



OPEN ACCESS

EDITED BY

Koen Smit,
HU University of Applied Sciences Utrecht,
Netherlands

REVIEWED BY

Philippine Waisvisz,
HU University of Applied Sciences Utrecht,
Netherlands
Stan Van Ginkel,
HU University of Applied Sciences Utrecht,
Netherlands

*CORRESPONDENCE

Yan Luximon,
✉ yan.luximon@polyu.edu.hk

RECEIVED 27 October 2025

REVISED 29 December 2025

ACCEPTED 15 January 2026

PUBLISHED 19 February 2026

CITATION

Wei X, Wang Y, Zhu A and Luximon Y (2026)
AIMERS: an AI-based MR scene design system
with human-centric perception optimization.
Front. Virtual Real. 7:1733259.
doi: 10.3389/frvir.2026.1733259

COPYRIGHT

© 2026 Wei, Wang, Zhu and Luximon. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

AIMERS: an AI-based MR scene design system with human-centric perception optimization

Xiaokang Wei¹, Yuqian Wang¹, Ao Zhu¹ and Yan Luximon^{1,2*}

¹School of Design, The Hong Kong Polytechnic University, Hong Kong, China, ²Laboratory for Artificial Intelligence in Design (AiDLab), Hong Kong, China

Visual realism is fundamental to convincing Mixed Reality (MR) experiences. However, current design workflows implicitly assume that physically-based rendering parameters naturally lead to perceptually realistic results. We begin from the opposite hypothesis: physically accurate parameters and users' perceived realism are often misaligned, leading to inconsistent visual fusion and significant design overhead. To address this problem, we present AIMERS, an AI- and perception-guided MR scene design framework. First, AI-based neural inverse rendering is used to automatically estimate lighting-independent material, geometry, and illumination properties from multi-view RGB inputs, removing the need for manual material calibration. We then introduce an interactive MR perceptual interface that allows users to adjust key realism parameters during immersive viewing, enabling us to capture perception-aligned preferences across scenes. By jointly analyzing physically derived parameters and perceptual data, we derive optimal parameter intervals that best match perceived realism across different scene categories. Controlled user studies reveal a consistent mismatch between physical correctness and human perception, and demonstrate that combining AI estimation with perception-guided adjustment leads to more coherent and convincing MR visual fusion. Overall, this work establishes a perception-aligned paradigm for MR scene design, bridging the gap between physical accuracy and human perception and providing practical guidance for building realistic MR applications.

KEYWORDS

artificial intelligence, inverse rendering, mixed reality, perceptual enhancement system, research methods, visual fusion

1 Introduction

Mixed Reality (MR) technology, encompassing both Augmented Reality (AR) and Virtual Reality (VR), enables seamless integration of physical and virtual elements – from inserting 3D-scanned human avatars into virtual environments to embedding digital objects in real-world settings. This transformative capability has revolutionized industries ranging from education and healthcare to architectural visualization and virtual production [Milgram and Kishino \(1994\)](#); [Rokhsaritalemi et al. \(2020\)](#). However, for MR content designers, creating these immersive experiences remains a complex process requiring expertise in 3D scanning, material authoring, and visual calibration. In particular, there is still a lack of principled guidance on how to configure rendering parameters so that virtual and real elements appear maximally realistic to human observers.

TABLE 1 Descriptive statistics of system physical variable value for Scene 1–3.

Parameter	Scene 1	Scene 2	Scene 3	Avg
Light_intensity	0.4	0.4	0.45	0.42
Light_type	0.5	0.5	1.0	0.67
Shadow_direction	0.0	0.0	0.0	0.0
Shadow_intensity	0.45	0.55	0.45	0.48
Cloth_basecolor	0.25	0.55	0.85	0.55
Cloth_roughness	0.15	0.30	0.45	0.30
Hair_basecolor	0.25	0.45	0.55	0.42
Hair_roughness	0.15	0.25	0.25	0.22
Pants_roughness	0.25	0.25	0.15	0.22
Shoes_roughness	0.55	0.55	0.25	0.45
Skin_basecolor	0.45	0.45	0.45	0.45
Skin_roughness	0.25	0.5	0.25	0.33

The primary bottleneck lies in achieving visual fusion – the perceptual coherence between real and virtual elements. Current designer workflows face two fundamental limitations: 1) Legacy 3D scanning pipelines preserve original scene lighting artifacts in scanned avatars, forcing designers to manually adjust material properties across different environments [Fan et al. \(2017\)](#); 2) Existing rendering engines and development systems prioritize physical accuracy over human perception, leaving designers to empirically test countless parameter combinations (lighting intensity, shadow softness, material roughness, etc.) through trial-and-error [Wei and Luximon \(2024\)](#). This disconnect between technical rendering outputs and human visual perception creates a “designer’s dilemma” – many MR artists face challenges in understanding which parameter configurations truly maximize perceived realism [Rokhsaritalemi et al. \(2020\)](#); [Kent et al. \(2021\)](#). Without an explicit model of how physical parameters relate to perceptual judgments, it is difficult to systematically reach an optimally realistic setting [Kyriltsias and Michael-Grigoriou \(2022\)](#).

Three critical technical barriers exacerbate these challenges. First, lighting and material inconsistencies between scanned assets and virtual environments persist due to baked-in illumination from source scans, making it unclear what the underlying, lighting-independent material properties should be. Second, the perceptual gap between physically accurate rendering and human visual preferences forces designers to develop heuristic adjustment strategies, which are often subjective and hard to generalize. Third, existing tools lack integrated solutions for scene-adaptive parameter analysis, making it difficult to compare physical parameters with user preferences and to reason about where perceptual optimality lies across different MR scenarios.

To address these issues and to better understand how optimal visual realism should be parameterized in MR, we present AIMERS, an AI-assisted MR scene design system that focuses on aligning physically grounded parameters with human perception. Our key idea is to first obtain reliable physical realism factors from real-world visual input, and then systematically capture how users adjust these

parameters in immersive MR environments, so that we can derive parameter ranges that correspond to high perceived realism. To operationalize this idea, AIMERS consists of three tightly connected components.

AI-Based Neural Inverse Rendering: Leveraging a multi-stage neural network framework with diffusion priors, we automatically reconstruct 3D avatars from multi-view images and recover physically-based material properties (diffuse albedo, roughness, and metallic), while eliminating the influence of environmental lighting. This provides a set of lighting-independent, physically grounded realism factors that can be consistently compared across different MR scenes and environments.

Human-Centric Perceptual Parameter Capture: We design an MR user perception system to capture perception-aligned realism parameters directly within immersive experiences. Specifically, we map over 40 technical parameters into 8 intuitive controls across three categories: Global Perception (environment light blending intensity and type, shadow direction and intensity), Material Adaptation (part-aware material adjustments such as skin, clothing, hair, pants, and shoes), and Scene Context (e.g., classroom, studio). These controls allow users to adjust visual settings based on their subjective sense of realism with real-time feedback, enabling us to record the parameter configurations that users consider visually most convincing.

Optimal Realism Parameter Ranges: Based on iterative adjustments collected from 20 participants across multiple MR scenarios, we statistically derive optimal parameter ranges (e.g., preferred intervals for light intensity or skin roughness) that maximize perceived visual fusion. Rather than treating realism as a single fixed setting, these ranges characterize where perceived realism is consistently high and reveal systematic deviations between physically derived parameters and perceptual preferences. The resulting parameter intervals serve as perception-aligned guidance for configuring MR scenes, grounded in both physical measurements and user studies.

Rigorous evaluations demonstrate that the realism parameter ranges obtained through AIMERS lead to higher visual fusion scores compared with purely physically derived configurations. Our analysis further confirms that physically accurate parameters and perceptually optimal parameters are not always aligned, and that the derived ranges capture stable, perception-consistent regions in the parameter space. By unifying inverse rendering with perceptual measurement, AIMERS helps resolve the “designer’s dilemma” from the perspective of realism quality: it clarifies how technical precision and human-centric visual fusion relate, and provides principled guidance for achieving convincing MR realism.

For the selected three scenarios, two types of classroom scenes are adopted as education is a vital application of MR technology and there are different classroom environments [Tang et al. \(2020\)](#); [Patel et al. \(2020\)](#). The other one, the MR studio scene, is chosen to cover virtual scenario diversity and observe how real-person hosts are perceived in a virtual broadcasting setting, as currently, some weather forecasts online or on televisions are given by real hosts in MR studios [Wang \(2024\)](#).

In summary, the main contributions of this work include:

- We utilize AI-driven methods to reliably extract physically grounded realism factors from real-world visual input.

- We introduce a user-centered system for capturing perception-aligned realism parameters in immersive MR environments.
- We formulate optimal visual realism parameter ranges that integrate physical accuracy with human perception.

2 Related work

2.1 MR visualization development

Mixed Reality is a class of simulators that combines both virtual and real objects to create a hybrid of the virtual and real worlds [Ohta \(1999\)](#). MR visualization acts as a bridge between virtual content and the real world, forming the core of the entire MR system. In recent years, advancements in graphic rendering algorithms and GPU hardware have made the blending of virtual objects with the real-world environment on MR devices increasingly natural and realistic. Enhancing light and shadow models necessitates the incorporation of temporal and dynamic effects, such as sudden changes in lighting and dynamic scenes, as well as the classification of algorithms tailored to different light paths [Marques et al. \(2018\)](#). [Gierlinger et al. \(2010\)](#) proposed a real-time rendering engine specifically tailored to the needs of MR visualization, which utilize image-based techniques for lighting and material acquisition allows for consistent integration of virtual objects into real-world environments. recommended using the fuzzy logic model to add soft shadows to a virtual object during embedding in a real scene. [Nasr Eddine and Junjun \(2019\)](#) used a holographic approach using georeferenced raster-based data integrated into a virtual world to enhance geographic visualization and data observation, with experiments conducted on the HoloLens to improve geographic edutainment. [Zhou and Zhou \(2023\)](#) proposed a mixed reality (MR) video fusion framework that dynamically projects video images onto 3D models as textures, utilizing remote rendering and browser-based implementation to overcome client limitations and reduce computational and bandwidth demands.

2.2 AI-driven inverse rendering

For inverse rendering, accurately Lighting estimation, particularly in indoor scenes, is a complex and essential task. Most current illumination estimation methods operate on single images, with a primary emphasis on integrating virtual objects into real images rather than making substantial alterations to the scene's illumination [Karsch et al. \(2011\)](#); [Garon et al. \(2019\)](#); [Zhan et al. \(2021\)](#). While traditional methods like a single environment map [Gardner et al. \(2017\)](#); [LeGendre et al. \(2019\)](#) and spherical lobes [Garon et al. \(2019\)](#) have been used, they often neglect spatial variations and high-frequency details in lighting. Recent innovations [Song and Funkhouser \(2019\)](#); [Srinivasan et al. \(2020\)](#) have attempted to improve 3D lighting representation, but still grapple with challenges like spatial instability and the lack of HDR information. [Li et al. \(2020\)](#) propose per-pixel spatially-varying spherical Gaussians (SVSG) lighting representation to capture high-frequency effects and demonstrate that SGs are superior to spherical harmonics (SH) for depicting lighting

details in indoor scenes. Neural-PIL [Boss et al. \(2021b\)](#) proposes a pre-integrated lighting network based on image-based lighting (IBL), showing better performances on conveying global illumination than SGs and SH. Hence, we utilize a neural HDR-radiance field to represent the IBL at any spatial point, thereby ensuring a more accurate and detailed depiction of indoor lighting scenarios with a focus on physically accurate HDR lighting prediction.

Material estimation in inverse rendering can be categorized into two levels: object level and scene level. Object-level estimation [Zhang et al. \(2021\)](#), [Zhang et al. \(2022\)](#); [Munkberg et al. \(2022\)](#); [Liang et al. \(2022\)](#); [Boss et al. \(2021a\)](#) focuses on individual objects, often in controlled or simplified environments. Object-level approaches are generally less complex, as they deal with fewer variables and more straightforward lighting conditions. In contrast, scene-level material estimation [Choi et al. \(2023\)](#); [Li et al. \(2023\)](#) is significantly more challenging due to the complexity and variability of entire scenes. This includes diverse lighting, multiple objects with different materials, and shadows.

The complexity of scene-level material estimation is further compounded by the choice between single-view and multi-view approaches. Single-view material estimation [Li et al. \(2020\)](#); [Gardner et al. \(2017\)](#), despite its simplicity and lower data requirements, often faces the ill-posed issue, where insufficient information leads to ambiguous or inaccurate estimations. This is particularly evident in complex scenes where a single viewpoint cannot capture the entirety of the scene's lighting and material properties. In contrast, multi-view material estimation [Choi et al. \(2023\)](#); [Zhang et al. \(2022\)](#), [Zhang et al. \(2021\)](#); [Munkberg et al. \(2022\)](#) leverages images from multiple viewpoints, providing a more comprehensive understanding of the scene. It can significantly reduce the ambiguity associated with single-view estimations, allowing for more accurate and reliable material property extraction. Our work utilizes multi-view images for material estimation, specifically addressing the challenges at the scene level.

2.3 Visual perception in MR

Since vision is crucial for perception, achieving the best fusion requires considering the overall impact of visual integration. Consequently, visual perception in MR systems has consistently been a primary research topic. [Fleming \(2014\)](#) proposed material perception plays a crucial role in the visual system by allowing us to effortlessly recognize and distinguish materials despite varying appearances. [Zhdanov et al. \(2019\)](#) explored virtual prototyping methods to assess and mitigate visual discomfort in AR, VR, and MR systems, addressing issues like vergence-accommodation conflicts and illumination differences. However, it may still face limitations in fully capturing the complexities of real-world visual perception. [Petikam et al. \(2018\)](#) study the relationship between real-world depth fidelity and visual quality in MR rendering, providing perceptual thresholds for various composition artifacts through user experiments. However, it is limited by its focus solely on depth information, without analyzing other rendering factors, which may constrain its applicability in real-world scenarios. [Potemin et al. \(2018\)](#) proposing a virtual prototyping approach to analyze visual perception problems in augmented and mixed

reality devices, comparing physically correct images with expected ones. However, it is limited by its reliance on physical results without further analyzing the impact on human visual perception. Du et al. (2024) proposed how the integration of subtle visual cues, such as shadows, lighting, textures, blur, and distortions, in MR interfaces can enhance user experience by making digital elements appear as natural components of the physical environment. However, it is limited by its focus on how virtual objects blend into real-world settings, without further exploring how real objects can be seamlessly integrated into virtual environments to improve overall MR fusion.

3 Methods

To address the challenge of achieving perceptually convincing visual realism in MR scenes, we develop a system that integrates AI-driven physical parameter extraction with perception-aligned user interaction. Our goal is to understand how rendering parameters should be configured so that virtual and real elements appear visually coherent to human observers. The overall pipeline of our mixed reality design system (AIMERS) as shown in Figure 1.

We first observe that directly placing 3D-scanned real elements into MR environments often preserves residual lighting and shadows from the source scene, which conflicts with the new environment and degrades perceived realism. To isolate physically meaningful scene attributes, we employ a deep learning-based neural inverse rendering approach to reconstruct the geometry and material properties of real elements while removing baked illumination effects (Section 3.2). This yields lighting-independent, physically grounded realism factors that serve as a baseline reference across MR scenes.

However, physically correct parameters alone do not necessarily correspond to users' perceived realism. To explicitly capture perceptual preferences, we design a human-centered MR interaction system that allows users to adjust realism-related parameters directly inside immersive environments. Over 40 technical rendering controls are mapped into 8 intuitive parameters covering global lighting, material behavior, and scene context. Users adjust these parameters with real-time feedback, enabling us to record the configurations they perceive as most visually realistic (Section 3.3).

Based on the realism settings collected from 20 participants across multiple MR scenarios, we then analyze the relationship between physically derived parameters and perceptually preferred values. Instead of predicting a single "optimal" configuration, we derive perception-aligned parameter ranges that consistently correspond to high visual realism. These ranges reveal where perceptual realism deviates from purely physically defined settings and provide principled guidance for MR scene configuration.

Finally, we evaluate the derived parameter ranges in representative MR applications, assessing visual fusion quality and alignment with user perception. Our results confirm that the perception-aligned intervals lead to more convincing visual realism than physically derived parameters alone, demonstrating the practical value of combining AI-based inverse rendering with user-driven perceptual measurement.

3.1 Neural inverse rendering for real-scanned avatar

When building MR systems, we typically embed 3D models of real-scanned elements into virtual 3D models. However, we have observed that since real-scanned objects often carry the lighting and shadow effects of the original scene, this residual light and shadow create disconnection in the new scene. Therefore, we consider how to eliminate the impact of the original lighting. Given that the material is not affected by lighting, we can obtain the material properties of the real elements to remove the interference of light and shadow. Nevertheless, compared to virtual objects (designed by artists), it is usually difficult to directly obtain the display material parameter values of real objects. Leveraging the recently developed Ref-GS technique Zhang et al. (2025), an AI-based inverse rendering approach built on differentiable neural rendering method, we achieve efficient PBR material and geometry reconstruction of avatars directly from multi-view images. This established method enables rapid decomposition of visual appearance into intrinsic material properties and geometric attributes, providing a practical solution for high-quality avatar creation in mixed reality applications.

Given a set of posed RGB images of a real-world human avatar, our goal is to accurately decompose the geometry, materials (Basecolor/Roughness) under unknown illumination, which can be used in re-render process with new environment light. Based on Ref-GS technique Zhang et al. (2025), we firstly use the 3D-GS Kerbl et al. (2023) to represent the avatar's 3 days Neural Radiance Field, which include global illumination and geometry property. And then, we use the additional Neural Network MLP to represent a material field. Meanwhile, to eliminate residual shadow and efficiently improve the quality of material, we introduce a diffusion-based material estimation prior as material regularization, the pipeline as shown in Figure 2.

After obtaining precise geometry, we utilized differentiable rendering techniques to optimize materials, include basecolor \hat{A} and roughness \hat{R} . The rendering equation can be written as Equation 1:

$$L_o(\hat{\mathbf{x}}, \omega_o) = \int_{\Omega^+} L_i(\hat{\mathbf{x}}, \omega_i) \left(\frac{\hat{A}}{\pi} + f_s(\hat{\mathbf{x}}, \hat{R}, \omega_i, \omega_o) \right) (\omega_i \cdot \hat{\mathbf{n}}_x) d\omega_i \quad (1)$$

where $\hat{\mathbf{n}}_x$ is normal at surface point $\hat{\mathbf{x}}$, ω_i is light incident direction, ω_o is view direction.

3.2 Human-centric perception adjustment system for MR

Visual realism in VR is often achieved by estimating physically based parameters such as lighting, materials, and geometry. Many VR design workflows assume that physically accurate rendering will naturally lead to realistic visual experiences. However, observations from realistic VR applications show that this assumption does not always hold. Even when physical parameters are correctly estimated, users may still perceive the scene as unrealistic.

This mismatch suggests that physical realism and human perceptual realism are not always aligned in interactive VR

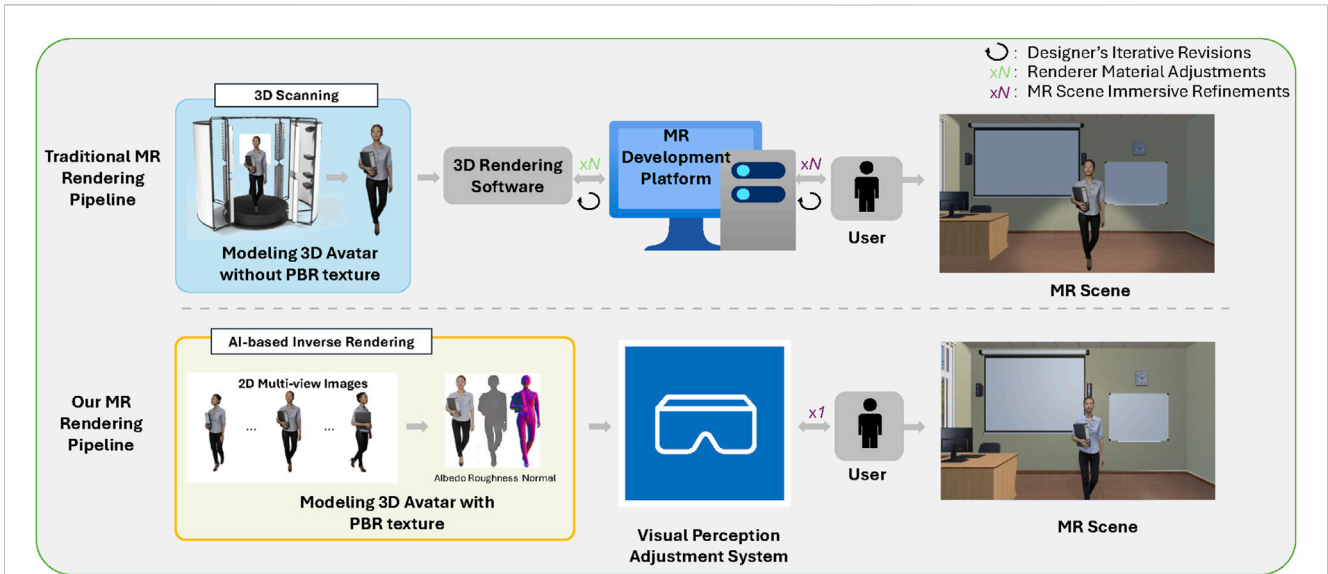


FIGURE 1 The overall pipeline of our mixed reality design system (AIMERS), combining an AI-based inverse rendering model and a human-centric visual perception adjustment system.

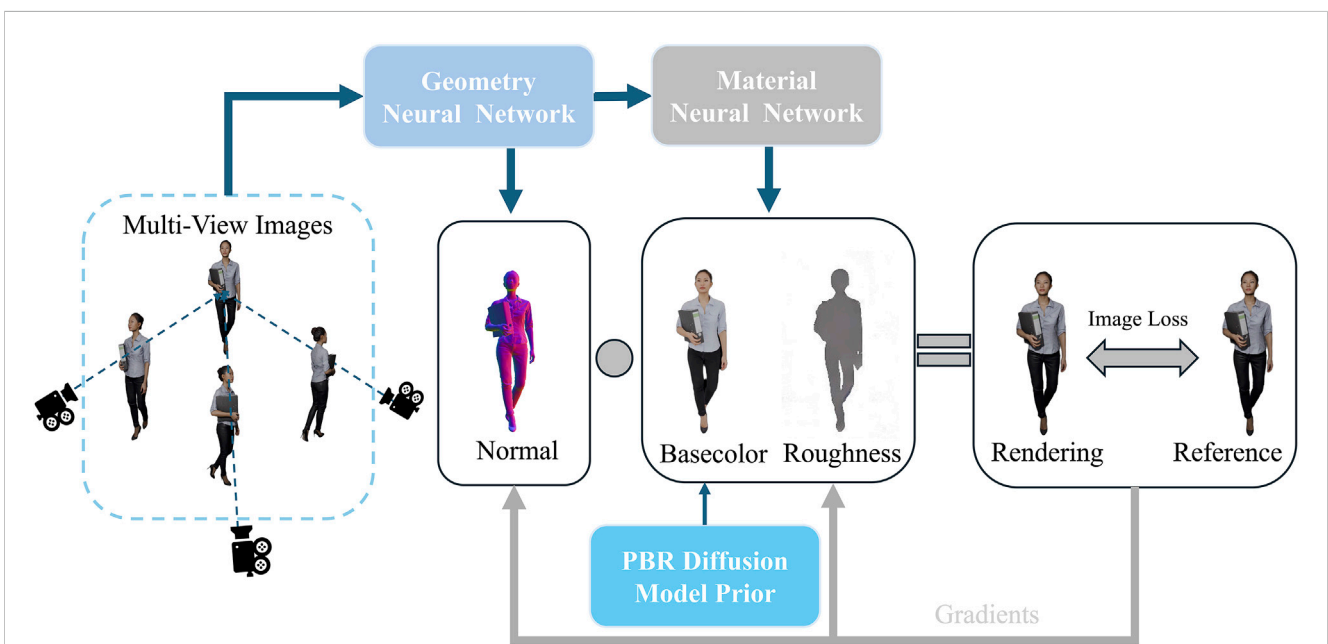


FIGURE 2 The inverse rendering pipeline consists of two phases: 1) geometry estimation: we input multi-view images from scanned avatar and obtain surface normal information by 3dgs-based neural rendering; 2) material estimation: we jointly optimize basecolor and roughness with diffusion prior.

environments. Based on this observation, this study is guided by the following hypothesis:

H1: Physically derived visual realism parameters do not always align with users' perceptual judgments of realism in realistic VR scenes.

Testing this hypothesis requires access to realism parameters as perceived by users during interaction. However, existing VR systems mainly support physically based rendering and rely on parameters

predefined by designers. They do not provide mechanisms for users to directly adjust realism-related parameters or to record perception-aligned values.

To address this gap, this study introduces a user-centered VR system that allows users to freely adjust visual realism parameters within realistic VR scenes. Through interactive adjustment, the system captures parameter values that reflect users' perceptual preferences rather than physical correctness alone. These perception-aligned

parameters enable direct comparison between physical and perceptual realism. This system includes global and local variables, and allows users to adjust different variable parameters to obtain the optimal parameter combination to achieve optimal immersion in MR scenes, as shown in Figure 3. User studies are conducted to compare physically derived parameters with user-adjusted parameters across three realistic VR scenarios: a Classroom, a Computer Room, and a Studio. These scenes represent common and socially relevant uses of realistic VR. The experimental results confirm the proposed hypothesis and provide the foundation for building an optimal visual realism parameter model in later studies.

Notably, inspired by Wei and Luximon (2024) and Disney's PBR model, we select an group significant variables for visual fusion adjustment.

Global Variables : To ensure the avatar and background blend seamlessly, we need to blend the lighting and shadows of the avatar with those of the background scene Hughes et al. (2004). Therefore, we designed the lighting and shadow variables of the environment as global control variables for blending. However, we assume there might be a gap between the physical rendering results and human perception in some variables. For example, the direction of the shadows on the rendered avatar may not necessarily match the optimal blending perception from a human perspective.

- **Light Types:** This is a discrete variable. And here, we simplify the types of lighting by using general illumination in MR, include point light, direction light and spot light.
- **Light Intensity:** This is a continuous variable. We consider that different lighting intensities can affect the visual blending effect, we selected a moderate range of lighting intensities and normalized it to a scale from 0 to 1. As the intensity increases, the ambient light becomes brighter.
- **Shadow Intensity:** This is a continuous variable. Although there is a linear relationship between shadow intensity and lighting intensity, our previous work has shown that human visual perception is quite sensitive to variations in shadow intensity under the same lighting conditions. Therefore, we designed shadow intensity as a separate variable to determine the range that best aligns with human perception of seamless blending.
- **Shadow Direction:** This is a discrete variable. Here, based on the earlier assumption that the shadow direction obtained from physical rendering may not be optimal for seamless blending, we designed a test range for shadow direction derived from the directly rendered shadows.

Local Variables : Since the avatar is essentially a complex composite, made up of multiple parts and different materials, and because user have varying levels of visual perception for different parts, we assume that the visual blending of different parts affects the overall blending effect in an MR scene. We divided the avatar into five parts: hair, skin, top, pants, and shoes. For each part, we designed two local variables: roughness and base color saturation. As pants and shoes are black and therefore the base color saturation is not adjusted for pants and shoes.

- **Roughness:** This is a continuous variable. Roughness describes the degree of smoothness of the material for the adjusted part.

In our MR environment, we consider that the roughness calculated through inverse rendering may not fully achieve optimal blending. Therefore, the default parameter value for roughness is set to the result from inverse rendering. We then adjust this parameter based on the physical results, ultimately aiming to find the optimal roughness range for each part that aligns with the majority of users' perception of seamless blending.

- **Base-color Saturation:** This is a continuous variable. Since determining the optimal color for each part of the avatar is very challenging, given the infinite possibilities for base colors like those of clothing, we opted to analyze the saturation of the base color to understand its impact on visual blending. This approach simplifies the analysis to some extent, while also providing valuable insights for designing the colors of avatars in the MR system.

Furthermore, we use Participatory Design Method to obtain the optimal MR fusion system for our research scene. By allowing multiple users to adjust the parameters of all variables to achieve their perceived optimal fusion effect, we can continuously refine our optimal fusion parameter range through this process. The optimal equation is shown in Equation 2:

$$D_{in} = \frac{\sum_{x \in [L,R]} 1}{(R-L)}, D_{out} = \frac{\sum_{x \notin [L,R]} 1}{1 - (R-L)}, \quad (2)$$

where L and R denote the left and right boundaries of the test interval, respectively. The term $\sum_{x \in [L,R]} 1$ represents the total number of data points within the interval $[L, R]$. For each data point x inside this interval, 1 is added. The denominator $(R - L)$ is the width of the interval $[L, R]$. Therefore, D_{in} is defined as the number of data points inside the interval $[L, R]$ divided by the width of the interval, which is the density of data points inside the interval, and D_{out} is external density.

We define the objective function $f(L, R)$ which quantifies the difference in data point densities inside and outside the interval $[L, R]$, and then we optimization goal is to find the optimal as shown in Equation 3, and the objective function is Equation 4.

$$f(L, R) = |D_{in}(L, R) - D_{out}(L, R)| \quad (3)$$

$$\delta^* = \arg \max_{\delta} f(P - \delta, P + \delta) \quad (4)$$

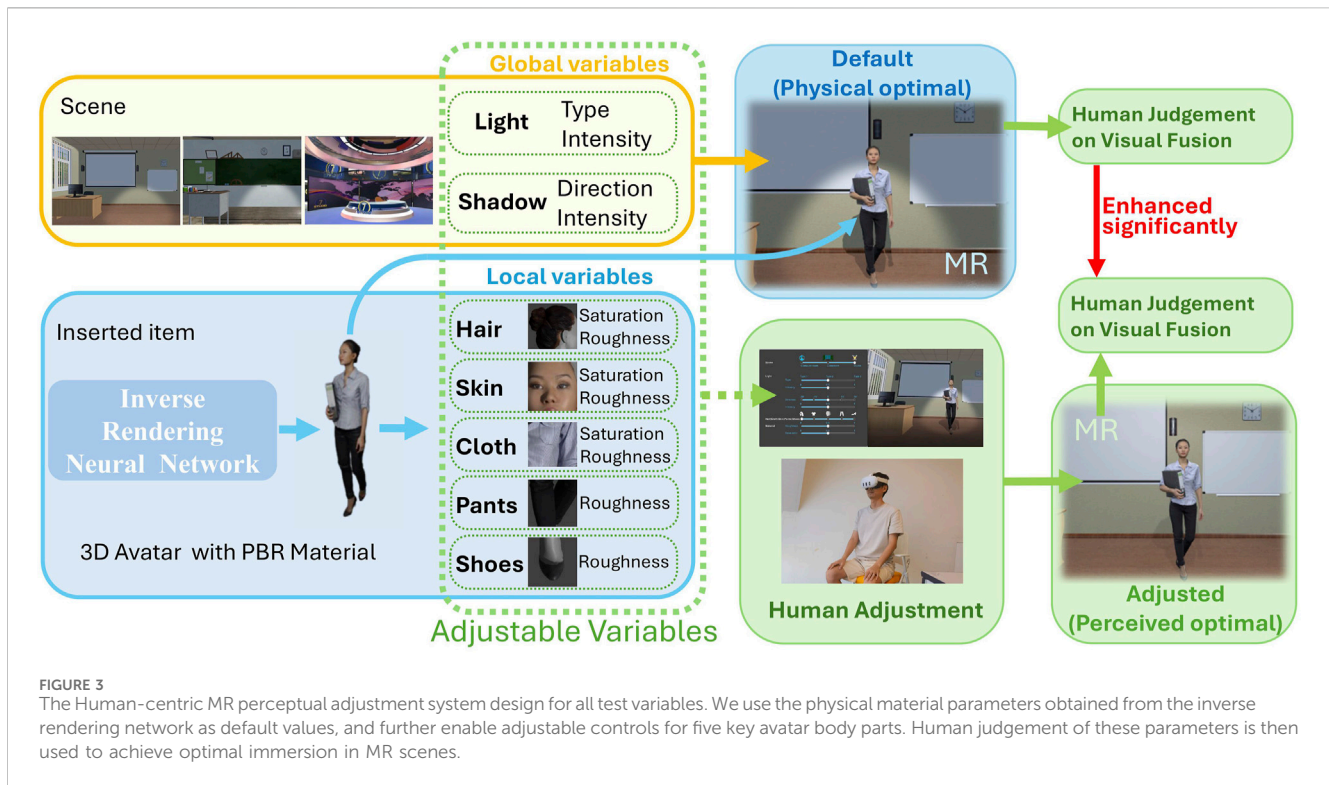
where P is the center of the densest region in the histogram, representing the peak of the distribution. δ is an offset used to explore intervals of different widths centered at P . By optimizing this objective $f(L, R)$, we can find the optimal offset that maximizes the density difference between inside and outside the interval δ^* , thereby determining the corresponding optimal interval $[L^*, R^*]$.

4 Experiments

4.1 Study 1: deriving physical visual realism factors for MR

4.1.1 Experiment setting

We use a two-phase training strategy to implement our inverse rendering method, which includes the geometry and radiance field,



and material estimation. Our input is multi-view images for real-scanned avatar, and output is normal, basecolor and roughness of the avatar. All experiments are conducted on a single NVIDIA GeForce RTX 3090 GPU. The full AI-based inverse rendering process requires approximately **5 hours** to complete. The details are as follows:

Stage 1: Geometry Estimation. We jointly optimize the geometry network F_d and radiance network F_c in this stage following VolSDF Yariv et al. (2021). We use two separate MLPs for F_d and F_c , each consisting of 4 layers of 256 hidden units with a rectified linear unit (ReLU) activation function. In addition, we encode the input surface position $\hat{\mathbf{x}}$ with 10 levels of periodic functions, respectively, before feeding them into our network. We optimize our geometry and HDR-radiance network for 250K iterations with a batch size of 1,024 in this stage, which takes about 3 h for an object. We design the optimize loss is \mathcal{L}_1 by Equation 5.

$$\mathcal{L}_1 = \mathcal{L}_{render} + \lambda_{eik} \mathcal{L}_{eik}, \tag{5}$$

where \mathcal{L}_{render} is $\|L_o - \hat{L}_o\|_1$, and we set weights $\lambda_{eik} = 0.1$. Here, L_o is ground truth rgb and \hat{L}_o is predicted rgb.

Stage 2: Material Estimation. In the section of material estimation, the Material MLPs include four fully connected layers, each with 512 hidden units and ReLU activation functions. Following these four layers, the Material feature network is divided into separate basecolor and roughness layers with 512 hidden units and Sigmoid activation. Specifically, basecolor layer outputs the basecolor term $\hat{A} \in \mathbb{R}^3$ and roughness layer outputs the roughness term $\hat{R} \in \mathbb{R}^1$ at surface position $\hat{\mathbf{x}}$. In this stage, we optimize our material model for 25K iterations with a

batch size of 256, which takes approximately 2 h for a single object. The optimize loss is L2 by Equation 6.

$$\mathcal{L}_2 = \mathcal{L}_{render} + \lambda_{basecolor} \mathcal{L}_{basecolor} + \lambda_{rough} \mathcal{L}_{rough}, \tag{6}$$

where $\mathcal{L}_{basecolor}$ and $\mathcal{L}_{roughness}$ is the MSE Loss between predicted value and diffusion prior value. We set weights $\lambda_{eik} = 0.1$, $\lambda_{normal} = 0.001$, $\lambda_{basecolor} = 0.0003$, $\lambda_{rough} = 0.001$ in our experiments.

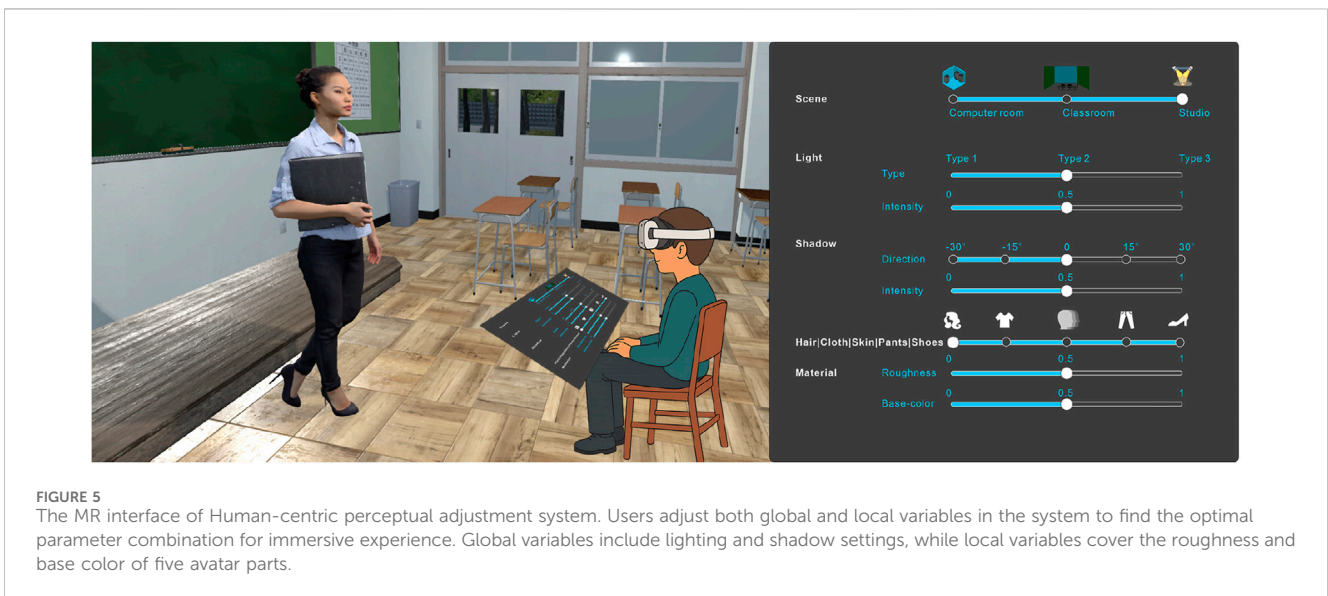
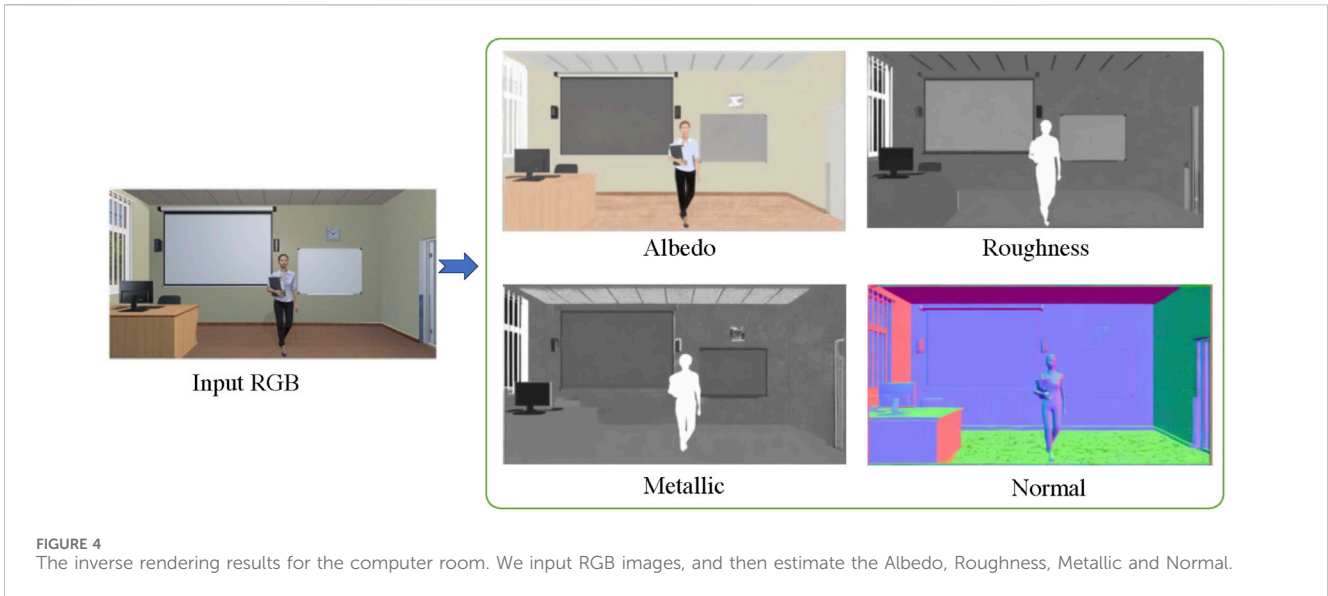
4.1.2 Result

After optimizing geometry and material neural network, we can get the rendering factor of the scene and avatar, which can delight and remove shadow for the real-scanned avatar from raw scene illumination effect. And we can use the clean basecolor and geometry to relighting in our MR system. We compare the effect between raw rgb input and recovered material in Figure 4.

4.2 Study 2: human-centric perceptual optimization for MR scene design

4.2.1 Experiment setting

To reduce reliance on designer experience and repeated testing in traditional MR scene design, we propose a human-centric system that helps designers efficiently create immersive experiences. Our approach further narrows the gap between physically rendered outputs and human perception in MR environments by introducing a parametric evaluation framework that systematically examines rendering factors through user-driven perceptual tuning. We define different variables in the



adjustment system, including global and local variables in **Figure 5**. Global variables are Light and Shadow, and local variables are Roughness and Base-color Saturation in different avatar body parts. All continuous variables are normalized to the range of 0–1. Discrete variables include the type of lighting and the direction of the shadows. We have selected three common types of lighting: Point Light, Direction Light, and Spot Light. Additionally, for the default shadow direction, we have adjusted by adding $\pm 15^\circ$ and $\pm 30^\circ$. And our system is developed based on Unity, and for different default types of light sources, we have selected three scenarios as test environments: a classroom, a computer room, and a studio. The classroom typically features a combination of indoor and outdoor lighting. The computer room is characterized by primarily indoor point lighting (here, we consider area lighting as a collection of multiple point sources), and the main light source in the studio is usually indoor spotlights. Furthermore,

we conducted a comparative time analysis between our optimized workflow and conventional development pipelines across three benchmark MR scenarios, which validate the efficiency of our system.

4.2.2 Questionnaire

The questionnaire is designed to collect users’ perceptual judgments of visual realism under different parameter settings and to validate **Hypothesis H1**, which states that physically derived realism parameters do not always align with human perception in realistic VR scenes.

The questionnaire consists of three core items, explicitly defined as Q1–Q3, which directly support the comparison between physically derived and perception-aligned realism settings. After completing the interaction for each scene, participants were asked to answer the following questions:

- Q1 (Physical Realism Rating): Participants rated the overall visual realism of the scene rendered under the Physical condition using a 7-point Likert scale, where 1 indicates very unrealistic and 7 indicates very realistic.
- Q2 (User-Tuned Realism Rating): Participants rated the overall visual realism of the scene rendered under the User-tuned condition, reflecting their preferred parameter configuration, using the same 7-point Likert scale.
- Q3 (Preference Judgment): Participants indicated which condition appeared more visually realistic by choosing between the Physical condition, the User-tuned condition, or no noticeable difference.

Questions Q1 and Q2 provide quantitative measures of perceived realism before and after user adjustment, while Q3 offers a direct qualitative comparison of perceptual preference between the two conditions. Together, these three items form the primary evidence used to evaluate the perceptual–physical mismatch in visual realism.

4.2.3 Participants

We recruited 20 participants (13 male, 7 female, ages 18–35) from our university. The recruitment was based on their knowledge of rendering and MR, which was screened using an eight-question quiz (provided in the [Supplementary Material](#)). Regarding MR and Rendering Software experience, half of the people have experience with professional rendering software, and 17 individuals have experience using MR-related equipment, seven of the participants use MR weekly. All participants had either normal vision or vision corrected to normal. For their time and contribution to the study, each participant was compensated with participant fee.

4.2.4 Data collection and procedures

By placing real avatars into different virtual scenarios, we have constructed a virtual reality experimental environment. Our user study followed a within-subjects design, with a usability testing method and Questionnaires. The study, on average, took less than 1 h. Participants were initially welcomed and reviewed a consent form. Then, we briefly introduced the study's objectives and procedural steps. The users wear the VR device, Quest 3, to observe and determine how to adjust the variables. Specifically, to simplify the process, the experimenters will adjust variable values based on the user's command until the user finds the optimal value. The details of this experiment is shown as follow:

Preparation and Training: We asked participants to adjust the chair height to a comfortable level and adjust the Quest 3 headset for the VR condition before they started the training session. We confirmed that the participant was comfortable and could see the content in all display environments clearly. And then, we first explain the concepts of visual fusion level and adjustable variables to the participants. Then, we will adjust these variables within the system to ensure that participants can perceive the feedback related to these variables before proceeding with the experiment.

System Adjustment Process: During the experiment, we test each scenario separately. Taking into account the relationship between global variables and local variables, the global variables may have an overall impact. Therefore, we first adjust the global

TABLE 2 Descriptive statistics of system User-tuned variable value for Scene 1–3.

Parameter	Scene 1	Scene 2	Scene 3	Avg
Light_intensity	0.5	0.5	0.4	0.47
Light_type	0.5	1.0	1.0	0.83
Shadow_direction	0.25	0.0	0.25	0.17
Shadow_intensity	0.45	0.55	0.55	0.52
Cloth_basecolor	0.25	0.65	0.85	0.58
Cloth_roughness	0.15	0.30	0.75	0.40
Hair_basecolor	0.25	0.40	0.65	0.43
Hair_roughness	0.15	0.35	0.25	0.25
Pants_roughness	0.25	0.25	0.25	0.25
Shoes_roughness	0.35	0.25	0.55	0.38
Skin_basecolor	0.45	0.65	0.65	0.58
Skin_roughness	0.25	0.85	0.25	0.45

variables, and then adjust the local variables of each body part. Additionally, we randomize the sequence of scenarios and test variables for each participant. We set default value is based on our physically inverse rendering model, and normalize our adjust value from 0 to 1. The physical variable value as shown in [Table 1](#). And the user-tuned variable value as shown in [Table 2](#).

4.2.5 Data analysis and results

In this section, we present the statistical analysis of our collected data, outline the strategies participants employed to manage the display space, and provide summarized qualitative feedback for each condition. We documented significance at levels of $p < 0.05$ (*). and all participants successfully completed the study tasks, resulting in no variance in accuracy metrics.

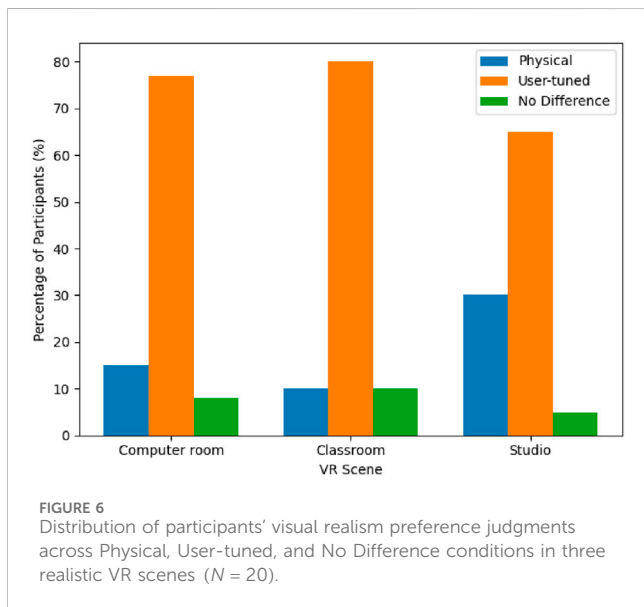
Preference Data Processing: [Table 3](#) summarizes participants' preference judgments between the Physical and User-tuned conditions across the three VR scenes. For all scenes, a larger proportion of participants reported that the User-tuned condition appeared more visually realistic than the Physical condition.

As illustrated in [Figure 6](#), participants across all three scenes showed a clear tendency to favor the User-tuned condition over the Physical condition, while the proportion of “No Difference” responses remained relatively low. Specifically, in Computer room, 77% of participants preferred the User-tuned condition, compared to 15% who favored the Physical condition, while 8% reported no noticeable difference. A similar pattern was observed in Classroom, where 80% of participants preferred the User-tuned condition and 10% preferred the Physical condition. In the Studio scene, the preference for the User-tuned condition was most pronounced, with 65% of participants selecting it as more visually realistic, compared to only 30% favoring the Physical condition.

Across all three scenes, the proportion of participants indicating no noticeable difference remained relatively low. These descriptive results suggest a consistent tendency for participants to favor perception-aligned parameter settings over physically derived

TABLE 3 Preference distributions between Physical and User-tuned conditions across VR scenes (N = 20).

Scene	Physical (%)	User-tuned (%)	No difference (%)
Computer room	15	77	8
Classroom	10	80	10
Studio	30	65	5



parameters. To determine whether these observed preference distributions reflect statistically significant deviations from random choice, formal statistical analyses were conducted and are reported in the following section.

Participants' preference judgments across the Physical, User-tuned, and No Difference conditions are summarized in Table 3. Across all three VR scenes, the User-tuned condition consistently received the highest proportion of selections, indicating a strong perceptual preference for user-adjusted realism parameters over physically derived values.

To examine whether these preference distributions deviated from random choice, chi-square goodness-of-fit tests were conducted for each scene. As reported in Table 4, the results show that preference distributions for Computer room, Classroom, and the Studio scene all significantly deviated from a uniform distribution across the three response options ($p < 0.001$). The corresponding effect sizes, measured by Cramér's V, ranged

from 0.55 to 0.70, indicating medium to strong effects. These results confirm that participants' realism judgments are not randomly distributed.

To further assess whether participants with a clear preference systematically favored perception-aligned settings over physically derived parameters, binomial tests were performed after excluding responses indicating "No Difference." The results demonstrate that, for all three scenes, participants selected the User-tuned condition significantly more often than the Physical condition. Strong preferences were observed in Computer room and Classroom ($p < 0.001$), while a statistically significant but more moderate preference was found in the Studio scene ($p = 0.02$).

Overall, these results provide consistent evidence that physically derived realism parameters do not fully align with human perceptual judgments in realistic VR scenes. Participants across different scenarios tended to prefer perception-aligned parameter configurations over physically computed values. These results support Hypothesis H1 and motivate the need to explicitly incorporate human perceptual factors in subsequent modeling and optimization of visual realism.

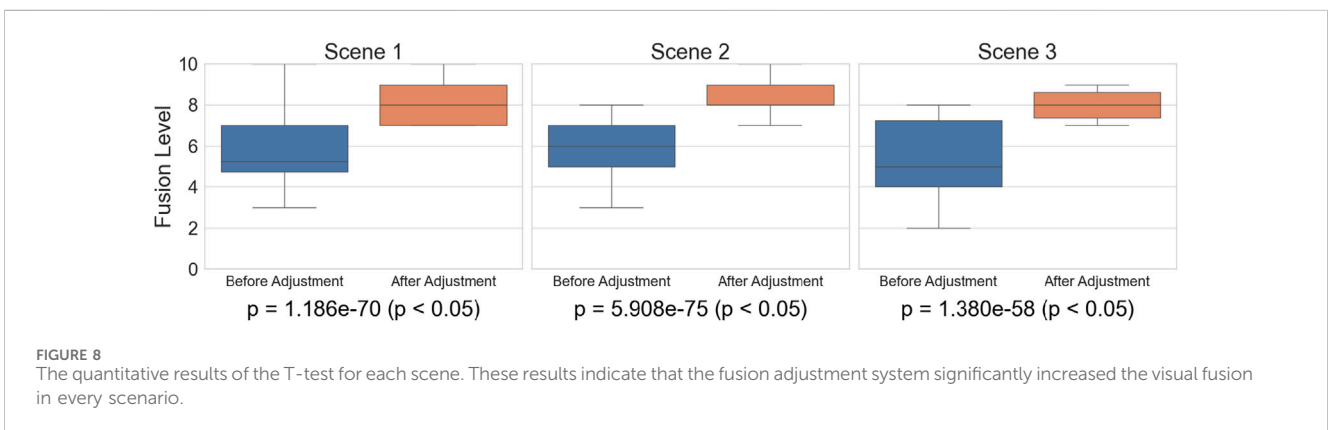
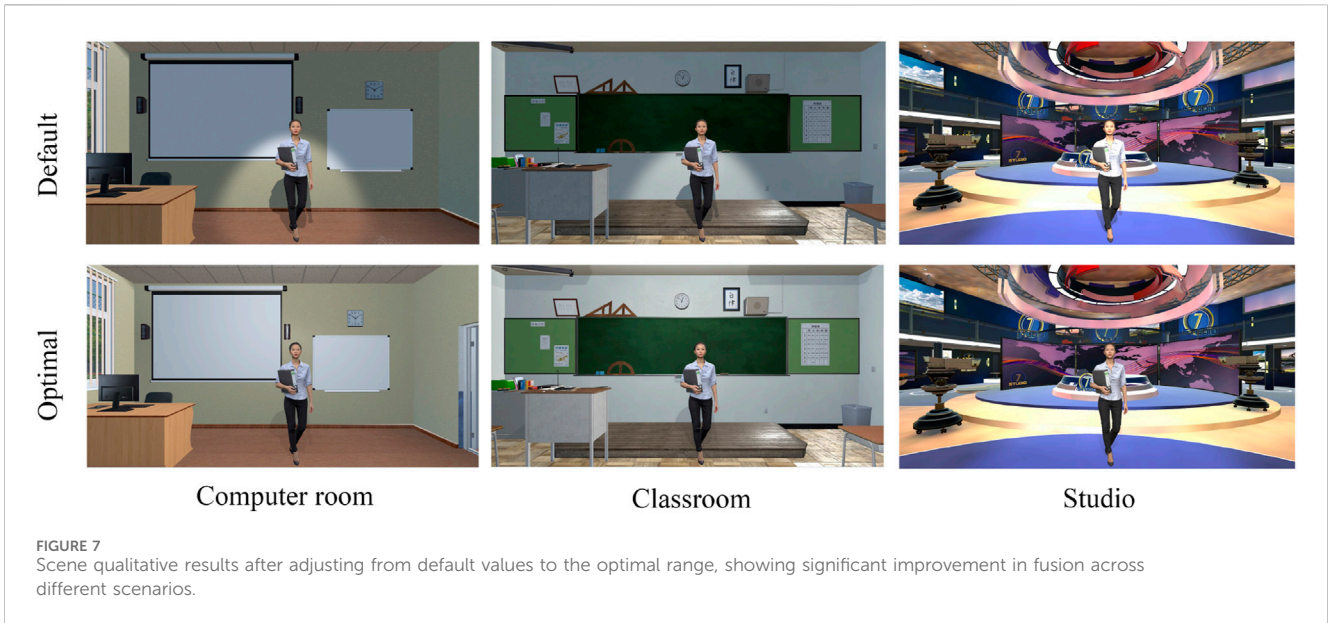
Statistical Analysis of MR Visual Fusion adjustment System: Our goal is to validate the effectiveness of our system between before and after Adjustment. Consequently, we formulated Null Hypothesis (H0) grounded in empirical findings from prior research and the testing conditions, and here, the H0 Hypothesis is no difference in the mean values before and after adjustment. We analyze the results for single scenarios and multiple scenarios separately. And here, we show the qualitative result after adjust in Figure 7.

- Single scenarios: we conducted T-tests on the data of all test participants to determine the significance in different scenarios. The results showed that the P-value for Scenario 1 was 1.186e-70; for Scenario 2, it was 5.907e-75; and for Scenario 3, it was 1.380e-58. The P-values for all scenarios were significantly less than 0.05, indicating that there are significant differences in the test data across these scenarios, as shown in Figure 8.

TABLE 4 Statistical analysis of participants' preference judgments between Physical and User-tuned conditions across VR scenes (N = 20).

Scene	χ^2	df	p-value	Cramér's V	Binomial p-value
Computer room	25.80	2	< 0.001	0.66	< 0.001
Classroom	29.40	2	< 0.001	0.70	< 0.001
Studio	18.20	2	< 0.001	0.55	0.02

Chi-square goodness-of-fit tests examine whether preference distributions across three response options (Physical, User-tuned, No Difference) deviate from a uniform distribution. Binomial tests compare User-tuned and Physical conditions after excluding "No Difference" responses. Statistical significance was assessed at $\alpha = 0.05$.



- Multiple scenarios: we also conducted an corresponding T-tests of the default and adjusted fusion scores for all three scenarios. The results showed that the P-value was 2.504e-196, which is definitely less than 0.05. This

indicates that the human adjustment makes a significant difference between the default parameters obtained through the physics-based inverse rendering method and the fusion data adjusted based on user

TABLE 5 Descriptive statistics of system variable ranges for Scene 1–3. The intersection range across all scenes. “None” means no shared range exists and each scene is set independently.

Parameter	Scene 1	Scene 2	Scene 3	Intersection
Light_intensity	(0.4, 0.6)	(0.4, 0.6)	(0.45, 0.55)	(0.45, 0.55)
Light_type	0.5	0.5	1.0	None
Shadow_direction	0.0	0.0	0.0	0.0
Shadow_intensity	(0.45, 0.55)	(0.45, 0.55)	(0.45, 0.55)	(0.45, 0.55)
Cloth_basecolor	(0.45, 0.55)	(0.45, 0.55)	(0.45, 0.55)	(0.45, 0.55)
Cloth_roughness	(0.15, 0.45)	(0, 0.30)	(0.45, 0.55)	None
Hair_basecolor	(0.25, 0.35)	(0.45, 0.55)	(0.45, 0.55)	None
Hair_roughness	(0, 0.25)	(0, 0.25)	(0.25, 0.35)	None
Pants_roughness	(0, 0.25)	(0.25, 0.35)	(0.15, 0.45)	None
Shoes_roughness	(0.45, 0.55)	(0.45, 0.55)	(0.25, 0.75)	None
Skin_basecolor	(0.45, 0.55)	(0.45, 0.55)	(0.45, 0.55)	(0.45, 0.55)
Skin_roughness	(0.15, 0.45)	(0.4, 0.6)	(0.25, 0.35)	None

perception in our adjustment system, as shown in Figure 9.

Optimal Range Estimation: Based on participants continuously adjusting system variables to achieve their optimal fusion, we expect to obtain the optimal parameter combinations for the test environment from these data. We first analyzed each scene separately. For the global and local variables of each scene, we used the method proposed in the previous section to perform the analysis and calculations. By optimizing the density difference inside and outside the interval, we obtained the optimal parameter ranges. Each parameter calculated for every scene is shown in the Table 5. We normalized the range of parameter values for all variables to 0–1. Meanwhile, we also analyzed the intersection range across all scenes and found that the range is the same for most of the scenes. More details in Appendix.

5 Findings and discussions

Our primary objective is to validate the effectiveness of our physics-based inverse rendering model and user perception-based MR system, in order to identify the optimal system parameter combinations that achieve the best possible integration. Although physics-based inverse rendering has rapidly advanced in achieving photorealistic quality in the implementation of MR systems, the physics-based rendering model does not entirely align with human perception of optimal visual integration in MR environments. Therefore, our research aims to develop an adjustable rendering system for MR that combines physical models with user perception experiences. By adjusting parameters based on feedback from multiple users, we seek to identify the MR system parameters that best meet user preferences, thereby achieving optimal integration in test scenarios. This will provide valuable reference for the development of existing MR system pipelines. Our

subsequent findings and discussions will focus on the interplay between physical rendering models and perceptual outcomes.

5.1 Does physically accurate rendering correspond to what users perceive as visually real?

The experimental results provide clear empirical evidence that this consistency does not always hold. Across all three scenarios, participants consistently favored perception-aligned parameter settings over physically computed values, supporting **Hypothesis H1**. These findings challenge the common assumption that physically accurate rendering alone is sufficient to achieve convincing visual realism in VR environments.

One possible explanation for this mismatch lies in the nature of human visual perception. While physically based rendering aims to simulate light transport and material interactions according to physical laws, human perception of realism is influenced by additional factors such as visual comfort, contextual expectation, and tolerance to physical inaccuracies. Users may therefore prefer parameter configurations that deviate from strict physical correctness but better match their subjective experience of what appears “natural” or “believable” in a given VR context.

The observed differences in preference strength across scenes further suggest that the perceptual–physical mismatch is context-dependent. In the Computer room and Classroom scenarios, strong preferences for the User-tuned condition indicate that users are particularly sensitive to realism-related parameters in structured indoor environments, where lighting consistency and material appearance strongly affect plausibility. In contrast, although the Studio scene also showed a significant preference for perception-aligned settings, the effect was comparatively more moderate. This may be attributed to greater perceptual tolerance in broadcast-style environments, where stylization and controlled lighting are more common and therefore less likely to violate user expectations.

Importantly, these findings highlight a fundamental limitation of current VR design workflows that rely exclusively on physically derived parameters. Without mechanisms to capture and incorporate user perceptual preferences, designers risk producing visually accurate yet perceptually unconvincing scenes. The results of this chapter demonstrate the necessity of treating perceptual realism as a first-class component in VR scene design rather than as a byproduct of physical simulation.

5.2 Is MR scene adjustable fusion system beneficial?

Our study results reveal that an MR system combining both physical and perceptual aspects can optimize users’ visual fusion in MR environments. We conducted evaluations across different scenarios and found that this approach is effective. During user testing, we had users observe the visual fusion between the inserted avatar and the environment. We found that, regardless of whether the MR environment was a classroom, computer room, or studio, users experienced a noticeable improvement in visual fusion before and after adjusting the MR system. This subjective feedback aligns

closely with findings from [Wei and Luximon \(2024\)](#), indicating that there is still a gap between users' perception of visual fusion in MR systems and the values derived from physics-based calculations. In comparison, we conducted a more in-depth exploration within the MR system, allowing users to make intuitive judgments, thereby providing more credible experimental results. Given the significant impact of ambient lighting on overall effectiveness, we selected test environments based on common lighting types. This allowed users to intuitively perceive the interaction effects between the inserted avatar and the MR environment, and based on their potential cognition of the scene and a certain type of lighting, we made our evaluations. During the testing process, we first had users observe the MR system designed with rendering factor values obtained through physics-based inverse rendering model. Users were then asked to judge the visual fusion in this state. The results showed that most users were dissatisfied with the current visual fusion, particularly noting that the global lighting variables and the clothing materials did not completely match the settings of the current scene, resulting in a noticeable sense of separation. We found that people tend to prefer point light sources as the optimal choice in classroom settings, while in studio environments, they lean towards spotlights. This reflects the differing lighting needs of various scenarios and highlights the sensitivity of human visual perception to these differences. It is understandable that classrooms typically require even lighting to ensure all students can clearly see the blackboard and other visual aids. In contrast, studios use spotlights to highlight specific objects or areas.

5.3 The optimal range of scenarios parameters

Another objective of our study is to identify the optimal parameters combination range for visual fusion in MR test scenarios. We conducted an optimal range analysis of all adjustable variables for each scenario according to [Section 4.2](#). Delving deeper into underlying factors, we explored the issue from two aspects. First, by comparing the optimal adjustment ranges for each scenario with the default values obtained from the physical-based inverse rendering model, we found that some variables were distributed near the default values, while others deviated. This indicates that solely relying on physical-based models to design MR systems is still insufficient to fully meet users' optimal perception of visual fusion. Specifically, most roughness-related variables, including `Cloth_roughness`, `Hair_roughness`, `Pants_roughness`, and `Skin_roughness`, have optimal ranges that are smaller compared to the default values in all scenes. This indicates that, from the users' perceptual standpoint, the avatar's overall roughness distribution in these test environments should appear rougher rather than more specular.

Secondly, we observed that the optimal ranges for the same variables may vary between different scenarios, while also sharing some commonalities. Although the default lighting types differ between computer bar and teaching room, most of the optimal ranges are similar since both scenarios are classroom types. The primary exceptions are the distributions of `Pants_roughness` and `Skin_roughness`, which show significant differences with almost no overlapping areas. This indicates that in the computer room and

classroom scenarios, the roughness of pants and skin requires special attention. In the computer room environment, users believe that lower values of `Pants_roughness` and `Skin_roughness` enhance the visual fusion. Meanwhile, compared to the other two scenarios, Studio scenario has fewer variables with overlapping optimal ranges due to its greater environmental differences. Only a few variables—`Light_intensity`, `Shadow_direction`, `Shadow_intensity`, `Cloth_basecolor`, and `Skin_basecolor` overlapping ranges, meaning they are less influenced by the environment and need only slight tuning for scene fusion.

5.4 Key factors influencing the degree of user fusion perception in MR

After participants completed the system adjustments, we also asked them to rank the importance of various variables. From the ranking data, we found that 77.78% of participants considered lighting to be the most important factor for global variables such as illumination and shadow. Regarding the impact of different body parts on the fusion effect of the avatar with the background environment, 77.78% of participants ranked skin as the most critical, while 50% considered shoes to have the least impact. Additionally, 44.44% ranked cloth as the second most important, with hair and pants having relatively similar weights in the middle rankings. Our findings are consistent with prior research work [Gonçalves et al. \(2023\)](#), both demonstrating that global lighting has the greatest impact on visual perception, followed by shadows. Meanwhile, users in MR environments perceive that the upper body of the avatar, including the skin and top garments, has a greater influence on their judgment of the fusion effect. This finding is also consistent with previous work [Van der Veer et al. \(2018\)](#).

6 Conclusion

Traditional MR scene design has largely depended on designer experience and repeated perceptual adjustments, making it difficult to reason about how rendering parameters should be configured to achieve convincing visual realism. In this work, we propose AIMERS, a perception-aligned MR realism framework that integrates neural inverse rendering with immersive perceptual parameter capture, with the goal of understanding where perceptual optimality lies in MR visual fusion.

We first utilize AI-based neural inverse rendering with diffusion-based priors, allowing us to reconstruct geometry and physically-based material attributes from real scenes while eliminating baked lighting artifacts. This provides lighting-independent baseline parameters for MR assets when inserted into new environments. Building on these physically grounded factors, we design a parameter adjustment interface derived from the physically-based rendering model and tailored for complex real-world avatars. The interface exposes global and local variables in a perceptually meaningful way, enabling users to explore and select visually convincing realism settings.

Through systematic perceptual experiments across multiple MR scenarios, we analyze the distributions of user-preferred configurations and derive optimal parameter ranges for visual

fusion rather than a single fixed solution. These ranges offer reliable design references and reveal systematic deviations between physically accurate and perceptually optimal settings. Our findings further show that global illumination and material properties of the upper body (especially roughness and base color) exert the strongest influence on perceived realism, while other variables have more limited contributions.

Overall, AIMERS reframes MR realism as a problem of aligning physical accuracy with human perception. By combining AI-based parameter extraction with perceptual measurement, our work provides principled guidance for configuring MR scenes and contributes empirical insight into how different rendering parameters shape the perception of visual fusion. We believe these results form a foundation for future research on perception-aware MR system design, inverse rendering, and visual realism modeling.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Faculty Research Committee (on behalf of PolyU Institutional Review Board). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

XW: Software, Investigation, Formal Analysis, Validation, Methodology, Writing – original draft, Data curation, Project administration, Visualization, Conceptualization. YW: Writing – review and editing, Validation, Conceptualization, Visualization. AZ: Validation, Writing – review and editing, Visualization, Data curation. YL: Formal Analysis, Resources, Visualization, Writing – review and editing, Supervision.

References

- Boss, M., Braun, R., Jampani, V., Barron, J. T., Liu, C., and Lensch, H. (2021a). “Nerd: neural reflectance decomposition from image collections,” in *ICCV*.
- Boss, M., Jampani, V., Braun, R., Liu, C., Barron, J. T., and Lensch, H. P. A. (2021b). “Neural-PIL: neural pre-integrated lighting for reflectance decomposition,” in *Proceedings of the 35th International Conference on Neural Information Processing Systems (NIPS '21)* (Red Hook, NY, USA: Curran Associates Inc.) Article 818, 10691–10704.
- Choi, C., Kim, J., and Kim, Y. M. (2023). IBL-NeRF: image-Based lighting formulation of neural radiance fields. *Comput. Graph. Forum.* 42 (7), doi:10.1111/cgf.14929
- Du, Y., Huang, X., and El-Zanfaly, D. (2024). “Subtle visual cues in mixed reality: influencing user perception and facilitating interaction,” in *Proceedings of the 16th Conference on Creativity and Cognition*, 556–560.
- Fan, S., Ng, T.-T., Koenig, B. L., Herberg, J. S., Jiang, M., Shen, Z., et al. (2017). Image visual realism: from human perception to machine computation. *IEEE Transactions Pattern Analysis Machine Intelligence* 40, 2180–2193. doi:10.1109/TPAMI.2017.2747150
- Fleming, R. W. (2014). Visual perception of materials and their properties. *Vis. Research* 94, 62–75. doi:10.1016/j.visres.2013.11.004
- Gardner, M.-A., Sunkavalli, K., Yumer, E., Shen, X., Gambaretto, E., Gagné, C., et al. (2017). Learning to predict indoor illumination from a single image. *ACM Trans. Graph.* 36 (6), 1–14. doi:10.1145/3130800.3130891
- Garon, M., Sunkavalli, K., Hadap, S., Carr, N., and Lalonde, J.-F. (2019). “Fast spatially-varying indoor lighting estimation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6908–6917.
- Gierlinger, T., Danch, D., and Stork, A. (2010). Rendering techniques for mixed reality. *J. Real-Time Image Process.* 5, 109–120. doi:10.1007/s11554-009-0137-x

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was supported by the Laboratory for Artificial Intelligence in Design (Project 3.1), the Innovation and Technology Fund of the Hong Kong Special Administrative Region, and Project P0050655 from the Non-PAIR Research Centers of The Hong Kong Polytechnic University. Their financial support is gratefully acknowledged.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frvir.2026.1733259/full#supplementary-material>

- Gonçalves, G., Melo, M., Monteiro, P., Coelho, H., and Bessa, M. (2023). The role of different light settings on the perception of realism in virtual replicas in immersive virtual reality. *Comput. and Graph.* 117, 172–182. doi:10.1016/j.cag.2023.10.021
- Hughes, C. E., Konttinen, J., and Pattanaik, S. N. (2004). The future of mixed reality: issues in illumination and shadows. 6–9.
- Karsch, K., Hedau, V., Forsyth, D., and Hoiem, D. (2011). Rendering synthetic objects into legacy photographs. *ACM Trans. Graph.* 30 (6), 1–12. doi:10.1145/2070781.2024191
- Kent, L., Snider, C., Gopsill, J., and Hicks, B. (2021). Mixed reality in design prototyping: a systematic review. *Des. Stud.* 77, 101046. doi:10.1016/j.destud.2021.101046
- Kerbl, B., Kopanas, G., Leimkühler, T., and Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.* 42 (4), 1–14. doi:10.1145/3592433
- Kyrlitsias, C., and Michael-Grigoriou, D. (2022). Social interaction with agents and avatars in immersive virtual environments: a survey. *Front. Virtual Real.* 2, 786665. doi:10.3389/frvir.2021.786665
- LeGendre, C., Ma, W.-C., Fyffe, G., Flynn, J., Charbonnel, L., Busch, J., et al. (2019). “DeepLight: learning illumination for unconstrained mobile mixed reality,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5918–5928.
- Li, Z., Shafei, M., Ramamoorthi, R., Sunkavalli, K., and Chandraker, M. (2020). “Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2475–2484.
- Li, Z., Wang, L., Cheng, M., Pan, C., and Yang, J. (2023). “Multi-view inverse rendering for large-scale real-world indoor scenes,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12499–12509.
- Liang, R., Zhang, J., Li, H., Yang, C., Guan, Y., and Vijaykumar, N. (2022). Spidr: Sdf-based neural point fields for illumination and deformation. *arXiv Preprint arXiv:2210.08398*.
- Marques, B. A. D., Clua, E. W. G., and Vasconcelos, C. N. (2018). Deep spherical harmonics light probe estimator for mixed reality games. *Comput. and Graph.* 76, 96–106. doi:10.1016/j.cag.2018.09.003
- Milgram, P., and Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* 77, 1321–1329.
- Munkberg, J., Hasselgren, J., Shen, T., Gao, J., Chen, W., Evans, A., et al. (2022). “Extracting triangular 3d models, materials, and lighting from images,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8280–8290.
- Nasr Eddine, A., and Junjun, P. (2019). “Geospatial data holographic rendering using windows mixed reality,” in *E-Learning and Games: 12th International Conference, Edutainment 2018, Xi’an, China, June 28–30, 2018 (Springer)*, 21–25.
- Ohta, Y. (1999). Mixed reality: merging real and virtual worlds
- Patel, S., Panchothiya, B., Patel, A., Budharani, A., and Ribadiya, S. (2020). A survey: virtual, augmented and mixed reality in education. *Int. J. Eng. Res. and Technol. (IJERT)* 9 (5), 1067–1072. doi:10.17577/IJERTV9IS050652
- Petikam, L., Chalmers, A., and Rhee, T. (2018). “Visual perception of real world depth map resolution for mixed reality rendering,” in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) (IEEE)*, 401–408.
- Potemin, I. S., Zhdanov, A., Bogdanov, N., Zhdanov, D., Livshits, I., and Wang, Y. (2018). Analysis of the visual perception conflicts in designing mixed reality systems. *Opt. Des. and Test. VIII (SPIE)* 10815, 181–194. doi:10.1117/12.2503397
- Rokhsaritalemi, S., Sadeghi-Niaraki, A., and Choi, S.-M. (2020). A review on mixed reality: current trends, challenges and prospects. *Appl. Sci.* 10, 636. doi:10.3390/app10020636
- Song, S., and Funkhouser, T. (2019). “Neural illumination: lighting prediction for indoor environments,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6918–6926.
- Srinivasan, P. P., Mildenhall, B., Tancik, M., Barron, J. T., Tucker, R., and Snavely, N. (2020). “Lighthouse: predicting lighting volumes for spatially-coherent illumination,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8080–8089.
- Tang, Y.-M., Au, K. M., Lau, H. C., Ho, G. T., and Wu, C.-H. (2020). Evaluating the effectiveness of learning design with mixed reality (mr) in higher education. *Virtual Real.* 24, 797–807. doi:10.1007/s10055-020-00427-9
- Van der Veer, A. H., Alsmith, A. J., Longo, M. R., Wong, H. Y., and Mohler, B. J. (2018). Where am i in virtual reality? *PLoS One* 13, e0204358. doi:10.1371/journal.pone.0204358
- Wang, Y. (2024). Application of virtual reality technology in video news reporting. *J. Electr. Syst.* 20, 160–166.
- Wei, X., and Luximon, Y. (2024). “Exploring factors influencing visual realism in augmented reality user experience,” in *International Conference on Human-Computer Interaction (Springer)*, 169–182.
- Yariv, L., Gu, J., Kasten, Y., and Lipman, Y. (2021). Volume rendering of neural implicit surfaces. *Adv. Neural Inf. Process. Syst.* 34, 4805–4815.
- Zhan, F., Zhang, C., Yu, Y., Chang, Y., Lu, S., Ma, F., et al. (2021). “Emlight: lighting estimation via spherical distribution approximation,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35, No. 4, 3287–3295. doi:10.1609/aaai.v35i4.16440
- Zhang, K., Luan, F., Wang, Q., Bala, K., and Snavely, N. (2021). “Physg: inverse rendering with spherical gaussians for physics-based material editing and relighting,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5453–5462.
- Zhang, Y., Sun, J., He, X., Fu, H., Jia, R., and Zhou, X. (2022). *Modeling indirect illumination for inverse rendering*. CVPR.
- Zhang, Y., Chen, A., Wan, Y., Song, Z., Yu, J., Luo, Y., et al. (2025). “Directional factorization for 2d gaussian splatting,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 26483–26492.
- Zhdanov, A. D., Zhdanov, D. D., Bogdanov, N. N., Potemin, I. S., Galaktionov, V. A., and Sorokin, M. I. (2019). Discomfort of visual perception in virtual and mixed reality systems. *Program. Comput. Softw.* 45, 147–155. doi:10.1134/s036176881904011x
- Zhou, Q., and Zhou, Z. (2023). Web-based mixed reality video fusion with remote rendering. *Virtual Real. and Intelligent Hardw.* 5, 188–199. doi:10.1016/j.vrih.2022.03.005