Virtual Reality (2026) 30:204
https://doi.org/10.1007/s10055-026-01433-z

ORIGINAL ARTICLE

Visual quality features in immersive virtual environments: a systematic review of their impact on user experience

Enlian Zhang[1] · Jin-Bin Im[1] · Kyung-Tae Lee[2] · Ju-Hyung Kim[1]

Received: 10 December 2025 / Accepted: 6 July 2026
© The Author(s) 2026

Abstract
Visual quality is a critical determinant of user experience in immersive virtual environments. Although many studies have focused on integrated measures of presence or immersion, few have examined how the distinct aspects of visual quality relate to specific experiential factors. This review decomposes the visual quality features into three domains: modelling, hardware and rendering, and motion and interaction. User experience in a virtual environment can be divided into perception, realism, immersion, and presence. A detailed manual review of 90 focused studies published between 2014 and 2025 is conducted, and their main findings are mapped according to these categories. This review highlights the importance of linking technical features to user-centred outcomes. It offers practical guidance for virtual reality and augmented reality applications in education, architecture, simulation, entertainment, and healthcare. Furthermore, this review identifies research gaps, including inconsistencies in the reported effects of hardware features and the limited integration of multimodal feedback in current studies. Future research should further explore adaptive and multimodal approaches to improve user engagement.

Keywords Virtual environment · Perception · Presence · Immersion · Realism · Visual quality

1 Introduction

The aim of immersive virtual environments (IVEs) has traditionally been the replication of physical reality—an ideal state of visual realism in which the virtual world becomes perceptually indistinguishable from the real one. As display technologies advance towards higher resolutions, wider fields of view, and physically based rendering, visual fidelity is often assumed to asymptotically approach real-world complexity. However, this continuous pursuit of high-fidelity graphics presents a challenge regarding the efficiency of visual enhancements. Determining whether increasingly expensive visual features consistently translate into meaningful experiential gains, or if they exceed users' perceptual thresholds and consequently offer diminishing returns, remains critical.

Resolving this uncertainty holds significant practical implications, particularly given the widespread adoption of IVEs in education (Cheng et al. 2024), training (Jensen and Konradsen 2018), healthcare (de Farias Macedo and Apolinario 2014), and entertainment (Wang et al. 2023b). In these domains, the conceptual foundation of the experience depends on enabling users to feel physically present without actual spatial displacement. This reliance on 'experiential immersion' makes the nuanced relationship between technical visual parameters and users' subjective responses particularly critical.

Technologies such as virtual reality (VR) and mixed reality (MR) enable users to engage in fully immersive or hybrid digital-physical spaces, enabling multisensory interactions and contextual engagement (Milgram and Kishino 1994; Slater and Wilbur 1997). Within these human-centred construct systems, four experiential dimensions frequently emerge as the key aspects of IVE experiences: perception, realism, immersion, and presence (Diemer et al. 2015; Jung and Lindeman 2021; van der Land et al. 2013). These constructs serve as conceptual lenses through which researchers

✉ Ju-Hyung Kim
kcr97jhk@hanyang.ac.kr

[1] Department of Architectural Engineering, Hanyang University, Seoul, Republic of Korea

[2] Department of Architectural Engineering, Kangwon National University, Chuncheon-si, Republic of Korea

Published online: 12 August 2026

understand how users interpret, emotionally engage with, and mentally represent virtual content. Although conceptually distinct, they are also deeply interrelated: perception initiates the user's interpretation of sensory input; realism evaluates the plausibility and coherence of the environment; immersion is often characterised by the objective sensorimotor contingencies of the system (Slater 2009), but it is also occasionally interpreted as psychological absorption (Witmer and Singer 1998); and presence captures the subjective sense of 'being there'. However, these constructs often exhibit overlapping and reciprocal influences, complicating efforts to delineate their relationships with system-level features.

Although a wide range of factors, including visual, auditory, and haptic stimuli, can influence user perception, realism, immersion, and presence in IVEs, visual input plays a particularly dominant role (Mohebbi et al. 2022). This is largely because of the reliance on head-mounted displays (HMDs) in most immersive systems, which place vision at the forefront of the user experience. Among the many design variables influencing these experiences, visual quality, which includes resolution, lighting, rendering techniques, and material realism, has been consistently cited as the primary determinant (McMahan et al. 2012; Mizuho et al. 2023). However, the existing literature has several critical limitations. First, research often assumes a direct correlation between higher visual quality and better user experience outcomes, without clearly specifying which experiential dimensions are affected or through what mechanisms. Second, studies are often application-specific and fragmented across domains such as gaming, architecture, and therapy, thus limiting the generalizability of the findings. Third, most analyses isolate visual quality as a technical variable, neglecting its perceptual and emotional implications, thereby reducing the complex experiential construct to merely a hardware setting.

These limitations indicate a broader problem: the need for an integrative, user-centred framework that connects visual quality features with psychological user experience constructs across diverse application contexts. A more nuanced and critical synthesis is required that considers both bottom-up sensory inputs—such as resolution, colour, and motion—and top-down cognitive and emotional processes—such as narrative framing, expectation, and familiarity. This perspective is essential for advancing IVE research not only towards higher realism but also towards greater experiential effectiveness.

To address these gaps, this review systematically analyses the influence of different types of visual quality features on users' experiences in IVEs, with a specific focus on four interrelated constructs: perception, realism, immersion, and presence. Rather than adopting a domain-specific or purely technical lens, this review draws on interdisciplinary studies to identify recurring mechanisms, methodological inconsistencies, and unresolved research challenges.

The following research questions guided this study:

RQ1: What is the relationship between visual quality and the experiential dimensions of perception, immersion, presence, and realism in IVEs?
RQ2: How do specific visual quality features influence perception, immersion, presence, and realism?
RQ3: To what extent have recent studies examined the role of visual quality in shaping perception, immersion, presence, and realism?

This review synthesises insights from both the technical and experiential domains to provide an evidence-based framework for optimising visual quality in immersive systems, supporting technological development as well as meaningful user engagement and behavioural outcomes in IVEs.

2 Visual quality and experiential factors

2.1 Perception

Perception is a complex higher-order process that extends beyond basic sensations by integrating, filtering, organising, and interpreting sensory inputs to construct subjective conscious experiences (Goldstein 2007). Perception is the fundamental mechanism by which users interpret and interact with digital content in IVEs. The perceptual process involves the transformation of external stimuli into neural signals, their subsequent interpretation by the brain, and behavioural responses that can affect further perception and interaction with the environment. Active sensorimotor interactions, such as hand tracking and haptic feedback, may further reinforce the users' sense of agency and embodiment. This enhances the illusion of perception in virtual environments.

Human perception is fundamentally multimodal, utilising vision, hearing, touch, proprioception, vestibular sensation, smell, and taste, which together shape experiences in both real and virtual worlds (Jerald 2016). Individual differences, such as age, sensory acuity, cognitive capacity, and prior experience with immersive technologies, substantially modulate users' perception, integration, and adaptation to multisensory information in IVEs (Mostafavi et al. 2023).

Visual perception, which depends on the activity of the cone and rod cells in the retina, enables the recognition of spatial relationships, colour, and object details. In IVEs, the content that users view through an HMD considerably impacts their perceptual experience, with parameters such

as resolution and field of view (FoV) critically affecting spatial awareness and immersion. These system-defined visual parameters shape the visual fidelity, perceived realism, and similarity between the virtual scene and the real world, serving as key mediators between the virtual input and the user's perceptual interpretation. Visual fidelity is determined by factors such as resolution, FoV, lighting, and material properties, which directly influence users' perceptions of spatial depth, object solidity, and environmental coherence (Bertheaux et al. 2024; Cheng et al. 2023; Luebke et al. 2002). Visual fidelity in IVEs can be analysed from both computational and rendering resolutions, shading (Luebke et al. 2002; Slater and Wilbur 1997), and architectural perspectives such as lighting and material properties (Bertheaux et al. 2024; Cheng et al. 2023).

Although vision dominates perception, multimodal integration further enriches the immersive experience. For example, auditory, tactile, and proprioceptive cues provide spatialised sounds, tactile feedback, and body perception, which can make people feel immersed. Although these additional cues generally enhance immersion, a mismatch between the visual and vestibular signals can lead to discomfort or motion sickness. To mitigate perceptual conflicts and cybersickness, designers often employ adaptive rendering strategies, comfort settings, or gradual exposure protocols (Cummings and Bailenson 2016; Jerald 2016).

Additionally, although less frequently addressed in VR, olfactory and gustatory cues can enhance the sense of realism and presence when integrated with other modalities (Jerald 2016). The integration and synchronisation of these sensory channels, as well as multisensory coherence, are crucial for creating convincing and comfortable virtual experiences (Cummings and Bailenson 2016). Therefore, designers should consider not only the quality of each sensory input but also its harmonious alignment to maximise realism and minimise perceptual dissonance (Mühlberger et al. 2001).

Perception in IVEs is not a passive process; instead, it is an active, self-justifying experiential state in which users perceive virtual objects and spaces as existing independently of their viewpoints, through the coordination of multiple perspectives (Church 2013). Notably, both bottom-up (sensory-driven) and top-down (cognitive-driven) processes shape perceptual experiences. Narrative context or prior knowledge can enhance perceived realism and presence, even when objective visual fidelity remains unchanged (Gorini et al. 2011). However, excessive visual detail or strong contrast may disrupt immersion by drawing attention to rendering artefacts or creating an unnatural emphasis (Mišiak et al. 2023; Nehmé et al. 2020). Perceptual harmony and multisensory alignment are often more important than maximising individual sensory fidelity, as mismatches can quickly lead to discomfort, perceptual dissonance, or cybersickness.

2.2 Realism

Building on the foundational role of perception, realism is a higher-order experiential quality that emerges when a virtual environment is perceived as credible, coherent, and consistent with the physical world. MR inherently integrates VR with the physical world, creating a unified experience. In most cases, the ideal user experience is one in which both domains merge into a single perceived reality, rendering it difficult for users to distinguish between the real and virtual elements. Similarly, in MR-based telepresence applications, reconstructed 3D avatars should be seamlessly integrated among themselves and with the surrounding IVEs. Achieving this level of realism typically requires maximising photorealistic rendering and reconstructing environmental elements, objects, and human representations.

Although visual fidelity and high-resolution reconstruction are often considered prerequisites for realism, research suggests that perceived realism is not solely contingent on graphical accuracy. Gilbert (2017) introduced the concept of authenticity, defined as the degree to which a virtual environment aligns with users' Bayesian expectations and situational goals. From this perspective, even low-fidelity environments can be perceived as highly realistic if they align with perceptual priors and suppress extraneous or distracting information. Thus, authenticity provides a cognitive expectation-based complement to immersion, offering a richer model for explaining presence and realism in IVEs.

Realism in IVEs can be categorised into three distinct dimensions: conceptual, felt, and perceptual realism. Conceptual realism refers to the cognitive evaluation of an environment's plausibility, logical consistency, and coherence with real-world events, requiring a deliberate assessment of its content. Felt realism is an intuitive and immersive experience in which users feel physically present, emotionally engaged, and captivated by their surroundings. This type of realism aligns with the illusion of non-mediation and represents a non-effortful perception (Type 1 realism). By contrast, perceptual realism concerns the conscious evaluation of visual fidelity, including the realism of textures, virtual characters, and environmental details, and requires an effortful assessment of form (Type 2 realism). This classification highlights that realism is not solely dependent on high-fidelity visuals but is also shaped by coherence, consistency, and emotional engagement, all of which contribute to the user's overall perception of realism (Christophers et al. 2024).

2.3 Immersion

Realism addresses the perceived authenticity and coherence of virtual environments, whereas immersion focuses on the degree to which users are perceptually and sensorially enveloped by their experiences. Overall, these constitute complementary dimensions of presence in IVEs. Immersion refers to the objective extent to which a VR system projects stimuli onto a user's sensory receptors in a manner that is extensive, matching the surroundings, vivid, interactive, and narratively engaging (Jerald 2016). Although immersion is predominantly defined as an objective system property in the computer science tradition (Slater 2004), psychological perspectives (Witmer and Singer 1998) conceptualise it as a subjective state of absorption. It can be enhanced through technological features, such as spatial audio, 3D vision, FoV, and head tracking, all of which could contribute to a stronger sense of immersion in a VE (Potter et al. 2022).

Immersion is a complex and multidimensional phenomenon involving several layers of neuropsychological engagement, including perception, attention, and emotion. Among various definitions of immersion, one of the most widely cited is that of Janet Murray, who characterised it as 'the pleasurable experience of being transported to an elaborately simulated place' and 'the sensation of being surrounded by a completely other reality that takes over all of our attention and our whole perceptual apparatus' (as cited in Zhang et al. 2017, p. 1).

Several efforts have been made to classify the types of immersion, which can be broadly divided into embodied immersion and emotional immersion. This distinction is based on the mediated nature of the self-other relationship, whether the user primarily experiences a sense of *physical presence* within the IVE or engages cognitively and emotionally with a virtual 'other' (Sherman and Craig 2004).

Embodied immersion emphasises physical perception and spatial presence, enabling users to 'enter' a virtual world through sensory and motor integration (Civitarese 2008). By contrast, emotional immersion focuses on cognition, imagination, and emotional engagement, highlighting the user's affective projection and psychological resonance with virtual characters or narratives (Ermi and Mäyrä 2005; Sherman and Craig 2004).

Regardless of the type of immersion, its realisation depends not only on technical means but also on the user's psychological commitment and active participation. Although technological capabilities are essential, fostering user engagement is equally important for achieving immersive experiences. Presence emerges when immersion is cognitively and emotionally internalised by the user. Immersion alone cannot determine a user's subjective psychological response; instead, it is the user's perception and interpretation of immersive stimuli that produce presence (Jerald 2016). Thus, although immersion is a prerequisite for compelling VR experiences, presence—the subjective sense of 'being there'—completes the immersive illusion.

2.4 Presence

Presence has been defined as a psychological state or subjective perception in which, although all or part of one's current experience is generated or mediated by technology, an individual's awareness of this mediation is diminished or absent. Nonetheless, the VR research community continues to seek a precise definition, as clarity is essential for designing effective immersive experiences (Jerald 2016).

In IVEs, presence represents the extent to which users feel physically or psychologically situated within a virtual environment (Botella et al. 1999; Schubert et al. 2001; Slater and Wilbur 1997). This sense of 'being there' connects VR experiences to emotional engagement and is thus regarded as a multidimensional construct.

Over the past several decades, the definitions of presence have evolved significantly. Lombard and Ditton (1997) described it as a perceptual illusion of non-mediation, whereas Slater & Wilbur (1997) referred to it as the experience of being engaged with representations of a virtual world. More recent literature defines it as the feeling of being in a perceptible external world that surrounds the self (Schuemie et al. 2001; Seo et al. 2017).

Theories of presence are typically divided into descriptive and structural models. Descriptive models categorise presence into dimensions such as spatial presence, social presence, and self-presence, emphasising the external and system-related factors that influence the experience. These models are typically assessed using subjective questionnaires, such as the Igroup Presence Questionnaire, which provides qualitative insights into factors such as image quality and FoV that contribute to a sense of presence (SoP; Gilbert 2017). By contrast, structural models explore the internal cognitive mechanisms that enable the experience of presence. For example, directing attention towards the IVE (Witmer and Singer 1998) and forming its coherent mental representations (Sheridan 1999) are essential processes for generating a SoP. In addition, recent research has explored the use of physiological signals, such as heart rate, electrodermal activity, and electroencephalography (EEG), for presence assessment (Egan et al. 2016).

2.5 Summary

In IVEs, perception initiates user interaction by transforming multisensory stimuli into structured experiential inputs. Realism acts as an interpretive layer that enables users to

evaluate the plausibility and coherence of an environment relative to their expectations. Immersion reflects the system's objective technical capacity to generate rich and encompassing sensory inputs across modalities. By contrast, presence emerges as the subjective state of 'being there', shaped by the dynamic interplay between sensory fidelity, environmental authenticity, emotional resonance, and narrative engagement. Except for immersion, which reflects an objective system-level property, perception, realism, and presence are inherently subjective constructs that vary across individuals and contexts, and may, in practice, arise simultaneously rather than sequentially. Although analytically distinct, these four constructs are functionally interdependent, with each influencing and reinforcing the others in the construction of an immersive experience (see Fig. 1). The organisation of Sect. 4 follows these analytical distinctions, though cross-construct influences are noted where empirically supported.

3 Materials and methods

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines were followed to ensure transparent and comprehensive reporting (Page et al. 2021). PRISMA provides a standardised checklist to improve the quality and reproducibility of systematic reviews and meta-analyses.

A comprehensive search was conducted to identify the most relevant studies that explicitly examined the relationships between visual quality and presence, perception, realism, and immersion. These studies were sourced from multiple academic repositories, including scientific journals, conference proceedings, and relevant books.

The methodology for selecting relevant studies involved three key steps: (1) identifying appropriate databases for the search, (2) defining search queries, and (3) incorporating additional studies from other sources.

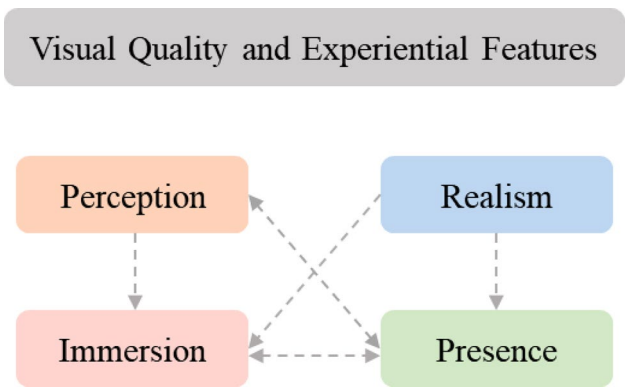

Fig. 1 Relationship between perception, realism, immersion, and presence

In the first step, databases indexed by the Web of Science, IEEE Xplore, and Scopus were used. The initial search retrieved 106 papers from Web of Science, 171 papers from IEEE Xplore, and 1780 papers from Scopus. After removing 239 duplicate entries, 1818 unique papers remained for further screening. The systematic literature search process is illustrated in Fig. 2.

The search was conducted using the following keyword combinations:

1. ('immersive virtual environment' OR 'virtual reality' OR 'mixed reality') AND
2. ('visual quality') AND
3. ('perception' OR 'immersion' OR 'presence' OR 'realism').

Subsequently, 126 papers written in languages other than English and 113 papers published before 2014 were excluded to ensure that the remaining studies covered the period from 2014 to 2025. Further filtering removed 110 papers that were not peer-reviewed journal articles, resulting in 1490 papers.

Title-based screening was conducted to exclude studies outside the research scope. The selection was further refined by assessing the relevance of keywords and filtering papers that were initially conference proceedings but were subsequently reclassified as journal articles. After this process, 138 papers were retained for final review. Finally, two papers that could not be accessed and 46 papers that were outside the scope of the research were excluded, resulting in a final set of 90 papers for analysis.

An analysis of the 90 included studies reveals the following temporal distribution. Figure 3 illustrates the annual distribution from 2014 to 2025: publication volume remained relatively low in the early period (2014–2018, 3–6 papers per year), followed by a gradual increase from 2019 onwards. A notable surge was observed from 2022, with the number of publications peaking at 15 in 2024, reflecting growing scholarly interest in visual quality and immersive experience research in recent years.

4 Relationships identified with a conventional manual review

This section categorises visual quality into three major dimensions: modelling, hardware and rendering, and motion and interaction. Each of these dimensions contributes to a user's immersive experience in IVEs across four experiential constructs: perception, realism, immersion, and presence. Although these constructs are discussed independently for clarity, they are interrelated. Improvements in visual quality

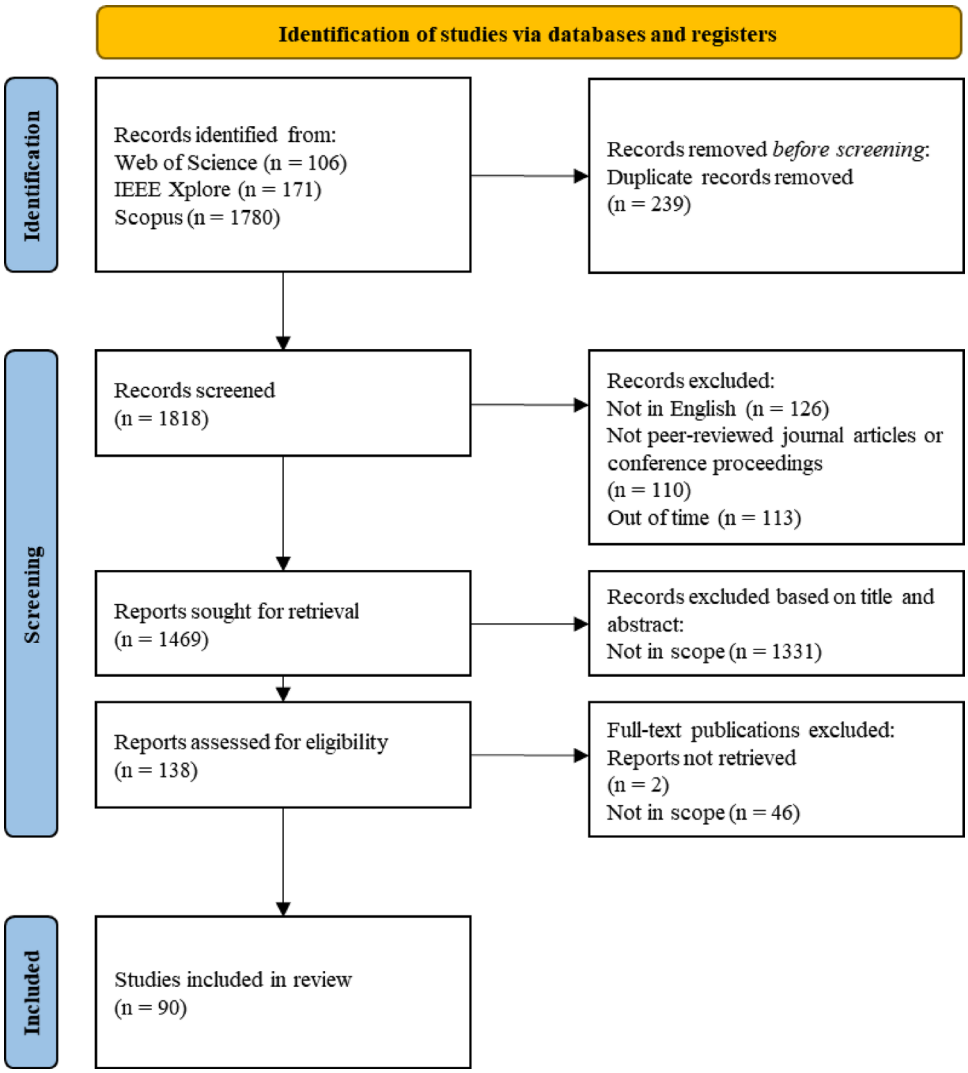

Fig. 2 PRISMA flow diagram

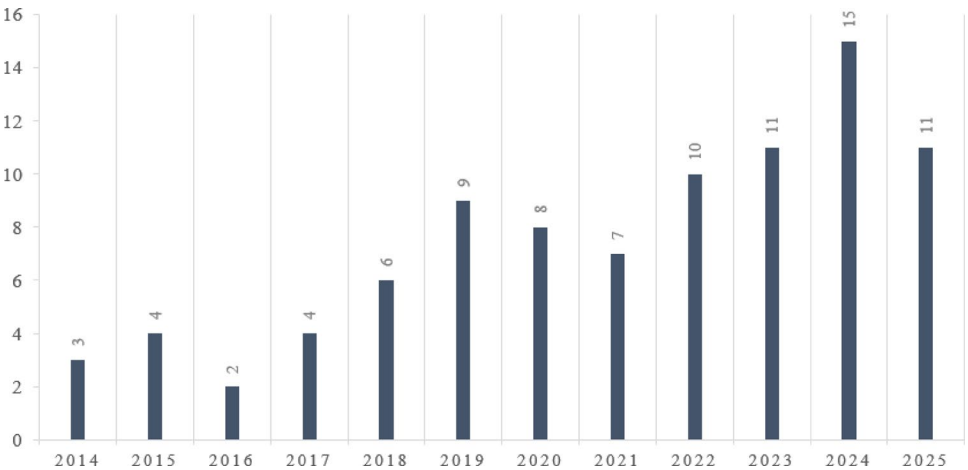

Fig. 3 Annual distribution of included studies (2014–2025)

typically propagate through the perceptual, cognitive, emotional, and spatial layers of experience. A detailed mapping of these relationships is presented in Fig. 4 and Tables 1, 2, and 3, which summarise the influence of each visual quality dimension on the four experiential constructs.

4.1 Modelling

4.1.1 Impact on perception

Many factors were included in the modelling segment. This review includes level of detail (LOD; Baigabulov and Ipalakova 2023; Hanel et al. 2014; Hu et al. 2018; Petikam et al. 2018), material properties (Bertheaux et al. 2024; Niu and Lo 2022), and spatial design (Luigi et al. 2015; Wang and Munakata 2024). Research has consistently shown that LOD directly shapes users' experience of virtual spaces: higher LOD enhances perceptual clarity and spatial understanding, whereas aggressive reduction leads to decreased realism and spatial awareness (Baigabulov and Ipalakova 2023; Hanel et al. 2014; Petikam et al. 2018). However, in MR applications, user perception studies have shown that natural visual compositions can be achieved using low-resolution depth maps (Petikam et al. 2018). Low LOD is sufficient to maintain visual perception, enabling coarse 3D geometry modelling to replace high-fidelity depth data without degrading visual quality. We can assume that the material properties also determine the LOD. Therefore, understanding the perceptions of extended reality (XR) users is crucial.

The LOD determines the shape and accuracy of an object, whereas the material determines the light, colour, and realism of the surface. Material properties such as smoothness, metalness, and specularity may significantly impact perception in IVEs by shaping the interaction of surfaces with light. Smoothness enhances surface polishing and clarity, whereas metalness increases environmental reflectivity, rendering objects appear more realistic (Bertheaux et al. 2024). Furthermore, variations in these material properties can lead to different emotional reactions: stronger specular reflections are associated with higher valence and arousal, whereas low metallurgy combined with high roughness is perceived as unnatural, occasionally evoking a slight sense of disgust (Mišiak et al. 2023). Consequently, material properties are perceived with greater accuracy and precision in VR environments than in 2D environments, particularly when evaluating metal roughness (Niu and Lo 2022).

Although object modelling is essential for conveying local realism, the overall composition of the virtual environment plays a crucial role in shaping users' perceptions. In addition to the realism of individual objects, the spatial arrangement and visual coherence of the environment influence users' senses of authenticity and immersion. Well-balanced

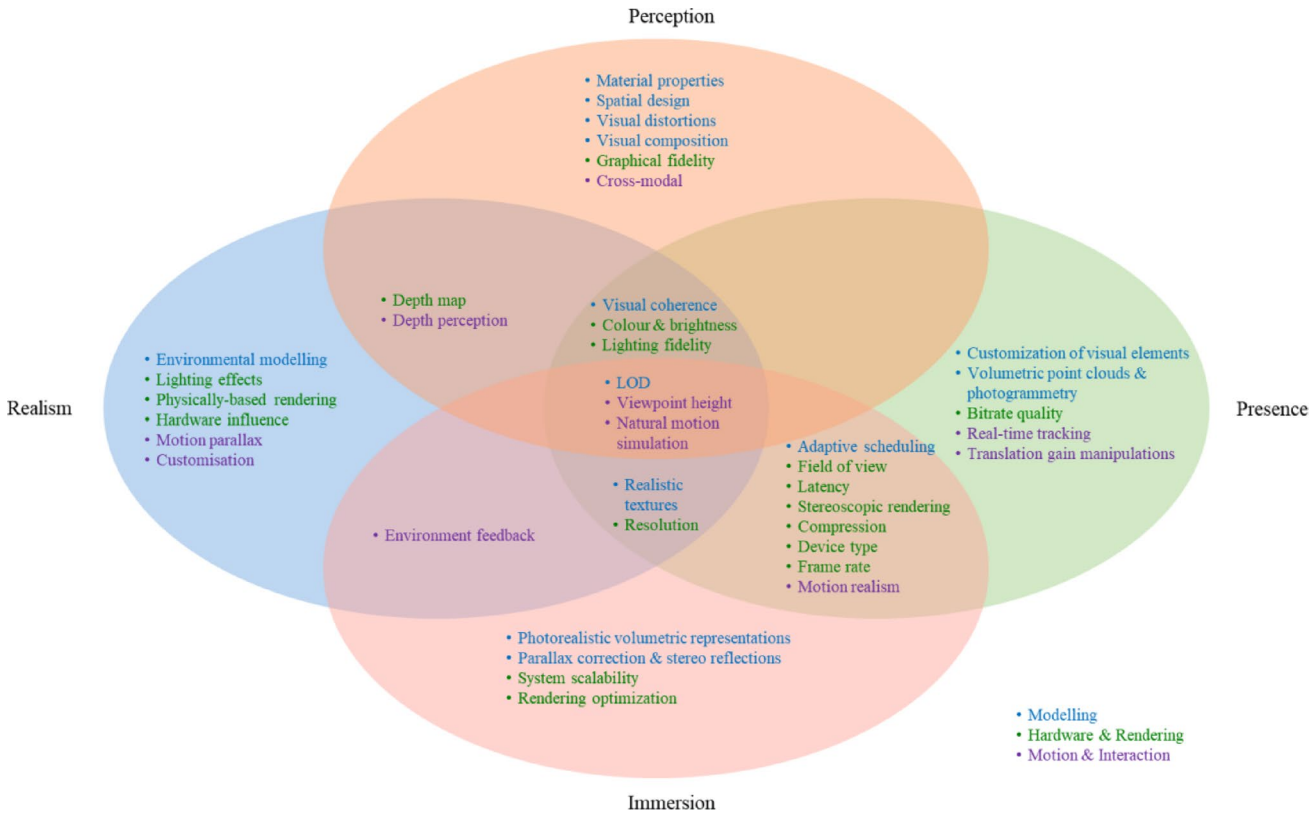

Fig. 4 Visual quality factors influencing perception, realism, immersion, and presence in virtual environments

Table 1 Modelling factors and their influence on perception, realism, immersion, and presence

Modelling factors	Perception	Realism	Immersion	Presence	Reference(s)
Level of detail	↑↓	↑	↑	↑	Baigabulov and Ipalakova (2023); Hanel et al. (2014); Hu et al. (2018); Petikam et al. (2018)
Material properties	↑↓				Bertheaux et al. (2024); Niu and Lo (2022)
Spatial design	↑				Li et al. (2022); Luo et al. (2022); Luigi et al. (2015); Wang and Munakata (2024); Zou et al. (2019)
Visual distortions	↓				Alfarasani et al. (2023); Stylidis et al. (2019)
Visual composition	↑↓				Wang et al. (2024b); Zhu et al. (2025)
Visual coherence	↑	↑		↑	Regenbrecht et al. (2017)
Environmental modelling		↑			Ackerman et al. (2022); Checa et al. (2021); Langa et al. (2022); Rhee et al. (2020); Zambrano et al. (2024)
Realistic textures		↑	↑↓	↑	Li et al. (2021); Mizuho et al. (2023); Warsinke et al. (2025)
Photorealistic volumetric representations			↑		Li et al. (2021); Menz et al. (2025)
Parallax correction and stereo reflections			↑↓		Mišiak et al. (2023)
Adaptive scheduling			↑	↑	Mekuria et al. (2015)
Customisation of visual elements				↑	Kapralos et al. (2014); Marino et al. (2021)
Volumetric point clouds and photogrammetry				↑↓	Nehmé et al. (2020)

↑ = positive effect; ↓ = negative effect; ↑↓ = mixed or context-dependent effects

compositions of vegetation, water, and buildings can enhance the sense of realism, whereas poorly designed spatial layouts may lead to a perception of unnaturalness (Wang and Munakata 2024). In addition to visual composition, the spatial proportions of architectural elements play a critical role in shaping perceptions. Key factors such as the window-to-wall ratio and floor area ratio shape the perceived openness or confinement of a space, thereby affecting users' comfort and perception of spatial scale. In addition, room-fusion techniques that seamlessly integrate different virtual spaces can further enhance perception by creating coherent and immersive environments. IVR has proven to be effective in replicating real-world sensory experiences, rendering it a valid tool for multisensory perception assessment in urban design contexts (Li et al. 2022; Luo et al. 2022; Luigi et al. 2015). Additionally, different visual components (such as greenery, architecture, water, and sky) can influence the user's subjective perception, including aesthetics, pleasure, sense of history, and cultural identity (Wang et al. 2024b). This aligns with the findings of Zhu et al. (2025), who demonstrated that spatial perception is more strongly influenced by ambient environmental context than by individual structural features. Although variations in frame geometry showed no significant effect on depth perception, the presence of peripheral visual information, such as surrounding spatial context and architectural cues, significantly enhanced perceived depth. This suggests that spatial perception in immersive environments is not solely determined by focal visual content, but is critically shaped by the broader contextual information available to the observer.

Nevertheless, in a virtual environment, some differences from reality will exist, which can be referred to as visual distortions. Visual distortions, such as noise and smoothing artefacts, can negatively impact perception by reducing the visibility of fine details (Alfarasani et al. 2023). In VR environments, users tend to be more sensitive to such distortions; as distortion levels increase, the perceived visual quality decreases accordingly. Noise artefacts are particularly noticeable on smooth surfaces, where they disrupt visual uniformity and reduce the perceived realism of a scene. Therefore, effective noise reduction and surface-smoothing

Table 2 Hardware and rendering factors and their influence on perception, realism, immersion, and presence

Hardware and rendering factors	Perception	Realism	Immersion	Presence	Reference(s)
Graphical fidelity	↑↓				Bahirat et al. (2018); Cheng et al. (2023); Debattista et al. (2017); Deng et al. (2022); Henriques et al. (2025); Jin et al. (2021); Ma et al. (2024); Mišiak et al. (2023); Mozhaeva et al. (2025); Waldow et al. (2024); Wang et al. (2023c); Wang et al. (2024c); Yamamoto et al. (2016); Zeng and Zhao (2023)
Depth map	↑	↑			Petikam et al. (2018); Zou et al. (2018)
Colour and brightness	↑↓	↑		↑	Abd-Alhamid et al. (2019); Chen et al. (2024); Costalunga et al. (2025); He et al. (2023); Kong et al. (2023); Mazikowski (2019); Mostafavi et al. (2023); Prachyabrued and Borst (2015); Wang et al. (2025)
Lighting fidelity	↑	↑		↑	Abd-Alhamid et al. (2019); Zhdanov et al. (2019)
Lighting effects		↑			Bugeja et al. (2014); Cheng et al. (2023); Mandala et al. (2025); Maniee et al. (2025); Xu et al. (2025)
Physically based rendering		↑			Rossoni et al. (2024)
Hardware influence		↑			Li et al. (2021)
Resolution		↑	↑	↑	Hameed et al. (2024); Chu et al. (2017); Elwardy et al. (2020); Li et al. (2021); Petikam et al. (2018); Pettersson et al. (2019); Zou et al. (2018); Zou et al. (2020)
System scalability			↑		Stotko et al. (2019)
Rendering optimisation			↑		Ma et al. (2024); Stotko et al. (2019)
Field of view			↑	↑	Hu et al. (2018); Zou et al. (2018)
Latency			↓	↓	Cheng et al. (2024); Cortés et al. (2024); Desai et al. (2017); Fei et al. (2019); Salmimaa et al. (2018); Stotko et al. (2019)
Stereoscopic rendering			↑	↑	Cortés et al. (2024); Salmimaa et al. (2018)
Compression			↓	↓	Salmimaa et al. (2018); Selzer et al. (2021)
Device type			↑↓	↑↓	Kim et al. (2020b)
Frame rate			↑	↑	Ma et al. (2024); Zou et al. (2018)
Bitrate quality				↑	Zou et al. (2018)

↑=positive effect; ↓=negative effect; ↑↓=mixed or context-dependent effects

techniques are essential for preserving visual clarity and enhancing the overall user experience in IVEs.

Finally, in the context of VR- and augmented reality (AR)-based product evaluation, perceived quality is widely adopted as a subjective indicator to assess user perceptions of various visual and geometric attributes. Stylidis et al. (2019) demonstrated that perceived quality-based metrics effectively capture perceptual differences related to display modality, resolution, and interactivity, thereby offering a useful tool for evaluating visual performance in immersive systems.

Table 3 Motion and interaction factors and their influence on perception, realism, immersion, and presence

Motion and interaction factors	Perception	Realism	Immersion	Presence	Reference(s)
Viewpoint height	↑↓	↑↓	↑↓	↑↓	Feng et al. (2023); Luan et al. (2024)
Natural motion simulation	↑	↑	↑	↑	Cho et al. (2025); Ingabo et al. (2024)
Cross-modal	↑				Allue et al. (2016); Malpica et al. (2020); Robotham et al. (2022)
Depth perception	↑	↑			Feng et al. (2023); Padilha and Teichrieb (2014); Padilha and Teichrieb (2015); Robotham et al. (2022)
Motion parallax		↑			Waldow et al. (2024)
Customisation		↑			Checa et al. (2021)
Environment feedback		↑	↑		Christophers et al. (2024); Luigi et al. (2015)
Motion realism			↑	↑	Calvi et al. (2017); Thompson et al. (2019)
Real-time tracking				↑	Regenbrecht et al. (2017)
Translation gain manipulations				↑↓	Selzer et al. (2022)

↑ = positive effect; ↓ = negative effect; ↑↓ = mixed or context-dependent effects

4.1.2 Impact on realism

Environmental modelling and scene construction quality significantly influence realism in both virtual and MR environments. Photorealistic elements, such as high-quality 3D models, detailed textures, and advanced lighting effects, significantly enhance the perception of realism (Checa et al. 2021). For instance, virtual forest environments that incorporate real-world geospatial data, detailed tree models, and dynamic environmental effects, such as wind and natural lighting, demonstrate heightened SoP and ecological authenticity (Ackerman et al. 2022). Similarly, in virtual museum settings, the integration of realistic 3D modelling, spatial audio, and interactive elements (such as avatars and dialogues) can improve user immersion and the perception of cultural authenticity (Zambrano et al. 2024).

Although photorealistic modelling and lighting significantly contribute to perceived realism, high graphical fidelity is not always a prerequisite. As Regenbrecht et al. (2017) demonstrated, realism can still be achieved in low-fidelity environments when visual and behavioural coherence are preserved. Low-fidelity styles, such as voxel-based simulations, can evoke a strong sense of realism and embodiment if the lighting, material responses, and interaction feedback remain internally consistent. Features such as appropriately scaled voxel sizes, simulated sensor noise, unified colour and material rendering, and perceptual anchors, including virtual mirrors, contribute to a spatially coherent experience. These findings underscore that the alignment between system design and user expectations—not visual detail alone—drives perceived authenticity in virtual environments. Corroborating this, Warsinke et al. (2025) demonstrated that reduced geometric detail did not significantly affect perceived realism in VR, whereas texture resolution remained the dominant factor (partial $\eta^2 = 0.709$), reinforcing the view that surface coherence, rather than structural complexity, drives realism judgements.

Realism in IVE modelling arises not only from visual fidelity but also from multilevel coherence, which encompasses the alignment of spatial structure, lighting logic, and user expectations. Although the virtual world maintains a consistent visual style, even with low graphical fidelity, users may still feel a sense of realism. By contrast, a photorealistic scene with inconsistencies in spatial or lighting logic may undermine the overall sense of realism.

4.1.3 Impact on immersion

In IVEs, high-quality 3D models, such as texture mapping, can significantly enhance user immersion, particularly in cultural or museum settings, where detailed models deepen user engagement and contextual understanding (Zambrano et al. 2024). This suggests that immersion is closely tied not only to visual richness but also to the fidelity with which visual elements convey meaning and support user tasks.

Although low video resolution significantly reduces immersion, increasing the resolution beyond a certain point yields diminishing returns (Potter et al. 2022). By contrast, spatial audio enhancements, particularly those involving head-tracked reverberation, continue to improve immersion even under low visual fidelity, highlighting the importance of cross-modal sensory compensation in immersive design.

In addition, the LOD models and adaptive scheduling techniques improve both perception and presence by maintaining frame rates while enhancing visual quality, which is crucial for reducing motion sickness and sustaining immersion (Hu et al. 2018). These methods highlight the importance of performance-aware visual quality strategies that dynamically balance load and realism. Similarly, photorealistic volumetric renderings can strengthen co-presence and overall immersion by enhancing environmental authenticity, particularly in collaborative or social IVEs, where spatial coherence is key (Li et al. 2021). In this context, encoding parameters such as texture resolution have been shown to substantially affect the perceived quality of volumetric content (Menz et al. 2025).

Although specific features, such as parallax correction and stereo reflections, can improve perceptual realism, their direct influence on immersion appears more nuanced and may depend on the interaction context or the user's familiarity with immersive technologies (Mišiak et al. 2023). This indicates that visual immersion arises not only from isolated graphical features but also from the integrated experience of coherent, responsive, and perceptually aligned environments.

4.1.4 Impact on presence

The quality of 3D models significantly influences the presence of IVEs (Hanel et al. 2014; Hu et al. 2018; Petikam et al. 2018), visual coherence (Regenbrecht et al. 2017), and the realism of elements (Mekuria et al. 2015). High visual fidelity can significantly enhance presence and influence memory performance (Mizuho et al. 2023). Studies have indicated that high-fidelity 3D models incorporating photorealistic representations, detailed textures, and optimised LOD enhance spatial presence and co-presence (Ackerman et al. 2022; Langa et al. 2022; Rhee et al. 2020). Spatial presence can be evaluated more scientifically by first enabling participants to experience the physical environment and then assessing the corresponding virtual environment (Zou et al. 2019). However, the visual coherence between virtual and physical elements may be more crucial than photorealism. For instance, consistent voxel environments and interactive elements, although rendered at a lower fidelity, can still sustain a strong SoP, particularly when perceptual alignment and task relevance are maintained (Regenbrecht et al. 2017).

In addition, LOD models and adaptive scheduling help maintain frame rates, enhance details, and support presence by ensuring smooth interactions (Mekuria et al. 2015). Customising visual elements such as brightness, colour schemes, and shading consistency also influences presence by reinforcing environmental beliefs (Kapralos et al. 2014; Marino et al. 2021). However, visual distortions, such as unnatural mesh simplifications, geometric deformations, and colour distortions, negatively impact presence perception (Nehmé et al. 2020).

Presence and co-presence can also be achieved when users engage in shared tasks or communicate within the same virtual space. As Li et al. (2021) highlighted, volumetric point clouds and photogrammetric environments allow for real-time visual updates from other users, enabling the display of realistic body shapes, clothing, and motion. This enhances the sense of social and spatial presence by fostering mutual recognition and embodied interaction. Their study also reported that, although users could clearly see others' volumetric representations, their own avatars often appeared blurry or less detailed. This asymmetry in self-versus-other representations can undermine self-presence and affect overall perception.

4.2 Hardware and rendering

4.2.1 Impact on perception

In IVEs, hardware capabilities and rendering techniques play critical roles in shaping user perceptions. The level of visual detail significantly affects how users experience virtual spaces (Bahirat et al. 2018; Cheng et al. 2023; Debattista et al. 2017; Deng et al. 2022; Mišiak et al. 2023; Waldow et al. 2024; Wang et al. 2023c; Yamamoto et al. 2016; Zeng and Zhao 2023). For example, Debattista et al. (2017) categorised visual quality into low, medium, and high-fidelity levels based on the complexity of lighting simulations. Their findings revealed that although higher-fidelity environments were consistently perceived as more realistic than low-fidelity ones, the difference in user experience between the medium- and high-fidelity conditions was not significant. This outcome suggests a point of perceptual saturation, where further improvements in graphical realism no longer yield corresponding gains in perceived quality. Nonetheless, lighting effects, realistic reflections, and dynamic shadows remain essential for enhancing depth cues and spatial coherence in IVEs.

Although fidelity directly affects perception, recent studies have emphasised that optimising the distribution of visual resources based on human vision may be more effective than uniform enhancement. Deng et al. (2022) and Jin et al. (2021) explored how rendering strategies informed by the human visual system can optimise performance without compromising perceptual realism. By leveraging the natural acuity fall-off between the foveal and peripheral visual fields, their proposed foveated rendering system delivers high-resolution imagery only in the user's gaze region. This gaze-contingent rendering significantly reduces computational latency, achieving near-ground-truth perceptual

quality at a minimal cost. Building on this principle, recent work has further extended perceptually-informed rendering into new domains. Mozhaeva et al. (2025) demonstrated that incorporating spatio-temporal models of peripheral vision into wide-angle VR video transmission can reduce bitrate by 60% without perceptual quality loss. Wang et al. (2024c) integrated visual acuity and contrast sensitivity models into neural radiance field rendering, achieving approximately 83 FPS while surpassing existing methods in both foveal and peripheral salient region quality. Overall, these studies confirm that aligning rendering strategies with human visual sensitivity, rather than uniformly enhancing fidelity, offers a more efficient pathway to high-quality immersive experiences. Based on the same principle, parallel advances have carried this logic into new rendering domains: Henriques et al. (2025) demonstrated that 3D Gaussian splatting can replace ray-traced peripheral vision in VR at significantly lower computational cost, achieving perceptually acceptable reconstruction of complex lighting effects; Ma et al. (2024) applied a similar logic to fluid simulation through ViP-Fluid, which allocates particle resolution based on both gaze proximity and dynamic visual salience, achieving up to 2.15 × speed-up over high-resolution baselines while maintaining perceptual quality. Beyond rendering strategies, perceptual alignment has also informed quality assessment methods. Kim et al. (2020a) proposed a method for deep learning-based VR image quality assessment, aiming to achieve a higher correlation with human perception in VR environments.

In addition to rendering efficiency, maintaining consistent depth cues is essential for perceptual stability in immersive environments. Inconsistent depth cues in VR, particularly the mismatch between motion parallax and binocular disparity, can affect users' perception. Motion parallax enhances the sense of three-dimensionality; however, its impact on visual errors depends on the structure and depth complexity of the object. A negative correlation was found between perceived depth and visual error, suggesting that stronger depth perception is associated with fewer distortions. However, objects with a high occlusion potential, such as a space shuttle, showed increased visual errors when the motion parallax was stronger. This highlights the importance of an object-specific design in maintaining perceptual coherence (Waldow et al. 2024).

Despite these challenges, lighting and shadow techniques remain indispensable tools for enriching depth cues and environmental realism, thereby enhancing spatial perception. Transparent surfaces and dynamic shadows further enrich immersion by providing realistic depth cues. Even when the environment lacks graphical fidelity, proper lighting and rendering techniques can substantially improve visual perception and realism (Cheng et al. 2023; Debattista et al. 2017; Deng et al. 2022; Waldow et al. 2024; Wang et al. 2023c; Yamamoto et al. 2016).

Although these techniques are widely applied in VR environments, the accuracy of depth maps becomes even more critical in MR, where virtual content must align precisely with the physical world. In MR compositions, higher depth map resolutions yield more accurate geometric reconstructions of real-world environments, effectively reducing visual artefacts. This improved accuracy enhances the alignment between virtual and real-world elements, particularly in complex interactions such as shadows and occlusions. Conversely, lower depth resolutions introduce noticeable artefacts that disrupt spatial coherence and negatively impact perceived realism (Petikam et al. 2018). Therefore, high-resolution depth maps are crucial for enhancing user perception and fostering a more immersive MR experience (Zou et al. 2020).

Visual quality factors, such as colour consistency, luminance, and contrast, can directly influence spatial presence and realism in immersive cave automatic virtual environment systems (Mazikowski 2019). In addition, colour and brightness settings influence both perception and emotional responses (Chen et al. 2024; He et al. 2023; Mostafavi et al. 2023; Prachyabrued and Borst 2015). Increased brightness, colour temperature, and luminance distribution contrast improve object visibility and assist users in navigating virtual spaces more efficiently (Abd-Alhamid et al. 2019). Beyond luminance and hue, colour saturation has also been shown to shape perceptual expectations and sensory evaluations, and its effects vary systematically by context. For instance, in food-related MR environments, higher saturation enhances perceived quality and taste intensity while reducing the perceived naturalness of certain stimulus categories (Costalunga et al. 2025). Moreover, illumination levels and correlated colour temperature (CCT) are linked to specific EEG band features in the frontal and parietal brain regions, where cooler colours evoke calmness and warmer colours stimulate excitement (Mostafavi et al. 2023). Wang et al. (2025) further demonstrated that in urban nightscape VR environments, medium-to-low colour temperatures (3000–4000 K) significantly enhanced aesthetic appeal compared to high colour temperatures (> 5000 K), with eye-tracking data confirming stronger visual attention under warmer lighting conditions. By fine-tuning these visual elements, designers can shape user experiences and create IVEs that are both engaging and comfortable.

Collectively, these findings illustrate that hardware and rendering strategies should exceed fidelity and incorporate perceptual alignment, system responsiveness, and emotional resonance to fully support immersive user experiences in IVEs.

4.2.2 Impact on realism

Technical factors, such as resolution (Pettersson et al. 2019; Zou et al. 2018), rendering quality (Petikam et al. 2018; Rossoni et al. 2024; Zhdanov et al. 2019; Zou et al. 2018), and hardware performance (Li et al. 2021), play pivotal roles in shaping perceptual realism. Depth resolution directly affects the seamless integration of virtual and real elements. Low-resolution depth maps introduce visual artefacts that disrupt the perception of realism and presence, whereas high-resolution maps reduce these distortions, leading to enhanced immersion and believability (Petikam et al. 2018). Similarly, studies on video resolution and compression have shown that higher resolutions improve perceived quality (Elwardy et al. 2020). However, excessive compression can degrade visual fidelity and negatively affect realism.

Small discrepancies in reflection rendering are particularly noticeable to users, necessitating high-precision rendering for maintaining realism (Yamamoto et al. 2016). In AR systems, the deformation and magnification of the reflected images play significant roles in visual perception. Compared to magnification, observers are notably more sensitive to lateral distortions in the reflected images. However, when the reflected objects are arranged asymmetrically, sensitivity to visual inconsistency decreases significantly. This study found that users could tolerate approximately 20% variation in size owing to magnification, whereas the tolerance for lateral displacement was relatively narrower, at approximately 1°. Under asymmetrical arrangements, this tolerance range is doubled, indicating that such configurations help mask the perceptual inconsistencies caused by reflection mismatches. To enhance the realism of XR applications, point cloud models should be rendered using physically based rendering techniques that incorporate full surface properties such as normal, roughness, and metallicity, and respond dynamically to ambient lighting through image-based lighting and shading (Rossoni et al. 2024). Thus, virtual content can be visually integrated into a real or synthesised scene in a physically credible manner.

Beyond resolution and rendering fidelity, lighting accuracy plays an equally critical role in establishing visual realism, particularly in environments where virtual and real-world elements should be seamlessly merged. To seamlessly integrate virtual objects into real-world scenes, analysing the illumination conditions of the physical environment and replicating similar lighting parameters for the virtual content are crucial. By reconstructing accurate real-world lighting and capturing elements such as shadows, reflections, and diffused light, virtual objects can be rendered with greater consistency, thereby reducing perceptual discrepancies between the real and virtual layers. This enhances visual coherence and strengthens the sense of realism (Zhdanov et al. 2019).

At the rendering level, lighting fidelity plays a central role in realism by ensuring physically plausible light behaviour. Global illumination techniques simulate the natural behaviour of light, such as diffusion, colour bleeding, and dynamic shadows, thereby rendering virtual environments more natural and immersive (Bugeja et al. 2014). Synchronisation between virtual and real-world illumination enhances depth perception and shape recognition, further strengthening the sense of realism (Cheng et al. 2023). In addition, advanced volumetric rendering techniques contribute to this effect by enhancing the perception of depth and spatial structure. In immersive environments, inconsistencies such as unnatural brightness, the absence of shadows, or incorrect shadow directions can break spatial logic and severely undermine the user's sense of realism. For instance, Mandala et al. (2025) found that an OLED-based HMD could not fully replicate the spectral power distribution of natural daylight, resulting in perceptual differences in colour temperature and brightness between real and virtual environments, which affected users' overall visual experience. Empirical evidence supports this: Xu et al. (2025) found no significant perceptual differences between VR and real-world lighting across multiple scenarios, whereas Maniee et al. (2025) further showed that CCT variations significantly influenced perceived visual comfort, with medium CCT (5000 K) yielding the highest satisfaction rates, together underscoring the importance of photometric consistency in achieving perceptual realism.

In addition to rendering, the choice of hardware impacts realism. HMDs generally provide higher levels of immersion and realism than traditional 2D screens because they deliver a more comprehensive FoV and stereoscopic depth perception (Li et al. 2021). However, hardware limitations, such as low display resolution or the absence of realistic dynamic simulations, can significantly hinder the overall sense of realism, even in VR systems that otherwise offer a strong SoP (Chu et al. 2017). In certain cases, the lack of dynamic realism has been identified as a critical factor that negatively affects the user experience (Pettersson et al. 2019), suggesting that future research should focus on improving hardware fidelity and long-term user experience.

Specifically, when the image resolution is insufficient or visible pixel granularity is present, users may find it difficult to clearly discern the interface elements or environmental textures. This visual degradation undermines the perception of continuity and consistency between the virtual and physical environments. Additionally, the absence of natural lighting cues, such as directional lighting, shadows, or reflections, can cause virtual objects to appear unnaturally 'suspended' or visually detached from their surroundings. In

this study, the virtual driving scene was characterised by its monochromatic tone and ambiguous light source direction, which contributed to a flat visual field lacking depth and spatial realism. Together, these shortcomings diminish the immersive quality of the environment and reduce the users' perception of realism.

4.2.3 Impact on immersion

User immersion in VR is strongly influenced by visual fidelity, which encompasses resolution, FoV, and overall graphics quality. These elements collectively shape engagement and absorption in the environment. High-resolution displays and wide-viewing-angle designs can enhance a user's visual absorption of a virtual environment, thereby improving immersion (Hameed et al. 2024). Compared to traditional 2D screens, HMDs provide a greater sense of immersion owing to their wider FoV and stereoscopic effects (Desai et al. 2017). Transmission quality, which includes bitrate, latency, and packet loss, also affects perceived video quality and, consequently, user immersion (Fei et al. 2019). In a 360° VR video, lower transmission quality can reduce both user comfort and immersion. However, Cortés et al. (2024) investigated delay thresholds in volumetric social XR communication and found that end-to-end delays beyond 900 ms significantly degraded user-perceived presence, interaction fluency, and social awareness.

In multiclient live telepresence, maintaining immersion critically depends on system scalability and low-latency streaming. Balancing visual quality and transmission delay is essential to provide a seamless and immersive user experience. To overcome these challenges, Stotko et al. (2019) proposed a system that incorporates reconstruction streaming optimisations, such as depth discontinuity filtering, voxel block allocation down-sampling, and marching cube-based voxel pruning, which collectively reduce latency and enhance visual fidelity. Similarly, Ma et al. (2024) addressed the quality-efficiency trade-off in real-time VR fluid rendering by introducing a perception-guided temporal acceleration strategy. At the rendering level, LOD optimisation helps maintain high frame rates while improving visual details, reducing motion sickness, and enhancing the user's SoP and immersion (Hu et al. 2018).

4.2.4 Impact on presence

Visual fidelity, particularly resolution and FoV, directly affects spatial and social presence. Studies have confirmed that higher video and audio quality improves realism and spatial presence, with a wider FoV and higher bitrates contributing to stronger presence (Baigabulov and Ipalakova 2023; Zou et al. 2018). However, beyond certain thresholds (e.g., 30 Mbps for the video bitrate), additional improvements yield diminishing returns.

A higher resolution enables finer visual details to approach the limits of human visual acuity, reinforcing both realism and immersion. Moreover, higher frame rates help minimise motion blur and latency, improve visual stability, and reduce motion sickness, further enhancing the SoP. Moreover, a wider FoV enhances the users' sense of spatial enclosure, making the virtual environment more immersive and strengthening the overall SoP.

Although visual fidelity sets the foundation for presence, it should be supported by stable transmission quality to maintain immersive continuity, particularly in streaming-based VR applications. Low-latency streaming, minimal compression artefacts, and high-resolution videos facilitate immersive continuity, whereas network lag, frame drops, and stitching errors can disrupt SoP (Cheng et al. 2024; Salmimaa et al. 2018). In 360° VR and MR systems, stereoscopic depth cues and smooth streaming can significantly improve perceived SoP, whereas aggressive compression and low-quality depth-based rendering (e.g., low depth-based index settings) may disrupt depth perception and weaken SoP (Selzer et al. 2021).

Beyond signal delivery, the type and performance of the display device further modulate a user's experience of presence. HMDs generally elicit a stronger SoP than traditional 2D screens; however, this advantage is highly dependent on device performance. Poor frame rates, motion blurring, and rendering limitations can impair SoP and lead to discomfort or motion sickness (Kim et al. 2020b). In addition to hardware specifications, environmental factors, such as lighting accuracy, play a minor yet important role in enhancing presence, particularly through improved realism and spatial coherence. Moreover, accurate lighting in VR, including realistic contrast and colour representation, supports SoP by improving environmental consistency (Kong et al. 2023; Mostafavi et al. 2023). Collectively, these interrelated technical elements shape the user's ability to feel 'present' within VR and MR environments.

4.3 Motion and interaction

4.3.1 Impact on perception

Depth perception is pivotal in shaping how users interact with and interpret IVEs (Padilha and Teichrieb 2014). Empirical studies have shown that depth cues, including motion parallax, stereoscopic disparity, and transparency, significantly affect users' distance judgements and spatial awareness in VR, with their effectiveness varying by scene complexity and interaction modality (Allue et al. 2016; Feng

et al. 2023; Luan et al. 2024; Malpica et al. 2020; Padilha and Teichrieb 2015; Robotham et al. 2022).

4.3.2 Impact on realism

Depth cues lay the perceptual foundation, whereas realism in virtual environments depends on users' physical interaction with and interpretation of motion dynamics. Motion parallax, the natural change in an object's position relative to the user's movement, is crucial for enhancing depth perception and three-dimensional realism. When accurately simulated, motion parallax reduces visual errors and strengthens spatial awareness, rendering the IVEs more realistic (Waldow et al. 2024).

Moreover, using real-world physical interfaces within IVEs—such as employing custom hardware, including crane controllers in safety training simulations—can increase immersion and user engagement, further enhancing the perception of realism (Checa et al. 2021).

Notably, the distinction between perceptual realism and felt realism offers insights into the users' experience of IVEs. Although perceptual realism is tied to visual fidelity and technical accuracy, felt realism is driven by emotional engagement and the conceptual coherence of the environment. Studies have demonstrated that even in abstract or abnormal virtual spaces, users can experience strong emotional responses and a heightened sense of realism if the environment elicits engagement and maintains logical consistency (Christophers et al. 2024). Supporting this, comparative studies have shown that IVR can produce perceptual responses closely aligned with real environments across visual, acoustic, and global dimensions, reinforcing its validity as a multisensory evaluation tool (Luigi et al. 2015).

4.3.3 Impact on immersion

Beyond visual and perceptual realism, the integration of motion and interaction is central to sustaining immersion. Previous studies have revealed that realistic lighting, dynamic motion simulations, and authentic interaction elements significantly enhance immersion. Interactive features such as 3D models, avatars, dialogue systems, and spatial audio work together to enhance users' perception of context and increase their engagement (Calvi et al. 2017). This can make users feel more immersed in this environment, and non-visual elements, such as sound design, contribute to a heightened sense of user immersion.

4.3.4 Impact on presence

As immersion intensifies user engagement, the experiential outcome is often reflected in the degree of presence experienced in a virtual space. Lighting, motion, and interaction realism influence presence, with younger users generally responding more positively to high-fidelity dynamic environments (Calvi et al. 2017). Higher realism in motion and physics simulations, such as realistic underwater caustics, fog effects, and wind interactions, enhances spatial presence by making the environment feel more natural (Thompson et al. 2019).

In addition, studies have confirmed that interactive virtual objects, dynamic window views, and real-world geospatial data significantly strengthen presence, as they align more closely with users' expectations of reality (Cho et al. 2025; Ingabo et al. 2024). Specifically, the movement of outdoor elements, such as trees swaying in the wind, passing cars, and pedestrians walking, contributes to the perceived quality of the window view, thereby influencing the SoP. Motion-based manipulation techniques, such as translation gains, can enlarge virtual spaces beyond the user's physical area. Although such manipulations might disrupt presence, empirical results indicate that presence can be preserved or even enhanced when physical movement is consistent and meaningful interaction is embedded in the task structure (Selzer et al. 2022). To maintain interactive real-time immersion, systems should achieve an average end-to-end latency of 40 ms and a frame rate of at least 60 Hz, which are thresholds below which perceptual breaks are minimised, and the SoP is sustained (Regenbrecht et al. 2017).

Beyond motion dynamics and latency thresholds, the spatial configuration of the virtual viewpoint also plays a critical role in shaping presence. Feng et al. (2023) confirmed that eye height in pre-captured panoramic environments significantly influences both distance perception and body location ownership, with higher eye heights leading to a marked decline in immersion responses. Building on this, Luan et al. (2024) extended these findings to dynamic real-walking contexts, demonstrating that a viewpoint height approximately 10 cm below the user's eye level provides optimal spatial presence during locomotion, and that even minor mismatches in viewpoint fluctuation, as small as 2 cm, can significantly disrupt spatial presence by inducing conflicts between visual and vestibular perception. In summary, motion and interaction are not merely supplemental features but fundamental components that shape how users perceive, believe in, and inhabit virtual environments.

5 Discussion

5.1 Relationships between visual quality and experiential constructs

This review focused on the association between visual quality and its influence on users through an interconnected system involving perception, realism, immersion, and presence. Rather than a simple linear cascade, it is a recursive, dynamic process. For example, a strong presence can reinforce perceptual stability and task engagement, whereas coherent realism facilitates deeper immersion, even under constrained graphical fidelity (Gilbert 2017; Regenbrecht et al. 2017).

Perception is the foundation and is shaped by sensory clarity, spatial composition, materials, lighting, resolution, and other factors (Cheng et al. 2023; Luebke et al. 2002). Enhanced perceptual input contributes to environmental credibility, leading to stronger conceptual, perceptual, and felt realism (Christophers et al. 2024; Gilbert 2017). Realism promotes deeper cognitive engagement and, when combined with consistent narrative and sensory feedback, increases immersion (Witmer and Singer 1998; Zhang et al. 2017). Presence, as a 'being there' psychological state, primarily stems from the fact that what users see and feel is sufficiently realistic, and concurrently, they are emotionally engaged and completely immersed in IVEs (Schubert et al. 2001; Slater and Wilbur 1997).

Although these four experiential dimensions theoretically exhibit a certain hierarchical structure, in practice, they form a dynamic, mutually influencing, and interactive system. For example, enhanced perception strengthens presence, realism deepens immersion, and immersion promotes perception. However, owing to the overlap and ambiguity in the definitions and applications of these concepts, challenges exist in their measurement and theoretical construction. Clarifying the overlap and relationships between these constructs is crucial for establishing a more comprehensive user experience model and developing more precise evaluation tools.

5.2 Effects of specific visual quality features on experience

Synthesising the reviewed literature reveals that the relationship between visual quality features and user experience is not linear. Although foundational visual quality is a prerequisite for immersion, our analysis identifies a critical 'Law of Diminishing Returns' in IVEs. In line with prior studies, our findings suggest that coherence across multiple perceptual cues, such as behavioural consistency, visual–behavioural alignment, and contextual plausibility, plays a critical role in shaping users' experiential responses. Under such conditions, even low-fidelity representations can evoke a strong sense of 'felt realism', without necessarily requiring high levels of graphical detail (Huang et al.2023; Fraser et al. 2024; Xiao et al. 2025).

A recurring finding is that human perception in VR has a saturation point; for instance, Debattista et al. (2017) and Regenbrecht et al. (2017) demonstrated that beyond a certain fidelity threshold, improvements no longer provide additional benefits and may even cause distraction. At this stage, allocating computational resources to increase pixel count or geometric complexity has become inefficient, as it contributes minimally to subjective presence. Conversely, excessive visual detail can generate 'cognitive clutter', distracting users from the narrative or current. Future research should seek to formalise such metrics, drawing on frameworks such as that proposed by Wang et al. (2024a), who examined specific perceptual quality metrics for identifying fidelity saturation points in VR systems.

In terms of modelling, excessive simplification of LOD reduces perceived realism, causing model distortion and making spatial structures difficult to understand, thereby weakening the user's immersive experience (Bertheaux et al. 2024; Niu and Lo 2022; Wang and Munakata 2024). In addition, increased visual distortion reduces perceived quality and immersion (Alfarasani et al. 2023). Extreme material properties and unreasonable spatial layout designs can significantly impair the realism and comfort of a virtual environment (Bertheaux et al. 2024; Niu and Lo 2022).

Beyond the aforementioned visual quality dimensions, cross-modal sensory compensation offers an additional strategy for enhancing immersion under hardware constraints. Potter et al. (2022) established that spatial audio rendering contributes independently to VR immersion alongside visual fidelity, suggesting that the two modalities can be leveraged in a complementary manner. Flores-Vargas et al. (2025) demonstrated experimentally that real-time spatial audio naturalisation significantly enhanced perceived presence and audio-visual integration even under low-resolution visual conditions, suggesting that the brain integrates multisensory inputs in ways that can partially offset visual limitations. This indicates that improving immersion in IVEs need not rely solely on advances in modelling, rendering and hardware, and motion and interaction—strategic investment in audio fidelity may achieve comparable experiential gains through perceptual substitution.

Finally, the review demonstrates that in IVE, dynamic system performance, such as low latency and high frame rates, is more critical than static visual fidelity. Although high-resolution textures and detailed shading enhance realism in static views, temporal stability, characterised by low latency, high frame rates, and consistent motion-to-photon

response, emerges as the dominant determinant of immersion and user comfort. Stotko et al. (2019) explicitly noted that interactive VR exploration requires high framerates and low latency to avoid motion sickness. Therefore, 'good' visual quality in IVEs is defined not merely by screenshot-level clarity but by the system's ability to maintain coherent, stable rendering during interactive motion. Over-prioritising rendering complexity at the cost of real-time performance compromises presence more severely than reduced geometric or texture detail.

Empirical evidence supports specific thresholds in this regard. Lower frame rates (e.g., 60–90 fps) have been associated with increased simulator sickness and perceived latency, whereas frame rates around 120 fps substantially mitigate these effects (Wang et al. 2023a; Wei et al. 2024). From a latency perspective, even small delays can disrupt perceptual coherence by desynchronising head movement and visual feedback, a primary contributor to cybersickness (Wagner et al. 2023). At a practical level, increased end-to-end latency has been shown to significantly prolong task completion time and reduce interaction accuracy in VR selection and steering tasks (Wei et al. 2025). To address these constraints, recent approaches have explored adaptive strategies that dynamically balance visual quality and system responsiveness, such as gaze-contingent rendering and latency-aware quality control mechanisms (Wu et al. 2025). Overall, these findings suggest that temporal stability should be considered a foundational prerequisite for immersive system design, optimised in parallel with, rather than subordinate to, visual fidelity. Notably, visual quality dimensions such as resolution exhibit diminishing returns beyond a certain threshold and trade off against refresh rate and frame rate, reinforcing the case for balancing temporal and visual factors rather than prioritising one over the other (Wang et al. 2022).

5.3 Current research gaps and future directions

Although research on visual quality in IVEs is growing, several critical gaps remain, limiting its theoretical integration and practical applications. The following points summarise the main limitations of this review and suggest directions for future research.

Most IVE studies are limited to specific application areas such as gaming, education, or architecture, whereas some have focused on the IVE experience. Currently, no unified cross-domain framework exists that integrates the technical, perceptual, and emotional dimensions of visual quality (Cummings and Bailenson 2016; van der Land et al. 2013). Moreover, most existing research has overlooked individual differences, such as age, sex, cognitive capacity, sensory acuity, and prior experience with immersive technology. These user-specific variables are likely to influence the perception and processing of visual quality (Mostafavi et al. 2023). To ensure an inclusive and user-centred IVE design, future studies should incorporate user profiling and adaptive rendering systems. For instance, user profiling based on individual behavioural characteristics and cybersickness sensitivity can enable systems to pre-configure appropriate rendering parameters for each user. At the rendering level, gaze-contingent techniques can be applied to enhance resolution within the current fixation region while reducing quality in peripheral areas, thereby aligning computational resources with the user's actual perceptual focus (Deng et al. 2022; Jin et al. 2021). Similarly, adaptive bitrate streaming systems can dynamically modulate visual quality in response to real-time network conditions, maintaining perceptual fidelity in the foveal region while reducing peripheral bandwidth demands (Wu et al. 2025). Overall, these approaches illustrate how user-centred adaptive frameworks can extend beyond conceptual advocacy to practical implementation, minimising latency while maximising immersive experience.

To address the lack of standardised measurement tools, future studies should adopt a structured triangulation approach that integrates subjective questionnaires with objective physiological measures. Specifically, we proposed a three-layer assessment framework: (1) subjective self-report measures, such as the IPQ or presence questionnaires, to capture users' conscious experiential states; (2) neurophysiological measures, including EEG and heart rate variability, to assess unconscious arousal and cognitive load responses to visual quality manipulations; and (3) behavioural measures, such as eye-tracking fixation patterns and reaction times, to provide task-level indicators of immersion and perceptual engagement (Egan et al. 2016). This multi-layered approach would enable researchers to cross-validate findings across measurement levels, reducing reliance on any single method and improving the robustness of conclusions about the relationship between visual quality and user experience.

Visual quality is often considered a static design attribute; however, in practice, it operates in a dynamic, interactive environment where user actions influence the system output. Further research is required on real-time feedback, adaptive rendering, and interactive personalisation to reflect how visual quality evolves during user interactions.

In summary, future research should move away from a hardware-driven perspective towards user-centred, adaptive, and cross-modal frameworks. Integrating perceptual psychology, HCI principles, and affective computing is crucial for developing visual quality systems that are not only realistic but also responsive, inclusive, and emotionally engaging.

6 Conclusion

This review investigated the effect of visual quality on users' experiential dimensions of IVEs. Through a systematic analysis of 75 studies, we first categorised visual quality features into three groups: modelling, hardware and rendering, and motion and interaction. We then evaluated each category with respect to three key experiential factors. The results obtained from the analysis provided structured evidence to answer these three research questions.

The findings confirm that visual quality affects users' interpretation, engagement, and responses to virtual content in three ways: modelling quality features, such as LOD, material, and spatial configuration, affect spatial realism and perceptual realism; hardware rendering quality features, such as resolution, FoV, and lighting fidelity, not only affect visual realism and comfort but also influence system performance and user confidence; and motion and interaction design quality affect depth perception, task involvement, and psychological presence. The best immersive experience can be achieved only when perceptual expectations are met, aesthetic effects are well maintained, and interaction speeds are fast enough to support responsive behaviour in virtual environments. This is not determined solely by high visual fidelity.

This review contributes to the field by linking technical attributes to human-centred experiential outcomes, enabling us to view visual quality as an enabler of user experience rather than merely a design parameter. These findings have important implications for participants and researchers in educational, architectural, simulation, entertainment, and medical VR/AR applications. However, the potential impact of these findings will be maximised if the limitations identified in this review are addressed in future studies.

Based on the synthesised evidence, we hypothesised that a virtual environment, which maintains material metallicity within an optimal range, avoids excessive video compression, refrains from oversimplifying LOD, and does not introduce excessive detail or contrast, will provide the most favourable user experience in terms of realism, immersion, and comfort.

This review covers the research published from 2014 to 2025, during a period of rapid technological advancement in this field. A limitation of this review is that the search strategy, which relied on keywords, paper titles, and abstracts, may have excluded some relevant studies. In particular, studies whose relevance was not explicitly reflected in their titles or keywords, or those that did not clearly report perceptual outcomes, may have been overlooked during the early screening stage.

This exclusion risk may have limited the inclusion of interdisciplinary research or studies that implicitly addressed perception, realism, immersion, or presence. Consequently, certain nuanced relationships or emerging perspectives may not have been fully captured. However, given that the screening criteria were systematically applied and focused on well-established terminology within the field, we believe that the core relationships identified in this review remain robust and representative of dominant research trends.

Author contributions Enlian Zhang conducted the data curation, analysis, investigation, visualization, and prepared the original manuscript draft. Jin-Bin Im, Kyung-Tae Lee, and Ju-Hyung Kim contributed to conceptualization, resources, validation, supervision, and manuscript review and editing. All authors reviewed and approved the final manuscript.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/.

References

Abd-Alhamid F, Kent M, Bennett C, Calautit J, Wu Y (2019) Developing an innovative method for visual perception evaluation in a physical-based virtual environment. BuildEnviron 162:10627 https://doi.org/10.1016/j.buildenv.2019.106278

Ackerman A, Cooper L, Malmsheimer R, Esch E, Volk T, Constantineau S, Gadoth-Goodman D (2022) Virtual reality visualization of sustainable management practices for forest carbon and climate change. J Digit Landsc Archit 7:392–400. https://doi.org/10.14627/537724038

Alfarasani DA, Lai YK, Rosin PL (2023) Subjective study of 3D mesh quality scores in virtual reality. ACM Int Conf Proc Ser 7–13. https://doi.org/10.1145/3603421.3603423

Allue M, Serrano A, Bedia MG, Masia B (2016) Crossmodal perception in immersive environments. Spanish Comput Graph Conf 16:1–7. https://doi.org/10.2312/ceig.20161307

Bahirat K, Lai C, McMahan RP, Prabhakaran B (2018) Designing and evaluating a mesh simplification algorithm for virtual reality. ACM Trans Multimed Comput Commun Appl 14(3s):1–26. https://doi.org/10.1145/3209661

Baigabulov S, Ipalakova M (2023) Virtual reality enabled immersive data visualization for data analysis https://ceur-ws.org/Vol-3680/S1Paper1.pdf

Bertheaux C, Zimmermann E, Gazel M, Delanoy J, Raimbaud P, Lavoué G (2024) Effect of material properties on emotion: a virtual reality study. Front Hum Neurosci 17:1301891. https://doi.org/10.3389/fnhum.2023.1301891

Botella C, Rey A, Perpiñá C, Baños R, Alcañiz M, Garcia-Palacios A, Villa H, Alozano J (1999) Differences on presence and reality judgment using a high impact workstation and a PC workstation. Cyberpsychol Behav 2(1):49–52. https://doi.org/10.1089/cpb.1999.2.49

Bugeja K, Debattista K, Spina S, Chalmers A (2014) High-fidelity graphics for dynamically generated environments using distributed computing. VS-GAMES Conf 1-8. https://doi.org/10.1109/VS-Games.2014.7012160

Calvi L, Santos CP, Relouw J, Endrovski B, Rothwell C, Sarà A, Lucrezi S, Palma M, Pantaleo U (2017) A VR game to teach underwater sustainability while diving. Sustain Internet ICT Sustain 1–4 https://doi.org/10.23919/SustainIT.2017.8379812

Checa D, Martínez K, Osornio-Rios RA, Bustillo A (2021) Virtual reality opportunities in the reduction of occupational hazards in industry 4.0. Dyna 96(6):620–626. https://doi.org/10.6036/10241

Chen K, Wan T, Matsuda N, Ninan A, Chapiro A, Sun Q (2024) PEAPODs: perceptual evaluation of algorithms for power optimization in XR displays. ACM Trans Graph 43(4):1–17. https://doi.org/10.1145/3658126

Cheng H, Xu C, Chen X, Chen Z, Wang J, Zhao L (2023) Realistic volume rendering with environment-synced illumination in mixed reality. ISMAR Adjunct 2023:423–428. https://doi.org/10.1109/ISMAR-Adjunct60411.2023.00090

Cheng R, Murat E, Yu LF, Chen S, Han B (2024) Understanding online education in metaverse: systems and user experience perspectives. IEEE VR 2024:598–608. https://doi.org/10.1109/VR58804.2024.00080

Cho Y, Karmann C, Andersen M (2025) Perception of window views in VR: impact of display and type of motion on subjective and physiological responses. Build Environ 274:112757. https://doi.org/10.1016/j.buildenv.2025.112757

Christophers L, Mulvaney P, Rooney B (2024) The felt realism of "unreal" environments: testing a dual awareness model of subjective realism. Int J Hum Comput Interact 40(24):8890–8908. https://doi.org/10.1080/10447318.2023.2291615

Chu J, Bryan C, Shih M, Ferrer L, Ma KL (2017) Navigable videos for presenting scientific data on affordable head-mounted displays. ACM Multimedia Syst Conf 250–260. https://doi.org/10.1145/3083187.3084015

Church J (2013) Possibilities of perception. OUP Oxford, Oxford

Civitarese G (2008) Immersion versus interactivity and analytic field. Int J Psychoanal 89(2):279–298. https://doi.org/10.1111/j.1745-8315.2008.00019.x

Cortés C, Viola I, Gutiérrez J, Jansen J, Subramanyam S, Alexiou E, Pérez P, García N, César P (2024) Delay threshold for social interaction in volumetric extended reality communication. ACM Trans Multimed Comput Commun Appl 20(7):1–8. https://doi.org/10.1145/3651164

Costalunga F, Bettelli A, Campanini ML, Mingardi M, Orso V, Gamberini L (2025) Taste through a different lens: the role of mixed reality color saturation in sensory and hedonic evaluations. ICVR 2025:312–321. https://doi.org/10.1109/ICVR66534.2025.11172406

Cummings JJ, Bailenson JN (2016) How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. Media Psychol 19(2):272–309. https://doi.org/10.1080/15213269.2015.1015740

Debattista K, Bashford-Rogers T, Harvey C, Waterfield B, Chalmers A (2017) Subjective evaluation of high-fidelity virtual environments for driving simulations. IEEE Trans Hum Mach Syst 48(1):30–40. https://doi.org/10.1109/THMS.2017.2762632

Deng N, He Z, Ye J, Duinkharjav B, Chakravarthula P, Yang X, Sun Q (2022) FoV-NeRF: foveated neural radiance fields for virtual reality. IEEE Trans vis Comput Graph 28(11):3854–3864. https://doi.org/10.1109/TVCG.2022.3203102

Desai K, Raghuraman S, Jin R, Prabhakaran B (2017) QoE studies on interactive 3D tele-immersion. IEEE Int Symp Multimed 130–137. https://doi.org/10.1109/ISM.2017.27

Diemer J, Alpers GW, Peperkorn HM, Shiban Y, Mühlberger A (2015) The impact of perception and presence on emotional reactions: a review of research in virtual reality. Front Psychol 6:26. https://doi.org/10.3389/fpsyg.2015.00026

Egan D, Brennan S, Barrett J, Qiao Y, Timmerer C, Murray N (2016) An evaluation of heart rate and electrodermal activity as an objective QoE evaluation method for immersive virtual reality environments. QoMEX 2016:1 https://doi.org/10.1109/QoMEX.2016.7498964

Elwardy M, Hu Y, Zepernick HJ, Chu TMC, Sundstedt V (2020) Comparison of ACR methods for 360° video quality assessment. ICSPCS 2020:1–10. https://doi.org/10.1109/ICSPCS50536.2020.9310071

Ermi L, Mäyrä F (2005) Fundamental components of the gameplay experience: analysing immersion. DiGRA Conf https://doi.org/10.26503/dl.v2005i1.119

Fei Z, Wang F, Wang J, Xie X (2019) QoE evaluation methods for 360-degree VR video transmission. IEEE J Sel Top Signal Process 14(1):78–88. https://doi.org/10.1109/JSTSP.2019.2956631

Feng Q, Shum HP, Morishima S (2023) Enhancing perception and immersion in pre-captured environments through learning-based eye height adaptation. ISMAR 2023:405–414. https://doi.org/10.1109/ISMAR59233.2023.00055

Flores-Vargas M, Bates E, McDonnell R (2025) First-person vocal auralisation in XR and its influence on perceived presence and audio-visual quality. ISMAR 2025:1586–1596. https://doi.org/10.1109/ISMAR67309.2025.00162

Fraser A, Branson I, Hollett R, Speelman C, Rogers S (2024) Do realistic avatars make virtual reality better? Examining human-like avatars for VR social interactions. Comput Hum Behav Artif Humans 2(2):100082. https://doi.org/10.1016/j.chbah.2024.100082

Gilbert SB (2017) Perceived realism of virtual environments depends on authenticity. Presence (Camb) 25(4):322–324. https://doi.org/10.1162/PRES_a_00276

Goldstein EB (2007) Cognitive psychology: Connecting mind, research and everyday experience. Wadsworth Publishing

Gorini A, Capideville CS, De Leo G, Mantovani F, Riva G (2011) The role of immersion and narrative in mediated presence: the virtual hospital experience. Cyberpsychol Behav Soc Netw 14(3):99–105. https://doi.org/10.1089/cyber.2010.0100

Hameed A, Möller S, Perkis A (2024) A holistic quality taxonomy for virtual reality experiences. Front Virtual Real 5:1434016. https://doi.org/10.3389/frvir.2024.1434016

Hanel C, Weyers B, Hentschel B, Kuhlen TW (2014) Interactive volume rendering for immersive virtual environments. IEEE 3DVis Workshop 73–74 https://doi.org/10.1109/3DVis.2014.7160104

He F, He Y, Sun L (2023) Gender differences in color perceptions and preferences of urban façades based on a virtual comparison. Build Environ 245:110907. https://doi.org/10.1016/j.buildenv.2023.110907

Henriques H, de Oliveira A, Oliveira E, Clua E, Trevisan D (2025) Comparing perceptual visual quality of hybrid neural and ray tracing in foveated rendering. J Interact Syst 16(1):328–342. https://doi.org/10.5753/jis.2025.5593

Hu Y, Zhu J, Li W, Zhang Y, Zhu Q, Qi H, Zhang H, Cao Z, Yang W, Zhang P (2018) Construction and optimization of three-dimensional disaster scenes within mobile virtual reality . ISPRS Int J Geo-Inf 7(6):215. https://doi.org/10.3390/ijgi7060215

Huang T, Li Y, Liang H (2023) Avatar type, self-congruence, and presence in virtual reality. Chinese CHI 2023:61–72. https://doi.org/10.1145/3629606.3629614

Ingabo SN, Sheng GQ, Chiang LN, Yen WC, Chan YC (2024) Is virtual reality an effective experimental tool for window view studies with dynamic content involved? LEUKOS 20(4):411–434. https://doi.org/10.1080/15502724.2024.2358791

Jensen L, Konradsen F (2018) A review of the use of virtual reality head-mounted displays in education and training. Educ Inf Technol 23(4):1515–1529. https://doi.org/10.1007/s10639-017-9676-0

Jerald J (2016) The VR book: human-centered design for virtual reality. ACM/Morgan & Claypool, San Rafael

Jin Y, Chen M, Goodall T, Patney A, Bovik AC (2021) Subjective and objective quality assessment of 2D and 3D foveated video compression in virtual reality. IEEE Trans Image Process 30:5905–5915. https://doi.org/10.1109/TIP.2021.3087322

Jung S, Lindeman RW (2021) Perspective: Does realism improve presence in VR? Suggesting a model and metric for VR experience evaluation. Front Virtual Real 2:693327. https://doi.org/10.3389/frvir.2021.693327

Kapralos B, Shewaga R, Ng G (2014) Serious games: Customizing the audio-visual interface. Virtual, augmented and mixed reality. Springer, Cham, pp 190–199

Kim HG, Lim HT, Ro YM (2020a) Deep virtual reality image quality assessment with human perception guider for omnidirectional image. IEEE Trans Circuits Syst Video Technol 30(4):917–928. https://doi.org/10.1109/TCSVT.2019.2898732

Kim W, Lee S, Bovik AC (2020b) VR sickness versus VR presence: a statistical prediction model. IEEE Trans Image Process 30:559–571. https://doi.org/10.1109/TIP.2020.3036782

Kong G, Chen P, Wang L, Chen S, Yu J, Chen Z (2023) Calibration of brightness of virtual reality light sources based on user perception in the real environment. J Build Eng 78, 107702. https://doi.org/10.1016/j.jobe.2023.107702

Langa SF, Montagud M, Cernigliaro G, Rivera DR (2022) Multiparty holomeetings: Toward a new era of low-cast volumetric holographic meetings in virtual reliaty. IEEE Access 10:81856–81876. https://doi.org/10.1109/ACCESS.2022.3196285

Li J, Subramanyam S, Jansen J, Mei Y, Reimat I, Ławicka K, Cesar P (2021) Evaluating the user experience of a photorealistic social VR movie. ISMAR 2021:284–293. https://doi.org/10.1109/ISMAR52148.2021.00044

Li Y, Yabuki N, Fukuda T (2022) Measuring visual walkability perception using panoramic street view images, virtual reality , and deep learning. Sustain Cities Soc 86:104140. https://doi.org/10.1016/j.scs.2022.104140

Luan H, Wang L, Gai W, Lv G, Luan X, Yang C (2024) Exploring the effects of viewpoint height & fluctuation on walking perception in panoramic tours. ISMAR 2024:574–583. https://doi.org/10.1109/ISMAR62088.2024.00072

Luebke D, Reddy M, Cohen JD, Varshney A, Watson B, Huebner R (2002) Level of detail for 3D graphics. Elsevier, Amsterdam

Luigi M, Massimiliano M, Aniello P, Gennaro R,Virginia PR(2015) On the validity of immersive virtual reality as tool for multisensory evaluation of urban spaces. Energy Procedia 78:471–476. https://doi.org/10.1016/j.egypro.2015.11.703

Luo J, Zhao T, Cao L, Biljecki F (2022) Semantic riverscapes: perception and evaluation of linear landscapes from oblique imagery using computer vision. Landsc Urban Plan 228:104569. https://doi.org/10.1016/j.landurbplan.2022.104569

Ma Q, Wu J, Fan R, Sun G, Shi X (2024) Vip-fluid: visual perception driven method for VR fluid rendering. ISMAR 2024:359–367. https://doi.org/10.1109/ISMAR62088.2024.00050

de Farias Macedo MC, Apolinario AL (2014) Improving on-patient medical data visualization in a markerless augmented reality environment by volume clipping. SIBGRAPI 149–156. https://doi.org/10.1109/SIBGRAPI.2014.33

Malpica S, Serrano A, Allue M, Bedia MG, Masia B (2020) Crossmodal perception in virtual reality. Multimed Tools Appl 79(5–6):3311–3331. https://doi.org/10.1007/s11042-019-7331-z

Mandala A, Koerniawan M, Indraprastha A, Mangkuto R, Wonorahardjo S (2025) Shaping sacredness by unveiling the impact of lighting in a church environment: comparing observations in real space and virtual reality. Build Environ 279:113028. https://doi.org/10.1016/j.buildenv.2025.113028

Maniee S, Zomorodian ZS, Tahsildoost M (2025) Assessing lighting design in urban open spaces: a virtual reality experiment. Build Environ 283:113362. https://doi.org/10.1016/j.buildenv.2025.113362

Marino E, Bruno F, Liarokapis F (2021) Color harmonization, deharmonization and balancing in augmented reality . Appl Sci 11(9):3915. https://doi.org/10.3390/app11093915

Mazikowski A (2019) Invenstigation of optical Optical properties of infitec and active stereo stereoscopic techniques for CAVE-type virtual reality systems. Metrol Meas Syst 26(1):139–151 https://doi.org/10.24425/mms.2019.126330

McMahan RP, Bowman DA, Zielinski DJ, Brady RB (2012) Evaluating display fidelity and interaction fidelity in a virtual reality game. IEEE Trans vis Comput Graph 18(4):626–633 https://doi.org/10.1109/TVCG.2012.43

Mekuria R, Cesar P, Doumanis I, Frisiello A (2015) Objective and subjective quality assessment of geometry compression of reconstructed 3D humans in a 3D virtual room . Appl Digit Image Process XXXVIII 9599:537–549. https://doi.org/10.1117/12.2203312

Menz W, Hartbrich J, Bhattacharya A, Rao R, Göring S, Raake A (2025) Impact of texture resolution, polygon count, quality level, and viewing distance on volumetric avatar visual quality. MetaCom 2025:165–172. https://doi.org/10.1109/MetaCom65502.2025.00037

Milgram P, Kishino F (1994) A taxonomy of mixed reality visual displays. IEICE Trans Inf Syst.

Mišiak M, Fuhrmann A, Latoschik ME (2023) Impact of reflection approximations on visual quality in virtual reality . ACM SAP 2023:1 https://doi.org/10.1145/3605495.3605794

Mizuho T, Narumi T, Kuzuoka H (2023) Effects of visual fidelity of virtual environment on presence, context-dependent forgetting, and source monitoring error . IEEE Trans vis Comput Graph 29(5):2607–2614. https://doi.org/10.1109/TVCG.2023.3247063

Mohebbi A, Amiri P, Kearney RE (2022) Identification of human balance control responses to visual inputs using virtual reality. J Neurophysiol 127(4):1159–1170. https://doi.org/10.1152/jn.00283.2021

Mostafavi A, Cruz-Garza J, Kalantari S (2023) Enhancing lighting design through the investigation of illuminance and correlated color Temperature's effects on brain activity: An EEG-VR approach. J Build Eng 75:106776. https://doi.org/10.1016/j.jobe.2023.106776

Mozhaeva A, Vlasuyk I, Potashnikov A, Alani N, Rogers S, Streeter L (2025) Enhancing wide-angle VR video transmission using human perception. IVCNZ 2025:1–6. https://doi.org/10.1109/IVCNZ67716.2025.11281650

Mühlberger A, Herrmann MJ, Wiedemann G, Ellgring H, Pauli P (2001) Repeated exposure of phobics to flightsin virtual reality. Behav Res Ther 39(9):1033–1050 https://doi.org/10.1016/S0005-7967(00)00076-0

Nehmé Y, Farrugia JP, Dupont F, Le Callet P, Lavoué G (2020) Comparison of subjective methods for quality assessment of 3D graphics in virtual reality. ACM Trans Appl Percept 18(1):1–23. https://doi.org/10.1145/3427931

Niu M, Lo CH (2022) Do we see rendered surface materials differently in virtual reality? A psychophysics-based investigation. Virt Real 26(3):1031–1045. https://doi.org/10.1007/s10055-021-00613-3

Padilha A, Teichrieb V (2014) Motion detection based ghosted views for occlusion handling in augmented reality. ISMAR 2014:291–292 https://doi.org/10.1109/ISMAR.2014.6948455

Padilha A, Teichrieb V (2015) Motion-aware ghosted views for single layer occlusions in augmented reality. ISMAR Workshops 2015:60–67. https://doi.org/10.1109/ISMARW.2015.20

Page MJ et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. https://doi.org/10.1186/s13643-021-01626-4

Petikam L, Chalmers A, Rhee T (2018) Visual perception of real world depth map resolution for mixed reality rendering. IEEE VR 2018:401–408 https://doi.org/10.1109/VR.2018.8446391

Pettersson I, Karlsson MA, Ghiurau FT (2019) Virtually the same experience?: Learning from user experience of in-vehicle systems in VR and in the field. DIS 2019:463–473. https://doi.org/10.1145/3322276.3322288

Potter T, Cvetković Z, De Sena E (2022) On the Relativeimportance of visual and spatial audio rendering on VR immersion. Front Signal Process 2:904866. https://doi.org/10.3389/frsip.2022.904866

Prachyabrued M, Borst CW (2015) Design and evaluation ofv isual interpenetration cues in virtual grasping. IEEE Trans vis Comput Graph 22(6):1718–1731. https://doi.org/10.1109/TVCG.2015.2456917

Regenbrecht H, Meng K, Reepen A, Beck S, Langlotz T (2017) Mixed voxel reality: presence and embodiment in low fidelity, visually coherent, mixed reality environments . ISMAR 2017:90–99. https://doi.org/10.1109/ISMAR.2017.26

Rhee T, Thompson S, Medeiros D, Dos Anjos R, Chalmers A (2020) Augmented virtual teleportation for high-fidelity telecollaboration . IEEE Trans vis Comput Graph 26(5):1923–1933. https://doi.org/10.1109/TVCG.2020.2973065

Robotham T, Singla A, Rummukainen OS, Raake A, Habets EAP (2022) Audiovisual database with 360° video and higher-order ambisonics audio for perception cognition, behavior, and QoE evaluation research . QoMEX 2022:1–6. https://doi.org/10.1109/QoMEX55416.2022.9900893

Rossoni M, Pozzi M, Colombo G, Gribaudo M, Piazzolla P (2024) Physically based rendering of animated point clouds for eXtended reality . J Comput Inf Sci Eng 24(5):054501. https://doi.org/10.1115/1.4063559

Salmimaa M, Kimmel J, Jokela T, Eskolin P, Järvenpää T, Piippo P, Müller K, Satopää J (2018) Live delivery of neurosurgical operating theater in virtual reality. J Soc Inf Display 26(2):98–104. https://doi.org/10.1002/jsid.636

Schubert T, Friedmann F, Regenbrecht H (2001) The experience of presence: factor analytic insights. Presence 10(3):266–281 https://doi.org/10.1162/105474601300343603

Schuemie MJ, Van Der Straaten P, Krijn M, Van Der Mast CAPG (2001) Research on presence in virtual reality: a survey . Cyberpsychol Behav 4(2):183–201 https://doi.org/10.1089/109493101300117884

Selzer MN, Ganuza ML, Castro SM (2021) High visual-quality scenes in low-cost virtual reality with collisions and irregular surfaces. IEEE Comput Graph Appl 41(4):40–51. https://doi.org/10.1109/MCG.2021.3068037

Selzer MN, Larrea ML, Castro SM (2022) Analysis of translation gains in virtual reality: the limits of space manipulation. Virt Real 26(4):1459–1469. https://doi.org/10.1007/s10055-022-00640-8

Seo Y, Kim M, Jung Y, Lee D (2017) Avatar face recognition and self-presence. Comput Hum Behav 69:120–127. https://doi.org/10.1016/j.chb.2016.12.020

Sheridan TB (1999) Descartes, heidegger, gibson, and god: toward an eclectic ontology of presence. Presence 8(5):551–559 https://doi.org/10.1162/105474699566468

Sherman WR, Craig AB (2004) Understanding virtual reality: interface, application, and design. Morgan Kaufmann, Burlington

Slater M (2004) How colorful was your day? Why questionnaires cannot assess presence in virtual environment . Presence 13(4):484–493. https://doi.org/10.1162/1054746041944849

Slater M (2009) Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. Philos Trans R Soc B 364:3549–3557. https://doi.org/10.1098/rstb.2009.0138

Slater M, Wilbur S (1997) A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environment. Presence 6(6):603–616

Stotko P, Krumpen S, Weinmann M, Klein R (2019) Efficient 3D reconstruction and streaming for group-scale multi-client live telepresence. ISMAR 2019:19–25. https://doi.org/10.1109/ISMAR.2019.00018

Stylidis K, Dagman A, Almius H, Gong L, Söderberg R (2019) Perceived quality evaluation with the use of extended reality . ICED 2019:1993–2002. https://doi.org/10.1017/dsi.2019.205

Thompson S, Chalmers A, Rhee T (2019) Real-time mixed reality rendering for underwater 360° video. ISMAR 2019:74–82. https://doi.org/10.1109/ISMAR.2019.00-24

van der Land S, Schouten AP, Feldberg F, van den Hooff B, Huysman M (2013) Lost in space? Cognitive fit and cognitive load in 3D virtual environments. Comput Hum Behav 29(3):1054–1064. https://doi.org/10.1016/j.chb.2012.09.006

Wagner P, Kim J, Allison R, Palmisano S (2023) Quantifying display lag and its effects during head-mounted display based virtual reality. SIGGRAPH Asia 2023 Posters:1–3. https://doi.org/10.1145/3610542.3626139

Waldow K, Decker L, Mišiak M, Fuhrmann A, Roth D, Latoschik ME (2024) Investigating incoherent depth perception features in virtual reality using stereoscopic impostor-based rendering. IEEE VRW 2024:921–922. https://doi.org/10.1109/VRW62533.2024.00257

Wang F, Munakata J (2024) Assessing effects of façade characteristics on perceived oppressiveness in high-rise window views via virtual reality. Build Environ 266:112043. https://doi.org/10.1016/j.buildenv.2024.112043

Wang J, Shi R, Xiao Z, Qin X, Liang H (2022) Effect of render resolution on gameplay experience, performance, and simulator sickness in virtual reality games. Proc ACM Comput Graph Interact Tech 5(1):1–15. https://doi.org/10.1145/3522610

Wang J, Shi R, Zheng W, Xie W, Kao D, Liang H (2023a) Effect of frame rate on user experience, performance, and simulator sickness in virtual reality. IEEE Trans vis Comput Graph 29(5):2478–2488. https://doi.org/10.1109/TVCG.2023.3247057

Wang M, Li YJ, Shi J, Steinicke F (2023b) SceneFusion: room-scale environmental fusion for efficient traveling between separate virtual environments. IEEE Trans vis Comput Graph 30(8):4615–4630. https://doi.org/10.1109/TVCG.2023.3271709

Wang Z, Gu X, Lu F (2023c) DEAMP: dominant-eye-aware foveated rendering with multi-oarameter optimization . ISMAR 2023:632–641. https://doi.org/10.1109/ISMAR59233.2023.00078

Wang J, Shi R, Li X, Wei Y, Liang HN (2024a) Omnidirectional virtual visual acuity: a user-centric visual clarity metric for virtual reality head-mounted displays and environments. IEEE Trans vis Comput Graph 30(5):2033–2043. https://doi.org/10.1109/TVCG.2024.3372127

Wang Y, Wang S, Pan Y, Li C, Chen C, Wang J (2024b) Immersive virtual reality and computer vision for heritage: visual evaluation

and perception of the industrial heritage sites along the Yunnan–Vietnam railway (Yunnan section). Herit Sci 12(1):1–15. https://doi.org/10.1186/s40494-024-01145-0

Wang Z, Wu J, Fan R, Ke W, Wang L (2024c) VPRF: visual perceptual radiance fields for foveated image synthesis. IEEE Trans vis Comput Graph 30(11):7183–7192. https://doi.org/10.1109/TVCG.2024.3456184

Wang X, Zhang Y, Li J, Zhang Z, Luo Y, Xue C (2025) Research on the aesthetic perception of color temperature in urban waterfront nightscape lighting interfaces: a virtual reality experiment approach. M2VIP 2025:166–173. https://doi.org/10.1109/M2VIP67511.2025.11165563

Warsinke M, Vergari M, Kojic T, Nikulin D, Möller S (2025) Investigating the effect of prior exposure and fidelity on quality and realism perception of VR digital twins. VRST 2025:1–9. https://doi.org/10.1145/3756884.3766019

Wei Y, Shi R, Batmaz A, Li Y, Huang M, Yang R, Liang H (2024) Evaluating and modeling the effect of frame rate on steering performance in virtual reality. IEEE Trans vis Comput Graph 31(9):5447–5461. https://doi.org/10.1109/TVCG.2024.3451491

Wei Y, Shi R, Xu K, Wang J, Gao B, Hui P, Yu L, Liang H (2025) Understanding the effect of latency on user performance of target selection in virtual reality. IEEE Trans vis Comput Graph. https://doi.org/10.1109/TVCG.2025.3634367

Witmer BG, Singer MJ (1998) Measuring presence in virtual environments: A presence questionnaire. Presence 7(3):225–240 https://doi.org/10.1162/105474698565686

Wu Z, Alhilal A, Tsui Y, Siekkinen M, Hui P (2025) EyeNexus: adaptive gaze-driven quality and bitrate streaming for seamless VR cloud gaming experiences. Proc ACM Netw 3(CoNEXT4):1–25. https://doi.org/10.1145/3768989

Xiao C, Yu D, Hornbæk K, Bergström J (2025) A concept at work: a review of motivations, operationalizations, and conclusions in VR research about presence. CHI 2025:1–21. https://doi.org/10.1145/3706598.3714279

Xu Z, Chang J, Han C, Wu H (2025) Perception of light environment in university classrooms based on parametric optical simulation and virtual reality technology. Buildings 15(15):2585. https://doi.org/10.3390/buildings15152585

Yamamoto S, Hosokawa N, Yokoya M, Tsumura N (2016) Acceptable distortion and magnification of images on reflective surfaces in an augmented reality system. Opt Rev 23(6):976–985. https://doi.org/10.1007/s10043-016-0282-y

Zambrano JLV, Cedeño AKA, Escobar JCP, Mera AFS (2024) Immersive virtual reality for the preservation of cultural heritage. ICEDEG 2024:1–6. https://doi.org/10.1109/ICEDEG61611.2024.10702094

Zeng H, Zhao R (2023) Perceptually-guided dual-mode virtual reality system for motion-adaptive display . IEEE Trans vis Comput Graph 29(5):2249–2257. https://doi.org/10.1109/TVCG.2023.3247097

Zhang C, Perkis A, Arndt S (2017) Spatial immersion versus emotional immersion, which is more immersive?. QoMEX 2017 https://doi.org/10.1109/QoMEX.2017.7965655

Zhdanov A, Zhdanov D, Potemin IS, Bogdanov N (2019) Analysis of the visual perception conflicts in the mixed reality systems with the real-world illumination parameters restoration. Digital Opt Technol 2019:148–161. https://doi.org/10.1117/12.2526169

Zhu H, Kong Y, Zhang H, Gu Z, Ohno R (2025) Effects of scenery frame on visual depth perception in classical Chinese gardens: a case study of the Lvyin Pavilion in Lingering Garden. Front Archit Res 14(2):402–415. https://doi.org/10.1016/j.foar.2024.08.009

Zou W, Yang F, Zhang W, Li Y, Yu H (2018) A framework for assessing spatial presence of omnidirectional video on virtual reality device. IEEE Access 6:44676–44684. https://doi.org/10.1109/ACCESS.2018.2864872

Zou W, Zhang W, Song J, Yang F, Le Callet P (2019) A novel method for assessing the spatial presence of omnidirectional videos: comparing virtual and real world. IEEE Access 7:183405–183413. https://doi.org/10.1109/ACCESS.2019.2920443

Zou W, Yang L, Yang F, Ma Z, Zhao Q (2020) The impact of screen resolution of HMD on perceptual quality of immersive videos. ICMEW 2020:1–6 https://doi.org/10.1109/ICMEW46912.2020.9105945

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