Improving Depth Perception in Immersive Media Devices by Addressing Vergence-Accommodation Conflict

Razeen Hussain, Manuela Chessa, *Member, IEEE,* and Fabio Solari

Abstract—Recently, immersive media devices have seen a boost in popularity. However, many problems still remain. Depth perception is a crucial part of how humans behave and interact with their environment. Convergence and accommodation are two physiological mechanisms that provide important depth cues. However, when humans are immersed in virtual environments, they experience a mismatch between these cues. This mismatch causes users to feel discomfort while also hindering their ability to fully perceive object distances. To address the conflict, we have developed a technique that encompasses inverse blurring into immersive media devices. For the inverse blurring, we utilize the classical Wiener deconvolution approach by proposing a novel technique that is applied without the need for an eye-tracker and implemented in a commercial immersive media device. The technique's ability to compensate for the vergence-accommodation conflict was verified through two user studies aimed at reaching and spatial awareness, respectively. The two studies yielded a statistically significant 36% and 48% error reduction in user performance to estimate distances, respectively. Overall, the work done demonstrates how visual stimuli can be modified to allow users to achieve a more natural perception and interaction with the virtual environment.

Index Terms—Depth-of-field, depth perception, immersive media, inverse blurring, reaching task, space-variant technique, vergence-accommodation conflict, virtual reality, Wiener deconvolution.

1 Introduction

The immersive media field has flourished over the past decade. This is largely due to the introduction of affordable commercial head-mounted displays (HMDs). Immersive media can take various forms, including virtual reality (VR), augmented reality (AR), mixed reality (MR), and 3D content. Modern consumer technology offers a very realistic representation, however, some perceptual issues still remain that subsequently lower the sense of immersion in HMDs [1].

Rendering in computer graphics has primarily focused on photo-realism, i.e., simulating images from idealized lenses without aberrations. However, these rendering techniques typically do not consider the imperfect optics of the human eye. This oversight is crucial because when humans view rendered images, they perceive them through their own eyes, which inherently have imperfections. Instead, rendering for perceptual realism focuses on simulating real-world viewing [2]. Addressing these imperfections in the rendering pipeline is essential for creating truly realistic visual experiences.

The human visual system utilizes a variety of cues to determine the size and distance of objects in their environment. Typical cues include disparity, motion parallax, occlusion, convergence, and accommodation [3], [4]. However, not all cues are used at all times. Which cues are being utilized are in essence determined by the distance to the objects. Human spatial reach can be divided into three circular egocentric regions, namely the personal space, action space, and vista space [5]. Objects within 2m are considered to be in personal space. Disparity, accommodation, and convergence are more prevalent in this case. From 2 to 30m, it is referred to as the action space, and occlusion and motion parallax are more dominant here. Distances beyond 30m are considered to be vista space. Only pictorial depth cues such as occlusion and relative size are used [6].

Over the years, several studies [7], [8], [9] have been conducted on AR/VR which suggest that users typically underestimate the distances to objects by around 25% [4]. This is significantly higher when compared to human performance in the real world where even when blind walking only 8% underestimation occurs [10]. There are many potential reasons for this difference. The weight of the HMD combined with limited field-of-view (FoV) is one potential reason [11]. In close surroundings, disparity also plays a role in this difference. Some studies have suggested that the inter-pupillary distance (IPD) setting also plays a crucial role [12].

When it comes to visual perception in immersive media systems, users tend to experience conflicting cues. Such contradictions not only give rise to many errors in object size and distance estimations but also affect immersion and make the users feel uncomfortable over long exposures. The most prevalent of such mismatches in modern AR/VR devices is the vergence-accommodation conflict (VAC) which is sometimes also referred to as accommodation-convergence mismatch. The conflict is most prominent when using an HMD.

- *R. Hussain, M. Chessa, and F. Solari are with the Department of Informatics, Bioengineering, Robotics and Systems Engineering, University of Genoa, Genoa, Italy.*
  E-mail: fabio.solari@unige.it

Manuscript received December 21, 2022; revised November 6, 2023.

1.1 Contributions

The primary objective of our work is to propose a software-based solution that addresses the issue of VAC in VR environments by improving depth perception. The focus is on understanding the discrepancies between the way humans perceive visual stimuli in the real world and in the virtual environment and adapting the visual stimuli, i.e., the stereoscopic images, to cater to this inconsistency. To achieve this objective, we present a novel technique that applies an inverse blurring approach to stereoscopic images. This approach eliminates the need for an eye-tracker, making it more accessible and practical for widespread use. By distorting the visual stimuli, we compensate for the accommodation blurring that occurs when objects are focused outside the focal plane, thus allowing the observer to obtain a more natural perception of the virtual contents.

To validate the effectiveness of our proposed solution, we conducted two user studies on depth perception by using an off-the-shelf VR device. The first user study focused on a reaching task, where participants interacted with virtual objects at different depths. The second user study involved a spatial awareness task, which assessed participants' ability to accurately perceive distances and relative positions of virtual objects in the environment. The results of both user studies provide empirical evidence of the effectiveness of our software-based solution in improving depth perception, thus compensating for VAC. Participants exhibited improved performance and a more accurate perception of depth and spatial relationships when viewing the stereoscopic images processed using our novel technique. These findings support the efficacy of our approach in addressing the VAC and enhancing the overall VR experience.

To summarize, our research makes the following key contributions:

- We propose a software-based solution to compensate for the discrepancies introduced by VAC in VR environments.
- We introduce a novel technique that applies an inverse blurring approach to stereoscopic images, without the need for an eye-tracker.
- Our technique compensates for accommodation blurring, allowing observers to obtain a more natural perception of virtual contents.
- We validate the effectiveness of our solution through two user studies, demonstrating improved performance and enhanced spatial perception in depth perception-based VR tasks.

By offering a practical and accessible solution to compensate for the perceptual discrepancies introduced by VAC, our work contributes to advancing the field of VR and enhancing the quality of user experiences in VR environments.

2 Background

This work addresses the convergence and accommodation mismatch by compensating for the accommodation blurring. In order to better understand the importance of the issue for immersive media, this section provides details about the concept, some solutions found in literature, and a brief background to the proposed approach.

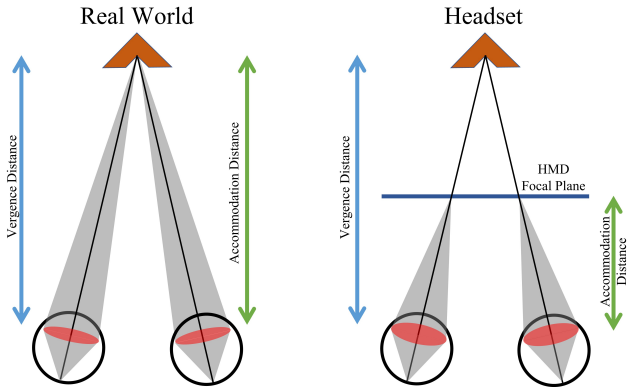

Fig. 1. Overview of the vergence-accommodation conflict. In natural viewing, eyes focus and converge at the same distance. Whereas, in stereo 3D viewing, the eyes focus and converge at different distances.

2.1 Convergence and Accommodation

When humans view objects in the real world, the eyes converge inwards while the ciliary muscles deform the lens. The former is referred to as convergence while the latter is called accommodation. This process occurs so that a sharp image is formed on the fovea. However, this is not always the case in immersive media setups. The image is shown at a fixed distance, i.e., the focal plane, while the depth of the virtual object varies with the content according to the disparity [13]. The basic geometry of this is shown in Fig. 1. The resulting conflict gives rise to several issues such as visual discomfort [14], reduction in performance [15], distortions of perceived depth [16], and reduced binocular image quality [17].

Oculomotor cues of consistent accommodation and convergence are related to the retinal cues of blur and disparity [3], i.e., retinal blur drives the accommodation response of the lens of the eye to focus at the desired depth whereas retinal disparity drives convergence [13]. Convergence and accommodation are two important cues prevalent in personal space and since most immersive media devices display stimuli to the user through a screen (focal plane) placed at a fixed distance, it is important that these two cues work cohesively. Accommodation is affected by several factors [18] such as natural blur [19], spatial features of the accommodation target [20], and foveal refractive error [21].

2.2 Addressing Vergence Accommodation Conflict

The issue of VAC in VR has gained significant attention from researchers in recent years. Several approaches and techniques have been proposed to address this conflict and enhance the overall visual experience in VR environments. In this section, we review some of the notable works in this domain.

Dynamic Depth-of-Field Techniques: Dynamic depth-of-field (DoF) techniques [22], [23] have been explored to mitigate VAC in VR. Such techniques involve dynamically adjusting the focus of virtual objects based on the user's gaze and the depth cues in the scene [24]. For instance, Duchowski et al. [25] and Rompapas et al. [26] introduced gaze-contingent rendering approaches that rendered virtual objects with varying levels of blur based on their depth

relative to the user's gaze. By simulating the natural focus behavior of the human eye, VAC was alleviated, resulting in improved visual comfort and immersion. Chakravarthula et al. [27] proposed a retinal speckle suppression algorithm that reduces the visibility of speckle artifacts that arise from coherent light interference in holographic displays. By adapting the rendering based on the viewer's gaze, the algorithm selectively applies spatial and temporal filtering to mitigate the speckle effect in the foveal region where visual acuity is highest.

Accommodation-Invariant Displays: Another line of research focuses on developing accommodation-invariant displays that decouple the accommodation and vergence responses of the eyes [28], [29]. Akeley et al. [30] introduced a multi-plane display that presents virtual scenes with different depths simultaneously, allowing users to perceive objects at varying depths without the need for constant accommodation changes. Foveated near-eye displays [31] have also been proposed that utilize custom microlens arrays to enable the display to allocate computational resources based on the viewer's gaze. Guzel et al. [32] introduced a system designed to correct visual prescriptions in VR headsets using perceptual guidance.

Light Field Displays: Light field displays use a combination of multiple images to create a 3D effect. They work by capturing multiple images of the same scene from different angles and then combining them to create a 3D effect. This allows the viewer to see the scene from different perspectives and provides a more immersive experience [33]. Huang et al. [34] proposed a light field display that presents multiple focal planes simultaneously. By providing depth information to the viewer without requiring accommodation adjustments, this approach can seemingly reduce VAC, although an in-depth analysis needs to be carried out to evaluate this. Furthermore, light field displays require a lot of processing power and are not yet widely available.

Varifocal Displays: Varifocal displays [35] have also emerged as a potential solution to mitigate VAC. These displays dynamically adjust the focal distance to match the viewer's gaze and simulate natural accommodation behavior. Akşit et al. [36] presented a varifocal near-eye display that employed liquid lenses to rapidly change the focal depth based on the user's gaze direction. This approach allowed users to experience sharp and clear virtual objects at different depths, reducing the discomfort caused by VAC. Chakravarthula et al. [37] proposed a system of auto-focus AR eyeglasses that utilizes eye-tracking technology and depth estimation algorithms to automatically adjust the focus of the eyeglasses based on the user's gaze. Similarly, Cholewiak et al. [38] proposed a method for rendering chromatic eye aberrations in VR displays by utlizing focus-adjustable lenses and gaze tracking to reproduce the natural relationship between accommodation and blur in HMDs. A prominent limitation of varifocal displays is that it requires precise eye-tracking. To compensate for the inaccuracies introduced by eye-tracking, Ebner et al. [39] proposed a gaze-contingent layered display that captures a focal stack within a fixed depth range at once and calculates display patterns using necessary focal stack images among the full focal stack.

These setups and techniques, although offering certain benefits, suffer from significant drawbacks that limit their practicality and adaptability to modern lightweight HMDs. One major drawback is the high computational cost associated with these setups. The complex algorithms and processing required to achieve their intended functionality demand substantial computational resources, which may be impractical or unfeasible for lightweight HMDs with limited processing power. Another limitation is the narrow field-of-view (FOV) associated with these setups. Additionally, these setups tend to be hardware-intensive, i.e., they rely on specialized hardware components or sensors that are not typically integrated into modern HMDs. The additional hardware requirements can make the setup cumbersome, less portable, and less developer-friendly, which may hinder widespread adoption.

2.3 Wiener Deconvolution-based Deblurring

When light rays enter the eyes through the cornea, they diffract to form a focused image on the retina. The diffraction pattern can be modeled as a point spread function (PSF). If this PSF is known, it is possible to identify the optical requirement of corrective lenses that are necessary to adjust the light rays entering the eyes. In the image processing domain, this operation can be analogously expressed by a deconvolution operator [40]. Convolution is the technique popularly used to apply filters to images. Deconvolution is the inverse process of convolution. Primarily, it is a computationally intensive process that can be used to recover the blurring in an image [41]. This process can also be referred to as inverse blurring or deblurring. Common deconvolution algorithms include inverse filtering, Wiener filtering, and iterative approaches such as the Lucy-Richardson algorithm. With the recent boom in the deep learning domain, many new algorithms have been proposed [42], [43], [44]. However, the iterative processes and the deep learning models are not viable solutions when it comes to AR/VR applications as fast processing is of utmost importance in order to update the scene in real-time and these methods have a very high processing and memory cost. For this reason, we developed our inverse blurring technique based on the Wiener deconvolution.

Generally, given an image i, the convolution operation with a blurring filter f can be defined as:

$$b = f * i + n \tag{1}$$

where $*$ is the convolution operator, n is the noise in the system and b is the resulting blurred image.

In the Fourier or frequency domain, Equation 1 can be written as:

$$B = FI + N \tag{2}$$

where B, F, I and N are the Fourier transforms of b, f, i and n respectively. Typically, this blurred image can be corrected through the inverse procedure:

$$I^{'} = \frac{B - N}{F} \tag{3}$$

However, this approach is not optimal as it amplifies the noise in the system. Instead, the Wiener deconvolution is considered more optimal for this type of task as it is insensitive to small variations in the signal power spectrum

[45]. The Wiener filter assumes that the image is modeled as a random process whose 2^{nd} order statistics along with noise variances are known. The image restoration model can be written as:

$$I^{'} = H_W B \tag{4}$$

where $I^{'}$ is the estimation of the original image and H_W is the Wiener filter.

Assuming the PSF is real and symmetric and the power spectrum of the original image and the noise is unknown, then the Wiener filter can be defined as:

$$H_W = \frac{H}{\mid H \mid^2 + \frac{1}{\lambda}} \tag{5}$$

where λ is the signal-to-noise ratio (SNR) and H is the estimate of the PSF of the blur.

The notion of incorporating deconvolution-based methods into immersive media is not new. It has been previously investigated in projection displays [46]. Image preconditioning is an important step in reducing the effects of out-of-focus projector blur. It can be done through blur kernel estimation algorithms [47], high-pass filters, and Wiener filters [48]. A limitation of these approaches in projection displays is the trade-off between ringing artifacts and contrast degradation.

Similarly, deconvolution has been applied to near-eye displays as well. Konrad et al. [49] introduced a computational method that decouples accommodation and convergence, allowing virtual objects to be presented with accurate focus cues regardless of their depth. The SharpView algorithm [40] utilizes Wiener deconvolution using a simplified Gaussian-based PSF estimation aimed at optical see-through AR devices with the aim of enhancing the visual quality of the virtual content. These works are mostly aimed at improving the visual acuity of the image as these works try to remove the effect of focus blur. On the other hand, our aim is to aid the occurrence of focus blur, which is absent in immersive devices due to constant accommodative distance for all virtual objects, for restoring a visual natural cue.

Furthermore, while other investigations on deconvolution-based techniques have primarily focused on visual clarity (e.g., evaluating the readability of texts [50], [51]), there is a notable research gap concerning the potential of such methods to enhance depth perception. To the best of our knowledge, prior studies have not explored the application of deconvolution-based techniques specifically for improving depth perception. Our work aims to bridge this gap by providing insights into how inverse blurring filters can improve depth perception.

Additionally, these works [40], [49] require real-time measurements of the user's eye measurements such as pupil size, focus distance, etc. to determine the PSF parameters. This requires specialized equipment that may not be suitable for consumer technologies. Moreover, these works were created and tested on custom hardware setups which are not easy to translate to off-the-shelf devices. On the other hand, our work aims to develop a general-purpose solution that can be easily integrated into consumer HMDs without the need for hardware modifications.

3 Proposed Technique

The proposed technique utilizes the Wiener deconvolution-based deblurring to alter the visual stimuli in VR systems. For a perfect solution, it is required to have the actual PSF of the user's eye. Estimating the PSF of human eyes is a challenging task due to the intricate nature of the visual system. In order to measure this PSF, various techniques such as wavefront sensing, retinal imaging, and psychophysical experiments can be employed. However, the exact shape of the PSF varies depending on the eye. For example, the PSF of a myopic eye would be different from that of a hyperopic eye. Moreover, PSFs slightly vary between different users, and configuring the system for each end-user is a tedious task and may limit the wide application of the technique.

When the true PSF is unknown, Chakravarthula et al. [27] suggested using an approximation of the PSFs. For out-of-focus distortions such as those that are naturally present in the human visual system, a circular PSF is considered a good approximation [52] since the PSF based on a thin lens model can approximate human eyes. Such a PSF can be defined by only one parameter, ρ which is the radius of the circle. Therefore, only two parameters (ρ and λ) need to be tuned to implement the Wiener deconvolution.

3.1 Parameter Tuning

The parameter ρ is dependent on the amount of blur present in the image while λ is a measure based on the noise present in the system. Since ρ is more significant to the restoration/deblurring process, it is recommended to tune it first.

The parameter ρ can be determined based on the distance between the user and the objects in view. To select the optimal value of ρ, the method exploits the depth-of-field (DoF). DoF is the distance between the nearest and farthest objects that are in acceptably sharp focus in an image. To estimate how much blur is naturally present in the system, the method uses the circle of confusion (CoC) concept from the field of optics [53] (see Fig. 2). When the lens is focused on an object at a distance of d_f, a circle with diameter C is imaged on the retina by an object placed at distance d_p. This diameter can be calculated using:

$$C = As \left| \frac{1}{d_f} - \frac{1}{d_p} \right| \tag{6}$$

where A is the aperture of the eye and s is the posterior nodal distance. So, the developed method computes the estimate of parameter ρ based on where the user is focusing.

It should be noted that the H matrix is basically a representation of the circular PSF in the Fourier domain with ρ being the egocentric radius (in pixels) corresponding to the CoC radius, represented by $\frac{C}{2}$ (in meters). When performing the actual convolution between the filter and screen image, the radius of the CoC is converted to ρ, scaling the values based on the dot matrix of the specific display used.

To obtain the optimal values of λ corresponding to each ρ value, a tuning process was carried out. A variety of virtual scenes containing virtual objects were created. They were blurred using a spatial blurring technique based on foveation and depth-of-field [54]. The blurring technique implements a hybrid approach to incorporate blurring in

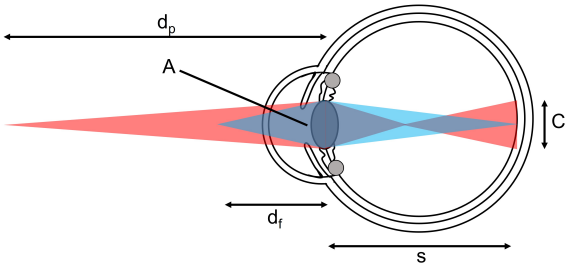

Fig. 2. Illustration of the circle of confusion concept. The point of fixation is at distance d_f. Point located at distance d_p forms a circle on the retina with diameter C. A denotes the aperture and s is the posterior nodal distance.

VR devices by selectively altering the level of blur based on the user's gaze and depth map while prioritizing rendering quality and sharpness in the central vision area. This mimics the natural focusing behavior of the human eye.

The inverse blurring filter H_W was applied to the resulting blurred images. This ensured that the original image and the values of parameter ρ are already known and the value of λ corresponding to each value of ρ can be determined.

To achieve this, the blurred images with known ρ were deblurred with different values of λ. To assess the quality of deblurring, image quality was measured for each image using metrics such as peak signal-to-noise ratio (PSNR) and mean structural similarity index measure (mean-SSIM) [55]. A cut-off value of 25 was chosen for PSNR. Similarly, 0.8 was chosen as the cut-off for mean-SSIM. Using these thresholds, the range of possible values for λ corresponding to each ρ was obtained. Visual information fidelity (VIF) [56] was also considered but not utilized in the final tuning process as it is developed based on natural scene statistics which do not always work well with artificial/virtual scenes.

To further fine-tune the parameters, the FovVideoVDP [57] metric was utilized which is an image/video quality metric that has been developed to evaluate quality based on the foveation that occurs in the human visual system. The FovVideoVDP metric takes into account the fact that the human eye is more sensitive to changes in sharpness and contrast in the central region of a visual stimulus than in the periphery and applies this knowledge to measure the amount of distortion that may be perceived by the viewer when viewing an image or watching a video. The metric considers various factors that affect visual quality, such as compression artifacts, motion blur, and noise. By utilizing the FovVideoVDP metric, the parameters of our system can be fine-tuned to ensure that the resulting images are of high quality and do not contain distortions that may be visible to the human eye. The threshold used for FovVideoVDP was 0.9. For fine-tuning, the effect was applied to the original image and not to the blurred image.

3.2 The Space Variant Inverse Blurring Technique

Eye trackers can help understand where the user is looking which can ultimately help estimate the PSF of the retinal blurring. While the popularity of eye-tracking technology in HMDs may be increasing, it is still not ubiquitous in the VR industry, and many devices on the market do not include this feature. To cater to a larger audience, the technique presented here does not use an eye-tracking system but rather makes use of the virtual scene depth map to calculate the parameter ρ. Real-time measurement of A is not possible in the absence of eye tracker, therefore, we approximate it based on scene brightness.

In most immersive devices such as VR/AR devices, the virtual content is shown on the display screen. Through optics, the focal distance is located at a fixed distance from the user. Typically, this value ranges between 1.5 to 2 meters depending on the setup of the HMD. For example, the HTC Vive Pro has a focus distance of 1.5m while Microsoft Hololens has a focus distance of 2m [58]. Instead of adjusting the PSF based on where the user is looking, the inverse blurring filter H_W is applied based on object distances to the focal plane, i.e., parts of the image are convolved with inverse filters for different PSF estimations that are the proper ones when the user is looking at a specific object. This way a space-variant approach is introduced: the inverse filters deform parts of the stereoscopic images to compensate for the accommodation blurring that develops when objects are focused outside of the focal plane. This enables the observer to see the virtual contents naturally, enhancing the immersive experience.

For optimal performance, immersive media devices require high processing power and resources. Computing the discrete Fourier transform (DFT) and inverse discrete Fourier transform (IDFT) of a large image in each frame to perform the inverse blurring can also be quite expensive. To overcome this, much of the processing can be done offline. For this purpose, it is proposed to compute the inverse blurring filter kernels offline and only convolve these filter kernels in real-time.

Using a shader in the linear color space, each pixel of the image is convolved with a different filter kernel. The choice of the inverse blurring filter kernel is dependent on the difference between the distance represented by the pixel in the depth map and the focal distance, i.e., the distance of each object obtained from the depth map to the focus plane. To this aim, filter kernels H_W are computed at different depth levels in the Fourier domain. Their equivalent filter kernels in the spatial domain are approximated.

When the user is using the device, the depth level of each object present in the user field-of-view is computed in real-time. The inverse blurring filter kernel representing the depth level of each pixel is extracted from the pre-computed data. Each pixel is convolved with the respective filter kernel determined by the depth level of the object representing the pixel. This way a space-variant implementation is achieved which does not require an eye tracker and can offer faster processing times since the computationally expensive elements of the technique are no longer performed in real-time. The overall process is illustrated in Fig. 3. The process is performed individually on each RGB channel. Although the implementation allows for each RGB channel to have a different set of parameters (ρ and λ), for the purpose of this work, the same values are used.

4 Reaching Experiment

In order to understand whether the proposed technique can help reduce the effects of VAC, a study on depth perception

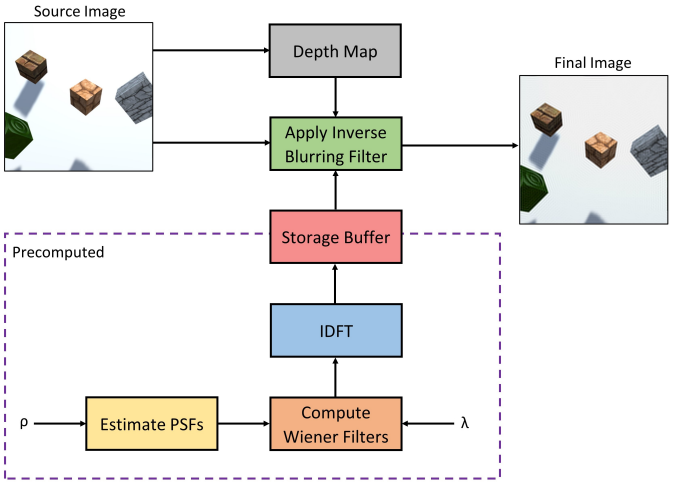

Fig. 3. Process flow of the proposed space variant inverse blurring method.

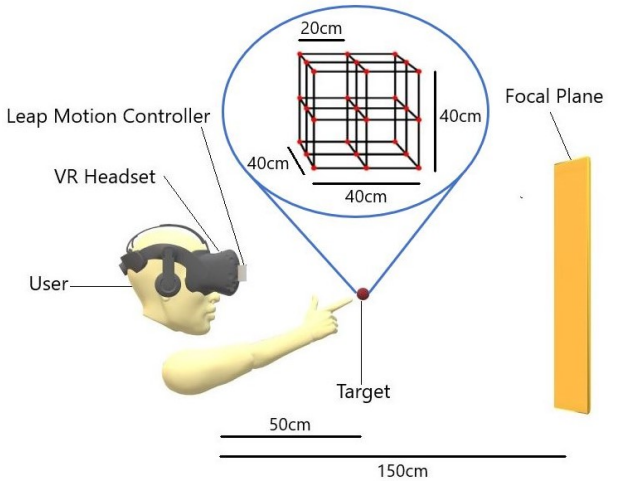

Fig. 4. Sketch of the reaching task. The target position is shown through a small ball that spawns at different locations representing the vertices of a 3x3x3 cubic grid. The 50cm target distance indicates the position of the center vertex.

was carried out. The task utilized was a reaching task in which the users were asked to reach a series of virtual positions using their right-hand index finger, thus exploring distances less than 1m.

4.1 Experimental Setup

The developed system was implemented using Unity operating on an Intel Core i7-9700K processor equipped with an NVIDIA GeForce 1080 graphics card. An HTC Vive Pro device that has a resolution of 1440 × 1600 pixels per eye and a 110°field-of-view was used for interacting with the user. A Leap Motion Controller by Ultraleap was used to track the position of the user's fingers. The Leap Motion Controller was attached to the headset using a camera mount.

A simplistic virtual 3D environment was created containing a spherical object of radius 1cm that spawned at different locations. The spherical object or ball acts as the target position that the user will have to try to reach (see Fig. 4). The users were asked to stand in an empty space. The user's viewpoint was centered, i.e., the head position of the user at the start of each session was reconfigured to act as the origin of the reference frame. For the purpose of the experiment, only the position of the right-hand index finger was tracked.

4.2 Procedure

Data were collected from 20 users (14 males and 6 females) with a mean age of 27.50 and a standard deviation of 6.84. All participants were volunteers and received no reward. The participants were all students, PhDs, and researchers at the University of Genoa and had to sign an informed consent. All users had normal to corrected-to-normal acuity. Users who normally wore corrective glasses or lenses wore them underneath the head-mounted display.

The target positions were vertices of a 3x3x3 cubic grid, where the length of each side was 40cm, i.e., each vertex was positioned 20cm from the adjacent vertex (see Fig. 4). Therefore, the total number of possible positions was 27. The cubic grid was placed 50cm from the user or 100cm from the focal plane. This distance ensured that all possible positions can be reached by the users. In each session, there were 54 trials, i.e, each position was repeated twice. The sequence of the target positions was randomly generated.

The user was asked to reach the target position in a blind-viewing configuration [9], [59], i.e., the stimulus was shown for a short duration (2s). After this, the stimulus was removed from viewing and the user was then asked to reach the position of the ball with their right-hand index finger. Once, they felt that they have reached the target position, they were asked to hold steady their finger and press a button on the HTC Vive Pro controller held in their left hand to register the position.

Two conditions were considered: normal viewing and inverse blurring viewing. In normal viewing, the stimuli were presented in full fidelity. This session acted as the control group to have a reference performance. The stimuli in the inverse blurring session were presented with our technique enabled. Four blur kernels were pre-computed, i.e., three corresponding to the three depth levels of the cubic grid and one corresponding to the wall in the background. The parameters used were based on the tuning process explained in Section 3.2. An example of how the visual stimulus varies between the two experimental conditions is shown in Fig. 5. All users underwent the experimental conditions in random order, i.e., half performed the normal session first and half performed the inverse blurring session first. This was done to ensure no bias was present in the experimental procedure.

For quantitative analysis, the finger positions were used. In order to also have a qualitative or subjective measure, a symptom questionnaire was used. The questionnaire used was developed by Hoffman et al. for their study on VAC [17] and later adapted by Shibata et al. for assessing discomfort in stereo display applications [14]. The questionnaire asked the users to rate their symptoms on a 5-point Likert scale, where 1 indicated no symptoms, 2 indicated mild symptoms, 3 indicated modest symptoms, 4 indicated bad symptoms, and 5 indicated severe symptoms. The questions

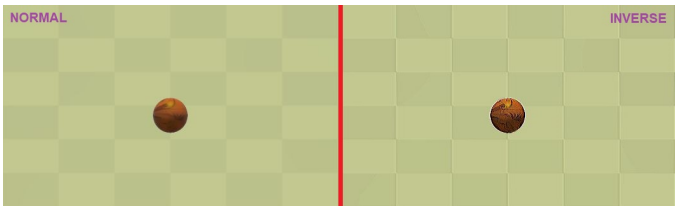

Fig. 5. Example stimuli showing the difference between the normal viewing and the inverse blurring viewing conditions. The ball has a stronger inverse blurring effect as compared to the wall in the background since the ball is further away from the focus plane.

TABLE 1
Mean average absolute errors and their standard deviations for the reaching task on the HTC Vive Pro. All values in cm. Values in bold represent statistically significant.

Error	Normal Viewing	Inverse Blurring
X	1.40 ± 1.51	1.04 ± 1.03
Y	1.31 ± 1.57	0.90 ± 0.84
Z	$\mathbf{3.54 \pm 2.56}$	$\mathbf{2.26 \pm 1.86}$
Euclidean distance	$\mathbf{4.45 \pm 2.78}$	$\mathbf{2.93 \pm 1.93}$

were:

- **Q1)** How tired are your eyes?
- **Q2)** How clear is your vision?
- **Q3)** How tired and sore are your neck, arm, and back?
- **Q4)** How do your eyes feel?
- **Q5)** How does your head feel?

The users filled out the symptom questionnaire after each of the two sessions. When both sessions were completed, the users were asked to fill out a session comparison questionnaire which was also adapted from the work of Shibata et al. [14]. In this questionnaire, the users were asked to rate their experience on a 5-point Likert scale where 1 indicated that the users preferred the first session and 5 indicated a preference for the second session. A rating of 3 indicated no preference between each session. The questions asked were:

- **Q1)** Which session was more fatiguing?
- **Q2)** Which session irritated your eyes the most?
- **Q3)** Which session gave you more headache?
- **Q4)** Which session did you prefer?

4.3 Data Analysis and Results

The error between the expected finger position and the perceived finger position was calculated. The mean errors along with their standard deviation are reported in Table 1. It can be seen that there is a small difference between the performance in the X (horizontal) and Y (vertical) planes. However, there is an improvement of around 1.28 cm in the Z (depth) plane. Error in the Euclidean space was also calculated and a decrease in the error can be noticed. A statistically significant difference has been found in the depth plane and in the Euclidean space ($p < 0.05$, Matlab t-test).

In order to understand, how the behavior is in each of the 3 depth planes. The mean error and standard deviation for each of the planes were plotted as shown in Fig. 6. The distances are measured from the user so the lower depth value indicates the ball is closer to the user. It can be seen that the error increases when the object is in the far field-of-view.

A similar analysis was also done in the horizontal and vertical planes. In the horizontal plane, value 0 indicates the position at the center of the display which is also the center of the user view. The error is higher when the target position was towards the left of the user. A potential reason for this could be that the user was asked to reach the locations with their right hand so the relative distance is higher. In the vertical plane, a similar trend can be seen. As the distance from the user increases, the error also increases.

The finger positions are shown in Fig. 7. It can be observed that with the normal viewing condition, the finger locations are more spread out or less compact as compared to the inverse blurring condition. To have a quantitative measure of this compactness or higher precision, we used the circle error probable (CEP) [60]. CEP is essentially a measure of precision and is defined as the radius of a circle, centered on the mean, whose perimeter includes 50% of the measured positions [61]. The computed radii for each position are reported in Table 2. The spherical error probable (SEP) [62] is also reported which is the same as CEP but for 3D data. The analysis demonstrates that with inverse blurring, the precision is higher since the radii are smaller.

TABLE 2
CEP (XZ; YZ; XY) and SEP (XYZ) radii for the measured finger positions. The data points corresponding to each plane are represented in the same order as in Fig. 7. XYZ radii represent the SEP. A smaller radius represents higher precision. The values are in cm.

	Normal			Inverse		
XZ	1.19	0.98	1.14	1.03	0.87	0.99
	0.81	0.67	0.90	0.58	0.61	0.66
	2.30	1.23	1.06	0.93	1.01	0.99
YZ	1.16	0.95	0.99	0.81	0.68	0.75
	0.77	0.77	1.22	0.70	0.64	0.66
	1.35	2.57	1.26	1.02	0.98	0.90
XY	1.17	1.28	1.03	0.98	0.96	0.96
	1.32	1.25	1.12	0.98	0.87	1.03
	1.39	1.15	2.24	0.99	0.90	1.19
XYZ	2.88			1.91		

TABLE 3
Mean time taken to perform the reaching task on the HTC Vive Pro.

Session	Time (s)
Normal	3.18 ± 1.01
Inverse Blurring	3.21 ± 0.55

The time taken to perform the task was also computed (see Table 3). This time does not include the 2s it took to display the stimuli. The users took similar times for each condition indicating that there was no influence on how much time they spent to reach the position. On average, each session lasted for 4.7 minutes.

Next, the subjective measures were analyzed. Fig. 8 shows the group means along with the standard deviations for the symptom questionnaire. The values are slightly lower for the inverse blurring condition, however, there is no statistically significant difference between the two conditions. More pronounced symptoms can be observed for Q3. This is due to the physical load of the task as many

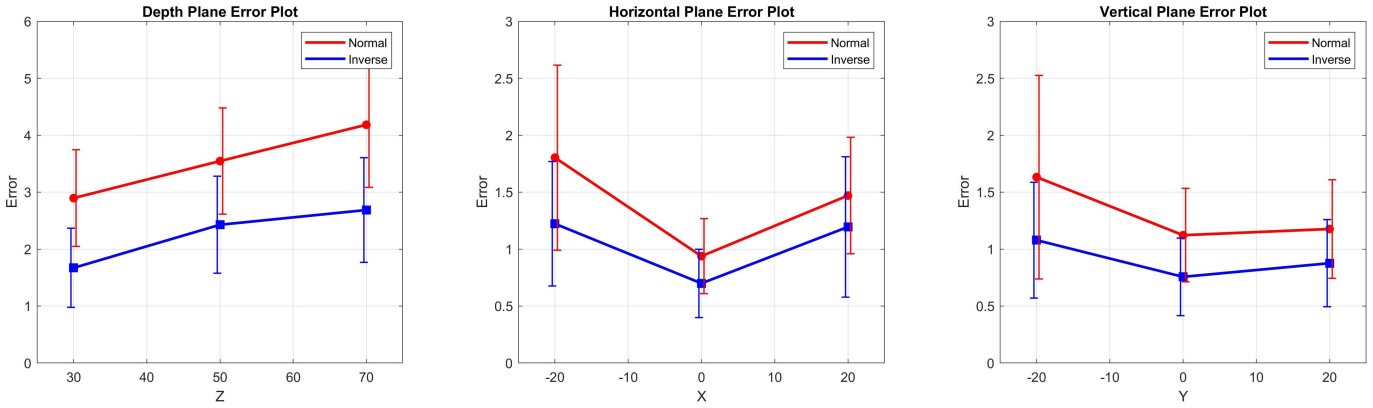

Fig. 6. Error plots and their standard deviations for the different distance levels (depth, horizontal and vertical). Values are in cm.

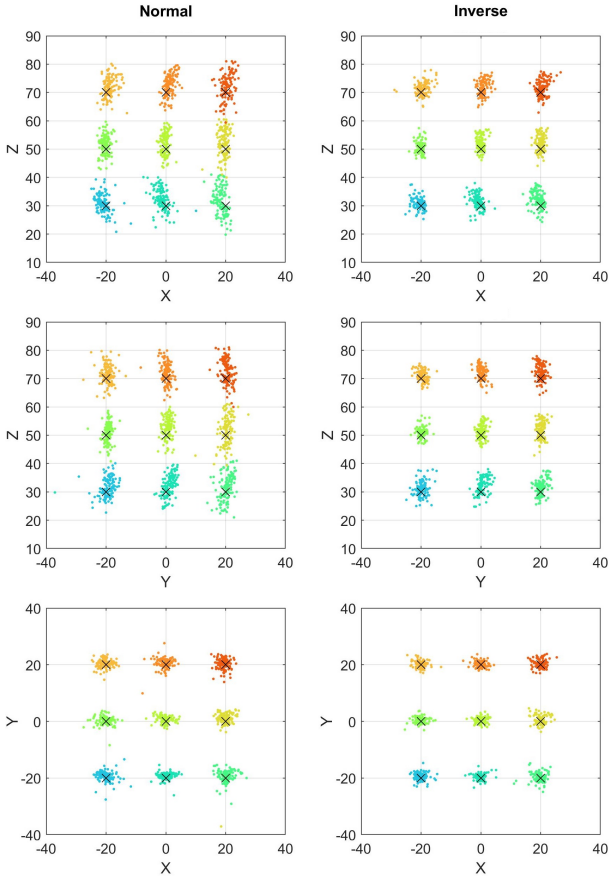

Fig. 7. Finger positions for the different target positions represented as a 2D plot. Values are in cm. (Please refer to Table 2 for quantitative measures of the compactness of the data points.)

users reported that their arms were tired after the sessions. The results for the session comparison questionnaire are also shown in Fig. 8. There is a slight preference for the session with the inverse blurring effect.

5 Spatial Awareness Experiment

A second experiment was conducted to verify the improvement in the performance with the proposed technique and further strengthen the findings of the reaching experiment,

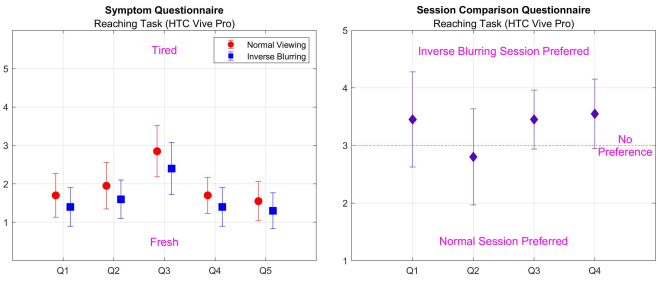

Fig. 8. Symptom questionnaire and session comparison questionnaire scores for the reaching task on the HTC Vive Pro.

by evaluating a different depth range, i.e., distances between 1m and 2m.

5.1 Experimental Setup

The basic experimental setup and the virtual environment were identical to the reaching experiment. In order to record the response of the users, the HTC Vive Pro controllers were used. Two virtual textured cubes of size 10x10x10 cm were placed equally distant from the center of the user view (one towards the left and the other towards the right). The distance between the cubes was 60cm in the horizontal plane and 0 cm in the vertical plane. Ten depth levels were created with 5 cm intervals. These depth levels spanned a range of 1.25m to 1.75m from the user. Since the PSF for defocus blur is identical regardless of whether the object is nearer or farther than the accommodative plane [2], we used six pre-computed filter kernels (five corresponding to the cubes levels and one corresponding to the wall in the background). A plus sign was placed at the center of the view (see Fig. 9).

5.2 Procedure

Data were collected from 18 users (12 males and 6 females) with a mean age of 25.89 and a standard deviation of 7.52. The conditions for the participants were identical to the reaching experiment. 9 of the participants in the spatial awareness study had also participated in the reaching experiment. In order to avoid carry-over effects or biases resulting from participants' prior experience with the reaching study, the spatial awareness study was conducted at a different

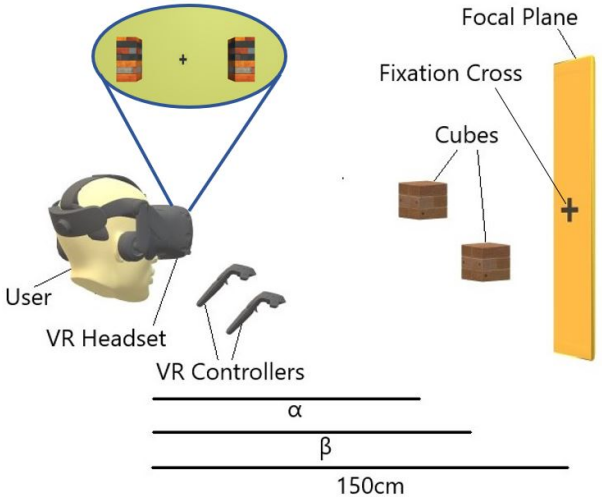

Fig. 9. Sketch of the spatial awareness task. The two cubes appear at different depth planes. These depths are denoted by α & β with a range from 125cm to 175cm. The plus sign at the center of the screen is placed at 150cm, i.e., at the focal plane.

time, with a sufficient interval between the completion of the reaching study and the commencement of this study.

The users were asked to wear the VR device whilst being seated on a chair. The users held an HTC Vive Pro controller in each hand. The controllers acted as the input source. The user pressed the trigger on the controllers to make the selection, i.e., if the user judged that the cube on the left was closer, they pressed the trigger held in the left hand and vice versa.

A scene containing the two cubes was shown (see Fig. 9). The depth level of each cube was randomly selected from the ten depth levels. Each user session lasted for 60 trials. The stimuli were shown for 800 ms. This time was chosen based on studies found in the literature which suggested that humans take around 500–800 ms to respond and fuse the stimuli depending on the distance [17], [63], [64]. When the stimuli disappeared, the users were asked to select which of the two cubes was closer to them. The users made the selection using the HTC Vive Pro controllers. The choice was forced, i.e., even if they perceived the two cubes at the same depth, they had to make a selection. This approach was based on the two-alternative force choice (2AFC) paradigm [65], [66].

Since the observed objects appear at different depths, users may use other depth cues such as relative size to estimate which object is closer. However, the purpose of the user study is to only study depth perception via accommodation and convergence. For this reason, the objects were scaled in such a way that they occupied the same number of pixels on the screen irrespective of their depth. This ensured that users only used accommodation and convergence to make their selections.

In each trial, before showing the stimuli, the users were asked to fixate on the plus sign. They were given 500ms to do this. This was done to ensure that the starting gaze condition was similar for all trials. The experimental conditions and protocol followed were identical to the reaching experiment, i.e., two conditions (normal viewing and inverse blurring viewing), randomized order, and post-session questionnaires.

TABLE 4
Group mean performance for the spatial awareness task on the HTC Vive Pro.

	Normal	Inverse Blurring
Correct	46.5 ± 4.9	51.8 ± 5.1
Incorrect	9.2 ± 4.1	3.3 ± 2.5

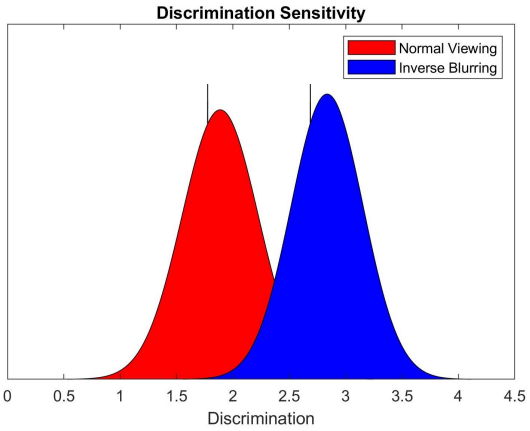

Fig. 10. Discrimination sensitivity plot on the HTC Vive Pro. The vertical bars are the true group means.

5.3 Data Analysis and Results

The number of correct and incorrect answers for all users was computed. The group means along with their standard deviations are summarized in Table 4. It can be observed that the error is much lower in the inverse blurring condition, indicating that the proposed technique lowers the conflict caused by accommodation and convergence in HMDs. It should be noted that in some trials, the two cubes appeared at the same depths. Those trials were considered neither correct nor incorrect.

To understand whether the results have statistical significance, the discrimination sensitivity can be computed for the 2AFC task [66]. The data for each user and condition was converted into discrimination $d^{'}$ [67]. A bootstrapping procedure was used to compute the group confidence levels on $d^{'}$ measurements [68]. Mean $d^{'}$ were computed for each user and condition from the original data re-sampled with replacement 8000 times. These bootstrapped distributions were then collapsed across observers to obtain group distributions for each condition. The group distributions were fitted over a Gaussian distribution from which the 2.5th and 97.5th quantiles were taken as the 95% confidence interval (CI).

Fig. 10 shows the discrimination for the two experimental conditions. A mean discrimination of 1.77 was observed for the normal viewing session whereas the discrimination increased to 2.68 when the inverse blurring condition was presented. The increase is statistically significant ($p < 0.05$, Matlab t-test). The results are summarized in Table 5.

The time it took the users to make the selection was also computed. This time does not include the 1.3s it took to display the stimuli. The mean time taken along with the

TABLE 5
Discrimination sensitivity for the two experimental conditions.

	Normal	Inverse Blurring
True Mean	1.77	2.68
Mean	1.88	2.83
95% CI	[1.55, 2.22]	[2.51, 3.15]

TABLE 6
Mean and standard deviation of the time taken to perform the spatial awareness task on the HTC Vive Pro.

Session	Time (s)
Normal	1.84 ± 0.67
Inverse Blurring	1.87 ± 0.28

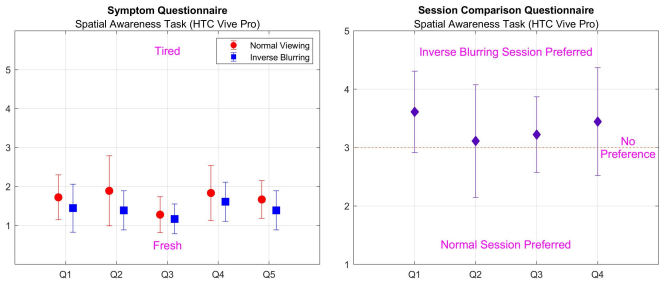

Fig. 11. Symptom questionnaire and session comparison questionnaire scores for the spatial awareness task on the HTC Vive Pro.

standard deviation is reported in Table 6. In both conditions, users took similar times to make their selections. On average, each session lasted for 3.2 minutes.

Next, the subjective measures were analyzed. Fig. 11 shows the group means along with the standard deviations for the symptom questionnaire. Similar to the reaching experiment, the values are slightly lower for the inverse blurring condition, however, there is no statistically significant difference between the two conditions. In the spatial awareness experiment, more pronounced symptoms for Q3 are no longer observed since the physical load of the task is quite low. The results for the session comparison questionnaire are also shown in Fig. 11. There is a slight preference for the session with the inverse blurring effect. After undergoing the two experimental sessions, many users highlighted that they found discriminating the object depths better with the proposed inverse blurring technique.

6 Discussion

For the inverse blurring filter, we used the Wiener filter. It was chosen because, unlike other superior deblurring algorithms, the computational load is less and it is insensitive to small variations in the signal power spectrum. Other approaches either use an iterative procedure that requires a high processing time resulting in an undesired low frame rate or they are based on deep learning models that have high memory requirements.

The shape of the PSF of the blur present in human eyes is similar to a disc filter (circular). However, the exact shape of the PSF varies depending on the eye. In the presented implementation, for a generalized solution, the circular PSF is used as an approximation of the PSF of the natural blur. If a more accurate PSF is used, the improvement in depth perception will be more (at least in theory). However, computing a more complex and accurate PSF will require additional resources and the system will lose generality.

The true SNR of the system is inherently difficult to determine accurately because it necessitates modelling the intricate workings of the human visual system. Furthermore, the SNR can significantly vary from one user to another due to individual differences in visual perception. Given these challenges, it is important to acknowledge that any estimation of the SNR will inevitably involve certain approximations. In our work, we utilize the Wiener filter, which incorporates an approximation of the true SNR to optimize its performance. However, it is essential to note that the SNR parameter in the Wiener filter is a constant value (denoted as λ) that needs to be tuned appropriately for each specific value of ρ, which represents the correlation between the input and output signals. This optimization procedure allows us to adapt the Wiener filter to different signal characteristics and noise levels, improving its performance in different scenarios. It is important to emphasize that the true SNR remains elusive due to the complex and subjective nature of human visual perception. Therefore, our approach acknowledges this limitation and focuses on optimizing the λ parameter within the Wiener filter to achieve the best possible results given the available information and the specific context of the problem.

In an ideal scenario, reproducing depth optically is most desirable, however, the nature of the majority of currently available HMDs necessitates reproducing depth through other means. In our system, we reproduce depth cues through deconvolution-based deblurring. This allows us to develop a general-purpose solution to improve depth perception. We recognize that the effectiveness of our method hinges significantly on the nuanced tuning of parameters of the Wiener deconvolution, and this aspect demands comprehensive exploration. The challenge lies in striking a balance between complexity and simplicity. While a simple model based solely on distance and a basic PSF approximation might seem appealing due to its straightforwardness, the richness of real-world visual perception often involves intricacies that demand a more nuanced approach. Our parameter tuning methodology was designed to cater to these complexities. By delving into a more detailed parameter space, we aimed to capture subtleties that might be overlooked in a simplistic model.

During the parameter tuning process, the choice of using artificial scenes was motivated by usage in the experimental studies. The conducted user studies used a VR environment; hence we use artificial scenes to tune the parameters. Furthermore, in other devices such as OST AR, the method would be applied to virtual objects only as well. However, it should be noted that we do not restrict the application to only artificial scenes. To have a more robust tuning process, some natural scenes may be included as well.

Currently, our method computes the filter kernels offline. This constrains the number of focus levels that can be addressed. In case, an object appears at a distance not placed at the pre-computed level (though many focus levels can be considered), then the closest level can be used. In the future, we plan to investigate how real-time calculation of the filters

using feedback from eye-trackers can aid in providing a more robust performance.

In most commercial HMDs, the focus plane is fixed with respect to the user position. This fixed distance served as a crucial parameter in the experimental setup as the target positions were referenced to the HMD allowing the effective pre-computation of the filter kernels. However, it is imperative to acknowledge the potential influence of this fixed distance on user perception, especially considering the varied configurations of different HMDs. The distance between the user's eyes and the display plane plays a pivotal role in shaping the visual experience, affecting factors such as perceived depth, clarity, and overall immersion. While our results provide valuable insights within the specified focus plane distance, the applicability of our findings to HMDs with different focal planes warrants careful consideration. As a future work, we aim to explore how the positioning of the focal plane impacts the efficacy of our deblurring technique. By conducting experiments across a range of display configurations, we can gain a more comprehensive understanding of the interplay between display distance and the effectiveness of our method. This nuanced exploration is essential for ensuring the generalizability and robustness of our approach, especially in the context of the diverse HMD landscape present in the immersive technology domain.

Although the quantitative measures assessed during the user studies show that the users performed the two tasks better with the proposed technique, the qualitative measures show no significant difference. From a subjective viewpoint, many users preferred the normal viewing condition. A potential explanation for this could be that humans are used to the aesthetics of normal viewing and any small deformations in the scene such as those introduced by the deblurring technique are often considered artifacts. Nevertheless, a slight preference for the sessions with the inverse blurring technique is shown in the comparison questionnaire. Hence, our technique enhances user performance while maintaining a comparable subjective experience.

During the experimental analysis, we mainly focused on how our deconvolution-based method affects depth perception in virtual environments. Our ability to correctly judge distances is affected by the mismatch between convergence and accommodation present in consumer HMDs. Our analysis mainly studied the improvement in depth estimation. We did not study how the accommodative response of the eye changes when deconvolution is applied or how the presence of optical correction affects the performance of our technique and the users' ability to interact with the virtual objects. Such further analysis which is planned in the future will further strengthen the usability of inverse blurring filters to compensate for the discrepancies introduced by VAC.

7 Conclusion

The aim of this work was to develop a technique for immersive media devices that incorporates inverse blurring distortions with the aim of mitigating VAC: here the inverse blurring counteracts the out-of-focus effect of virtual objects that are looked/focused out of the focal plane, thus restoring a natural vergence-accommodation effect. For this purpose, the Wiener deconvolution was used as the deblurring filter.

The inverse blurring filter was computed based on two parameters, λ and ρ which are the signal-to-noise ratio and radius of the circular approximation of human PSF, by performing a parameter tuning procedure. The filter is applied to the stereoscopic images using a novel space variant technique that does not require an eye-tracking integration to determine the optimal PSF. Thus, the developed technique can be incorporated into any immersive media device.

In order to understand how distortions to visual stimuli created by our technique can affect depth perception, two user studies exploring different depth ranges were carried out. The first study was based on a reaching task where users were asked to reach different positions in the personal space with their right-hand index finger while the second user study was based on a spatial awareness task where users had to indicate which of the two objects appeared closer to them. Experimental analysis showed that a statistically significant improvement of 36% and 48% was achieved respectively in depth perception. These values indicate the percentage difference in the mean error in user performance for the reaching task and the percentage difference in the mean discrimination sensitivity for the spatial awareness task between the two experimental conditions.

To conclude, the work offers insight into how inverse blurring can be used to compensate for the discrepancies introduced by the vergence-accommodation conflict in modern HMDs by improving depth perception as evidenced by the two evaluated tasks. By altering the visual stimuli based on techniques inspired by the human physiological system, we can bridge the gap between the real-world visual experience and its virtual counterpart. Possible future works will include testing the developed system on a variety of other devices comprising immersive media.

Funding

This work has been partially supported by the Interreg Alcotra projects PRO-SOL We-Pro (n. 4298) and CLIP E-Santé (n. 4793).

Acknowledgments

The authors would like to thank all the people who voluntarily participated in the user studies.

References

[1] B. Gerschütz, M. Fechter, B. Schleich, and S. Wartzack, "A review of requirements and approaches for realistic visual perception in virtual reality," *Proceedings of the Design Society: International Conference on Engineering Design*, vol. 1, no. 1, p. 1893–1902, 2019.

[2] S. A. Cholewiak, G. D. Love, and M. S. Banks, "Creating correct blur and its effect on accommodation," *Journal of Vision*, vol. 18, no. 9, pp. 1–1, 09 2018.

[3] S. Reichelt, R. Häussler, G. Fütterer, and N. Leister, "Depth cues in human visual perception and their realization in 3D displays," in *Three-Dimensional Imaging, Visualization, and Display 2010 and Display Technologies and Applications for Defense, Security, and Avionics IV*, B. Javidi, J.-Y. Son, J. T. Thomas, and D. D. Desjardins, Eds., vol. 7690, International Society for Optics and Photonics. SPIE, 2010, pp. 92 – 103.

[4] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert, "The perception of egocentric distances in virtual environments - a review," *ACM Comput. Surv.*, vol. 46, no. 2, 12 2013.

[5] J. E. Cutting and P. M. Vishton, "Chapter 3 - perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth*," in *Perception of Space and Motion*, ser. Handbook of Perception and Cognition, W. Epstein and S. Rogers, Eds. San Diego: Academic Press, 1995, pp. 69–117.

[6] S. O. Daum and H. Hecht, "Distance estimation in vista space," *Attention, Perception, & Psychophysics*, vol. 71, pp. 1127–1137, 2009.

[7] E. Klein, J. E. Swan, G. S. Schmidt, M. A. Livingston, and O. G. Staadt, "Measurement protocols for medium-field distance perception in large-screen immersive displays," in *2009 IEEE Virtual Reality Conference*, 2009, pp. 107–113.

[8] P. E. Napieralski, B. M. Altenhoff, J. W. Bertrand, L. O. Long, S. V. Babu, C. C. Pagano, J. Kern, and T. A. Davis, "Near-field distance perception in real and virtual environments using both verbal and action responses," *ACM Trans. Appl. Percept.*, vol. 8, no. 3, 8 2011.

[9] G. Ballestin, M. Chessa, and F. Solari, "A registration framework for the comparison of video and optical see through devices in interactive augmented reality," *IEEE Access*, vol. 9, pp. 64 828–64 843, 2021.

[10] B. G. Witmer and J. Wallace J. Sadowski, "Nonvisually guided locomotion to a previously viewed target in real and virtual environments," *Human Factors*, vol. 40, no. 3, pp. 478–488, 1998.

[11] L. E. Buck, M. K. Young, and B. Bodenheimer, "A comparison of distance estimation in hmd-based virtual environments with different hmd-based conditions," *ACM Trans. Appl. Percept.*, vol. 15, no. 3, jul 2018.

[12] R. S. Renner, E. Steindecker, M. MüLler, B. M. Velichkovsky, R. Stelzer, S. Pannasch, and J. R. Helmert, "The influence of the stereo base on blind and sighted reaches in a virtual environment," *ACM Trans. Appl. Percept.*, vol. 12, no. 2, 3 2015.

[13] G. Kramida, "Resolving the vergence-accommodation conflict in head-mounted displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 7, pp. 1912–1931, 2016.

[14] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks, "The zone of comfort: Predicting visual discomfort with stereo displays," *Journal of Vision*, vol. 11, no. 8, pp. 11–11, 07 2011.

[15] G. Maiello, M. Chessa, F. Solari, and P. J. Bex, "Simulated disparity and peripheral blur interact during binocular fusion," *Journal of Vision*, vol. 14, no. 8, pp. 13–13, 07 2014.

[16] S. J. Watt, K. Akeley, M. O. Ernst, and M. S. Banks, "Focus cues affect perceived depth," *Journal of Vision*, vol. 5, no. 10, pp. 7–7, 12 2005.

[17] D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence–accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision*, vol. 8, no. 3, pp. 33–33, 03 2008.

[18] P. Sanz Diez, A. Ohlendorf, F. Schaeffel, and S. Wahl, "Effect of spatial filtering on accommodation," *Vision Research*, vol. 164, pp. 62–68, 2019.

[19] F. A. Vera-Diaz, R. L. Woods, and E. Peli, "Shape and individual variability of the blur adaptation curve," *Vision Research*, vol. 50, no. 15, pp. 1452–1461, 2010.

[20] M. Day, L. S. Gray, D. Seidel, and N. C. Strang, "The relationship between object spatial profile and accommodation microfluctuations in emmetropes and myopes," *Journal of Vision*, vol. 9, no. 10, pp. 5–5, 09 2009.

[21] J. Xu, Z. Zheng, B. Drobe, J. Jiang, and H. Chen, "The effects of spatial frequency on the accommodation responses of myopes and emmetropes under various detection demands," *Vision Research*, vol. 115, pp. 1–7, 2015.

[22] Y. Kimura, S. Manabe, A. Kimura, and F. Shibata, "Representing virtual transparent objects on ost-hmds considering accommodation and vergence," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2020, pp. 694–695.

[23] J. March, A. Krishnan, S. J. Watt, M. Wernikowski, H. Gao, A. Ö. Yöntem, and R. K. Mantiuk, "Impact of correct and simulated focus cues on perceived realism," in *ACM SIGGRAPH Asia*, 2022.

[24] R. Hussain, M. Chessa, and F. Solari, "Modelling foveated depth-of-field blur for improving depth perception in virtual reality," in *2020 IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS)*, 2020, pp. 71–76.

[25] A. T. Duchowski, D. H. House, J. Gestring, R. I. Wang, K. Krejtz, I. Krejtz, R. Mantiuk, and B. Bazyluk, "Reducing visual discomfort of 3d stereoscopic displays with gaze-contingent depth-of-field," in *Proceedings of the ACM Symposium on Applied Perception*, ser. SAP '14. New York, NY, USA: Association for Computing Machinery, 2014, p. 39–46.

[26] D. C. Rompapas, A. Rovira, A. Plopski, C. Sandor, T. Taketomi, G. Yamamoto, H. Kato, and S. Ikeda, "Eyear: Refocusable augmented reality content through eye measurements," *Multimodal Technologies and Interaction*, vol. 1, no. 4, 2017.

[27] P. Chakravarthula, Z. Zhang, O. Tursun, P. Didyk, Q. Sun, and H. Fuchs, "Gaze-contingent retinal speckle suppression for perceptually-matched foveated holographic displays," *IEEE Transactions on Visualization & Computer Graphics*, vol. 27, no. 11, pp. 4194–4203, nov 2021.

[28] P. J. Bos, L. Li, D. Bryant, A. Jamali, and A. K. Bhowmik, "28-2: Invited paper: Simple method to reduce accommodation fatigue in virtual reality and augmented reality displays," *SID Symposium Digest of Technical Papers*, vol. 47, no. 1, pp. 354–357, 2016.

[29] N. Padmanaban, R. Konrad, T. Stramer, E. A. Cooper, and G. Wetzstein, "Optimizing virtual reality for all users through gaze-contingent and adaptive focus displays," *Proceedings of the National Academy of Sciences*, vol. 114, no. 9, pp. 2183–2188, 2017.

[30] K. Akeley, S. J. Watt, A. R. Girshick, and M. S. Banks, "A stereo display prototype with multiple focal distances," *ACM Trans. Graph.*, vol. 23, no. 3, p. 804–813, aug 2004.

[31] K. Akşit, P. Chakravarthula, K. Rathinavel, Y. Jeong, R. Albert, H. Fuchs, and D. Luebke, "Manufacturing application-driven foveated near-eye displays," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 5, pp. 1928–1939, 2019.

[32] A. Güzel, J. Beyazian, P. Chakravarthula, and K. Akşit, "Chroma-correct: Prescription correction in virtual reality headsets through perceptual guidance," 2022.

[33] D. Lanman and D. Luebke, "Near-eye light field displays," *ACM Trans. Graph.*, vol. 32, no. 6, nov 2013.

[34] F.-C. Huang, K. Chen, and G. Wetzstein, "The light field stereoscope: Immersive computer graphics via factored near-eye light field displays with focus cues," *ACM Trans. Graph.*, vol. 34, no. 4, jul 2015.

[35] P.-Y. Laffont, A. Hasnain, P.-Y. Guillemet, S. Wirajaya, J. Khoo, D. Teng, and J.-C. Bazin, "Verifocal: A platform for vision correction and accommodation in head-mounted displays," in *ACM SIGGRAPH 2018 Emerging Technologies*, ser. SIGGRAPH '18. New York, NY, USA: Association for Computing Machinery, 2018.

[36] K. Akşit, W. Lopes, J. Kim, P. Shirley, and D. Luebke, "Near-eye varifocal augmented reality display using see-through screens," *ACM Trans. Graph.*, vol. 36, no. 6, nov 2017.

[37] P. Chakravarthula, D. Dunn, K. Akşit, and H. Fuchs, "Focusar: Auto-focus augmented reality eyeglasses for both real world and virtual imagery," *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 11, pp. 2906–2916, 2018.

[38] S. A. Cholewiak, G. D. Love, P. P. Srinivasan, R. Ng, and M. S. Banks, "Chromablur: Rendering chromatic eye aberration improves accommodation and realism," *ACM Trans. Graph.*, vol. 36, no. 6, nov 2017.

[39] C. Ebner, S. Mori, P. Mohr, Y. Peng, D. Schmalstieg, G. Wetzstein, and D. Kalkofen, "Video see-through mixed reality with focus cues," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 5, pp. 2256–2266, 2022.

[40] K. Oshima, K. R. Moser, D. C. Rompapas, J. E. Swan, S. Ikeda, G. Yamamoto, T. Taketomi, C. Sandor, and H. Kato, "Sharpview: Improved clarity of defocused content on optical see-through head-mounted displays," in *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, 2016, pp. 173–181.

[41] T. C. O'Haver, *A Pragmatic Introduction to Signal Processing: With Applications in Scientific Measurement*. CreateSpace Independent Publishing Platform, 2016.

[42] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Learning to deblur," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 7, pp. 1439–1451, 2016.

[43] D. Ren, K. Zhang, Q. Wang, Q. Hu, and W. Zuo, "Neural blind deconvolution using deep priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6 2020.

[44] K. Zhang, W. Luo, Y. Zhong, L. Ma, B. Stenger, W. Liu, and H. Li, "Deblurring by realistic blurring," in *Proceedings of the IEEE/CVF*

Conference on Computer Vision and Pattern Recognition (CVPR), 6 2020.
[45] J. L. Starck, E. Pantin, and F. Murtagh, "Deconvolution in astronomy: A review," *Publications of the Astronomical Society of the Pacific*, vol. 114, no. 800, p. 1051, oct 2002.
[46] D. Iwai, S. Mihara, and K. Sato, "Extended depth-of-field projector by fast focal sweep projection," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 4, pp. 462–470, 2015.
[47] L. Zhang and S. Nayar, "Projection defocus analysis for scene capture and image display," in *ACM SIGGRAPH 2006 Papers*, ser. SIGGRAPH '06. New York, NY, USA: Association for Computing Machinery, 2006, p. 907–915.
[48] M. Brown, P. Song, and T.-J. Cham, "Image pre-conditioning for out-of-focus projector blur," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, 2006, pp. 1956–1963.
[49] R. Konrad, N. Padmanaban, K. Molner, E. A. Cooper, and G. Wetzstein, "Accommodation-invariant computational near-eye displays," *ACM Trans. Graph.*, vol. 36, no. 4, jul 2017.
[50] T. Cook, N. Phillips, K. Massey, A. Plopski, C. Sandor, and J. Edward Swan, "User preference for sharpview-enhanced virtual text during non-fixated viewing," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 1–400.
[51] M. S. Arefin, "[dc] sharpview ar: Enhanced visual acuity for out-of-focus virtual content," in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2021, pp. 731–732.
[52] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. Pearson, 2016.
[53] R. T. Held, E. A. Cooper, J. F. O'Brien, and M. S. Banks, "Using blur to affect perceived distance and size," *ACM Trans. Graph.*, vol. 29, no. 2, pp. 19:1–19:16, Apr. 2010.
[54] R. Hussain, M. Chessa, and F. Solari, "Mitigating cybersickness in virtual reality systems through foveated depth-of-field blur," *Sensors*, vol. 21, no. 12, 2021.
[55] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
[56] H. Sheikh and A. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no. 2, pp. 430–444, 2006.
[57] R. K. Mantiuk, G. Denes, A. Chapiro, A. Kaplanyan, G. Rufo, R. Bachy, T. Lian, and A. Patney, "Fovvideovdp: A visible difference predictor for wide field-of-view video," *ACM Trans. Graph.*, vol. 40, no. 4, 7 2021.
[58] B. C. Kress and W. J. Cummings, "Optical architecture of HoloLens mixed reality headset," in *Digital Optical Technologies 2017*, B. C. Kress and P. Schelkens, Eds., vol. 10335, International Society for Optics and Photonics. SPIE, 2017, p. 103350K.
[59] R. A. Weast and D. R. Proffitt, "Can i reach that? blind reaching as an accurate measure of estimated reachable distance," *Consciousness and Cognition*, vol. 64, pp. 121–134, 2018, visual Experience and Guidance of Action: A Tribute to Bruce Bridgeman.
[60] T. Davis, "Confidence region radius," 2008, retrieved August 12, 2022. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/10526-confidence-region-radius
[61] P. A. Zandbergen, "Positional accuracy of spatial data: Non-normal distributions and a critique of the national standard for spatial data accuracy," *Transactions in GIS*, vol. 12, no. 1, pp. 103–130, 2008.
[62] M. Kleder, "Sep - an algorithm for converting covariance to spherical error probable," 2004, retrieved August 12, 2022. [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/5688-sep-an-algorithm-for-converting-covariance-to-spherical-error-probable
[63] G. Heron, W. N. Charman, and C. Schor, "Dynamics of the accommodation response to abrupt changes in target vergence as a function of age," *Vision Research*, vol. 41, no. 4, pp. 507–519, 2001.
[64] J. Tucker and W. N. Charman, "Reaction and response times for accommodation," *American journal of optometry and physiological optics*, vol. 56, no. 8, p. 490—503, 8 1979.
[65] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen, "The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks." *Psychological review*, vol. 113, no. 4, p. 700, 2006.
[66] G. Maiello, M. Chessa, F. Solari, and P. J. Bex, "The (in)effectiveness of simulated blur for depth perception in naturalistic images," *PLOS ONE*, vol. 10, no. 10, pp. 1–15, 10 2015.
[67] T. Wickens, "Elementary signal detection theory," *Elementary Signal Detection Theory*, 10 2001.
[68] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, 1st ed. Chapman and Hall/CRC, 1994.

Razeen Hussain is currently a Research Fellow with the Department of Informatics, Bioengineering, Robotics, and Systems Engineering, University of Genoa, Italy. He received his PhD in Computer Science from the University of Genoa, Italy, and is an alumnus of the Erasmus Mundus EMARO+ cohort. His research interests include natural perception and interaction in VR and AR and computer vision applied to user interfaces.

Manuela Chessa (Member, IEEE) is currently an Associate Professor with the Department of Informatics, Bioengineering, Robotics, and Systems Engineering, University of Genoa, Italy. Her research interests include the study of biological and artificial vision systems, the development of bio-inspired computer vision models, natural human–machine interfaces based on virtual, augmented, and mixed reality, and perceptual and cognitive aspects of interaction in VR and AR.

Fabio Solari is currently an Associate Professor with the Department of Informatics, Bioengineering, Robotics, and Systems Engineering, University of Genoa, Italy. His research interests include computational models of visual perception, the design of bio-inspired artificial vision systems, the study of perceptual effects of virtual and augmented reality systems, and the development of natural human–computer interactions in VR, AR, and mixed reality.