VisionEnhancer - Mixed Reality App for Visual Accessibility

Jeremy Chew[1], Reto Gerber[1], Carlo Hartmann[1], Susanna Di Vita[1], Bessie Dominguez-Dager[1], Zuria Bauer[1]
{jechew, rgerber, carloh, sdivita } @ethz.ch
[1]ETH Zurich

Abstract

VisionEnhancer is a mixed reality (MR) application for Microsoft HoloLens 2 that enhances accessibility for individuals with visual impairments, including color vision deficiency (CVD) and low vision. It integrates dynamic color correction, real-time magnification, and Optical Character Recognition (OCR) with text-to-speech functionality. The system leverages Daltonization for personalized color adjustments and a Flask-based server pipeline for asynchronous OCR and text processing, enabling real-time performance. Extensive user studies demonstrate its potential to improve accessibility and usability in educational and professional settings. VisionEnhancer represents a concrete step toward realizing the potential of MR technology in creating adaptable, user-centered assistive tools.

1. Introduction

Over 1.4 billion people worldwide experience visual challenges, ranging from color vision deficiency (CVD) to low vision and myopia, significantly impacting their ability to interact with the environment, read text, or distinguish colors in daily life. Traditional assistive technologies, such as corrective lenses or magnification devices, often provide limited solutions, failing to address the dynamic and context-dependent needs of individuals with these impairments. Recent advances in mixed reality (MR) technologies, exemplified by devices like the Microsoft HoloLens 2, present an opportunity to bridge this gap by offering real-time, adaptive visual augmentation solutions tailored to individual needs.

The Vision Enhancer project seeks to leverage these advancements by developing a mixed reality application that integrates customizable features for both CVD and low vision users. The application combines dynamic color correction, real-time magnification, and Optical Character Recognition (OCR) with text-to-speech functionality to empower users to navigate their environments more effectively. By building on foundational methods such as Daltonization for color adjustment and asynchronous API-based OCR pipelines, the project aims to create a versatile and user-centered tool for enhancing visual accessibility. This work is particularly inspired by the team's personal experiences of visual challenges in academic settings, which drive a shared commitment to accessibility and inclusivity.

Mixed reality solutions are uniquely positioned to transform the landscape of visual accessibility. Unlike traditional assistive tools, MR applications can dynamically adapt to changing environmental contexts, offering users a seamless and intuitive interface for interacting with their surroundings. By integrating advanced computational techniques and leveraging the immersive capabilities of MR devices, the Vision Enhancer project aspires to set a new standard for assistive technologies, addressing not only functional needs but also enhancing user confidence and independence.

The remainder of this paper outlines the methodological framework of the Vision Enhancer project, evaluates its performance through user studies, and discusses the implications of integrating MR technologies for visual assistance. The following sections provide an overview of the related research landscape, highlight technical considerations, and set the stage for evaluating the project's outcomes.

2. Related Works

The Vision Enhancer project builds upon a growing body of research in augmented and mixed reality applications for visual augmentation. Prior studies, such as Chroma (10), have demonstrated the potential of wearable devices to address color blindness by employing real-time recoloring algorithms. Similarly, Zhao et al. introduced Foresee (12), a customizable head-mounted vision enhancement system that addresses low vision through features like magnification and edge enhancement. These works underscore the feasibility and impact of leveraging wearable technologies to assist individuals with visual impairments.

Color Vision Deficiency (CVD) solutions have benefited from methods like Daltonization, which adjusts colors in images to make them more distinguishable for users with different types of color blindness, including deuteranopia, protanopia, and tritanopia. Jefferson and Harvey's seminal works (5; 6) laid the groundwork for accommodating

CVD users through computational techniques, while recent advancements, such as Choudhry's implementation of live video recoloring (2), have expanded the scope to real-time applications. However, challenges in computational efficiency and scalability persist, particularly for applications requiring real-time processing on wearable devices.

For individuals with low vision, magnification tools have proven effective in enhancing visual clarity. Huang's augmented reality sign-reading assistant (4) highlighted the value of integrating OCR for text recognition, while van der Aa et al. (11) and Kasowski et al. (7) demonstrated the broader potential of extended reality (XR) technologies to improve accessibility for visually impaired users. However, these studies also noted significant barriers, including device ergonomics, latency, and usability challenges, which need to be addressed for widespread adoption.

Mixed reality applications for accessibility benefit from user-centered design principles, as emphasized by frameworks like those proposed by Langlotz et al. (8) and Schulz et al. (9). These frameworks advocate for iterative development processes that involve end-users throughout, ensuring that solutions align with their needs and preferences. In line with these principles, Vision Enhancer incorporates extensive user feedback to refine its features and evaluate its impact through structured usability and usefulness studies.

By integrating insights from these foundational works, the Vision Enhancer project aims to advance the state of the art in MR-based visual augmentation. Through innovative combinations of color correction, magnification, and OCR functionalities, the project aspires to provide a comprehensive solution that addresses the diverse challenges faced by individuals with visual impairments, setting a precedent for future developments in this domain.

3. Methodology

The VisionEnhancer app consists of 3 main features: Image magnification, color blindness filters and OCR with text-to-speech. In this section each feature and the methods used to implement them are discussed.

3.1. Image Magnification

For the zoom a copy of the camera input is set into the scene and adjust the view on there. The scene then has a webcam-shader where the zoom can be adjusted, by digitally magnifying the view. The zoom has a range of 1x-4x.

3.2. Color blindness filters

The color blindness filters also make use of the webcam shader. The implementation contains 3 different filters one for each type of color blindness. Using array multiplication the RGB-values are convert from the scene to LMS-space, multipled with the appropriate filter for the corresponding color blindness and then converted back to RGB. The LMS-space is constructed as follows:

$$\begin{bmatrix} 17.8824 & 43.5161 & 4.11935 \\ 3.45565 & 27.1554 & 3.86714 \\ 0.0299566 & 0.184309 & 1.46709 \end{bmatrix} \begin{bmatrix} r \\ g \\ b \end{bmatrix} = \begin{bmatrix} l \\ m \\ s \end{bmatrix}$$

To then adjust the values to the corresponding color blindness the LMS-matrix is multiplied with the respective matrix. The matrices for each color blindness are as follows: 1 for Protanope, 2 for Deuteranope and 3 for Tritanope.

$$\begin{bmatrix} 0.0 & 2.02344 & -2.52581 \\ 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \tag{1}$$

$$\begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.494207 & 0.0 & 1.24827 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \tag{2}$$

$$\begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.494207 & 0.0 & 0.0 \\ -0.395912 & 0.801109 & 0.0 \end{bmatrix} \tag{3}$$

The last step is to transform the LMS values back to RGB. Similar to above this is achieved by matrix multiplication. The conversion matrix is as follows:

$$\begin{bmatrix} 0.0809444479 & -0.130504409 & 0.116721066 \\ -0.0102485335 & 0.0540193266 & -0.113614708 \\ -0.000365296938 & -0.00412161469 & 0.693511405 \end{bmatrix}$$

From this the new RGB values are computed for the corresponding colorblindness. The values of the scene are then set to the new values.

3.3. OCR and Text-to-Speech

The OCR and Text-to-Speech functionality is implemented using the Google Cloud Vision and OpenAI GPT APIs, orchestrated through a Flask server that handles image processing and text-to-speech conversion. The pipeline operates as follows:

1. The images captured through the HoloLens are sent as base64-encoded data to the Flask server.

2. The server utilizes Google Cloud Vision API to detect text in the image. The image data are processed using a structured JSON request and the detected text is extracted from the API response.

3. The detected text is sent to the OpenAI GPT model with a custom prompt for additional processing designed to optimize speech output. Technical elements such as URLs, citations, and complex mathematical symbols are simplified or removed, and long paragraphs are restructured for clarity of speech (eg. once

extensive mathematical content and technical jargon are detected in the text, the server dynamically adjusts the response to either simplify the content or flags it as unsuitable for auditory output).

4. The processed text is returned to the Unity app for playback using Text-to-Speech, enabling visually impaired users to receive auditory feedback of textual information in their environment.

The server handles exceptions and provides clear auditory feedback in case of processing failures.

4. Results and User Study

In this section the project results and the user study are presented. The primary goals outlined in the project proposal were successfully achieved. The final application provides a suite of tools to potentially level the playing field for color blind or visually impaired students or other users. The project displays the still mostly untapped potential of Mixed Reality devices to aid in such cases. Thanks to the feedback of color blind users during the development process the authors were able to that they are functioning and where improvements were necessary. The user tests have shown that color blind were able to successfully identify the numbers on an Ishihara test sample. Furthermore using the Hololens 2 Webcam, users are able to magnify potentially hard to see areas like a distant blackboard. Lastly, a pipeline of state-of-the-art vision model API calls allows users to have text read aloud, addressing challenges faced by both color-blind and low-vision users. This feature successfully combines MR and machine learning, demonstrating the promising potential of this integration.

4.1. Study Design

To evaluate the app, the authors conducted a user study using the technology acceptance model (TAM) as designed by Davis (3). This allowed to test the following two aspects to gain important insights on the quality of the application:
1. Evaluate the app's usability by computing a usability score based on system usability scale (SUS) type questions as proposed by Brooke (1).
2. Gain qualitative feedback on the app's usefulness based on the participants opinion from affected (color blind or low vision) and unaffected users.
A total of 15 participants were recruited for the study. Participants completed a 16-question survey after using the application. The first four questions addressed potential external variables that could influence responses. These variables included:

- Usage setting: Whether the participant used the app on Demo Day (9 participants) or in another context (6 participants).
- Participants age: Tab. 1 provides participant age categories.
- Participants visual impairments: The study included 2 color-blind and 1 visually impaired participant.
- Familiarity with Mixed Reality Devices: Fig. 1 illustrates the participants' familiarity levels.

Category	Count	Percentage (%)
< 18	0	0.00
18-25	6	37.50
25-35	6	37.50
> 30	4	25.00

Table 1. Distribution of Study Participants by Age Category

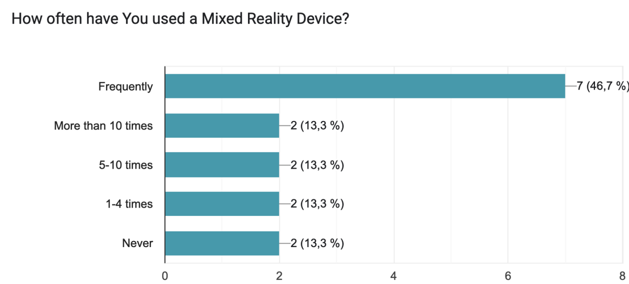

Figure 1. Distribution of Study Participants' familiarity with Mixed Reality Devices

4.2. Usability

To understand how usable the app is, the authors designed 10 to the app tailored likert questions. The participants were prompted with a statement and were to respond with a score from 1 meaning strongly disagree to 5 meaning strongly agree. The Questions are split in two equal sized groups once formulating a positive sentiment and the second a negative sentiment. They were also reordered to group in similar themed questions which would allow users less mental load of jumping between scenarios. The following Questions were defined:

1. I thought the app was attractive and/or easy to use
2. I found the magnification, filtering, and read aloud features to be well integrated
3. I felt very confident using VisionEnhancer
4. I think that colorblind or low vision users would like to use this app in lecture settings
5. I would imagine that most colorblind or low-vision users would learn to use this app very quickly

6. I found the app unnecessarily complex, given its purpose of aiding low-vision user
7. I needed to learn a lot of things before I could get going with the app
8. I think that I would need technical assistance from someone to use this app effectively
9. I thought there was too much inconsistency in how the app worked across its features
10. I found the app to be very difficult and/or frustrating to use

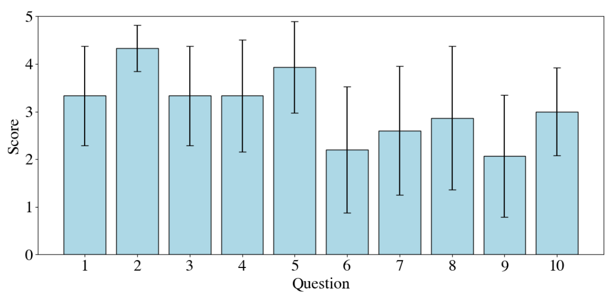

Figure 2. Survey Results: Average Score and Standard Deviation per Question addressing Usability Scores. The chart shows mean scores for questions 1–10, with error bars representing standard deviations.

Fig. 2 displays the resulting average score and standard deviation of the gathered responses. To compute the usability score the standard formula is applied to normalize the score in the range of 0 to 100:

$$S_u = 2.5 \cdot \left(\sum_{i=1}^{5} (r_i - 1) - \sum_{i=6}^{10} (5 - r_i) \right) \quad (4)$$

where r_i is the raw score of the i-th question.
The average system usability score across all users totals to 61.58 which is slightly below the adopted threshold of 68 for a system to be considered to have good usability.

4.3. Usefulness

To evaluate usefulness, participants provided subjective feedback based on their experience and anticipated impact. This assessment involved a subset of the usability questions, supplemented with additional questions tailored to this metric. The reason behind reusing the questions was that they overlapped and that the questionnaire was shorter to keep the participants' frustration lower and achieve more complete and thoughtful responses. The questions were posed such that not affected users could still give their subjective take which produced more data The following questions were used:

4. I think that colorblind or low vision users would like to use this app in lecture settings (same as 4)
5. I would imagine that most colorblind or low-vision users would learn to use this app very quickly (same as 5)
11. I would find this app useful for visually impaired users
12. I imagine this app could improve the performance of visually impaired students

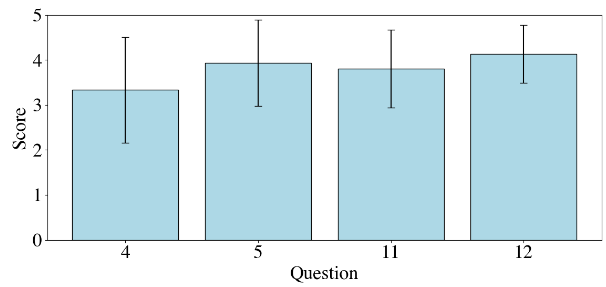

Figure 3. Survey Results: Average Score and Standard Deviation per Question addressing Usefulness Scores. The chart displays mean scores for questions 4, 5, 11, and 12, with error bars indicating standard deviations.

Fig. 3 shows the average score and standard deviation for each question from all participants. The overall sentiment was positive especially for question 12 with an average response of 4.13 regarding the potentional to increase performance of color blind or low vision students.

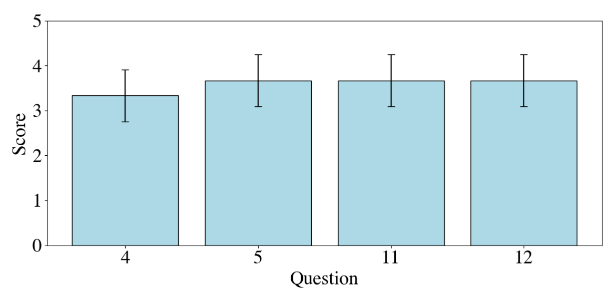

Figure 4. Average score and standard deviation per question for the usefulness score of the target users. The chart shows mean scores for questions 4, 5, 11, and 12, with error bars representing standard deviations. These results highlight variability in user feedback on the system's perceived usefulness.

Since the apps target group was color blind and low vision users the authors wanted to isolate the results of these participants. This is displayed in Fig. 4. It shows that affected users responded more neutral to the questions adressing the usefulness. The authors reasoned that not affected users were more optimistic whereas affected users might have reservations visualizing themselves using the device frequently.

5. Discussion

The results of the user study provide several interesting insights into the potential and limitations of using Mixed Reality devices like HoloLens 2 for vision enhancement applications. The discussion is structured around two key aspects: usability and usefulness.

5.1. Analysis of Usability Results

The system usability score of 61.58 falls slightly below the standard threshold of 68 for good usability, suggesting room for improvement in the application's interface and interaction design. Breaking down the usability scores reveals specific areas of concern:

First, Question 2 received the highest average score (4.3), indicating that users found the core features well-integrated. However, the relatively high standard deviation of 1.27 in Questions 6-10 (negative sentiment questions) suggests inconsistent user experiences, particularly regarding the system's complexity and learning curve.

To investigate potential causes the external variables were included to give a more refined view. Isolating the usability score based on familiarity with mixed reality devices lead to the following discoveries. Users who administered frequent usage of mixed reality devices result in an average usability score of 61.07. This is slightly lower than the overall usability score from which the authors refute the hypothesis that technical familiarity affects the usability. Surprisingly the group with the largest usability score of 71.25 was the two users that reported to have never used a mixed reality device before. The authors argue that the sample size was too sparse and that the novelty aspect of the technology caused the participants optimism especially regarding the questions adressing how the targeted users would benefit.

The lower scores on Questions 7 and 8, related to learning requirements and technical assistance, highlight the need for improved onboarding and user guidance. This aligns with previous research on Mixed Reality applications in educational settings [citation needed], where initial user training has been identified as a critical success factor.

5.2. Target User Perception

An interesting finding emerged from comparing the usefulness ratings between affected and unaffected users. The overall sentiment was positive (average score 4.13 for performance improvement potential), and a higher average usability score of 65.41. The increase is not significant but also shows that the target users were not put off from the app's usability. But affected users provided more conservative ratings regarding the usefulness. This disparity might be attributed to several factors:

1. Practical considerations: Affected users may have better insight into the practical challenges of using the device in real educational settings.

2. Experience with assistive technology: Users with visual impairments likely have experience with other assistive technologies, providing a more nuanced comparative perspective.

3. Device ergonomics: The current weight and form factor of the HoloLens 2 might raise concerns about long-term comfort during extended use.

6. Conclusions and Future Works

Several limitations should be considered when interpreting our results. First, the small sample size of affected users (n=3) limits the generalizability of their specific feedback. Second, the testing environment (primarily during demo day) may not fully reflect real-world usage conditions. Finally, the novelty effect of Mixed Reality technology might have influenced user responses, particularly among participants less familiar with such devices.

These limitations, while significant, do not diminish the promising potential demonstrated by the application but rather suggest areas for future research and development. A key point that would need to be addressed is that some users reported that having to pay attention to their head position is unnatural. This opens the question if there is a need for a different techonlogical approach or if users would get used to it.

In conclusion the app VisionEnhancer successfully showed how Mixed Reality and Machine Learning can be leveraged to enhance the vision of color blind or visually impaired or move the visual input to a different sense like auditory perception.

6.1. Future Works

Our findings suggest several directions for improvement: first, the usability scores indicate a need for simplified interaction patterns and more intuitive feature access. Future iterations should focus on reducing the learning curve identified in our study. The authors emphasize the importance of tightly including the target users in the development feedback loop. The disparity between affected and unaffected user ratings also suggests the need for more extensive testing in actual educational settings, particularly focusing on long-term usage scenarios. Finally, while the current implementation successfully demonstrates the concept, optimization of the color correction and the user experience could improve the real-world applicability.

7. Acknowledgments

The authors would like to thank the supervisors for their crucial guidance, the user study participants for their time and valuable feedback, as well as all visitors at the demo day for their inspiring interactions and discussions.

References

[1] John Brooke. Sus: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, 11 1995. 3

[2] Sami Ali Hussain Ali Choudhry. Live video and image recoloring for color vision deficient patients. Master's thesis, University of Windsor, 2021. 2

[3] Fred D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3):319–340, 1989. 3

[4] Jonathan Huang. An augmented reality sign-reading assistant for users with reduced vision. *PLOS ONE*, 14(1):e0210630, 2019. 2

[5] Luke Jefferson and Richard Harvey. Accommodating color blind computer users. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '06, pages 40–47, New York, NY, USA, 2006. ACM. 1

[6] Luke Jefferson and Richard Harvey. An interface to support color blind computer users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1535–1538, New York, NY, USA, 2007. ACM. 1

[7] Justin Kasowski et al. A systematic review of extended reality (xr) for understanding and augmenting vision loss. *Journal of Vision*, 23(1):1–20, 2023. 2

[8] Tobias Langlotz, Jonathan Sutton, and Holger Regenbrecht. A design space for vision augmentations and augmented human perception using digital eyewear. In *Proceedings of the 2024 ACM CHI Conference on Human Factors in Computing Systems*, CHI '24, pages 1–15, New York, NY, USA, 2024. ACM. 2

[9] Christoph Schulz, Nils Rodrigues, Marco Amann, Daniel Baumgartner, Arman Mielke, Christian Baumann, Michael Sedlmair, and Daniel Weiskopf. A framework for pervasive visual deficiency simulation. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 1–10. IEEE, 2019. 2

[10] Enrico Tanuwidjaja, Derek Huynh, Kirsten Koa, Calvin Nguyen, Churen Shao, Patrick Torbett, Colleen Emmenegger, and Nadir Weibel. Chroma: A wearable augmented-reality solution for color blindness. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '14, pages 799–810, New York, NY, USA, 2014. ACM. 1

[11] Hilde P. A. van der Aa et al. Performance on clinical outcomes, activities of daily living, and user experience on head-mounted displays for people with vision impairment. *Ophthalmic and Physiological Optics*, 2024. 2

[12] Yuhang Zhao, Sarit Szpiro, and Shiri Azenkot. Foresee: A customizable head-mounted vision enhancement system for people with low vision. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '15, pages 239–249, New York, NY, USA, 2015. ACM. 1