https://doi.org/10.2352/EI.2023.35.10.HVEI-222

Real-Time Image Processing with Augmented Reality Glasses for Mobile Low-Vision Users

Yang Cai, University of California, San Diego, California, USA
Yanqiao Wang, Carnegie Mellon University, Pittsburgh, PA
Paul B. Freeman, Allegheny General Hospital, Pittsburgh, Pennsylvania, USA

Abstract

We present an unobtrusive wearable assistive technology for people with vision disabilities of central vision loss (CVL) and low contrast sensitivity (LCS) to be used in a mobile situation. Our technology includes a real-time overlay of an enhanced image with tunable magnification, exposure, and digital filtering on augmented reality (AR) glasses. The optical waveguide glass is wireless with simultaneous central vision enhancement with mobility capability. It enables the wearer to continue to use peripheral vision and is fail-safe from electronic disruption such as power outages. Our preliminary tests with low-vision patients show that the AR glass improves patients' vision without disrupting mobility, for example, from 20/200 to 20/20 on average and patients were able to navigate through an office building environment.

Background

According to the Global Estimates of Vision Loss in 2020, globally, 1.1 billion people were living with vision loss [1]. This number will escalate rapidly as the population ages, implying a growing barrier to a wide range of activities requiring visual information access.

Low vision is a condition in which moderate visual impairment is visual acuity worse than 20/60 (6/18) in the better-seeing eye and cannot be fully corrected or improved with regular eyeglasses, contact lenses, or surgery [2]. Low vision is associated with two major types of visual impairments: central vision loss (CVL) and peripheral vision loss (PVL), with low contrast sensitivity (LCS) impacting both. There are a number of conditions that can cause both central vision loss and contrast sensitivity loss and can occur congenitally as well as being acquired over time. For seniors, a primary disease affecting the macula is age-related macular degeneration (AMD) affecting up to 8 million Americans. Low contrast sensitivity is also common in the aging population when the visual signals become blurry. For younger patients, conditions like foveal hypoplasia can impact central vision and contrast sensitivity.

Optically, bioptic telescopic spectacles (BTS) are a traditional method to assist visually impaired people to see better at a distance while still being able to move about in the environment. However, bioptic telescopic spectacles are evident, consisting of either monocular or binocular telescopes mounted in, or on, a pair of spectacles. They can be used for many distances viewing activities, including driving (in some cases) by people with visual acuity that is not sufficient to qualify for an unrestricted driver's license. However, BTSs alone do not fully improve visual contrast sensitivity.

Rapidly growing digital wearable technologies bring new dimensions to assisted vision systems. Currently, head-mounted digital devices do not look like "normal" eyeglasses and can interfere with navigating through the environment, as they eliminate the wearer's ability to use simultaneous central-peripheral vision integration for orientation and travel. The first commercially available head-mounted electro-optical system, the Visionics Low Vision Enhancement System (LVES), was a weighty "helmet" (992.2g), a battery-powered video display system for magnification and contrast enhancement, but the system housing eliminated normal peripheral vision and simply allowed for a field of view limited to the video display. Over the years, systems like IrisVision, E Sight, and NuEyes, have come down in size and weight, are used by visually impaired patients for magnification and enhanced contrast using live image processing [3], and while they offer some peripheral awareness, do not allow for simultaneous central-peripheral vision processing, and are different in appearance from a pair of "normal" glasses. Additionally, there are other head-mounted or hand-held digital devices that are nonvisual, like OrCam [4] which provide auditory feedback of specifics of the visual environment but are mostly useful for individuals with profound central vision loss.

The objective of our study is to develop inconspicuous lightweight, see-through glasses with enhanced image overlay that enable low-vision users to move about freely and safely while using a magnified image and contrast enhancement to identify detail in the environment. Ultimately, the goal is to enable the user to feel more normal in public.

Challenges in Augmented Reality

Augmented Reality is an interface that articulates a physical world with digital content, including visual, auditory, and haptic signals. A typical AR system incorporates three elements: sensing the environment, registering digital content to the environment, and interacting dynamically. In our case, the AR glasses are designed to overlay the enhanced image on the see-through near-eye display (NED). AR is a part of the metaverse family, where virtual reality delivers virtual content; AR, on the other hand, combines real-time sensor data into the physical world and allows the otherwise invisible to be visible, for example, a heads-up display of enhanced images, thermography, a projected vein finder, or an ultrasound flashlight [5]. The spectrum of Extended Reality (XR), including VR, MR (mixed reality, overlaying virtual objects on the real-world scene), and AR is illustrated in Figure 1.

In this framework, currently available AR glasses typically block the user's peripheral vision and are not pass-through or "fail-safe" designs. For example, when the battery dies, the user won't be able to see anything from the head-mounted system. They are designed for stationary use, not to be used while moving about in the environment.

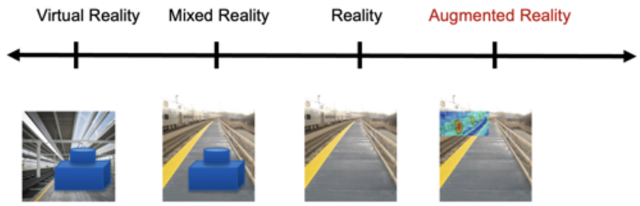

Figure 1. The spectrum of Extended Reality (XR), including VR, AR, and MR

There are many challenges in designing a pair of inconspicuous usable see-through AR glasses for low-vision users. First of all, the glasses need a fast video processing speed (> 30 fps). Second, ideally, the system should provide a certain minimum field of view (>30^o). Third, it should have enough resolution (1024 x780 pixels or higher). Fourth, it needs compact signal processing and computer vision algorithms because of the limitations of the onboard processor and power consumption. Finally, it is desirable to appear like a pair of regular glasses and to be lightweight.

Optically, there are a few options: the "birdbath" design or the optical waveguide design [7]. The latter is able to turn the projected light at a nearly 90-degree angle onto the lens and it enables it to make relatively thinner and lighter glasses. This allows for normal-looking glasses. However, optical waveguides are usually more complex, expensive, and have less resolution. They also have the rainbow artifact at a certain viewing angle.

Figure 2. The light paths of an optical waveguide component

Image Enhancement Space

There are many dimensions for image enhancement. For assisting low-vision users, we focused on three essential functions: magnification, exposure, and filters. Magnification or zooming is the most important function because many low-vision users need magnification to see objects in the distance.

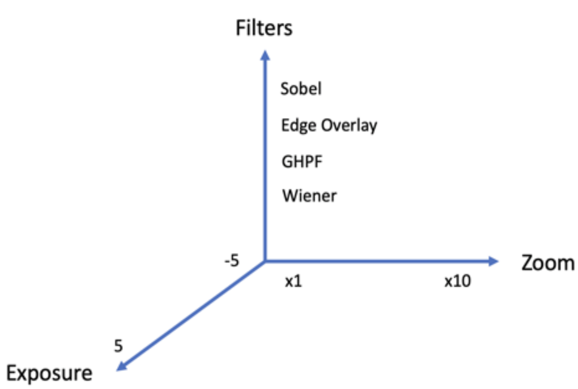

Figure 3. The image enhancement with zoom, exposure, and filters.

It is naturally the first step to simulate a bioptic telescope lens with a digital zooming function in the system. Because many AR glasses use a wide-angle lens, it is necessary to magnify the image to the 1:1 image size baseline before further magnification up to 10 times. The typical magnification range for a bioptic telescope lens is around 2.5x to 6x.

The second enhancement is to adjust the exposure of the camera. The range of the adjustment can be scaled from -5 (very dark) to 5 (very bright). This is a manual adjustment in addition to the automatic exposure adaptation function. It is helpful for observing a target in extreme lighting conditions such as dimmed lighting or a computer screen, etc.

The third enhancement is to use a library of digital filters. Digital zoom is very convenient to implement. However, it brings blurry or pixelated images that are hard to see. Digital filters are necessary to process the magnified images in real time and overlay the filtered images onto the lens. Furthermore, digital filters are designed to highlight significant objects, such as text with contours and gradient contrast for clarity. Currently, we developed a Sobel convolution filter, edge overlay, Gaussian high-pass filter, and Wiener filter.

We found that a combination of the three enhancement functions could help a low-vision user enhance remaining vision. The challenge is that it takes a lot of time and patience to optimize the settings. Users have individualized visual limitations based on their eye condition, and thus their individual preferences. The AR system is able to store the user's default settings and allow the user to adjust in real time. This is very useful in mobile and dynamic situations.

Image Amplification

Image amplification is necessary for many low-vision users. Digital zooming is inevitable when optical zooming reaches its limit. From a user experience point of view, zooming too much would cause confusion – the user might get lost about where to look, and the pixelated image could get blurry. By default, for our system, we set the zoom at a 1.5X ratio and let the user adjust the zooming "on the fly." Because of the limitation of the onboard processor, we use linear interpolation in the resizing process. Figure 4 shows the zooming effect on the AR glasses.

Figure 4. Image magnification is necessary for many low-vision users

Exposure Adjustment

Digital cameras are very sensitive to various lighting environments. In addition to the automatic exposure adjustment, manually fine-tuning is necessary for improving the clarity of the image. This adjustment also enables the negative images or halo effect around the text. Figure 5 shows the view before the exposure adjustment and after. Combined with zooming and exposure adjustment, the user can have a baseline for the further digital filtering processes.

Figure 5. Camera exposure adjustment in addition to the automatic control

Convolutional Filter (Sobel)

Convolutional filters are commonly used in computer vision. They are customizable matrices to be convoluted over pixels in the image. The convolutional filter algorithms are simple to implement but many of them are computationally expensive, which is a challenge to embedded systems that have very limited computational resources. The good news is that there are choices of filters for accomplishing the same task, with a cost of quality. In this study, we use the 3 by 3 Sobel convolutional filter. It yields a decent edge enhancement quality with a modest computational workload. Figure 6 shows the matrix of the Sobel filter and the comparison of the original and filtered image.

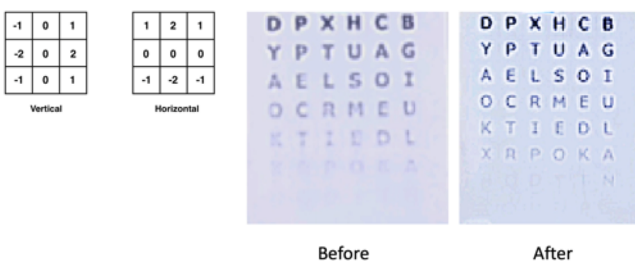

Figure 6. The Sobel filter matrix and the resulting image

Our tests showed that the Sobel filter is able to run at 30 fps on the Snapdragon processor and enhance the edge of the text. The chart on the right-hand side of Figure 6 is a contrast sensitivity test chart. The readable letter indicates the level of contrast sensitivity of the subject. The screen capture shows that the overall image clarity and contrast are enhanced with the convolutional filter.

Edge Overlay Filter

After edge detection, we want to overlay the contour of text on top of the original one. To highlight the contour further, we apply binary morphology to close the contour [8].

$$A \cdot B = (A \oplus B) \ominus B \tag{1}$$

where $\oplus$ and $\ominus$ denote the dilation and erosion, respectively.

From Figure 7, we can see the edges around letters are highlighted with smooth contours. The thickness of the contours can be adjusted from the binary morphologic parameters.

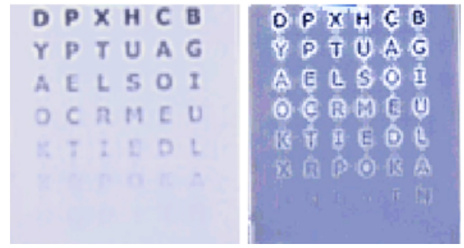

Figure 7. The original image and the result of edge overlay filter

Gaussian High Pass Filter (GHPF)

A high-pass filter removes low-frequency components of an image. The simplest way to create a high-pass filter is to subtract a low-pass filtered image from the original image [9]. Assume the Gaussian lowpass filter is:

$$H(u,v) = e^{-D^2(u,v)/2D_0^2} \tag{2}$$

where $D(u,v)$ is the distance from the center of the frequency rectangle. D_0 is the cutoff frequency. Its maximum value is 0.607. Let us inverse the point spread function (PSF). We have:

$$H(u,v) = 1 - e^{-D^2(u,v)/2D_0^2} \tag{3}$$

Figure 8 illustrates the relationship between the GLPF and GHPF in terms of point distribution functions.

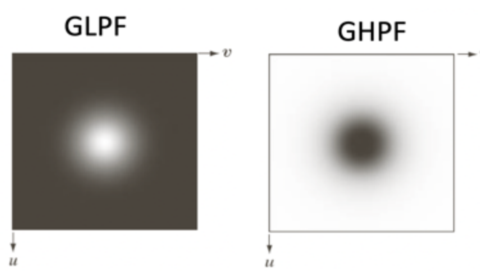

Figure 8. The relationship between the GLPF and GHPF

The GHPF algorithm enhances the sharpness of the image, especially the shape-defining edges that are helpful for the user to separate the foreground from the background. However, the GHPF also introduces high-frequency noises as well. We have to balance the sharpness and high-frequency noises to obtain an optimal solution in real time. From our lab test results, we found that the GHPF improves contrast sensitivity. The first two images in Figure 9 show the difference after applying the GHPF. The right image in Figure 9 also shows the result of the filter on the AR glass.

The GHPF yields smooth enhancement around the edge of the text. It is the best filter as voted by the testers, including patients. The second most used was the edge overlay filter. Perhaps, our vision system already has these filters so that the neural system responds well to the spatial high-frequency signals along the edges of an object, so the ancient hunters could distinguish animals from the background.

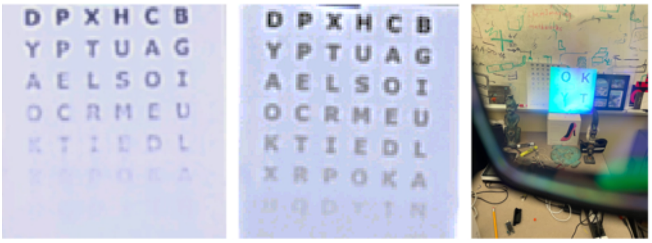

Figure 9. The original Contrast Sensitivity Test Chart streamed from the AR glass (left), the after GHPF (middle) captured from the laptop, and the GHPF results from the AR glass captured by a phone (right)

Wiener Filter

The motivation for exploring the Wiener filter is to reduce the out-of-focus (OOF) noises. It is to compute a statistical estimation of an unknown noisy signal using a related signal as an input and filter that known signal to produce the estimate as an output [10].

$$H_w = \frac{H}{|H|^2 + \frac{1}{SNR}} \tag{4}$$

where H is a frequency response of the point spread function (PSF). In order to recover an out-of-focus image by Wiener filter, it needs to know the SNR and R of the circular PSF.

From Figure 10, we can see the Wiener Filter increases the clarity of the details, e.g. the barcode lines. However, the Wiener filter is very computationally expensive and it also creates ripple-like artifacts.

Figure 10. The original digitally zoomed image (left) and after Wiener Filter (right)

Gesture Language for Real-Time Adjustments

An intuitive user interface (UI) is necessary for real-time adjustments. We use the touchpad on the side of the glasses to enable the user to select settings. We define the following gesture language:

- swapping forward/backward: zooming in or out
- swapping upward/downward: changing exposure
- long-tapping: switch filters
- double tapping: exit the app

The three basic gestures: swapping horizontally and vertically, and long-tapping take care of three basic image enhancement functions. There are more advanced gesture vocabularies that we could use such as two figures' movement horizontally and vertically. However, it would increase the mental workload of the user. In this study, we offload the advanced settings to a connected laptop or an Android phone. Figure 11 shows a user controlling the AR glasses with his finger on the touchpad to control the zooming ratio and exposure.

Figure 11. The user adjusted the zooming and exposure of the camera with his figure on the touchpad

Operation Modes

The AR system can operate in three modes: 1) wireless model, in which the user can wear the AR glasses without attaching to any cables; 2) remote control model, in which the settings can be remotely entered from a smartphone or a laptop; 3) live casting mode, in which the AR glasses can stream the live video to a phone or a laptop.

Live casting in a testing or clinical environment is critical because the developers (and ultimately eye care professionals) need to know where the user is looking and even what app is being used. This can save tremendous time during clinical trials and with user training See Figure 12.

Figure 12. The live casting mode enables the eye care practitioner to know where the patient is looking at and what settings on the AR glasses.

Clinical Lab Test Results

We tested our AR system at the Low Vision Clinic at Allegheny General Hospital, Pittsburgh, PA. So far, we have tested four patients, ranging from 24 years old to 100 years old. We tested both visual acuity and contrast sensitivity for both eyes. Due to the limitations of the AR glasses hardware that has only one display on the right eye, we tested the right eye when the user wore the AR glasses. In addition, the AR glasses do not have prescription lenses to correct for refractive errors, so we tested under two conditions: wearing both AR glasses and the patient's own prescription glasses or wearing the AR glasses only.

The visual acuity and contrast sensitivity were tested in the clinic with the standard projected charts on the wall and the reading chart in the hand.

The results of the visual acuity tests show that all subjects' acuity improved despite age. On average, the acuity could be improved from 20/200 to 20/20. The younger users had much more improvement than the senior subject.

We hypothesized that the AR glasses might improve contrast sensitivity as well. However, to our surprise, their results did not change much. This is likely because the sample subjects' contrast sensitivity was in the normal range. Table 1 summarizes our clinical test results.

We then conducted an informal mobility test with three younger persons (all had normal baseline mobility), who were able to read signs and nameplates on the wall while simultaneously walking around the hospital without assistance while wearing the AR glasses. This demonstrated that the glasses do not interfere with normal indoor mobility.

Table 1. Acuity and contrast sensitivity test results

No.	Age	Acuity (before)	Acuity (after)	Contrast Sensitivity
1	35	20/80 (both eyes)	20/25 (R) with AR glasses	within a normal range
2	30	20/200 (L) 20/100 (R)	20/40 (R) with AR+P glasses; 20/20 (R) with AR glasses	within a normal range
3	24	20/200 (both eyes)	20/40 (R) with AR+P glasses, 20/20 with AR	within a normal range
4	100	20/80 (both eyes)	20/40 (R) with AR glasses	N/A

Usability Study

We found that our users liked the AR glasses because they look like a normal pair of tinted glasses, and if they lose power, the screen does not go black. We also discovered that knowing where the user is looking is critical for clinical evaluation and training, therefore, live casting from the glasses to a laptop via USB cable or wireless was a great help during the testing and calibration. In the future, two displays and two cameras would be desirable as some users prefer using the left eye while others prefer the right eye. Additionally, prescription lenses should be available for those low-vision users who require them.

Conclusions

From our informal clinical testing, we found subjects with low vision preferred optical waveguide-based AR glasses over other low vison devices because they felt more "normal" when wearing the glasses.

The real-time image enhancement showed significant improvement in simultaneous integration of *visual acuity* and *mobility over other devices*. However, we have not found strong cases of improvement in patient's contrast sensitivity yet, probably because the majority of our subjects' contrast sensitivity was within a normal range.

This device is still in its very early stage of development. Future work would include a dual display and dual camera design, energy-efficient algorithms, better deblur filters and patient-specific design.

References

[1] https://www.iapb.org/learn/vision-atlas/magnitude-and-projections/global/
[2] https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment
[3] A.D. Deemer, Bradley CK, Ross NC, Natale DM, Itthipanichpong R,. Werblin FS, Massof RW, Low Vision Enhancement with Head-mounted Video Display Systems: Are We There Yet?, Optom Vis Sci. 2018 September; 95(9): 694–703. doi:10.1097/OPX.0000000000001278
[4] OrCam Device, WikiPedia: https://en.wikipedia.org/wiki/OrCam_device
[5] G. Stetten, Sonic Flashlight. https://medrobotics.ri.cmu.edu/node/128455
[6] Y. Cai, Extreme Reality Interfaces, Proceedings of AHFE-2022, July, NYC, 2022
[7] Optical waveguide: https://en.wikipedia.org/wiki/Waveguide_(optics)
[8] https://en.wikipedia.org/wiki/Closing_(morphology)
[9] https://newion.uwinnipeg.ca/~sliao/Courses/7205/Week07B.pdf
[10] https://en.wikipedia.org/wiki/Wiener_filter

Author Biography

Yang Cai is the founder and head of Visual Intelligence Studio and a Research Scientist at Calit2, UC San Diego, CA, USA. He is the author of 6 AI books, including the monograph "Instinctive Computing (Springer-London, 2017)

Paul B. Freeman is an optometrist and Director of the Low-Vision Clinic at Allegheny General Hospital, Pittsburgh, PA, USA.

Acknowledgment

This project is in part sponsored by DoT University Transportation Center (UTC) at Carnegie Mellon University and NIST Public Safety Communication Research program. The authors would like to thank Stan Caldwell, Lisa Kay Schweyer, Scott Ledgewood, and Professor Mel Sigel for their support. We would like to thank graduate students Haocheng Zheng, and intern Ahmed Othman for their programming.