Mask Balancing: Perception-Driven Dynamic Visibility Enhancement for Occlusion-Capable Optical See-Through Head-Mounted Displays

Yan Zhang, Rundong Chu, Qingtai Dong, Xiaodan Hu, Keyao You, Zijie Zhou, Zixuan Guo, Hangyu Zhou, Kiyoshi Kiyokawa, and Xubo Yang

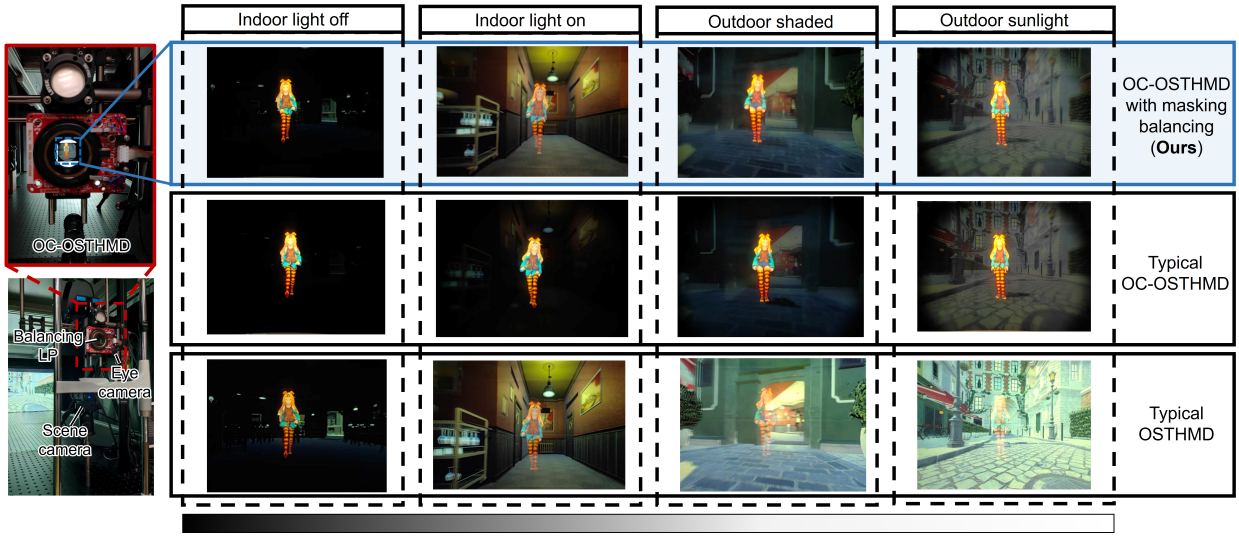

Fig. 1: We propose mask balancing for tackling the poor see-through transparency in existing OC-OSTHMDs. The figures in the most left column show the prototype of our system. It is built by upgrading a linear polarizer (LP) at the entrance pupil of OC-OSTHMD with a rotation mount. Then, the LP works as a balancing element to dynamically adjust the transmittance of the real scene and virtual images via deviating it cross-angle to the built-in polarizing beam splitter. The proposed method works in a perception-driven way. The visibility of users to the real scene and virtual images is tested at run-time, triggering the optimal rotation of LP. The right figures present the experimental results of OC-OSTHMD with mask balancing to typical devices. The real background is projected by the projector to simulate the illuminance rise from the indoor light off (~ 8 cd/m^2) to the outdoor sunlight (~ 414 cd/m^2) scenes. Among the three methods, OC-OSTHMDs with mask balancing show the optimized visibility for both the real and virtual images, which also agrees with the result of the user study. We believe the proposed method makes a solid improvement to OC-OSTHMDs.

Abstract— The poor transparency of occlusion-capable optical see-through head-mounted displays (OC-OSTHMDs) deteriorates the visibility of the real scene, hindering the practical application of the devices. Previous works mitigate the issue by upgrading the transmittance of the spatial light modulator (SLM). However, the strategy soon reaches a limit because further optimization requires improving the transmittance of all optical elements, e.g., lenses and beam splitters. Moreover, pixelated occlusion usually relies on polarizing the real scene light, inevitably cutting the input optical power by half. To overcome this limitation, we propose a mask balancing method that improves real-scene brightness through polarization blending. Specifically, the s-polarized component, which passes through the optical system to provide occlusion-capable vision, is blended with the p-polarized component, which bypasses the system to preserve the raw view of the real scene. The blending is realized by simply modulating the cross-angle between a polarizing beam splitter and a linear polarizer, benefiting the robustness and versatility of the proposed method. We introduce a perception-driven blending approach, where the cross-angle is optimized in real-time to balance the visibility of the real scene and the texture and lighting of the virtual object. A benchtop prototype is built. A user study with 12 participants is conducted to quantify the visibility threshold of the texture and lighting of virtual objects. Then, a user study with 12 participants proves that the proposed method improves the visibility of the real scene while keeping a good appearance of the virtual object. We believe the proposed method is an important step toward developing practical solutions for OC-OSTHMDs.

Index Terms—Augmented reality, mutual occlusion, optical see-through head-mounted displays, perception-driven optimization

1 Introduction

Occlusion-capable optical see-through head-mounted displays (OC-OSTHMDs) provide a promising solution to tackle the limitations of semi-transparent virtual images in typical augmented reality (AR)

- *Yan Zhang, Rundong Chu, Qingtai Dong, Keyao You, Zijie Zhou, Zixuan Guo, Xubo Yang (corresponding author) are with Shanghai Jiao Tong University, 200240, China. E-mail: {yan-zh | churundong | dqtiger24 | youkeyao | chou_tzu_chieh | zixuanguo | yangxubo}@sjtu.edu.cn.*
- *Xiaodan Hu is with Graz University of Technology, Graz 8010, Austria. E-mail: xiaodan.hu@tugraz.at.*
- *Hangyu Zhou is with Zhejiang University of Science and Technology, 310023, China, and also with Shanghai Jiao Tong University as a postdoctor. E-mail: zhouhy0571@sjtu.edu.cn.*
- *Kiyoshi Kiyokawa is with Nara Institute of Science and Technology (NAIST), Ikoma, Nara 6300192, Japan. E-mail: kiyo@is.naist.jp.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

eyeglasses. By modulating the transparency of incident light from the real scene on a pixel basis, OC-OSTHMDs are demonstrated to enhance AR graphics with more realistic depth cues, optimal visibility, vivid color appearance, etc. [16, 17].

A typical OC-OSTHMD consists of a real-scene focusing optical system and a spatial light modulator (SLM). The former works to converge the beam from the real scene into a light spot at the intermediate image plane, where the latter is controlled to block the light precisely [16]. Benefiting from the works in the past decades, OC-OSTHMDs have achieved impressive progress. Recent optical system allows the device to provide higher image quality [30], wider field of views (FOVs) [34, 35, 38], varifocal features [6, 7, 9, 26], smaller form-factor [18, 30, 37], and better color fidelity [33]. Meanwhile, advances in resolution and optical efficiency of SLMs facilitate the device occlude the real scene more precisely and with less optical power loss [18]. OC-OSTHMDs are approaching the practical application.

Nevertheless, existing OC-OSTHMDs still face the challenge of suboptimal transparency. The real scene light is linearly polarized in most devices for implementing occlusion, varifocal features, etc., causing a 50% loss of real scene luminance. Moreover, SLMs in OC-OSTHMDs cause a reduction of $\sim 70\%$ of the illuminance [37]. In addition, recent compact designs for OC-OSTHMDs modulate the real scene light to propagate through the polarizing beam splitter (PBS) repeatedly. Considering the repetition of propagation by 3 times [7, 31], and the PBS has a maximum transmittance of $\sim 90\%$ in general, the optical system further reduces the propagated light by 27%. Furthermore, real scene light in OC-OSTHMDs needs to propagate through an optical combiner for merging with virtual images, where an extra loss of $10\% - 50\%$ occurs. In practical applications, the wavelength sensitivity of polarizing optical elements (POEs) results in non-uniform transparency within the visible spectrum, dimming the real scene considerably. As a consequence, the maximum transparency of an OC-OSTHMD is $< 20\%$ in theory, and the reported value is even worse in a few prototypes [37, 38]. Users' visibility is deteriorated by the poor transparency of OC-OSTHMDs, restricting the indoor application of the device.

To tackle this issue, we introduce mask balancing for OC-OSTHMDs in this paper. As shown in Fig. 3, the proposed method dedicatedly adds the passed vision of the real scene (in s polarization) into the masked vision (in p polarization). Thus, the transparency of the real scene in the blending vision surpasses the upper boundary in typical OC-OSTHMDs. The blending ratio is dynamically modulated to guarantee that both the real scene and the virtual image appear clearly. We propose a perception-driven approach for optimizing the modulation, which is triggered by the user's visibility to the real scene and the virtual image comprehensively. A benchtop prototype is built, where the blending is driven by modulating the cross-angle of the linear polarizer and the PBS, as shown in Fig. 1. Then, user studies with 12 and 24 participants are conducted to quantify the perception of texture and lighting in virtual images, giving a threshold of 0.537 and 0.443 Weber contrast for texture and lighting recognition in OC-OSTHMDs, respectively. The perception of the real scene is evaluated by using the band-pass RMS contrast and the human contrast sensitivity function. We validate the proposed method in the prototype. The final user study was run with 12 participants. The result proves that OC-OSTHMDs with our mask balancing method outperform typical OSTHMDs and OC-OSTHMDs.

The primary contributions of this paper are highlighted as follows:

1) We propose the mask balancing method to improve real-scene visibility in OC-OSTHMDs. The method surpasses the see-through transparency of typical OC-OSTHMDs by blending the masked vision with the passed vision in a perception-driven approach.

2) We evaluate human perception of virtual texture and lighting in OC-OSTHMDs. Texture experiments revealed significant frequency-dependent differences, while lighting experiments indicated frequency-related trends consistent with expectations.

3) We built a benchtop prototype to validate the mask balancing method in OC-OSTHMDs. The prototype keeps the original optical structure of OC-OSTHMDs while integrating with a rotation mount, benefiting the proposed method with excellent versatility.

2 Related Work

2.1 Occlusion-Capable Optical See-Through Displays

To overcome the ghost-like appearance of virtual imagery in optical see-through AR, Kiyokawa et al. [17] first introduced OC-OSTHMDs, demonstrating pixel-wise occlusion with a 4-f relay system using an SLM. Their ELMO-4 prototype successfully supported mutual occlusion between virtual and real objects, but suffered from a bulky form factor, narrow FOV, and reduced see-through brightness [16].

Following this pioneering work, subsequent studies advanced OC-OSTHMDs along several directions. Wide-FOV systems were realized using paired mirrors [34, 35, 38] or conical optics [4, 15], expanding the view to over 143° [13, 38]. Varifocal occlusion has been explored with tunable or mechanically shifted lenses [6, 7, 9, 26], and compact designs leveraged prisms [30], folded optical paths [37], pinhole arrays [10], lens-arrays [32], or integrated DMD/LCoS engines with time-multiplexed occlusion and virtual imagery [7, 14, 31]. Light attenuation displays allow occlusion to be separately applied on each color channel, achieving better color reproduction [8, 12].

Despite these advances, transparency remains a fundamental challenge. Polarization-based architectures typically halve real-scene luminance, while SLMs and multiple passes through polarizing beam splitters further attenuate the light. Additional losses occur at optical combiners and polarizing elements, often leading to overall transparency well below 20% in practice. Such degradation restricts usability, particularly in indoor environments. To address this, research has shifted from transmissive LCD-based modulators toward reflective SLMs such as LCoS and DMD, which reduce polarization losses and improve see-through brightness. More recently, photochromic materials have been investigated as an alternative, eliminating the need for polarizers and enabling lighter optical architectures [2, 24]. However, these remain limited by slow switching speed, UV safety concerns, and insufficient pixel-level occlusion accuracy.

2.2 Perception-Driven Optimization for OSTHMDs

Optimizing the visual experience of OST-HMDs often requires considering not only optical design but also the perceptual mechanisms of the human visual system.

Mori et al. proposed BrightView [23], which imperceptibly dims incident light using liquid crystal shutters so that virtual objects appear brighter without users noticing real-world dimming. Their psychophysical study shows that exploiting visual adaptation can improve brightness consistency beyond hardware limits.

Hu et al. studied soft-edge occlusion with a single transparent LCD, where defocus causes blur and leakage [11]. They found that human perception tolerates blur differently from cameras and varies individually. Based on this, they proposed a user-preference-based expansion strategy to achieve perceptually complete occlusion.

Liu et al. presented a hybrid foveated depth of field rendering method, combining visual acuity eccentricity, depth of field, and chromatic aberration to guide rendering [22]. By aligning computation with perceptual sensitivity, their system reduces workload while preserving visual quality and alleviating vergenceaccommodation conflict.

Together, these studies demonstrate the effectiveness and specificity of perception-driven optimization for OST-HMDs: (1) the same visual stimuli can be perceived differently across individuals, (2) leveraging perceptual characteristics enables improvements such as reduced computation, surpassing hardware limitations, and lowering form factor, and (3) human vision tolerates certain imperfections, which can be beneficial for OST-HMD design.

3 Framework of Mask Balancing

In this section, we introduce the principle of mask balancing in three parts. First, we provide an overview of the system design. Then, we explain the optical basis that enables transmittance modulation. Finally, we define the visibility metrics for real and virtual images, which serve as the foundation for optimizing the balance between them.

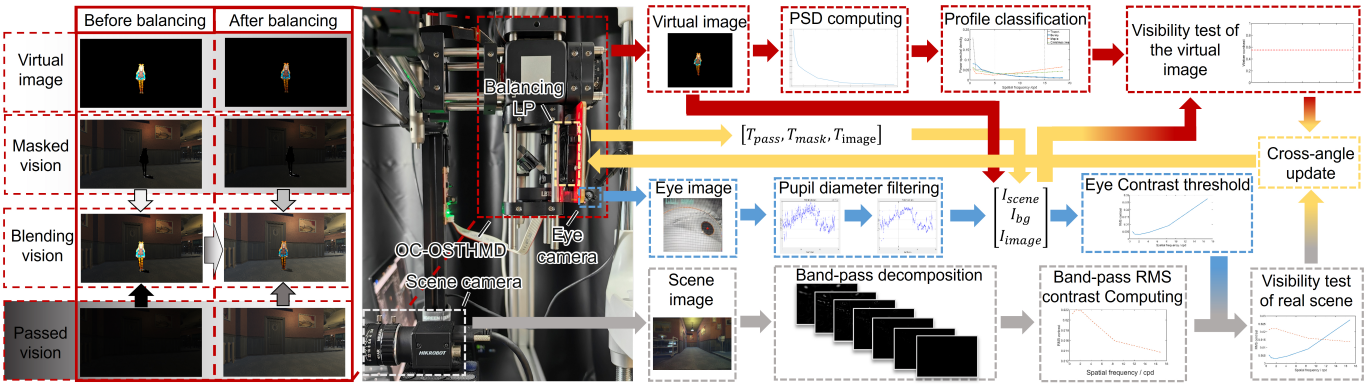

Fig. 2: Overviews of the proposed mask balancing method. Except for the OC-OSTHMD, an eye camera and a scene camera are used to capture the pupil diameter and the scene image at run-time. The captured pupil diameter is used to compute the illuminance information of the scene, and then to estimate the contrast sensitivity of users. The scene image goes into a band-pass analysis, and the band-pass RMS contrast is compared to the estimated eye contrast to test the visibility of users to the real scene. Meanwhile, Weber contrast is implemented to evaluate the visibility of the virtual image, and the visibility test is done between the evaluated value and the threshold value from the user study. Once the visibility test fails, the balancing LP works as the key component to adjust the transmittance of the system to the passed vision, masked vision, and virtual images, leading to an optimal visibility for both the real and the virtual image.

3.1 System Overview

The system overview of mask balancing is depicted in Fig. 2. In addition to a typical OC-OSTHMD (red rectangle in the middle), the system is attached with three add-ons. The eye camera (blue rectangle in the middle) works for capturing the pupil diameter of users, which is used to compute the scene illuminance I_{scene} observed by the users in real time. Due to the noise in the estimation data and the fluctuating nature of pupil size, the captured pupil diameter is filtered to omit the subtle variation, while keeping the oscillation caused by pupil reflex [19]. At the same time, a scene camera (gray rectangle in the middle) captures the real background perceived by the user. The scene image is decomposed into a stack of band-pass images by using the Peli contrast pyramid, and a contrast spectrum is computed to quantify the RMS contrast for each band-pass [25] We then determine the target band-pass for the visibility test of the real scene, and the visible threshold is estimated by implementing Barten's contrast sensitivity function (CSF) [1]. The testing result works as a trigger for updating the system balancing status, where the optimal rotation angle of the linear polarizer (LP) is set to the rotation mount. Since a PBS locates before the balancing LP, the modulation of their cross-angle impacts the transmittance $[T_{pass}, T_{mask}, T_{image}]$ of the passed real scene, masked real scene, and the virtual image thus changing the brightness of the corresponding parts in the perceived AR scene.

Images in the left solid rectangle of Fig. 2 illustrate how the blending vision is optimized by mask balancing. Once the visibility test for the scene image fails, the system recognizes the sub-optimal illuminance of the ambient environment, updating the cross-angle of the balancing LP to increase T_{pass} while reducing T_{mask} and T_{image}. Benefiting from the logarithmic nature of human brightness perception, even a slight increase in illuminance from the passed vision yields a noticeable improvement in the brightness of the real scene in the blending vision. This comes at the cost of a minor degradation in the visibility of the virtual image, as illustrated in the right column.

Meanwhile, the system runs a parallel thread to evaluate the visibility of the virtual image. The test works for recovering the balancing LP to keep a clear appearance of the virtual images when the ambient environment is well-lit. As shown in the top-right row of Fig. 2, an image frame is extracted to compute the power spectral distribution (PSD), and the image is tagged as either low spatial frequency or high spatial frequency according to the PSD profile. The visibility of the virtual image is quantified by comparing its average Weber contrast to the threshold, which is evaluated from user studies. In contrast to the strategy used for the real scene, the balancing LP is triggered to decrease T_{pass} and increase T_{mask} and T_{image} once the visibility test of the virtual image fails. Thus, the virtual image displayed in the blending vision is brightened by the improved illuminance to present vivid graphics.

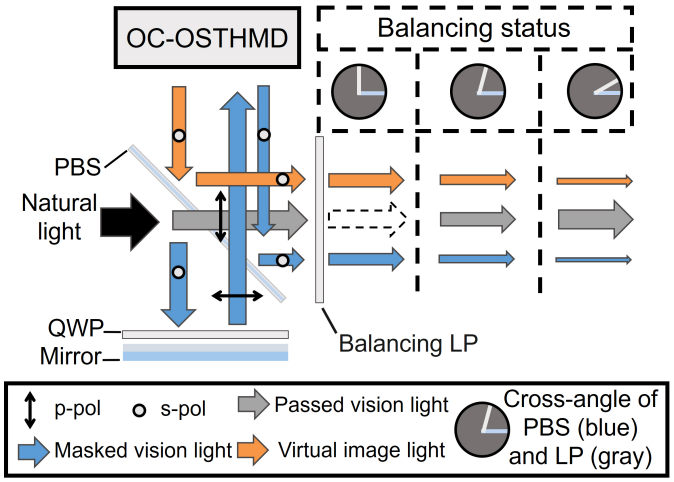

Fig. 3: The optics basis of mask balancing. The transmission of the passed vision, masked vision, and the virtual image are modulated by setting the cross-angle between the PBS and the balancing LP.

As a consequence, the system realizes the dynamic enhancement for the visibility of the real scene while keeping the virtual image clearly perceived. Since eye-tracking and scene cameras are readily available in commercial OST-HMDs, and PBS/LP elements are already common in recent compact OC-OSTHMDs [7,33], mask balancing requires only minimal hardware modification. It is therefore compatible with most optical designs of OC-OSTHMDs, offering excellent versatility.

3.2 Optics Basis of Mask Balancing

Fig. 3 illustrates the optical principle of mask balancing. The light from the real world (black arrows) is converted into p-polarized and s-polarized components, imaging the passed vision and masked vision, respectively. By nature of polarization, the p- and s-polarized components each carry half of the optical power of the incident light. However, the passed vision (gray arrows) is usually cut off by the orthogonal LP and PBS to make the users perceive solid virtual objects. The masked vision (blue arrows) provides the desired real image in AR, as it supports mutual occlusion of virtual content with the real background. Although it gives an appealing graphics effect, the masked

vision suffers a considerable loss of optical power from the optical system. The see-through transparency by the masked vision could be as low as $\sim 7\%$ [36, 37], severely limiting the users visibility of the ambient environment. Meanwhile, the users wearing OC-OSTHMDs perceive the light from the virtual display. The virtual image light (orange arrows) is reflected by the PBS to the users' eyes, thus having the same polarizing direction as the masked vision light. Due to the short optical path in OC-OSTHMDs and the high luminance (> 1000 nits) of commercial OLEDs, the virtual image light often exceeds the brightness of the real image provided by the masked vision.

Mask balancing adjusts the cross-angle between the balancing LP and the PBS to control the transmittance of the three visual components, aiming to maintain clear visibility of both real and virtual images. Assume a cross-angle of ϕ, the transmittance for the passed vision, the masked vision, and the virtual image can be formulated as:

$$T_{pass}^{(\varphi)} = \cos^2\varphi \quad , \\ T_{mask}^{(\varphi)} = \sin^2\varphi \cdot T_{HMD}, \quad T_{image}^{(\varphi)} = \sin^2\varphi \cdot T_{display}, \tag{1}$$

where T_{HMD} and $T_{display}$ are the intrinsic transmittance of the OC-OSTHMD for the masked vision and the virtual image exiting from the optical system, and the modulation of cross-angle ϕ leads to the transmittance products by the squared sine function.

3.3 Visibility Metrics for Real and Virtual Images

To estimate the optimal transmittance for the virtual image and the real scene to appear explicitly in the blending vision, we first define the metrics for quantifying their visibility. For the real scene, we choose the root mean squared (RMS) contrast to account for the visibility. Considering the complex contents in the real world, we further utilize the Peli pyramid to decompose the image into bands with ascending spatial frequency in an octave manner [25]. Fig. 4 shows the 7 band-pass images based on the captured real scene, and their RMS contrast is depicted in the right figure, revealing that scene visibility decreases as spatial frequency increases. The pyramid also indicates that most scene information lies in bands below 10 cpd (level 5). By further considering the tendency of eye's contrast sensitivity in the medium spatial frequency range, we choose the 4th band-pass (4.7 cpd) as the metric to quantify the visibility of the real scene. The metrics can also be adopted with higher band-passes when the OC-OSTHMD user is conducting tasks with precise operation, e.g., mechanical maintenance, supporting the potential for task-dependent mask balancing.

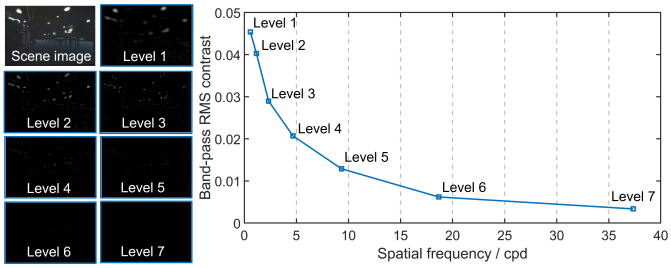

Fig. 4: The band-pass images decomposed from the captured real scene. The RMS contrast for each band-pass is depicted in the right figure, where the spatial frequency of the band-pass follows an Octave manner from 0.6 cpd to 37.4 cpd.

Unlike the real scene, virtual objects in AR are relatively sparse and less complex. Therefore, we apply the Weber contrast as the visibility metric for the virtual images. Assuming a virtual image has the mean gray scale P_{mean} after gamma correction, the Weber contrast of the virtual image is derived as:

$$C_{image} = \frac{P_{mean} I_{max} \cdot T_{image}^{(\varphi)}}{I_{bg} \cdot T_{pass}^{(\varphi)}} \tag{2}$$

where I_{max} is the maximum luminance of the virtual images displayed by the panel, and I_{bg} is the background illuminance of the real world. In practice, we measured the illuminance at the exit pupil of OC-OSTHMDs by setting the virtual display to all-white. Thus, $I_{max} \cdot T_{display}$ in Eq. 2 can be treated as a constant since it is solely determined by the hardware attributes of the OC-OSTHMD.

Although the visibility of virtual objects is not quantified in a band-pass manner, the Peli pyramid is used to derive the power spectral density of virtual objects when it is initialized. The PSD is computed by dividing the band-pass magnitude by the total energy summed from all bands. Fig. 5 depicts the PSD curves of 4 common virtual objects, where Bunny and Teapot exhibit lower power in the high-frequency range due to the simple geometry and texture. In contrast, Maple and Christmas tree show increased PSD beyond 4.7 cpd, yielding profiles distinct from Bunny and Teapot. Therefore, we conduct an adaptive visibility test for virtual objects in the mask balancing method. Section 4 explains how different visibility thresholds are evaluated for virtual objects by the user study.

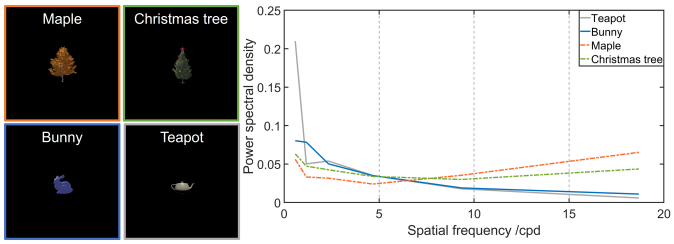

Fig. 5: The power spectral density (PSD) of typical virtual objects. According to the profile of PSD, we classify virtual objects into low spatial frequency targets (such as the teapot and bunny) and high spatial frequency targets (such as the maple and Christmas tree), implementing an adaptive threshold for the visibility test.

4 Visibility Threshold for Optimal Balancing

To realize mask balancing in practice, it is necessary to establish quantitative visibility thresholds for both real and virtual content. This section first introduces the band-pass visibility test for the real scene based on contrast sensitivity modeling. We then describe two user studies that evaluate thresholds for virtual texture and lighting perception. Finally, we integrate these results into a dynamic balancing strategy that adapts transmittance according to real-time visibility conditions.

4.1 Band-Pass Visibility Test for Real Scene

Band-pass RMS contrast is an intrinsic feature of the scene image. We therefore define the real-scene visibility threshold as the contrast sensitivity required for users to recognize the band-pass information. We adopted the contextual contrast sensitivity function by Triantaphillidou et al., where the critical contrast sensitivity S_c can be defined as:

$$S_c \geq \frac{1}{C_R \cdot \sqrt{1-k}} \tag{3}$$

where k is a scene-dependent factor, but can be safely set to 0.02 in most scenarios, as suggested by Triantaphillidou et al. [28]. C_R is the band-pass RMS contrast derived by the peli pyramid.

We then apply the well-known Barten's contrast sensitivity function to compute S_c in Eq. 3 [1]. The function provides a precise model to estimate eye contrast sensitivity in terms of spatial frequency, stimulus size, and scene illuminance. For system validation, we use the spatial frequency at the 4th band-pass in Bartens CSF, and substitute the target size with the OC-OSTHMDs FOV, measured as 13.7° diagonally.

4.2 Texture Visibility Estimation for Virtual Images

As aforementioned, we implement an adaptive threshold for estimating the visibility of virtual images, which depends on the spatial frequency of the perceived objects. To establish such thresholds, we designed user studies targeting two perceptual aspects of virtual content. In this subsection, we evaluate texture visibility, while the next subsection

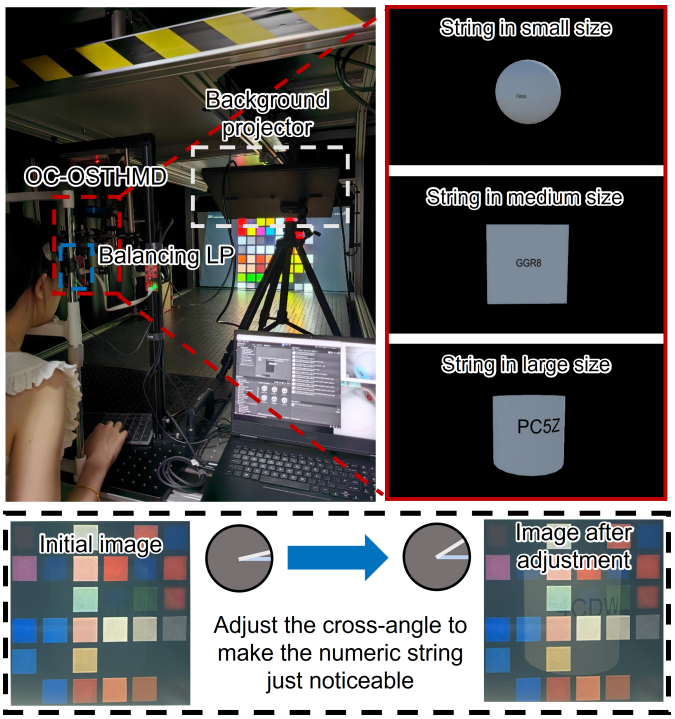

Fig. 6: The apparatus for the texture visibility experiment. The strings in various dimensions simulate the texture at different spatial frequencies on virtual objects. Each string is randomly composed of 1 number and 3 letters. The task for participants is to manually adjust the cross-angle of balancing LP by starting at 5°, confirming the angle that makes the number just noticeable, then reporting the number verbally. The figures in the bottom row show the initial AR scene observed by the participant at the beginning of a trial, then, the participant increases the cross-angle to make the numeric string just visible.

(Section 4.3) extends the evaluation to lighting effects. We deliberately exclude mutual occlusion from these tests, as it would overly restrict the balancing range for optimizing real-scene visibility. Instead, we focus on texture and lighting, which are less affected by occlusion but more easily washed out when the background is brightened. Both studies incorporate spatial frequency to validate the visibility tests.

All procedures involving human subjects in this research were approved by the Shanghai Jiao Tong University Institutional Review Board for Human Research Protection. All participants provided informed consent prior to participation.

4.2.1 Apparatus

The experiment setup is shown in Fig. 6. The OC-OSTHMD displays virtual images while the projector presents a partial color checker as the real background. The projector is set to display the real background with 4 levels of maximum illuminance at 29, 121, 358, 626 cd/m^2 to simulate conditions ranging from a dim public area to a typical office environment. The background is cropped from the digital color checker, and we set different cropped areas for each illuminance setting, where the mean pixel intensities within the perceived region by users are 0.25, 0.24, 0.25, and 0.2, respectively. All the virtual images displayed by the OC-OSTHMD have similar illuminance around 15.6 cd/m^2. We compute Weber contrast for the text by using Eq. 2, where the background illuminance is treated as the maximum value times mean pixel intensity. The text string is composed of 1 number and 3 letters, placed in a random sequence. We set the strings into 3 dimensions. The smallest string spans only 0.685° in the users' vision. According to our pilot experiment, further downsizing the string dimension would make it difficult for people with normal vision to recognize the number, even with the real background fully occluded. Then, the dimensions of the other 2 levels are set ascendingly in the octave manner, so the spatial frequency halves in the frequency domain. A keyboard is provided to participants to adjust the balancing angle. The left and right arrows make a rough adjustment of 5° counter-clockwise and clockwise, and the up and down arrows make a fine adjustment of 1° counter-clockwise and clockwise. A Pupil Labs Core eye tracking camera is used to capture the users' pupil diameter at 120 Hz. The experiment project is developed by using Unity engine (version 2022.3.28f1), and all devices are driven by a portable laptop with an Intel i9-14900HX CPU, an NVIDIA GeForce RTX 4080 Laptop GPU, and 64 GB of RAM.

4.2.2 Participants

We recruited 12 participants (6 females, 6 males; mean age = 26.62, SD = 3.59) through university campus forums and advertisements. Regarding familiarity with AR, ten participants had previous experience with AR headsets. Among them, three reported using VR weekly, five used it monthly, and two had only occasional use. For their expertise in graphics rendering, the group included one expert, six participants with extensive experience, three participants with only basic knowledge, and two participants with no relevant experience. Half of the participants determined the threshold angle using the strategy of selecting the angle at which they first recognized the number, while the other half confirmed the threshold angle by repeatedly adjusting until they just clearly recognized the number.

4.2.3 Procedure

Upon arrival, participants were fitted with prescribed contact lenses for vision correction and completed the Mars contrast sensitivity test. Their left eye was occluded, and they were seated comfortably for eye-tracking calibration. They were then presented with numeric strings of three different sizes, rendered as textures on a basic AR model. The balancing LP was initially set to 5°, making the AR model barely visible and the numeric strings fully hidden.

We adopted the experiment design from the text legibility evaluation experiment by Gabbard et al, [5]. Participants were instructed to adjust the balancing LP angle using the keyboard to modify the AR images brightness until the numeric strings became just perceptible. Once the numbers were visible, participants verbally reported them and pressed the spacebar to proceed to the next trial. The experiment included three string sizes and four background luminance levels, with each condition repeated three times, for a total of 36 trials. All trials were presented in randomized order, and each numeric string appeared only once throughout the experiment.

4.2.4 Result and Discussion

We analyzed 427 valid samples from 12 participants (36 text strings) using a linear mixed-effects model (LMM) with random intercepts for participant and text string. The model treated string size (high, mid, low) as the main fixed factor, and used Satterthwaite approximation for inference. This approach accounts for the repeated-measures nature of the task, where each participant adjusted the balancing LP angle multiple times across conditions.

Results showed a clear effect of string size (see Fig. 7). In the raw data, Weber contrast values increased with spatial frequency: for the high-frequency strings, Q1 = 0.094, Median = 0.254, Q3 = 0.537; for the mid-frequency strings, Q1 = 0.058, Median = 0.088, Q3 = 0.175; and for the low-frequency strings, Q1 = 0.057, Median = 0.071, Q3 = 0.135. These distributions illustrate that participants required substantially higher contrast to recognize the smallest strings, whereas medium and large strings were comparably easy to identify. Estimated marginal means (EMM) (95% CI) from the LMM were High = 1.219 [0.309, 2.128], Mid = 0.053 [-0.849, 0.954], and Low = 0.050 [-0.853, 0.953]. Here, EMMs are computed as model-predicted threshold contrasts with random effects for participant and text string set to zero, and with covariates such as pupil diameter held at their mean value, while the remaining factors are averaged over. Under this definition, the relatively large EMM for the High-frequency condition indicates that this hardest condition is most sensitive to the model-based normalization, whereas the near-zero EMMs for Mid and Low suggest that the required contrast is low and essentially similar in these easier conditions (consistent with

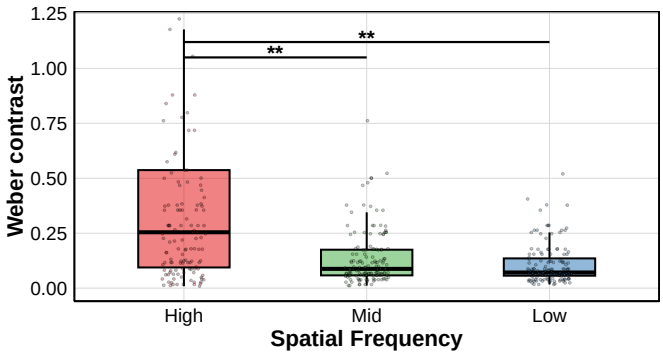

Fig. 7: Weber contrast by spatial frequency in texture visibility estimation. Colored boxplots and jittered points show the distribution of raw data. Horizontal brackets and significance markers indicate pairwise comparisons based on EMM from the LMM (Holm-adjusted).

the non-significant MidLow contrast). Pairwise contrasts confirmed High vs. Mid ($\Delta = 1.166$, $p = .0037$) and High vs. Low ($\Delta = 1.169$, $p = .0037$) as significant, but Mid vs. Low was negligible ($p = .993$). These findings align with the task design, where participants needed more contrast to recognize the smallest numeric strings, whereas medium and large strings were comparably easy to identify.

All other factors, including gender, VR/AR experience, graphics experience, strategy choice, and pupil diameter, had no measurable effect on thresholds. A robustness check using per-participant median thresholds replicated the frequency effect. Variance component analysis suggested minimal individual differences (ICC ≈ 0.04), indicating that the effect of string size was highly consistent across participants.

4.3 Lighting Visibility Estimation for Virtual Images

4.3.1 Apparatus

The setup for the lighting visibility experiment was similar to that of the previous study. Instead of showing strings on the basic model to participants, we prepared three sets of objects with complex geometry and texture, as shown in Fig. 8. Each set contained two objects that have a similar PSD profile. Fig. 5 depicts the classified profiles for Teapot, Bunny, Maple, and Christmas tree, respectively. In addition, we also included a robot and a sofa with a medium PSD profile for the experiment, as they represent the virtual character and furniture in typical AR applications. All virtual objects were duplicated into pairs: the original retained full lighting effects (e.g., specular reflection and shadows), while the duplicate included only the Albedo. To compensate for the color shift caused by removing lighting effects, the Albedo of the objects was adjusted so that each pair appeared in a similar color.

4.3.2 Participants

We recruited 24 participants (12 females, 12 males; mean age = 24.67, SD = 3.06) using a similar approach as in texture visibility estimation. Eighteen participants had prior experience with AR headsets, with five using VR weekly, ten monthly, and three only occasionally. Regarding graphics rendering expertise, the group included one expert, nine participants with extensive experience, nine with basic knowledge, and five with no relevant experience. One quarter of the participants (n = 6) determined the threshold angle at the first angle at which they could no longer distinguish the left from the right virtual objects, while the remaining three quarters (n = 18) determined the threshold by repeatedly adjusting and confirming the just-not-distinguishable angle.

4.3.3 Procedure

Upon arrival, participants underwent the same preparation as in the previous experiment, including vision correction and eye-tracking calibration. Two virtual objects were presented simultaneously in the participants' central vision, one on the left and one on the right, with lighting applied to only one side. The experimenter manually adjusted the lighting to ensure that the two images appeared comparable. At the

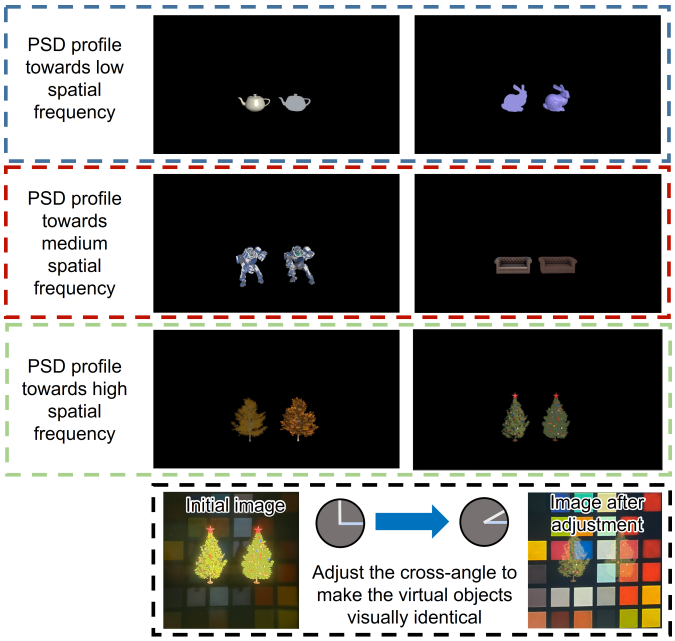

Fig. 8: The virtual objects used in the lighting visibility experiment. The bottom row shows the images initially observed by the participant at the beginning of a trial, where Christmas trees have distinguished appearances by on/off lighting effect (bottom-left). Then, the participant reduces the cross-angle from 90° to make the paired tree visually identical (bottom-right). In addition, to make the virtual objects more vivid, we render the cast shadow onto the projected real background in terms of the projector's orientation. The OC-OSTHMD visualizes the cast shadow by displaying it on the LCoS.

start of each trial, the balancing LP was set to 0°, allowing participants to clearly observe both objects and easily distinguish the lighting effects. Participants were instructed to adjust the angle using the keyboard until the two objects appeared visually identical. Once satisfied that the objects matched, they pressed the spacebar to proceed to the next trial. The objects were presented across four background luminance levels, identical to those in texture visibility estimation. Each combination was repeated twice, with left-right positions balanced, resulting in a total of 24 trials.

4.3.4 Result and Discussion

Same as texture visibility estimation, we analyzed 1145 valid samples from 24 participants (6 virtual objects $\times$ 4 luminance levels $\times$ 2 repetitions) using an LMM with random intercepts for participants. The objects treated spatial frequency (high, mid, low) as the main fixed factor, and the Satterthwaite approximation was used for inference.

Results showed no significant effect of spatial frequency on contrast thresholds (see Fig. 9). In the raw data, quartiles of Weber contrast were High: Q1 = 0.046, Median = 0.120, Q3 = 0.430; Mid: Q1 = 0.042, Median = 0.132, Q3 = 0.370; and Low: Q1 = 0.080, Median = 0.162, Q3 = 0.443. EMM were High = 0.302 [-1.93, 2.54], Mid = 0.298 [-1.93, 2.53], and Low = 0.779 [-1.46, 3.02]. Although these differences were not statistically significant, the trend was consistent with expectations: low-frequency models, being harder to distinguish, tended to yield slightly higher thresholds than high- or mid-frequency models. This trend may be explained by the fact that lighting effects themselves contain high-frequency information. Even on low-frequency objects, highlights such as specular reflections introduce localized regions of higher illuminance, making them more distinguishable despite their overall smoother geometry.

Other factors, including gender, VR/AR experience, graphics experience, strategy choice, and pupil diameter, showed no measurable effect on thresholds. A robustness check using per-participant median thresholds yielded the same null result. Variance component analysis

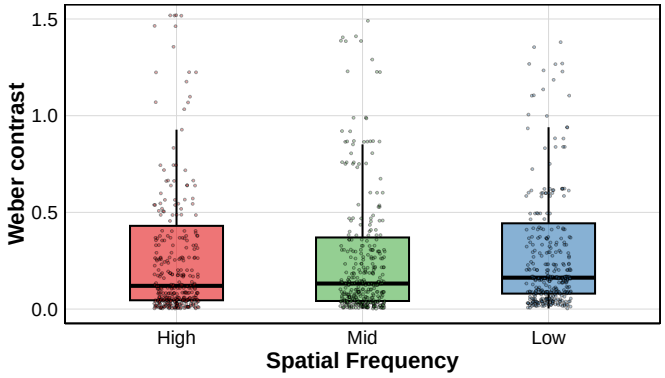

Fig. 9: Weber contrast by spatial frequency in lighting visibility estimation. Colored boxplots and jittered points show the distribution of raw data.

suggested minimal individual differences (ICC ≈ 0.042), indicating high consistency across participants.

4.4 Dynamic Balancing Strategy

When the visibility test fails for either the real scene or the virtual image, the system triggers the rotation mount to reset the cross-angle between the balancing LP and the PBS, as shown in Fig. 2. Since the real scene and the virtual image are evaluated by different metrics, improvement of visibility for either causes an opposite effect on the other. Hence, the transmittance is adjusted dynamically to keep a balanced illuminance for both sides.

4.4.1 Balancing for the Real Scene

Assuming the sub-optimal contrast sensitivity S_{raw} under the scene illuminance of I_{scene}, a target illuminance I_{target} needs to be provided for making the user's visual performance meet the desired contrast sensitivity S_c of the real scene. Thus, I_{target} follows the relationship with I_{scene}:

$$\sqrt{\frac{I_{target}-I_{scene}}{I_{scene}}} = \frac{S_c}{S_{raw}}, \tag{4}$$

which is given by Weber's law. Except for the scotopic vision, the contrast sensitivity of the human visual system strictly follows Weber's law until it reaches the saturated point [27]. Considering that the balancing for the real scene always triggers when I_{scene} is insufficient for supporting good contrast sensitivity for users, it is safe to utilize Eq. 4 for estimating the target scene illuminance. By substituting Eq 4 into Eq. 3 to link the human contrast sensitivity with band-pass RMS contrast, we derive the target scene illuminance I_{target} as:

$$I_{target} = I_{scene} \cdot (1 + \frac{S_c^2}{C_R^2 \cdot (1-k)}) \tag{5}$$

The illuminance of the entire scene I_{scene} can be computed by using the pupil reflex model with the real-time tracked pupil diameter.

Then, the cross-angle between the balancing LP and the PBS is triggered to reach the target illuminance. The threshold angle ϕ_{thres} is formulated as:

$$\phi_{thres} = \arcsin\sqrt{\frac{I_{target}-I_{bg}}{I_{bg}+(T_{mask}-1)\cdot I_{max}T_{display}}} \tag{6}$$

where I_{bg} is the real background illuminance, which can be computed by subtracting the vision-averaged pixel intensity of the virtual image and dividing by $(T_{pass}+T_{mask})$.

4.4.2 Balancing for the Virtual Image

As introduced in Section 3.3, we adopted different visibility thresholds by classifying the virtual image in terms of its PSD. For the virtual object whose PSD shapes are similar to the Maple and Christmas tree in Fig. 5, which keeps noticeable energy when the spatial frequency exceeds 4.7 cpd, we treat it as a high spatial frequency target; otherwise, it is classified as a low spatial frequency target. We consider the classification according to the results from the two user studies that indicate a significant difference in texture visibility between high and low-to-medium spatial frequency. Since there is no significant difference across spatial frequencies in lighting recognition, we decided to utilize the upper quartile limit of contrast (0.443) as the visibility threshold for the low spatial frequency target for maximizing the display quality. Similarly, we set the threshold as 0.537 based on the upper quartile limit of Weber contrast for texture recognition for the high spatial frequency target. However, the classification does not operate automatically in the current prototype. We manually choose the threshold according to the PSD profile of the displayed objects. We will integrate the classification feature into the system in the future.

When the system detects that the Weber contrast of the presented virtual image drops below the threshold, it triggers the cross-angle of the balancing LP to improve the transmittance T_{image} while suppressing T_{pass} for the passed vision. The threshold angle can be computed by using Eq. 2.

5 Prototype of Mask Balancing

We constructed a benchtop prototype to demonstrate the feasibility of mask balancing in practice. The following subsections detail the optical and computational configurations of the system, including the integration of real-time band-pass analysis and eye-tracking. We further describe the pupil-diameter filtering methods adopted to stabilize the noisy input from the eye tracker, which proved critical for reliable visibility estimation.

5.1 Prototype Configurations

We built the prototype based on the OC-OSTHMD introduced by Zhang et al., [33]. We replaced the beamsplitter for combining the virtual image with the masked vision with $T:R=2:8$. We used the aspherized achromatic lenses (#49663) with a focal length of 35mm to improve the image quality. As a result, the optical system provides a FOV of 13.7° and transmittance $T_{mask}=0.0662$ for the masked vision. Meanwhile, the prototype supports an eye relief of 15mm (measured from the balancing LP) and an eyebox of 20mm, allowing the participants to rotate their eyeballs freely during the experiment. The rotation mount used for driving the balancing LP is ELL14 from Thorlabs. Although the mount supports a customized speed for rotating, we set it to run at full speed in the prototype.

The Unity project of the prototype consists of two threads, where the band-pass analysis of the real scene runs separately. It takes advantage of implementing the scene image decomposition by a customized time step, releasing the computation load. A main thread works for the visibility test by integrating the outputs from the scene analysis thread and the eye tracking camera. Thread synchronization is achieved through global variables. The initial cross-angle of the balancing LP is set when the main thread receives the first band-pass data of the real scene. Due to the fluctuating nature of human eyes, we implement filters on the estimated pupil size from the eye-tracking camera. An empirical threshold is set to detect the strong oscillation of the pupil size, which indicates a noticeable change in the ambient lighting. The visibility test is called as long as the pupil reflex exceeds the threshold. At that point, either the real scene visibility or virtual image visibility test is triggered, which then dominates the modulation of the balancing LP.

5.2 Pupil Diameter Filtering

Due to measurement errors in the eye tracker, the recorded pupil diameter often exhibits strong fluctuations. Therefore, filtering is necessary to obtain stable and reliable data. We experimented with two filtering methods: the Moving Average filter and the Kalman filter. Prior to filtering, we applied a preprocessing step to the pupil diameter data by Kret et al., primarily to eliminate isolated samples [19].

The Moving Average filter uses a fixed-length window to compute the mean of the data within the window. This method effectively sup-

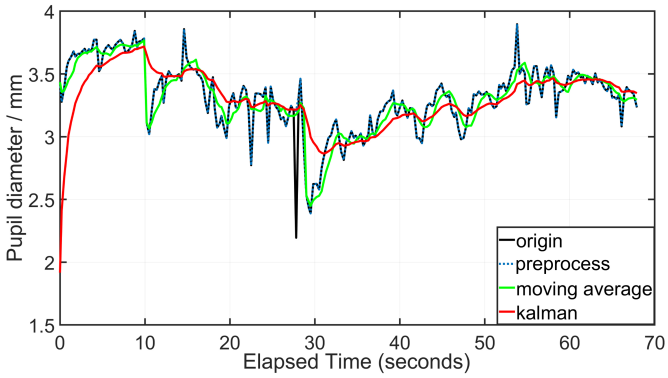

Fig. 10: The recorded pupil diameter of subject 9 during the final user study (black curve). Due to the highly fluctuating diameter input, we implement a pre-process to eliminate isolated samples. We tried the Moving Average filter (green curve) and the Kalman filter (red curve) for processing the diameter data in real time. As a result, the Kalman filter shows the most stable performance.

presses random noise but introduces a delay when the signal changes. To mitigate this delay, we incorporated a threshold-based jump detection: if the difference between a new data point and the current window average exceeds the threshold, the window is reset from that point. While this approach allows transitions to be detected, distinguishing between noise and actual transitions relies entirely on the threshold value, which must be carefully chosen.

The Kalman filter, in contrast, is a well-established method for filtering in dynamic systems. It estimates the system state by recursively predicting and updating based on statistical models. This makes it more robust to noise while also enabling rapid tracking of changes, since gradient information is preserved. After experimentation, we ultimately adopted the Kalman filter as it provided superior performance.

6 Method Demonstration

While previous studies established perceptual thresholds under controlled conditions, it remains necessary to validate the effectiveness of mask balancing in realistic AR scenarios. To this end, we designed a final user study in which participants viewed dynamic scenes across multiple lighting environments. By comparing our approach with conventional AR and static OC-OSTHMDs, we assess how mask balancing supports both real-scene visibility and virtual-scene clarity in practice.

6.1 Apparatus

As shown in Fig. 1, we conduct a final user study to demonstrate the dynamic performance of the proposed method. The apparatus is modified from the one used for the former user study, where a scene camera is integrated to run the band-pass analysis of the real scene. The entire scene is rendered by Unity. The open source assets, including Amazon Bistro and Unity-chan, are utilized to build the animation. We split the entire scene into the background, the character, and the mask. The background and the character are displayed by the projector and OC-OSTHMD, respectively. Therefore, misalignment of the optical axis between the scene camera and the OC-OSTHMD influences the system performance marginally, since it can capture the full real background for band-pass analysis. Masks, as well as the cast shadow of the character to the background, are displayed by the LCoS of the OC-OSTHMD. The scene is designed to switch lighting conditions 3 times. The animation starts with the indoor light off (~ 8 cd/m^2), and then switches to the indoor light on (~ 135 cd/m^2). With the character exiting the room, the background changes to outdoor shaded (~ 287 cd/m^2). Finally, the animation stops at outdoor sunlight (~ 414 cd/m^2). In addition, to compare the user experience between the proposed method and the typical AR Displays, we simulate an OSTHMD by setting the cross-angle ϕ as 45°, and disabling the occlusion feature of the prototype.

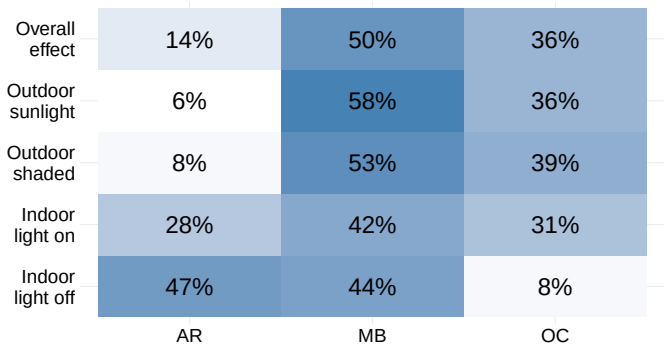

Fig. 11: Proportion of participant preferences for the three methods (AR, MB, OC) across indoor, outdoor, and overall conditions. Each cell shows the percentage of participants who selected the corresponding method.

6.2 Participants

We recruited 12 participants (6 females, 6 males; mean age = 25.62, SD = 2.75) using a similar approach as in the texture visibility experiment (Section 4.2). 11 participants had prior experience with AR headsets, with 3 using VR weekly, 6 monthly, and 2 occasionally. In terms of graphics rendering expertise, the group included 1 expert, 6 participants with extensive experience, 3 with basic knowledge, and 2 with no relevant experience.

6.3 Procedure

The overall procedure was similar to texture visibility experiment, except that participants did not manually adjust the balancing LP. The same AR scene was presented using three different display methods: conventional AR without occlusion (AR), our mask balancing strategy (MB), and a conventional static OC-OSTHMD (OC). Comparisons were conducted three times in a pairwise manner. For each comparison, participants viewed the same scene sequentially with two different methods and completed a questionnaire after each session. Each scene included 4 lighting conditions: indoor (lights off, lights on) and outdoor (shaded, sunlight). Participants were asked to indicate their preferred method for each lighting condition based on the clarity of the virtual objects relative to the real scene, as well as their overall preference for the display method.

6.4 Result and Analysis

Fig. 11 summarizes the participants pairwise preferences for three display methods (AR, OC, and MB) across indoor and outdoor conditions. The results show clear context-dependent trends. Under indoor light-off condition, AR was most favored (47%), while MB (44%) followed closely and OC (8%) was rarely chosen. This outcome is expected, since AR displays are unaffected by ambient illumination and can maintain sharp virtual imagery even in the absence of environmental light. In contrast, conventional OC-OSTHMDs suffer from transparency degradation indoors, as the real-scene light is linearly polarized, which leads to substantial luminance loss and reduced scene visibility.

Under indoor light-on condition, preferences became more balanced, with MB (42%) slightly ahead of OC (31%) and AR (28%). As ambient illumination increases, AR imagery becomes more transparent and thus less prominent, while MB can adaptively balance real-scene and virtual-scene luminance, ensuring stable perception.

In outdoor environments, MB consistently dominated. In outdoor sunlight condition, MB reached 58%, clearly outperforming OC (36%) and AR (6%). In outdoor shaded condition, MB again led with 53% versus 39% for OC and only 8% for AR. In the overall evaluation, MB was chosen by half of the participants, followed by OC (36%) and AR (14%). These results verify the robustness of the mask balancing method across diverse lighting environments: it adaptively modulates the blending ratio between real-scene and virtual imagery, thereby preserving both visibility and clarity where conventional AR or OC systems fail.

7 Limitations and Future Work

7.1 The degradation of occlusion quality

Because the visibility of the real scene is prioritized, and the transmittance modulation of real and virtual images is inherently coupled, some degradation of occlusion quality is inevitable in OC-OSTHMDs using the proposed method. Importantly, our design goal is not to preserve the physical occlusion (that is, fully blocking the light from the real scene) at all times, but to maintain perceptually functional occlusion under different real-world illumination conditions, where severely darkened real scenes can be more detrimental to usability and safety than moderate occlusion leakage. Nevertheless, the limitation may be alleviated by advances in micro-OLED luminance, which would make it easier to keep an optimal Weber contrast of the perceived virtual image (40 times brighter than the real background to produce the perceptually realistic mutual occlusion [21]) while preserving higher real-scene brightness. Meanwhile, the proposed method can move the priority towards virtual images when the OC-OSTHMD user performs tasks requiring accurate depth sensing (e.g., architecture design) in a static and safe environment [3]. One of the future works for applying the mask balancing method in practice is designing task-oriented balancing strategies. What is more, color compensation techniques are proven to make the virtual image appear less transparent even in typical OSTHMDs, integrating the technique into OC-OSTHMD with mask balancing is promising to considerably mitigate the degradation of occlusion quality under sub-optimal illumination [20].

7.2 Unnoticeable Mask Balancing

As mentioned in section 5.1, we rotate the balancing LP at full speed (430° per second), making the transmittance balanced as soon as possible. However, a couple of participants reported displeasure about the sudden drop in the scene brightness when the system works to improve the visibility of virtual images. Therefore, it is necessary to keep the illuminance change in an unnoticeable manner [23], realizing more user-friendly modulation.

7.3 Binocular Mask Balancing

Since the prototype only supports monocular vision, the visible threshold for the real scene and the virtual image does not take the performance of binocular vision into account. It is suggested that the human visual system has noticeable variation, including contrast sensitivity and pupil reflex, between monocular and binocular vision [1, 29]. Therefore, the binocular visibility threshold needs to be evaluated in OC-OSTHMDs for practical applications in the future.

8 Conclusion

We propose mask balancing that enhances the visibility of the real scene in OC-OSTHMDs. The proposed method achieves the improvement by modulating the cross-angle between the built-in LP and PBS of OC-OSTHMDs, and the modulation is leveraged by a perception-driven approach. We evaluate the visibility of the real scene and virtual images by using band-pass RMS contrast and Weber contrast. Visibility tests are conducted by comparing the metrics with human contrast sensitivity and a critical Weber contrast thresholds from user studies. The cross-angle is dynamically controlled in terms of test results. According to the final user study with our prototype, the proposed method outperforms typical OSTHMDs and OC-OSTHMDs in the dynamic illumination scene.

Acknowledgments

This work was funded by the Shanghai Pujiang Program (23PJ1406800).

References

[1] P. G. Barten. Formula for the contrast sensitivity of the human eye. In *Image Quality and System Performance*, vol. 5294, pp. 231–238. SPIE, 2003. 4, 5, 10

[2] M. Chae, K. Bang, Y. Jo, C. Yoo, and B. Lee. Occlusion-capable see-through display without the screen-door effect using a photochromic mask. *Optics Letters*, 46(18):4554–4557, 2021. 3

[3] H.-L. Chi, S.-C. Kang, and X. Wang. Research trends and opportunities of augmented reality applications in architecture, engineering, and construction. *Automation in construction*, 33:116–122, 2013. 10

[4] D. Dunn, C. Tippets, K. Torell, P. Kellnhofer, K. Akşit, P. Didyk, K. Myszkowski, D. Luebke, and H. Fuchs. Wide field of view varifocal near-eye display using see-through deformable membrane mirrors. *IEEE transactions on visualization and computer graphics*, 23(4):1322–1331, 2017. 3

[5] J. L. Gabbard, J. E. Swan, and D. Hix. The effects of text drawing styles, background textures, and natural lighting on text legibility in outdoor augmented reality. *Presence*, 15(1):16–32, 2006. 6

[6] T. Hamasaki and Y. Itoh. Varifocal occlusion for optical see-through head-mounted displays using a slide occlusion mask. *IEEE transactions on visualization and computer graphics*, 25(5):1961–1969, 2019. 3

[7] W. Han, J.-W. Lee, J.-Y. Shin, M.-H. Choi, H.-R. Kim, and J.-H. Park. Varifocal occlusion in an optical see-through near-eye display with a single phase-only liquid crystal on silicon. *Photonics Research*, 12(4):833–853, 2024. 3, 4

[8] Y. Hiroi, T. Hiraki, and Y. Itoh. Factoredsweeper: Optical see-through display integrating light attenuation and addition with single spatial light modulator. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 61–70. IEEE, 2024. 3

[9] Y. Hiroi, T. Kaminokado, S. Ono, and Y. Itoh. Focal surface occlusion. *Optics Express*, 29(22):36581–36597, 2021. 3

[10] X. Hu, C. Ebner, Y. Zhang, K. Kiyokawa, and A. Plopski. X-Mask: Improving Soft-Edge Occlusion in Optical See-Through Displays with Cross-Shaped Pinholes . In *2025 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 677–686, Oct. 2025. doi: 10.1109/ISMAR67309.2025.00077 3

[11] X. Hu, Y. Zhang, A. Plopski, Y. Itoh, M. Perusquía-Hernández, N. Isoyama, H. Uchiyama, and K. Kiyokawa. Perception-driven soft-edge occlusion for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 2024. 3

[12] Y. Itoh, T. Langlotz, D. Iwai, K. Kiyokawa, and T. Amano. Light attenuation display: Subtractive see-through near-eye display via spatial color filtering. *IEEE transactions on visualization and computer graphics*, 25(5):1951–1960, 2019. 3

[13] H. J. Jang, J. Y. Lee, G. W. Baek, J. Kwak, and J.-H. Park. Progress in the development of the display performance of ar, vr, qled and oled devices in recent years. *Journal of Information Display*, 23(1):1–17, 2022. 3

[14] Y.-G. Ju, M.-H. Choi, P. Liu, B. Hellman, T. L. Lee, Y. Takashima, and J.-H. Park. Occlusion-capable optical-see-through near-eye display using a single digital micromirror device. *Optics letters*, 45(13):3361–3364, 2020. 3

[15] K. Kiyokawa. A wide field-of-view head mounted projective display using hyperbolic half-silvered mirrors. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 207–210, Nov 2007. 3

[16] K. Kiyokawa, M. Billinghurst, B. Campbell, and E. Woods. An occlusion capable optical see-through head mount display for supporting co-located collaboration. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003. Proceedings.*, pp. 133–141. IEEE, 2003. 3

[17] K. Kiyokawa, Y. Kurata, and H. Ohno. An optical see-through display for mutual occlusion with a real-time stereovision system. *Computers & Graphics*, 25(5):765–779, 2001. 3

[18] B. Krajancich, N. Padmanaban, and G. Wetzstein. Factored occlusion: Single spatial light modulator occlusion-capable optical see-through augmented reality display. *IEEE transactions on visualization and computer graphics*, 26(5):1871–1879, 2020. 3

[19] M. E. Kret and E. E. Sjak-Shie. Preprocessing pupil size data: Guidelines and code. *Behavior research methods*, 51(3):1336–1342, 2019. 4, 8

[20] T. Langlotz, M. Cook, and H. Regenbrecht. Real-time radiometric compensation for optical see-through head-mounted displays. *IEEE transactions on visualization and computer graphics*, 22(11):2385–2394, 2016. 10

[21] J. Liu, A. Jindal, C. Mantel, S. Forchhammer, and R. K. Mantiuk. How bright should a virtual object be to appear opaque in optical see-through ar? In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 676–684. IEEE, 2022. 10

[22] J. Liu, C. Mantel, and S. Forchhammer. Perception-driven hybrid foveated depth of field rendering for head-mounted displays. In *2021 IEEE Inter-*

national Symposium on Mixed and Augmented Reality (ISMAR), pp. 1–10. IEEE, 2021. 3

[23] S. Mori, S. Ikeda, A. Plopski, and C. Sandor. Brightview: Increasing perceived brightness of optical see-through head-mounted displays through unnoticeable incident light reduction. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 251–258. IEEE, 2018. 3, 10

[24] C.-W. Ooi, Y. Hiroi, and Y. Itoh. A compact photochromic occlusion capable see-through display with holographic lenses. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 237–242. IEEE, 2023. 3

[25] E. Peli. Contrast in complex images. *Journal of the Optical Society of America A*, 7(10):2032–2040, 1990. 4, 5

[26] K. Rathinavel, G. Wetzstein, and H. Fuchs. Varifocal occlusion-capable optical see-through augmented reality display based on focus-tunable optics. *IEEE transactions on visualization and computer graphics*, 25(11):3125–3134, 2019. 3

[27] J. Rovamo, J. Mustonen, and R. Näsänen. Modelling contrast sensitivity as a function of retinal illuminance and grating area. *Vision research*, 34(10):1301–1314, 1994. 8

[28] S. Triantaphillidou, J. Jarvis, A. Psarrou, and G. Gupta. Contrast sensitivity in images of natural scenes. *Signal Processing: Image Communication*, 75:64–75, 2019. 5

[29] A. B. Watson and J. I. Yellott. A unified formula for light-adapted pupil size. *Journal of vision*, 12(10):12–12, 2012. 10

[30] A. Wilson and H. Hua. Design and prototype of an augmented reality display with per-pixel mutual occlusion capability. *Optics express*, 25(24):30539–30549, 2017. 3

[31] A. Wilson and H. Hua. Design of a pupil-matched occlusion-capable optical see-through wearable display. *IEEE Transactions on Visualization and Computer Graphics*, 2021. 3

[32] Y. Yamaguchi and Y. Takaki. See-through integral imaging display with background occlusion capability. *Applied optics*, 55(3):A144–A149, 2016. 3

[33] Y. Zhang, S. Hong, W. Qian, K. You, H. Zhou, K. Kiyokawa, and X. Yang. Color correction for occlusion-capable optical see-through head-mounted displays by using phase-modulation. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 268–276. IEEE, 2025. 3, 4, 8

[34] Y. Zhang, X. Hu, K. Kiyokawa, N. Isoyama, N. Sakata, and H. Hua. Optical see-through augmented reality displays with wide field of view and hard-edge occlusion by using paired conical reflectors. *Optics letters*, 46(17):4208–4211, 2021. 3

[35] Y. Zhang, X. Hu, K. Kiyokawa, N. Isoyama, H. Uchiyama, and H. Hua. Realizing mutual occlusion in a wide field-of-view for optical see-through augmented reality displays based on a paired-ellipsoidal-mirror structure. *Optics Express*, 29(26):42751–42761, 2021. 3

[36] Y. Zhang, X. Hu, K. Kiyokawa, and X. Yang. Add-on occlusion: Building an occlusion-capable optical see-through head-mounted display with hololens 1. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 1003–1004. IEEE, 2023. 5

[37] Y. Zhang, X. Hu, K. Kiyokawa, and X. Yang. Add-on occlusion: Turning off-the-shelf optical see-through head-mounted displays occlusion-capable. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2700–2709, 2023. 3, 5

[38] Y. Zhang, N. Isoyama, N. Sakata, K. Kiyokawa, and H. Hua. Super wide-view optical see-through head mounted displays with per-pixel occlusion capability. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 301–311. IEEE, 2020. 3