VRT: Virtual Reality Toolkit

Goyal, Agam
agoyal25@wisc.edu

Mahajan, Ishaan
imahajan@wisc.edu

Jagtap, Mihir
mjagtap@wisc.edu

Aggarwal, Shivansh
saggarwal24@wisc.edu

Abstract

Vision impairment severely impacts quality of life among adult populations, and people with vision impairment often have lower rates of workforce participation as well as work productivity. We plan to implement a Virtual Reality (VR) software stack that would help people with low vision in specific, to identify objects, see images and read better. Through our VR stack, we provide the ability to change the size of text by magnification, contrasting colors in images, and by generating captions for these images. To mitigate dealing with issues related to reading these captions, we will also implement a text-to-speech framework for these captions. This stack is only a software stack, and can be embedded into VR headsets or special glasses to help people suffering from these kind of problems. With the advent of Virtual Reality and the advances being made towards the Metaverse, we hope our work will help people with low vision see better and provide them with a more inclusive tech-space in the future.

Introduction

According to WHO, around 2.2 billion people suffer from some kind of vision problems, making day-to-day life difficult in many situations [1]. A major part to sense any kind of object, or an image is through vision. This is a major problem in our opinion, as it denies certain individuals equal opportunities, and also possesses danger in some cases. For example, driving at night, for people with astigmatism might be risky in certain scenarios, or people having color blindness might not be able to distinguish colors of different fluids, in some cases distinguish blood discharge in certain fluids. Our stack will help solve this problem, and mitigate the risks and potential hazards that come with certain vision problems. We plan to make this possible for everyone by providing tools and mechanisms to comprehend an image/object as clearly as possible, and if not possible in some scenarios, we plan to provide subtitles and using Natural Language Processing, enable text to speech.

Purpose

According to CDC (Center for Disease Control and Prevention), there were 4.2 million Americans aged 40 years or older who suffered some kind of vision impairment in 2012 [2]. This number is expected to double by 2050 which depicts the severity of this issue. People with vision impairment face challenges in their day-to-day life which include reading text, cooking, shopping, and a number of mobility issues which we are unaware of. Many of them have revealed that low contrast sensitivity has led them to walk slowly and cover only short distances [4]. This has been further proven by an analysis by Nabila Jones which states that low vision people shop predictably. It was hard for them to buy a product if its location was changed in the grocery store [3]. Furthermore, there have been reports which states that 6.2% of children younger than 18 have vision condition [2]. Students with vision impairment have trouble focusing on objects like blackboard and books which is one of the reasons behind the high college dropout rate. The motivation behind developing a VR or glasses is to assist these people with performing daily tasks and seeing things clearly.

Approach to Solving the Problem

Leading companies have developed state-of-art applications which are supporting people with impaired visions. Microsoft has presented Seeing VR, a set of 14 tools that enhance VR application for people with low vision by providing visual and audio augmentations [5]. Meta is utilizing Automatic Alternative Text (AAT) to utilize object recognition to caption photos which helps visually impaired individuals [6].

Our approach to solve the problem has four main components: Magnification, Contrast, Object Captioning, and Text-to-Speech.

- **Magnification** is the most commonly used in vision enhancement to enable people with impaired vision to see details. The field view of the camera is adjusted so that magnification level is adjusted.
- **Contrast**: Increased luminance contrast is important for visually impaired people. We plan to enhance the brightness difference between each pixel and the average of its adjacent pixels. We plan to integrate high contrast color schemes [7] which will be best for people with low vision.
- **Object Captioning**: We plan to recognize objects/situations in the field view and include captions with appropriate magnification and contrasting which will help them recognize the objects.
- **Text-to-Speech**: Integrating Text-to-speech with object captioning system will help visually impaired people understand the field view and surrounding environment with better rate.

We plan on to work on an existing approach, as in making the use of VR, but we plan on giving this a new direction. VRs are mainly used to provide entertainment benefits, however, we will be using it to help people.

Final Goals and Evaluation

Our final goal for this project is to successfully implement the Virtual Reality stack fully, and make it scalable for deployment. For evaluation, we hope to get some user feedback from classmates and instructors to understand where we can improve our implementation of the software.

This is an **updated timeline** of how we will go about the project:

Timeline	
Task	Deadline
1. Research about various low vision problems	15^{th} October
2. Work on creating contrasting algorithms	28^{th} October
3. Text Extraction	5^{th} November
4. Mid-term report	10^{th} November
5. Image Captioning and Magnification	22^{th} November
6. Text-to-Speech	30^{th} November
7. Web page	4^{th} December
8. Presentation	6^{th} December

Midterm Progress Report

After submitting the project proposal, we started with our research about various problems faced by low-vision people. Two of the most prominent ones were the difficulty they face in being able to view low-contrast images, and in reading text. We thus decided to address these issues first, and started out with implementing image contrasting algorithms and a basic text-to-speech engine.

Image Contrasting:

For a start, we worked on three different ways of implementing image contrasting:

Contrast Stretching:

In contrast stretching, we take the existing intensity values in the image and "stretch" them to fit the entire range of potential values of 0 to 255.

We do this by a simple technique of taking the individual pixel intensity values, and subtracting the minimum intensity values and then dividing by the difference in the maximum and minimum intensity values in the image. Finally, to get them back to the usual range of intensities, we multiply this value back by 255.

Here are some results that we obtained:

(a) Before Contrasting

(b) After Contrasting

Figure 1: Contrast Stretching - Example 1

(a) Before Contrasting

(b) After Contrasting

Figure 2: Contrast Stretching - Example 2

Histogram Equalization:

Histogram modeling techniques (e.g. histogram equalization) provide a sophisticated method for modifying the dynamic range and contrast of an image by altering that image such that its intensity histogram has a desired shape. Histogram equalization employs a monotonic, non-linear mapping which re-assigns the intensity values of pixels in the input image such that the output image contains a uniform distribution of intensities (i.e. a flat histogram). This technique is used in image comparison processes (because it is effective in detail enhancement) and in the correction of non-linear effects introduced by a display system [8].

We are trying the color enhancement of low light areas. Contrast enhancement of color images is typically done by transforming an image to a color space that has image intensity as one of its components. One such color space is L*a*b*. Use color transform functions to convert the image from RGB to L*a*b* color space, and then work on the luminosity layer 'L*' of the image. Manipulating luminosity affects the intensity of the pixels, while preserving the original colors.

Histogram Equalization performs histogram equalization. It enhances the contrast of images by transforming the values in an intensity image so that the histogram of the output image approximately matches a specified histogram (uniform distribution by default).

Adapt Histogram Equalization performs contrast-limited adaptive histogram equalization. Unlike Histogram Equalization, it operates on small data regions (tiles) rather than the entire image. Each tile's contrast is enhanced so that the histogram of each output region approximately matches the specified histogram (uniform distribution by default). The contrast enhancement can be limited in order to avoid amplifying the noise which might be present in the image.

Here are the results by using Histogram Equalization and by using Contrast-limited Adaptive Histogram Equalization (CLAHE):

(a) Before

(b) After

Figure 3: Histogram Equalization - Example 1

(a) Before

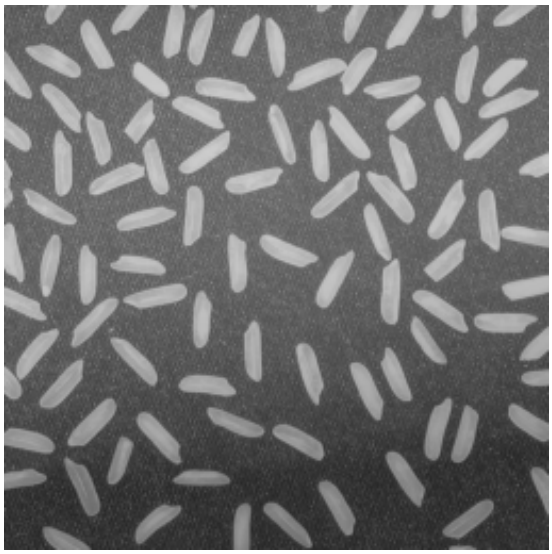
(b) After

Figure 4: Contrast-limited Adaptive Histogram Equalization (CLAHE) - Example 1

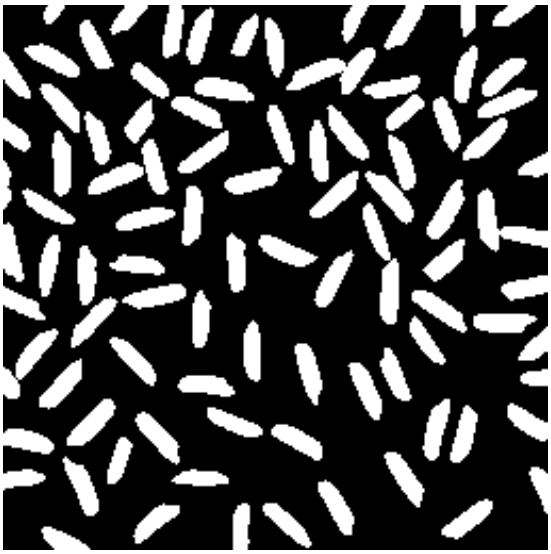
Correct Non-uniform Illumination:

To enhance an image as a pre-processing step before analysis, we correct the nonuniform background illumination and convert the image into a binary image to make it easy to identify foreground objects (individual grains of rice). We can then analyze the objects, such as finding the area of each grain of rice, and then can compute statistics for all objects in the image.

We get the following result for the Correct Nonuniform Illumination:



(a) Original Image



(b) After Contrasting - individual grains of rice

Figure 5: Correct Non-uniform Illumination - Example 1

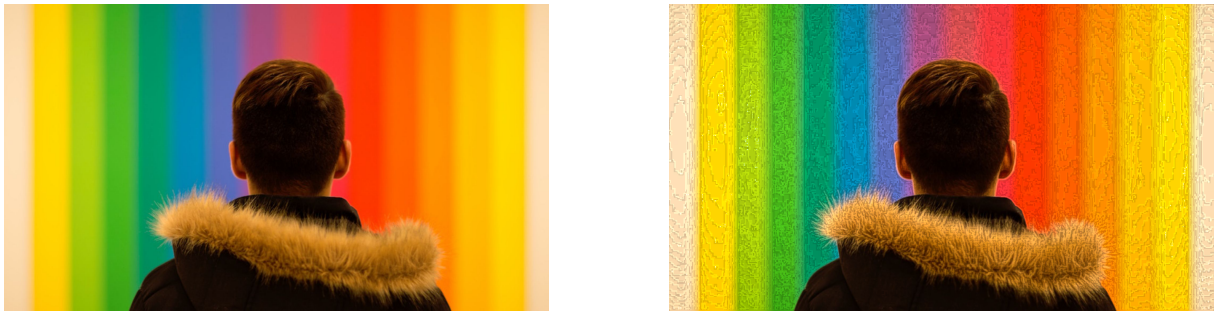
Novel Approach

One of the techniques that we tried to develop for image contrasting is using brightness differential which is a parameter to check the range of the brightness of input image. It is defined as the difference between the darkest and lightest pixel of image dividing the result by lightest pixel value and multiplying it with 100. According to survey, images with high value contrast are easy to see for low vision people [9]. Our intuition is to change pixels' lightness or value if the neighboring pixels' brightness differential is less. We are convoluting a n x n box with stride 2(n-1) so that each pixel is visited only once. We are dividing the pixels at indices in

image corresponding to box into two groups based on their value. Pixels which are closer to the minimum of the values are placed in one group and the others are placed in second group with maximum value. Next, we are decreasing the value of pixels in first group and increasing the value of pixels in second group by multiplying by a fixed number. We are performing this step if and only if the brightness differential of maximum and minimum value pixels of an iteration of box is less than some threshold. Finally, after contrasting the image we are removing the salt and pepper noise using median filter that may have been added during the process.

(a) Before Contrasting (b) After Contrasting

Figure 6: Novel Approach - Example 1



(a) Before Contrasting (b) After Contrasting

Figure 7: Novel Approach - Example 2

Here, as we can see in the result images the background of the text is changing according to the color of the text. The text written in white is making the background of the text darker so that it becomes easier to read. Similarly in the second image, the algorithm is drawing boundaries between colors so that it becomes easier for the people with low vision to perceive different colors.

Text-to-Speech:

The first step to get this done was to extract the text, if present, from images, and use them to convert it to audio.To do this, we researched about different ways, and finally decided to go ahead with Optical Character Recognition (OCR) technique. OCR is a machine learning framework, that helps detect characters. To get started with, we used the 'Tesseract' library in python that helps implement OCR.

The current problems we are facing with the above approach is working with the 'Tesseract' library. The library is really useful, however, there a few setup issues we are working on debugging.

The approach from here will be to work on simple images at first, and then work with images with 'noise', i.e., images where extracting text will be difficult. After getting the text, we will save it in a csv file, and then use that csv file for audio conversion.

References

[1] Vision Impairment and Blindness. 14 Oct. 2021, www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment.

[2] "Fast Facts of Common Eye Disorders." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 9 June 2020, https://www.cdc.gov/visionhealth/basics/ced/fastfacts.htm.

[3] Jones, Nabila, et al. "An Analysis of the Impact of Visual Impairment on Activities of Daily Living and Vision-Related Quality of Life in a Visually Impaired Adult Population." British Journal of Visual Impairment, vol. 37, no. 1, 2018, pp. 50–63., https://doi.org/10.1177/0264619618814071.

[4] Szpiro, Sarit, et al. "Finding a Store, Searching for a Product." Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2016, https://doi.org/10.1145/2971648.2971723.

[5] Zhao, Yuhang, et al. SEEINGVR: A Set of Tools to Make Virtual Reality More Accessible to ... https://www.microsoft.com/en-us/research/uploads/prod/2019/01/SeeingVRchi2019.pdf.

[6] Dingman, Hayden. "How Facebook Is Using AI to Improve Photo Descriptions for People Who Are Blind or Visually Impaired." Tech at Meta, 2 Nov. 2021, tech.fb.com/artificial-intelligence/2021/01/how-facebook-is-using-ai-to-improve-photo-descriptions-for-people-who-are-blind-or-visually-impaired.

[7] Lewis, Veronica. "Choosing High Contrast Color Schemes for Low Vision." Veroniiiica, 25 Aug. 2020, veroniiiica.com/2019/10/25/high-contrast-color-schemes-low-vision.

[8] Robert Fisher, Simon Perkins, Ashley Walker and Erik Wolfart, 2003, https://homepages.inf.ed.ac.uk/rbf/HIPR2/histeq.htm

[9] Fallatah, Samaher, et al. "An Investigation of the Appropriate Level(s) and Ratio of Value Contrast for Partially Sighted Individuals." SAGE Open, vol. 10, no. 2, 2020, p. 215824402092403., https://doi.org/10.1177/2158244020924031.