Proceedings of the 56th Hawaii International Conference on System Sciences | 2023

How can Extended Reality Help Individuals with Depth Misperception?

Gholamhossein Kazemi
University of South Eastern Norway
Gkaze@usn.no

Karen Stendal
University of South Eastern Norway
Karen.Stendal@usn.no

Abstract

Despite the recent actual uses of Extended Reality (XR) in treatment of patients, some areas are less explored. One gap in research is how XR can improve depth perception for patients. Accordingly, the depth perception process in XR settings and in human vision are explored and trackers, visual sensors, and displays as assistive tools of XR settings are scrutinized to extract their potentials in influencing users' depth perception experience. Depth perception enhancement is relying not only on depth perception algorithms, but also on visualization algorithms, display new technologies, computation power enhancements, and vision apparatus neural mechanism knowledge advancements. Finally, it is discussed that XR holds assistive features not only for the improvement of vision impairments but also for the diagnosis part. Although, each specific patient requires a specific set of XR setting due to different neural or cognition reactions in different individuals with same the disease.

Keywords: Extended Reality, XR, Vision Impairment, Depth Misperception.

1. Introduction

Information technology (IT) has strongly affected the health care sector, in some cases in life-changing ways (Riva, 2002). For example, damaged retinal neurons cause irreversible sight deficiencies and affect vision, however, Extended Reality (XR) technology is thought to help overcome some aspects (Danciu et al., 2011). Increased understanding of the human body mechanisms, enable improvement of the technologies that interact with individuals. In this regard, XR is being explored to investigate its implications for visionary health issues.

Scholars have studied XR and its applications since its first appearance in academia and industry in the 1950s, and the number of research regarding the applications of XR in health has been growing in the last few years (Berryman, 2012). Nonetheless, these studies mostly cover Computer-Assisted Surgery, Three-Dimensional Imaging, and Computed X-Ray Tomography (Eckert et al., 2019). It reveals that despite the upcoming trend of actual uses of XR in the treatment of patients, some areas are less explored to the extent that there is still a gap in how XR can improve depth perception in patients, especially patients with monocular depth misperception. It is not properly explored that how different settings of XR technology can help individuals with different depth misperceptions and needs. Thus, the research question is as follows:

What XR techniques and technologies can improve which visionary depth misperceptions?

A structured literature review is conducted to investigate this research question. This study underpins XR and visionary issues in general and narrows down the scope of the investigation to how XR can help individuals with depth misperception. Accordingly, XR and its applications and advances will be reviewed in detail along with the explanation of human visionary system mechanisms and its different deficiencies that can occur to anyone.

2. Concepts

In this section the concepts of this literature review are presented to create a better understanding of the field.

2.1. Immersive technology, VR, AR, and MR

Amazingly, the very first endeavors for creating an immersive environment refer to the 17th century when museums and theatres used the reflection of glasses and mirrors to create an immersive experience in the real world. However, the first computer-generated experience of combining reality and virtuality took place in 1963 at MIT and the first prototype was created in 1968 at Harvard University (Billinghurst et al., 2015).

Scholars have defined immersive technology as technology which fades the boundaries between the

URI: https://hdl.handle.net/10125/103026
978-0-9981331-6-4

real world and the virtual world by bringing the sense of immersion to the user via developing realistic virtuality (Suh & Prophet, 2018). Immersive technology embraces the concepts of Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), which are embedded perfectly in the reality-virtuality continuum suggested by Milgram and Kishino (1994). However, the new term Extended Reality (XR) has become an umbrella term covering all the concepts (Çöltekin et al., 2020).

Starting with VR, its origin refers to the 1960s and many definitions of it have emerged over time that all highlight three main features of immersion, perception of presence in the environment, and interaction with the environment. VR includes real-time interactive 3D graphic objects and illudes the sense of participation to the user in a synthetic environment using 3D imaging, stereoscopic Head Mounted Displays (HMDs), and multi-sensors of sound and tracking (Cipresso et al., 2018).

AR is considered a newer technology, although it also has been growing since the 1960s and its different emerged definitions all highlight three main features of combining real and virtual objects in a real environment, real-time interactivity, and registration of virtual and real objects to each other (Cipresso et al., 2018). AR devices include the same elements of 3D imaging, stereoscopic HMDs, and audio and olfactory sensors equipped with body tracking sensors. Quality of experience and sense of realism and presence in AR is dependent on the same factors as in VR (Cipresso et al., 2018). Nonetheless, AR requires more accuracy since it modifies reality and renders virtual objects in real-time (Billinghurst et al., 2015).

Eventually, MR is placed between the two extremes of VR and AR and is clarified as a mix of reality and virtuality; however, it is not restricted to only a visualized interaction between reality and virtuality. MR also encompasses both virtual and real elements that allow data contextualization, interactive real-time virtual objects, and spatially mapped 3d contents (Liberatore & Wagner, 2021).

2.2. XR Evolution

Although the very fundamental features of XR are still the same, great developments and changes have taken place during all these years that have both improved the experience of XR and diversified the core of the technology into different branches that each have their own characteristics. It all started with the "Sensorama" and "Ultimate Display" projects in the 1960s where the former was a motorcycle running simulator and the latter was a university project. These immersive systems used HMD, audio, olfactory, haptic, smell, wind, and interactive graphics to enhance the reality of the experience. Years later, head tracking techniques were added to update virtual images based on the position of the user. In the 1970s the University of North Carolina improved the concept of immersion by facilitating the interaction of two or more users in a 2D virtual space at the same time because of capturing the body's movement and projecting it into an artificial reality (Cipresso et al., 2018).

In the 1980s the U.S. Air Force customized the HMDs to adjust them to their "Super Cockpit" project that further created a basis for modern aircraft helmets, and simultaneously the National Aeronautics and Space Agency (NASA) was building their own HMD from cheap Liquid Crystal Displays (LCDs) for their Virtual Interface Environment Workstation (VIEW) project (Billinghurst et al., 2015). Immersive systems were growing fast and becoming commercialized in the late 1980s when stereoscopic-displaying and mechanical arm tracking devices emerged and developed image stability and movement tracking. Over the years until 2000, eye tracking systems, GPS-based systems, and Mobile AR systems were introduced and elevated the immersive systems more than ever (Cipresso et al., 2018). XR was born in research labs and then spread over different fields of aerospace, defense, industry, education, gaming, and medicine, where big companies like BMW and Boeing used it to improve their manufacturing steps and automotive systems, and health care institutions used it for 3D anatomical visualizing and remote robotic assistance (Aslan et al., 2019). In the last decade, big tech companies like Google and HTC have also been developing their own devices like Google Glass, Google HoloLens, HTC Vive, etc. (Cipresso, et al., 2018).

To acquire a more detailed approach towards the technological evolution of XR, its elements and their progression will be explored in the following. Each XR system consists of two general categories of devices, namely input and output devices. The former gathers the information provided by the user, environment, sensors, etc. and sends it to the processor of the XR, while the latter gets the visual, auditory, olfactory, haptic, etc. content from the processor and provides it to the user (Kim et al., 2020). In other words, input devices enable the user to interact with the virtual environment, and the output devices enable the users to sense what is happening in the virtual environment (Cipresso et al., 2018). Different combinations of input and output devices provide diverse XR systems that each can serve a different goal, satisfy a unique taste, or even treat a special disease (Carmigniani et al., 2011).

2.3. Human Binocular Vision

Binocular vision is a fundamental feature of every typical individual, involving the alignment of the eyes and integration of the 2D perceived images from each eye to create a 3D perception of the peripheral environment, although 3% of the population lack binocular functionality or are not able to benefit it fully. In binocular vision, each eye receives a slightly different image that together form retinal disparities that provide cues for stereopsis or stereoscopic vision (Candy & Cormack, 2021). Stereopsis is categorized into static and dynamic; for example, a moving object has different velocities of movement in each eye's image that are identified as dynamic disparities, while with temporal latency of images received by each eye still the brain can judge depth efficiently with the help of the central nervous system. Oppositely, static stereopsis happens when the observer and the environment have no movement and the image is fixed (Gonzalez & Perez, 1998).

As illustrated in Figure 1, to reach a stereoscopic vision, both eyes create nine potential and two ghost matches from different images to render a 3D image. The main visual signals that are used in the stereoscopic functionality of the eyes to enhance the 3D perception of human vision are disparity, texture, shadow, blur, size, occlusion, shading, motion, and perspective. Each of these signals, solely or together, create cues for visual perception of the environment; however, unfortunately, biological computations responsible for interpreting these signals in the visual cortex are still not discovered (Welchman, 2016).

Figure 1 - Human Binocular Vision Match Points and Signals (Welchman, 2016)

3. Review Process

This literature review has conducted a structured search for prominent scholarly articles based on the guidelines introduced by Webster and Watson (2002). Accordingly, Google Scholar, IEEE, Web of Science, and PubMed were selected to cover both technology and health fields sufficiently. Due to the topic and research question, keywords were identified as "Extended Reality", "Virtual Reality", "Augmented Reality", Literature Review", "Depth Perception", Stereopsis", "Stereoscopic", "Vision", "Health", "Medicine", and "Therapy". Since the topic is a combination of XR technology intertwined with vision, the MeSH index was also considered to increase the accuracy of the search process.

MeSH or Medical Subject Headings is a controller for indexing health articles in PubMed that is introduced by the United States National Library of Medicine (National Center for Biotechnology Information, 2021). In this regard, keywords of "Vision", "Depth Perception", "Stereopsis", and "Stereoscopic" were searched both normally and Mesh-indexed. The whole process consists of 56 times of searches including different combinations of keywords with the operators of "AND" and "OR" in the mentioned databases. Google Scholar, IEEE, Web of Science, and PubMed were searched respectively, 17, 13, 15, and 11 times.

Citation of the articles and authority of the journals were considered as the main factor in the selection. After three rounds of filtration, 199 articles and proceedings were eligible based on the considered inclusion and exclusion criteria. The first round of filtration started with reading the title of the articles to check if it includes either XR, AR, VR, MR, Binocular Vision, Depth Perception, or Review keywords. In the second round the publication year, citation, and authority of the article were considered to increase the validity and authority. In this regard, publications older than 20 years were selected only if they were cited more than 100 times or have been published in a journal or conference with H-Index more than 47. Ultimately, the abstract and conclusion of the articles from the second round were read to indicate the relevance and eligibility of the publication. In this respect, publications must either included definition and description of the keywords, were a review article of the topics close to the interest of this study or were exactly pointing the assistive features of XR in helping people with depth misperception.

The average citation of all the articles is 152.9 and the average citation for the articles from 2020 up to now is 5.7. It is crystal clear that research around the combination of XR and Vision is increasing greatly since 33% of the publications belong to the last 6 years and 14% of them belong to 2020 up to now. Though, only a very few numbers of them are concentrating on depth perception, and even less are concentrated on how XR can help people with depth misperception.

4. Literature Findings

In this section, the taxonomies of XR and its technological elements will be explored with an emphasis on depth perception mechanisms. Afterward, the newest achievements and findings of

XR and depth perception will be scrutinized to distinguish the possible assistive features of XR for improving depth misperception in individuals.

4.1. Extended Reality Systems

Input devices of any XR system consist of a combination or full set of gloves, wireless wristbands, smartphones, trackers, sensors, speech recognizers, etc., and the output devices can be a combination or full set of displays, aural and olfactory gadgets, haptics, glasses, etc. (Cipresso et al., 2018). Based on the interest of this study and considering the research question, this paper only focuses on trackers, visual sensors, and displays. These three elements of XR systems are directly influencing users' visual experience and can potentially change this experience for individuals with visual impairments, hence these elements will be explored deeper in the following.

Human cognition is inspired by vision, hence displays as the means of visualization are a critical element of XR. Displays have developed greatly since their first uses in black and white televisions up to their recent uses as Cave Automatic Virtual Environment (CAVE) or in HMDs. CAVEs are rooms covered fully with displays to deliver the most possible sense of visualized immersion, and HMDs are wearable displays that are categorized into three main types of non-see-through, video-see-through, and optical-see-through (Gerschütz et al., 2019). Displays placed in XR systems can be video-based, optical-based, projection-based, or eye-multiplexed, like glasses (Billinghurst et al., 2015).

In XR systems whose origin stems from AR, video-based systems use cameras and visualize a digital illusion of seeing the real environment on video-see-through displays. In video-see-through displays, the fully computerized visualizations can deliver a detailed and accurate experience. However, in video-see-through displays, the incorrect occlusion of virtuality and reality is a challenge that further can be solved with depth signals from the real world being put in methods like mask objects or depth cameras, and another challenge is the lighting and color differences between the real world and the processed video (Billinghurst et al., 2015).

On the other hand, video-see-through displays are gap-free, or in other words, there is no temporal gap between the rendered virtuality and the real-world image, and they provide a more flexible Field of View (FOV) virtuality with the help of wide-lens cameras. After all, the most important things to consider using a video-see-through display are, respectively, required computational power, resolution enhancement, distortion avoidance, latency elimination, and eye placement. These challenges stem from having an indirect virtual image of the world that is captured through a camera and is being rendered at the same time (Billinghurst et al., 2015).

Furthermore, optical-see-through AR systems use beam splitters like half mirrors or combined prisms in their displays to capture the real environment image through these splitters and combine it with the reflected video of the image on the display. Displays can be fully or semi-transparent equipped with optical combiners. Optical-see-through systems have a huge advantage over video-see-through in resolution reality, distortion-freeness, eyes placement, and latency-freeness because it delivers a direct view of the real world with less computational power required (Billinghurst et al., 2015).

However, registration of virtual objects to the real world is a challenge that further can be solved with calibration. Since calibration is heavily dependent on the spatial relationship between the eyes and the display, accurate eye tracking is required to enhance reality-virtuality alignment, but, in a dynamic environment, tracking objects itself leads to a temporal delay. Also, depth calculation is more challenging in optical-see-through systems. Finally, a fixed amount of transparency itself can affect lighting and colors when the user is indoors vs outdoors (Billinghurst et al., 2015). In XR systems whose origin stems from VR, non-see-through displays provide a fully immersive digital version of the world and deliver a proper FOV without relying on the user's view of the real environment (Yin et al., 2021).

Last, projection-based systems project virtually computerized videos on real surface objects, and eye-multiplexed systems provide both virtual video and rendered-free video of the real world for the user and let the user mentally combine them himself or herself (Billinghurst et al., 2015). Depending on the XR setting, displays can be put at a far or close distance to the user, hung to the user, or even worn by the user. Displays in a far or close distance are mostly used in desktop applications, hung or hand-held displays are used in mobile situations, and worn or HMD displays are used to immerse the user from his or her point of view (Kim et al., 2020).

Another influential element of an XR setting that can inspire human vision is tracking. Any XR setting can benefit from different combinations of tracking techniques and technologies. Tracking is an inevitable part of XR settings that can be used in hand, head, or full-body movement tracking, GPS location tracking, eye tracking, physical object tracking, etc. (Billinghurst et al., 2015). Tracking techniques and technologies use sensors and video cameras and estimate viewpoint poses (Kim et al., 2018). For example, magnetic trackers measure the polarization and orientation of a receiver within the field of its magnetic effective area to calculate the pose. They are fast, light, and small, and occlusion and optical disturbance do not affect them; however, magnetic materials and electromagnetic fields cause deficiencies in the calculation, and the farther the receiver the lower the resolution.

Another instance is vision-based tracking which uses Infrared sensors, Visible Light sensors, and 3d Structure sensors to determine the camera pose by data collected from these optical sensors. (Billinghurst et al., 2015). In this regard, infrared sensors are one of the initiatives of their kind that are scalable, precise, robust, and use the light emitted from the objects to detect them; however, it is expensive and highly dependent on light sources, and reflective surfaces and objects can mislead their calculation. Moreover, Visible Light trackers are optical sensors that use techniques like Fiducial, Naturel Feature, and Model-Based. In this regard, Fiducial techniques use artificial landmarks or papers to track the environment with the help of detecting and comparing marks placed in the environment. Natural Feature techniques combine recent improvements in cameras and image processors with algorithms like SIFT, SURF, BRIEF, ORB, BRISK, and FREAK to increase tracking capabilities (Billinghurst et al., 2015).

Model-based is another example of Visible Light tracking that is based on known 3D structures. Two examples of model-based tracking techniques are, namely Simultaneous Localization and Map Building (SLAM) which was first used in navigating robots and is suitable for large unknown environments, and Parallel Tracking and Mapping (PTAM) which separates environment mapping and camera posing processes and is suitable for small environments (Billinghurst et al., 2015). The 3D structured tracking uses depth signals to create a high-quality three-dimensional model of the real environment, and Inertial tracking uses the Inertial Measurement Units (IMU) sensors like accelerometers, gyroscopes, and magnetometers to calculate the relative orientation and velocity of the object. Inertial tracking is range limitless, and interference-free in the exposure of magnetic, acoustic, optical, and RF sources; however, relying on velocity in performance reduces the accuracy of position and orientation over time (Billinghurst et al., 2015).

Last but not the least, location tracking is benefiting GPS with an accuracy of fewer than 3 meters that can be reduced to centimeters combined with other sensors and techniques. Ultimately, considering these features, strengths, and weaknesses, a Hybrid tracking model, which is a combination of different technologies and techniques, seems to be more flexible for any customized purpose (Billinghurst et al., 2015). Tracking is not only limited to the environment or objects. Motion tracking, including head and body tracking, voice tracking, and eye tracking are other applications of trackers. Head tracking is feasible via using HMDs, voice tracking is doable via using microphones, and eye tracking is possible via capturing eyes' movements like in gaze points (Kim et al., 2020). Eye-tracking information helps visualization updates with different approaches, namely optical, electroocular, and electromagnetic. Optical ones reflect the eyes' surface in gaze situations and are more commercialized, electroocular ones use electrooculogram via skin electrodes to commeasure generated corneoretinal by retinal epithelium within the eyes, and electromagnetic ones use voltage fluctuations of magnetic coined lenses placed in the eyes in a gaze situation (Gobbetti & Scateni, 1989).

4.2. Vision Impairments

From the health perspective, the focus of this study is on depth misperception in individuals with vision impairments. A 2020 study indicates that over 285 million individuals are visually impaired consisting of 39 million blind and 246 million low vision impaired. To assess one's vision, factors like the visual field, near and distance vision, near and distance contrast, and color vision should be considered. Thanks to technology, this assessment and possible impairment therapies are available via using XR (Gopalakrishnan et al., 2019).

Binocular vision, as mentioned before, is an important factor for depth perception. However, some individuals either suffer from binocular depth misperceptions or monocular depth misperceptions. In this regard, some of these disorders will be explained in the following. One of the binocular disorders is suppression, which is simultaneously the cause and result of amblyopia. It blocks stereoscopic vision and prevents from learning to see in stereo and is treatable easily in most cases.

Strabismus is another disorder that can be totally different from patient to patient. Eyes are not aligned, and this misalignment results in binocular depth misperception because these patients cannot use the binocular features of their vision system, although they see through both eyes. Strabismus in patients with albinism leads to disfunction of primary visual cortex vergence (Backus et al., 2017).

In addition, diplopia which results in seeing every image perceived by the eyes in double. Amblyopia decreases the acuity of vision and causes loss of binocular depth perception in long run. Stereo deficiency or lack of binocular function alignment may lead to lack of stereoacuity in short distances while these patients may have no problem with large disparities. Stereo deficiency can be a result of fixation disparity dysfunction, suppression, or inability to extract binocular disparities from the stimuli (Backus et al., 2017). Convergence insufficiency is another disorder that causes changes to binocular vergence, and the patient needs to refixate it by using depth perception cues (Backus et al., 2017).

Eye health and functionality are comprehensible via checking, respectively, ocular health, refractive status, strabismic versus non-strabismic binocularity, sensory fusion, motor fusion, near versus distance stereopsis, accommodation, and oculomotor. Last but not the least, people who lack binocular vision have difficulties in depth estimation, mobility, and orientation even in their daily routines. Binocular vision offers a 180° FOV that in monocular vision it reduces 20°. This can change the daily life of individuals since they need to estimate depth, distance, contrast, acuity, etc. every day in every task they perform (Mhaske et al., 2020).

4.3. Depth Perception

This section scrutinizes the depth perception process in XR settings and in human vision. Starting with human vision, the eyes and brain relate to multiple back-and-forth neural routes that deliver the neural mechanisms responsible for depth perception, 2D images are captured from the environment with the retina to construct a 3D estimation of the world using 3D cues of disparity, texture, shadow, blur, size, occlusion, motion, perspective, and shading (Welchman, 2016).

The signals are like modules that can work solely, in a linear combination, or in a complex relationship with each other to enhance the acuity of depth perception perceived by the brain. These modules have outputs that can affect each other's input, in other words, there is a scenario that indicates that there is no separable module in depth perception mechanism and instead they are all in a quasi-independent situation with limited interactions with each other that together shape the depth perception. Although it has been confirmed different elements of the vision apparatus are more sensitive to different depth signals, and different parts of the brain are more integrated into processing different depth signals, still a full understanding of the neural mechanisms is impossible due to lack of knowledge about the computational logic behind these mechanisms (Welchman, 2016).

On the other hand, depth perception in XR systems tries to mimic human mechanisms to enhance the perception of depth. It is a two-way relationship, where the increase in knowledge about human vision mechanisms helps XR technology to mimic depth perception better, and the better the XR technology can estimate depth the more it can help vision enhancement in individuals (Krajancich et al., 2020). XR systems apply stereo rendering algorithms to model the human visual system; however, there is still a gap in getting equal to reality due to the technology, knowledge, and practicality limitations. In this regard, XR systems are trying to enhance their depth perception algorithms with different techniques and technologies, namely Gaze-contingency, Calibration, Pupil steering, Disparity manipulation, etc. (Krajancich et al., 2020).

Moreover, depth perception enhancement is relying not only on depth perception algorithms, but also on visualization algorithms, display new technologies, computation power enhancements, and vision apparatus neural mechanism knowledge advancements. To have a better understanding of the most recent discoveries influencing depth perception and visualization, in the next step some examples will be explained.

4.4. Recent Discoveries

Since the focus of this study is on depth perception, related advancements in displays, sensors, and vision, will be explained deeper in context. In this regard, a 2019 study defines accommodation-convergence, the field of view, depth of view, pupil position, and interaction as significant requirements of realistic perception. For example, they indicate the importance of interpupillary distance determination in adjusting the visualization to the users' eyes and indicate that field of view should be increased since a normal human can cover about 190° to 290° and most of the HMDs have 110° FOV (Gerschütz et al., 2019).

A 2020 study indicates that eye tracking and gaze tracking are one of the most influential factors in rendering optimization that leads to visualization enhancement. There is a difference between eye-tracking and gaze-tracking, where gaze-tracking includes the users' head movements will be tracked in addition to the eye's movement. They propose perception-based rendering which is highly dependent on the psychological aspects of human vision (Matthews et al., 2020).

Perception-based rendering can be done with the help of foveated rendering, multi-rate shading, and AI-powered foveation. Foveated renderings that can reduce the vertical resolution decreases the response time and eliminate latency in rendering, multi-rate shading manipulates contrast and reduces the resolution for low-contrast regions that opens space for other optimizations like motion-related shading to can tackle the challenge of rendering in motion situations, and AI-powered foveation uses Deep Neural Networks (DNN) to sparse the pixels of a video or image with different densities over the temporal image with the highest density in a foveal region (Matthews et al., 2020). They also propose Eye-Dominance-Guided Foveated Rendering and Combining Shading Rate Images.

Another 2021 study draws out the importance of technology advancement in vision enhancement with a focus on displays. This study indicates that display designing steps should include consideration of the neural coupling between accommodation and vergence to develop disparity detection in a wider range of distances (Candy & Cormack, 2021). A 2021 study exploring the recent developments in XR devices has categorized them into vision interactive, motion interactive, haptic and force feedback, and physiological signal interactive devices.

Another scholarly article of 2020 indicates Gaze-contingent and Eye-Aware HMD calibration and Pupil Steering as helpful advancements in XR systems. The former uses ocular parallax algorithms and gathers not only eligible but also non-eligible information regarding distances between centers of rotation and projection. This helps ordinal depth perception in monocular vision. And the latter is highly dependent on the precise calibration of users' eyes while removing pupil expansion in the pupil steering process. Instead, they propose to track the users' eyes and steer the exit pupil of displays optically towards the users' eyes (Krajancich et al., 2020).

Finally, a 2020 study that has surveyed the monocular depth perception, indicates that advances in Convolutional Neural Networks (CNN) algorithms can elevate monocular depth perception. These deep-learning algorithms apply to depth judgments made by XR settings; however, their daily application is not accessible yet due to the high computational power required and they still need a lot more modification to become as precise as acceptable (Khan et al., 2020).

5. Discussion

This study discusses the necessity of its origin based on the gap that is detected both in the application and literature about the assistive features of XR in vision enhancement. Starting with the applications, most of the XR applications in the medical field that stem from VR are focused on physical disabilities, sensory impairment, cognitive disabilities, autism, learning disabilities, attention deficit, behavioral disorders, and traumatic injuries (Jeffs, 2009). Other medical application examples that stem from AR are cardiac intervention, bone tumor resection surgery, sinus surgery, and spinal surgery (Ha & Hong, 2016).

It is revealed that there is a gap in addressing assistive applications of XR in enhancing depth perception experience in patients suffering from depth misperception. This gap is addressable by applying advanced knowledge of trackers, visual displays, and sensors in the application exploration of conceptual XR settings that directly influence the user's visual experience and depth perception.

For example, one of the most recent articles exploring XR applications in the medical field, indicates that virtual surgery, operation planning, physiological diagnosis, education, training, mental illness treatment, limb pain treatment, surgery training, digital data storage, depression reduction, and body movement tracking are the known applications of XR systems in the medical field (Javaid & Haleem, 2020). However, the potential assistive features of trackers in elevating the experience of depth perception in patients with vision impairment is not mentioned. Hence, this study aimed to explore the gaps by investigating possible applications of XR for vision enhancement. According to the research question, which is "What XR techniques and technologies can help which visionary depth misperception?", this study scrutinized the concepts of XR and vision, surveyed the evolutionary trends regarding these motifs, classified vision impairments, explained the depth perception, and extracted the recent findings of previous scholars.

According to the literature, different types of VR, AR, and MR are included in the XR term, hence this study chose XR term to encompass all the concepts at the same time to enable a holistic investigation of how such a technology can help vision impairments. In other words, different XR settings have different characteristics, and simultaneously vision impairments are caused by different reasons that may be addressable by one specific type of system that is only achievable via a holistic approach.

Further, considering the human visual apparatus and its fundamental logics of 3D construction, influential elements of XR that participate in the visual perception were extracted and explored in detail that revealed how important displays, sensors, and trackers are for vision impairments. After clarification of influential factors, their classification and trend were explored and different examples of XR settings that use various combinations of these elements were extracted from the literature. Also, prevalent vision impairments were diagnosed, and their cause and roots were extracted from the literature.

Based on these findings from the literature, this study suggests that every patient with vision impairment should, first, find the cause and characteristics of his or her visionary problem. This step is necessary because even the same disease of stereopsis deficiency in two different persons can be a consequence of different levels of eye convergence. This applies to all visionary diseases. Then the patient must be aware that the technology may affect each patient differently based on their psychological perception and cognition of the visualizations. Not only does the vision apparatus of two persons with the same disease suffering from the same problem, nor the cognition part of their brain reacts similarly to the same treatment. This clarification can be done with a consent form or contract between the patient and the medical staff participating in the diagnosis or treatment process. This way, not only the diagnosis and treatment process will be individualized but also the patients' rights regarding the medical rights and information-related rights in the context of XR technology will be considered well in advance.

Next, both the diagnosis and the treatment parts can be done with the help of XR technology. Trackers as an influential element of XR can help track the movement of the eyes and head of the patient to assess the pupil, visual cortex, eye muscle, etc. responses to being exposed to different visualizations. As mentioned before, discovered 3D-perception signals that stimulate the vision apparatus are categorized into 9 modules, and with the help of XR and manipulation of these signals through different visualizations and with the help of eye trackers and high-quality cameras that scan eyes, the diagnosis of a wide range of eye impairments become possible. Then, after checking the cause of the vision impairment, tailored XR settings can be set up to start the process of treatment.

As previously discussed, displays with different features can be combined with different cameras, sensors, and trackers to set up an XR system that can directly point to the deficiency and stimulate it. Features of XR are not limited to its hardware, there are also numerous algorithms that optimize the trackers and sensors in different ways that can have different implications for any patient. For example, as mentioned before, SIFT, SURF, BRIEF, BRISK, etc. are all natural feature tracking algorithms that have different characteristics. One is more detailed with higher computational power, one is faster and good for mobile situations, and one is in between and suitable for handheld devices. Considering the capabilities of different tracking systems, it is crystal clear that based on the vision impairment each of these algorithms can have different levels of efficiency.

Not only there is a gap in the literature for potential conceptual designs of XR settings that enhance depth perception experience in patients with visual impairments, but also there is a gap for the potential hypothetical applications of XR in diagnosing and treating depth misperception in patients with visual impairments. For example, to help a sportsman with vision impairments who needs accuracy and detail in depth perception, maybe the best is to choose the SURF algorithm which is faster than SIFT and is rotation-and-scale-invariant. Or for example, to help a patient with binocular vision dysfunction, maybe the best is to choose an HMD device that can adjust the interpupillary distance automatically by tracking the eyes of the user. What this study is trying to convey is that not only the patient and his or her disease and its characteristics are important, but also the context, the cost, the wearability, and the priorities should be considered. Maybe a patient is willing to use AR for the treatment process, and another one prefers VR. Where there is a flexibility of XR setting combination, the needs and priorities of the patient should be considered since still the understanding of the computational logics behind the human vision is not complete and many scholars include psychological factors in the equation as well.

To put it in a nutshell, XR holds assistive features not only for the improvement of vision impairments but also for the diagnosis part. According to the vast number of possible XR systems that can be made with different combinations of displays, algorithms, sensors, trackers, cameras, etc., and because even two persons with the same disease may have different neural problems or cognition reactions, it is impossible to name a specific set of XR system suitable for a specific visual impairment. Ultimately, the technology and the understanding of the human visual mechanisms are growing fast these days and the costs are reducing. This sheds light on the future joints of XR systems and human visual apparatus in both ways. XR Systems and human apparatus can help each other grow in functionality and conception of the environment in a two-way direction.

6. Ethics

This study is a joint between technology and health that includes the participation of medical staff, technology, and patients in the context of diagnosis or treatment processes. Regardless of this joint relationship, every patient and doctor relationship are confidential due to patients' rights to anonymity and patients' information shall not be revealed other than with the consent of the patient and only for the process of diagnosis and treatment (Carmigniani et al., 2011).

It is achievable by providing a consent form that shares patients' rights with them and asks them about their consent in using, analyzing, sharing, or revealing data in different levels of information classification. It should be considered that every step in technological advances is meant to simplify the life of the users not conflicting with it (Carmigniani et al., 2011). Unfortunately, there are no internationally defined standards for XR systems, considering this issue, and according to the origin of XR that is based on information visualization and information collection from the environment and the user, having no standard of privacy or ethical codes may raise concerns for the patients as the focal participant of such applications (Berryman, 2012).

As XR technology gains popularity more and more every day, ethical concerns grow as well on the other hand. For example, advances in GPS tracking, eye tracking, and motion tracking are concurrent with a rich dataset of information being gathered from the user and his or her activities that can be processed with different algorithms that can dig deep into the recognition of patients' location and characteristics with good accuracy (Venkatesan et al., 2021).

Also, the side effects of being exposed to XR systems in a medical process also raise concerns. Namely, motion sickness, the strain on the ocular system, degraded limb and postural control, reduced sense of presence, and negative response development are confirmed that can happen in exposure to XR systems. Hence, dependent on the medical process, whether it is the diagnosis, treatment, etc., considering the personal characteristics of the participants, health institutions benefiting from XR systems should detect any adverse effect in advance and put the patient and his or her rights as their top priority (Riva, 2002).

After all, not only the medical process participants but also the vendors and technology developers must consider the vast applications of what they are developing and possible security breaches that may happen in that special context. The best can be a combination of customized security, privacy, and ethical features for any XR application that is well scrutinized and understood by the participants through security checks, privacy policies, and ethical consents.

7. Conclusion and Future Work

This study conducted a structured literature review to find out how XR technology can help individuals with vision impairments. After three rounds of filtration, 199 articles were extracted for further investigation of the abstract and conclusion. The history of XR was explored and concepts of XR and human apparatus vision were extracted from the literature and investigated in detail. This study focused deeply on XR influential elements like displays, trackers, and visual sensors, and on vision influential modular signals and the way they work together.

After extraction of influential factors, their evolutional trend was investigated and different types of displays, devices, trackers, and sensors were explained in detail. Moreover, prevalent vision impairments were detected in the literature and their cause and effects were listed in tables. Ethics and privacy were explored from the perspective of the patient and discussions were made around all the findings from the article. The research question was answered generally since it is impossible to name a specific set of XR settings that can be helpful for different patients with the same disease. However, it was mentioned that XR can help the diagnosis and treatment processes of vision impairments.

Limitations of this study lay in the field of lack of medical knowledge, time constraints, lack of practical experiments, and real treatment cases. Another limitation that stems from lack of medical knowledge is the generality of the research that is limited to basic vision impairments and technological solutions.

Future work can be focused on the practical part of these findings by gathering patients with the same disease to test different combinations of XR settings in the process of diagnosis and treatment. Also, future work can be narrowed down to vision impairments in a specific context. For example, vision impairments in sportsmen who need to enhance depth perception while maintaining agility, accuracy, and speed.

8. References

Aslan, D., Çetin, B. B. & Özbilgin, İ. G. (2019). An Innovative Technology: Augmented reality based information systems. *Procedia Computer Science, 158*, 407-414. https://doi.org/10.1016/j.procs.2019.09.069

Berryman, D. R. (2012). Augmented reality: A review. *Medical References Services Quarterly, 31(2)*, 212-218. https://doi.org/10.1080/02763869.2012.670604

Billinghurst, M., Clark, A. & Lee G. (2015). A Survey of Augmented Reality. *Foundations and Trends in Human-Computer Interaction 8(2-3)*, 73-272. https://doi.org/10.1561/1100000049

Candy, T. R. & Cormack, L. K. (2021). Recent understanding of binocular vision in the natural environment with clinical implications. *Progress in Retinal and Eye Research*. https://doi.org/10.1016/j.preteyeres.2021.101014

Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E. & Ivkovic, M. (2011). Augmented reality technologies, systems and applications. *Multimedia Tools and Applications, 51*, 341-377. https://doi.org/10.1007/s11042-010-0660-6

Cipresso, P., Giglioli, I. A. C., Raya, M. A. & Riva, G. (2018). The past, present, and future of Virtual and Augmented Reality research: A network and cluster analysis of the literature. *Frontiers in Psychology, 9,* 1-20. https://doi.org/10.3389/fpsyg.2018.02086

Çöltekin, A., Lochhead, I., Madden, M., Christophe, S., Devaux, A., Pettit, C., Lock, O., Shukla, S., Herman, L., Stachoˇn, Z., Kubíˇcek, P., Snopková, D., Bernardes, S. & Hedley, N. (2020). Extended Reality in spatial sciences: A review of research challenges and future directions. *Geo-Information, 9(7).* https://doi.org/10.3390/ijgi9070439

Danciu, M., Gordan, M., Vlaicu, A. & Antone, A. (2011). A survey of augmented reality in health care. *Electronics and Telecommunications, 52(1),* 13-22. Retrieved from http://users.utcluj.ro/~atn/papers/ATN_1_2011_3.pdf

Eckert, M., Volmerg, J. S. & Friedrich, C. M. (2019). Augmented reality in medicine: Systematic and bibliographic review. *JMIR mHealth and uHealth, 7(4).* https://doi.org/10.2196/10967

Gerschütz, B., Fechter, M., Schleich, B. & Wartzack, S. (2019). *A review of requirements and approaches for realistic visual perception in virtual reality.* International Conference on Engineering Design ICED 2019, Delft, Netherlands. https://doi.org/10.1017/dsi.2019.195

Gobbetti, E. & Scateni, R. (1998). Virtual Reality: Past, present, and future, *Virtual Environments in Clinical Psychology and Neuroscience: Methods and Techniques in Advanced Patient-Therapist Interaction* (3-20). IOS. Retrieved from http://www.p-arch.it/handle/11050/682

Gonzalez, F. & Perez, R. (1998). Neural mechanisms underlying stereoscopic vision. *Progress in Neurobiology, 55(3),* 191-224. https://doi.org/10.1016/S0301-0082(98)00012-4

Gopalakrishnan, S., Jacob, C. E. S., Kumar, M., Karunakaran, V. & Raman, R. (2019). Comparison of visual parameters between normal individuals and people with low vision in a virtual environment. *Cyberpsychology, Behavior, and Social Networking, 23(3),* 171-178. https://doi.org/10.1089/cyber.2019.0235

Ha, H. G. & Hong, J. (2016). Augmented reality in medicine. *Hanyang Medical Reviews, 36(4),* 242-247. https://doi.org/10.7599/hmr.2016.36.4.242

Javaid, M. & Haleem, A. (2020). Virtual reality applications toward medical field. *Clinical Epidemiology and Global Health, 8(2),* 600-605. https://doi.org/10.1016/j.cegh.2019.12.010

Jeffs, T. L. (2009). Virtual reality and special needs. *Themes in Science and Technology Education, 2(1-2),* 253-268. http://earthlab.uoi.gr/theste/index.php/theste/article/view/34

Khan, F., Salahuddin, S. & Javidnia, H. (2020). Deep learning-based monocular depth estimation methods – A state-of-the-art review. *Sensors, 20(8).* https://doi.org/10.3390/s20082272

Kim, K., Billinghurst, M., Bruder, G., Duh, H. B. L., Welch, G. F. (2018). Revisiting trends in augmented reality research: A review of the 2nd decade of ISMAR (2008-2017). *IEEE Transactions on Visualization and Computer Graphic, 24(11),* 2947-2962. https://doi.org/10.1109/TVCG.2018.2868591

Kim, Y. M., Rhiu, I. & Yun, M. H. (2020). A systematic review of a virtual reality system from the perspective of user experience. *International Journal of Human-Computer Interaction, 36(10),* 893-910. https://doi.org/10.1080/10447318.2019.1699746

Krajancich, B., Kellnhofer, P. & Wetzstein, G. (2020). Optimizing depth perception in virtual and augmented reality through Gaze-contingent stereo rendering. *ACM Transactions on Graphics, 39(6),* 1-10. https://doi.org/10.1145/3414685.3417820

Liberatore, M. J., & Wagner, W. P. (2021). Virtual, mixed, and augmented reality: a systematic review for immersive systems research. Journal of Virtual Reality, 25, 773-799. https://doi.org/10.1007/s10055-020-00492-0

Matthews, S., Uribe-Quevedo, A. & Theodorou, A. (2020). *Rendering optimizations for virtual reality using eye tracking.* 2020 22nd Symposium on Virtual and Augmented Reality, Porto de Galinhas, Brazil. https://doi.org/10.1109/SVR51698.2020.00066

Mhaske, A., Nikam, A., Bhandari, S. & Mantri, S. (2020). *Smart glasses to assist monocular vision to estimate depth.* 2020 IEEE International Conference of Innovation in Technology (INOCON), Bengaluru, India. https://doi.org/10.1109/INOCON50539.2020.9298311

Milgram, P. & Kishino, F. (1994). A Taxonomy of Mixed Reality Visual Displays. *IEICE Transactions on Informations and Systems, E77-D*(12), 1321-3129. https://search.ieice.org/bin/summary.php?id=e77-d_12_1321

National Center for Biotechnology Information. (2021). *MeSH.* NCBI. https://www.ncbi.nlm.nih.gov/mesh/

Riva, G. (2002). *Virtual reality for health care: The status of research, 5(3),* 219-225. https://doi.org/10.1089/109493102760147213

Suh, A. & Prophet, J. (2018). The state of immersive technology research: A literature analysis. *Computers in Human Behavior, 86,* 77-90. https://doi.org/10.1016/j.chb.2018.04.019

Webster, J. & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly, 26*(2), xiii-xxiii. https://www.jstor.org/stable/4132319

Welchman A. E. (2016). The human brain in depth: How we see in 3D. *Annual Review of Vision Science, 2,* 345-376. https://doi.org/10.1146/annurev-vision-111815-114605

Yin, K., He, Z., Xiong, J., Zou, J., Li, K. & Wu, S. T. (2021). Virtual reality and augmented reality displays: Advances and future perspectives. *Journal of Physics: Photonics, 3(2).* https://iopscience.iop.org/article/10.1088/2515-7647/abf02e/pdf