences for most comparisons ($p < .001$), except between Flashlight-Opaque and Flashlight-Transparent. Over all, there were more switches for FullScene-Opaque and FullScene-Transparent than for Flashlight-Opaque and Flashlight-Transparent ($p < .001$).

## 4.4 Rate of Correctly and Falsely Identified Differences

Evaluating the rate of correctly identified differences, a value of 1 represents 100% of the total possible differences (30), meaning all differences were correctly identified. We found values ranging from $M = 0.25, SD = 0.14$ (Gesture-FullScene-Opaque) to $M = 0.71, SD = 0.21$ (Button-FullScene-Transparent) (see Figure 2d). We found a significant interaction effect between the Area of Effect and Opacity factors ($\chi^2 = 6.2811, p < 0.05$), while all main effects and further interactions were not significant. Post-hoc tests showed that FullScene-Transparent had a significantly higher percentage of correctly identified objects compared to FullScene-Opaque ($p < 0.001$) and Flashlight-Transparent($p < 0.05$). No significant difference was found for other combinations.

Evaluating the number of wrongly identified objects, we found values ranging from $M = 2.43, SD = 2.53$ (Gesture-Flashlight-Opaque) to $M = 6.28, SD = 4.21$ (Button-Flashlight-Transparent) (see Figure 2e). The analysis showed a significant interaction effect between the Area of Effect and Opacity factors ($\chi^2 = 6.2811, p < 0.05$), while all main effects and further interactions were not significant. Post-hoc tests revealed a significantly lower number of wrongly identified objects for FullScene-Transparent than for FullScene-Opaque($p < 0.001$) and Flashlight-Transparent ($p < 0.05$). No significant difference was found for other combinations.

### 4.5 Semi Structured Interview

All participants said that participating in the study "was an interesting experience" (P7) or that "It was very fun" (P9). Participants found the button interaction more intuitive and reliable than the swipe gesture ("The buttons are more convenient and easier to use." - P8), while the swipe gesture was perceived as inconsistent or harder to use than the buttons ("Sometimes when I swiped on the interactions it didn't register quite right" - P12). Certain combinations of interactions and visualizations were favored, particularly the Flashlight-Opaque and FullScene-Transparent visualizations paired with the Button interaction. The Flashlight-Opaque visualization was preferred over Flashlight-Transparent, as "the half-transparent [...] flashlights were a bit more difficult" (P18). The Flashlight-Opaque visualization was appreciated for its ease and enjoyment in finding differences, and "was very easy to become familiar with and fun to use" (P22). The combination of FullScene-Transparent helped avoid repeated scanning, with Participant 17 stating, "You saw all of the differences right away, so it was just a matter of clicking on them." Targeting small or hidden objects seemed to be challenging.

## 5 Discussion

### 5.1 Buttons increase usability compared to gestures, at least for a small number of scenes.

Our results indicate that buttons outperformed gestures in terms of usability, as reflected by higher SUS scores, fewer scene switches, and participant feedback. We observed that some participants experienced unreliabilities with the gesture system, which likely contributed to their confusion and the increased number of (maybe accidental) scene switches in the gesture condition. These issues may stem from challenges in understanding the gesture interface or problems with our implementation, both of which could disrupt the overall user experience. Interestingly, we found no significant differences in NASA-TLX scores between the two interaction methods. This indicates that while gestures were less reliable and intuitive, they were not necessarily more mentally or physically demanding for participants. It is important to note that this study compared interaction methods across only three scenes, and the results might differ in scenarios involving a larger number of scenes. With an increasing number of scenes, the physical space available for buttons may become a limiting factor, potentially impacting their usability. This highlights the potential scalability advantage of gesture-based systems, provided their reliability and intuitiveness can be improved. The better performance of buttons highlights the importance of designing intuitive and reliable interaction methods. Difficulties with gestures emphasize the need to improve their responsiveness and clarity.

### 5.2 Full-Scene provides a more immediate overview, while Flashlight helps to better distinguish details.

There was no significant difference between the Flashlight and FullScene conditions for the TLX or SUS scores. The Flashlight condition was perceived as engaging and enjoyable, which is beneficial because it helps ensure users enjoy using the system. However, the FullScene-Transparent performed better than the Flashlight-Transparent condition. Overlaying multiple pieces of information can create visual clutter, as seen in 2D environments, but here, FullScene-Transparent still showed a significant improvement compared to the baseline (FullScene-Opaque). The FullScene-Transparent visualization can provide an immediate overview without the need to scan the room "by hand", while the Flashlight conditions offer a visual cue when quickly hovering over an object and moving away, which might help users to notice changes, as evidenced by the good performance of Flashlight-Opaque. The Flashlight conditions resulted in fewer scene switches compared to the FullScene conditions, which may be explained by this visual cue. Both conditions (FullScene and Flashlight) offer different advantages, and there is no clear better option when considering this variable alone.

### 5.3 Opacity has an effect on the Area of Effect Visualizations

Results indicated that the Opacity has a significant impact when paired with the Area of Effect variable. In general, FullScene-Opaque resulted in the most scene changes, higher workload (NASA-TLX), lower SUS scores, and fewer correctly identified differences compared to FullScene-Transparent. This was expected and is likely due to the need for constant scene switching without additional assistance, with the FullScene-Opaque condition serving as a baseline. Pairing FullScene with Transparent improves the experience, mitigates many of the drawbacks, and leads to fewer errors. These results suggest that while interaction modality did not affect the identification accuracy, the combination of the Area of Effect and Opacity factors had a significant impact on performance. The effect of the Area of Effect dimensions appears to be influenced by the level of transparency (Opacity). While pairing the Transparent condition with the FullScene condition led to improvements, this was not the case for the Flashlight condition. The combination was perceived as unpleasant and performed worse. Therefore, the effect does not carry over to the Flashlight condition. This may be because the overall intensity of visual changes when hovering over objects is more pronounced in the FLASHLIGHT-OPAQUE visualization, which is why the effect does not transfer to the FLASHLIGHT-TRANSPARENT condition. Although FLASHLIGHT-TRANSPARENT did not perform as effectively, it still resulted in more correctly identified differences compared to the condition with no support (FULLSCENE-OPAQUE).

## 5.4 Limitations and Future Work

Future work should refine gesture interactions for better accuracy and reliability. It would be valuable to explore the scalability of both button and gesture-based interactions in more complex environments and with larger datasets. Furthermore, there is room to investigate the effects of varying levels of transparency and blending in combination with a different-sized area of effect, which could help identify the most optimal configurations for improving performance, particularly in complex scenarios. Improving visual feedback, such as experimenting with more pronounced visual effects, could help users better identify changes, improving accuracy and user experience. The small number of scenes tested in our experiment may not fully represent the challenges in more dynamic environments. Expanding the number of scenes would allow us to assess interaction methods in more varied and complex contexts. Lastly, this study focused on button and gesture interactions, but exploring other modalities, like voice commands or haptic feedback, could provide further insights.

# 6 Conclusion

We conducted a controlled experiment to explore the impact of different interaction and visualization techniques on the accuracy, usability, and perceived workload of spotting differences between scenes in VR. We varied the three independent variables—Interaction Modality, Area of Effect, and Opacity —resulting in 8 conditions. Our results showed that button interactions outperformed gestures in usability, at least for a small number of scenes. The FULLSCENE provides a more immediate overview, while the FLASHLIGHT view helps to better distinguish details. OPACITY influenced performance when paired with AREA OF EFFECT. While transparency improved the experience in the FULLSCENE condition, it did not translate well to the FLASHLIGHT condition. Future work should focus on refining gesture interactions, exploring the scalability of both modalities in larger environments, and investigating the effects of different levels of blending. Expanding the number of scenes and incorporating other feedback mechanisms could further enhance the user experience in immersive VR systems.

# References

[1] European Commission. 2024. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[2] Bruno Fruchard, Arnaud Prouzeau, Olivier Chapuis, and Eric Lecolinet. 2019. Leveraging Body Interactions to Support Immersive Analytics. In *CHI 2019 - The ACM CHI Conference on Human Factors in Computing Systems - Workshop on Interaction Design & Prototyping for Immersive Analytics*. Glasgow, United Kingdom, 10 pages. https://hal.science/hal-02095993

[3] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D. Hansen, and Jonathan C. Roberts. 2011. Visual comparison for information visualization. *Information Visualization* 10, 4 (2011), 289–309. doi:10.1177/1473871611416549

[4] Bethany Growns and Kristy A. Martire. 2020. Human factors in forensic science: The cognitive mechanisms that underlie forensic feature-comparison expertise. *Forensic Science International: Synergy* 2 (2020), 148–153. doi:10.1016/j.fsisyn.2020.05.001

[5] Chris Harrison, Hrvoje Benko, and Andrew D. Wilson. 2011. OmniTouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) *(UIST '11)*. Association for Computing Machinery, New York, NY, USA, 441–450. doi:10.1145/2047196.2047255

[6] Chris Harrison, Shilpa Ramamurthy, and Scott E. Hudson. 2012. On-body interaction: armed and dangerous. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction* (Kingston, Ontario, Canada) *(TEI '12)*. Association for Computing Machinery, New York, NY, USA, 69–76. doi:10.1145/2148131.2148148

[7] Chris Harrison, Desney Tan, and Dan Morris. 2010. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '10)*. Association for Computing Machinery, New York, NY, USA, 453–462. doi:10.1145/1753326.1753394

[8] Sandra G. Hart. 2006. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 50, 9 (2006), 904–908. doi:10.1177/154193120605000909 arXiv:https://doi.org/10.1177/154193120605000909

[9] L. Holmquist and Christopher Ahlberg. 1997. Flip Zooming: A Practical Focus+Context Approach to Visualizing Large Data Sets. (1997), 763–766.

[10] L. Inglada Galiana, L. Corral Gudino, and P. Miramontes González. 2024. Ethics and artificial intelligence. *Revista Clínica Española (English Edition)* 224, 3 (2024), 178–186. doi:10.1016/j.rceng.2024.02.003

[11] Kyungyoon Kim, John V. Carlis, and Daniel F. Keefe. 2017. Comparison techniques utilized in spatial 3D and 4D data visualizations: A survey and future directions. *Computers & Graphics* 67 (2017), 138–147. doi:10.1016/j.cag.2017.05.005

[12] Kyungyoon Kim, Bret Jackson, Ioannis Karamouzas, Moses Adeagbo, Stephen J Guy, Richard Graff, and Daniel F Keefe. 2015. Bema: A multimodal interface for expert experiential analysis of political assemblies at the Pnyx in ancient Greece. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 19–26.

[13] Gierad Laput, Robert Xiao, and Chris Harrison. 2016. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) *(UIST '16)*. Association for Computing Machinery, New York, NY, USA, 321–333. doi:10.1145/2984511.2984582

[14] Pranav Mistry, Pattie Maes, and Liyan Chang. 2009. WUW - wear Ur world: a wearable gestural interface. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI EA '09)*. Association for Computing Machinery, New York, NY, USA, 4111–4116. doi:10.1145/1520340.1520626

[15] Stefan Nolle and Gudrun Klinker. 2006. Augmented reality as a comparison tool in automotive industry. In *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE, Santa Barbara, CA, USA, 249–250. doi:10.1109/ISMAR.2006.297829

[16] Alex Pang and Michael Clifton. 1995. Metaphors for visualization. In *Visualization in Scientific Computing '95*, Riccardo Scateni, Jarke J. van Wijk, and Pietro Zanarini (Eds.). Springer Vienna, Vienna, 1–9.

[17] Carl Orge Retzlaff, Srijita Das, Christabel Wayllace, Payam Mousavi, Mohammad Afshari, Tianpei Yang, Anna Saranti, Alessa Angerschmid, Matthew E. Taylor, and Andreas Holzinger. 2024. Human-in-the-Loop Reinforcement Learning: A Survey and Position on Requirements, Challenges, and Opportunities. *J. Artif. Int. Res.* 79 (April 2024), 57 pages. doi:10.1613/jair.1.15348

[18] David Schroeder, Fedor Korsakov, Carissa Mai-Ping Knipe, Lauren Thorson, Arin M Ellingson, David Nuckley, John Carlis, and Daniel F Keefe. 2014. Trend-centric motion visualization: Designing and applying a new strategy for analyzing scientific motion collections. *IEEE transactions on visualization and computer graphics* 20, 12 (2014), 2644–2653.

[19] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143–146. doi:10.1145/1978942.1978963

[20] Lei Zhang, Ashutosh Agrawal, Steve Oney, and Anhong Guo. 2023. VRGit: A Version Control System for Collaborative Content Creation in Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 36, 14 pages. doi:10.1145/3544548.3581136