Color Pen Reader

By Team Tao

Oluwadarasimi Adeniyi (Dara)

Haoran Wu (Steven)

Michael Syrov (Michael Mordkovich)

Kaixian Zhu (Ellis)

Faculty Advisor: Dr. Anand Santhanakrishnan

Capstone project proposal submitted in partial fulfillment of

the requirement of the Degree of Bachelors of Science

in the Department of Electrical and Computer Engineering

in the School of Engineering and Computing Sciences

of New York Institute of Technology

Spring 2018

Table of Contents

Abstract

Dyslexia affects 5-10% of the general population, with some estimates going up to 17%. This has a profound effect on productivity of the labor force, as well as creating economic loss as tens of thousands of dollars are spent by families to accommodate their dyslexic children. However, even after their formative years, dyslexic people struggle to hold down a job due to their disability. Furthermore, about 3.4 million people in the US alone are blind.

One of the best investments any dyslexic person can make for their children is a purchase of the pen reader. It is a product for dyslexic, visually impaired, or otherwise handicapped people that reads written text aloud upon scanning text. It utilizes an Optical Character Recognition engine implemented in software to take raw input from the scanner at the front of the pen and cleans it up until it recognizes a character of a certain font. The pen reader then compiles words and sentences out of these characters and outputs and reads them aloud. It can also read aloud the meanings of certain words so that the user can understand unfamiliar words. It can also translate between languages. However, the pen reader's limitations include an inability to read colored text on colored backgrounds. This can get in the way as a lot of jobs depend on reading text whether or not it is of a different color, or that it is on a colored background.

Our main objective is to explore the use of different filters and image processing techniques to improve the pen reader's performance in reading colored text on colored backgrounds, or to design our own pen reader which would be able to read such text.

1. Executive Summary

The problem we would like to address is the inability of the pen reader to recognize colored text on a colored background. A pen reader is a device that scans lines of text and uses Optical Character Recognition (OCR) technology [3][8] to recognize words and read them out loud to the user. Currently, there are some pen readers with limited color reading ability, but they fail reliably on certain colors and certain backgrounds due to their limitations. We have explored the possibility of the creation of a modular attachment or a modification to the commercially available pen reader that can enable the pen to read text no matter what the color is, and found it was impossible. Therefore, we have designed our own pen reader that can take colored input. To do this, we designed a filter module in software using a Raspberry Pi computer to make colored text input readable to available open-source software like Tesseract and festival.

2. Opportunity

Our project will help handicapped people by letting them read more text they cannot otherwise read. Around 5-10%, with some estimates going as far as 17%, of the general population of the US is dyslexic[1]. Around 3.4 million people in the US are legally blind[2]. There aren't very many things that are written in black text on a white background except for new books and special labels. Many things, such as product packaging, old yellowed books, logos, and others have their text stylized in a special font or color or printed on a unique background. The pen reader on its own is designed to take different fonts into account, but the problem is that it cannot read color, and so these people have more trouble getting through their workday than the average worker does. By creating a new pen reader that can filter out color in text, we can lift that barrier to dyslexic, blind, or workers with other reading disabilities and allow them to work in jobs previously unavailable to them. Any job where you have to sort things by label or involves other form of text (package sorter, warehouse worker, etc) is filled with this kind of work. Dyslexic students and researchers would also benefit from this technology by making it easier to study from old, yellowing books. It will also help those with disabilities with navigation. There is also an opportunity to reduce the amount of money spent on special education for dyslexic people, as well as an opportunity to reduce the amount of stress these workers experience, improving their productivity.

3. Proposed Plan of Action

3.1. Project Components

We have designed a new open-source pen reader with an image processor that filters out the raw information from the pen reader's scanner and transforms it into a form that is acceptable for the pen reader to process. The image processing occurs on a Raspberry Pi while scanning takes place with both a Raspberry Pi and a Trinket M0 microprocessor working in tandem. We have attempted to create a modular attachment but found it was impossible after doing more research and contacting the company that manufactures the pen we have studied. The robotic arm portion of the project has been cancelled due to these changes in the project requirements.

We achieve this by first scanning the text with a Raspberry Pi camera and combining the set of input images. We then apply a Gaussian filter to the input first to mitigate noise. After the Gaussian filter stage, we use a differential filter (edge detector) in software in order to sense sharp changes in color and intensity between pixels and create a black and white image based on that. This will create an outline of the text scanned. We have further added a post-processor to the system that detects letters and detects the kind of letter it is (with no holes like an L, with one hole like an O, or with two holes like a B) and fills those edges in with black color so the pen reader is able to process it. Processing of the filtered text is performed by Tesseract, an open-source OCR engine and the output speech is created by festival, an open-source text-to-speech converter.

We started with a high level design of the hardware and software and determined the inputs, outputs, and the requirements of each part in the project. After this, we continued to refine each high-level part into a series of low level requirements and designs in somewhat of a waterfall approach. One of the main components in the open-source pen reader's design is the OCR engine and the text to speech software. Open-source software exists for both of these – we have chosen Tesseract for the OCR engine and festival for the text to speech conversion because Tesseract is the best open-source OCR engine available currently and festival does its job well. Because these components have well-defined outputs and inputs, and since the output of festival is the only required output of our pen reader, we have chosen to pursue the refinement of the design backwards; that is, we have decided to design and implement the pen reader from the output to the input.

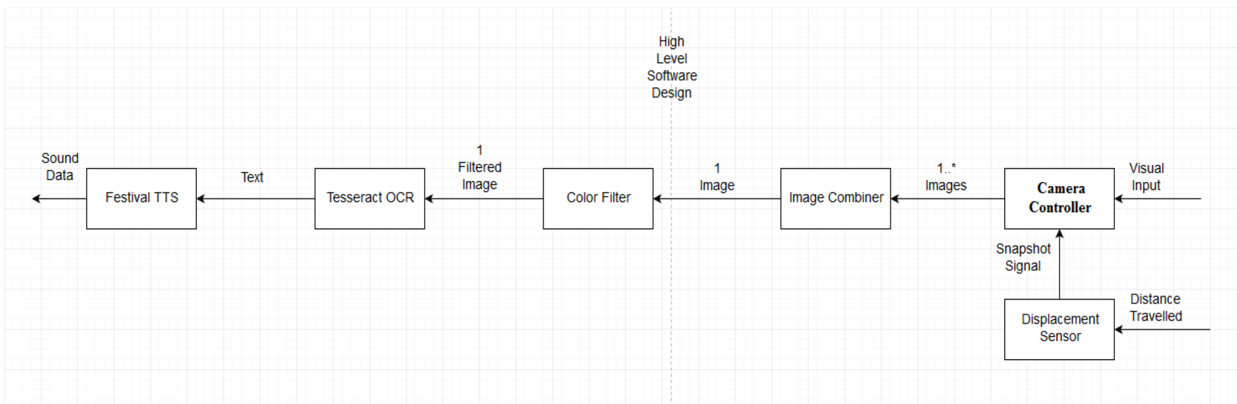

Figure 1: High-level Software Design

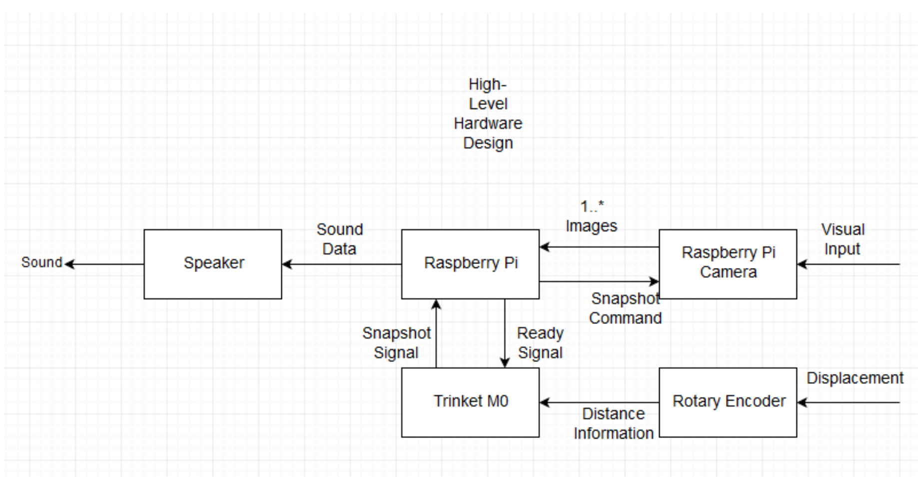

Figure 2: High-level Hardware Design

The camera is controlled by Raspberry Pi, which listens to two signals isScanning and CaptureInterrupt sent by microprocessor Trinket M0. When isScanning signal is low, the camera module is off-work and when isScanning signal is high, the camera will take one photo every time the captureInterrupt signal's rising edge comes. The camera will go back to off-work mode when isScanning signal goes back to low from high and the next stage will execute.

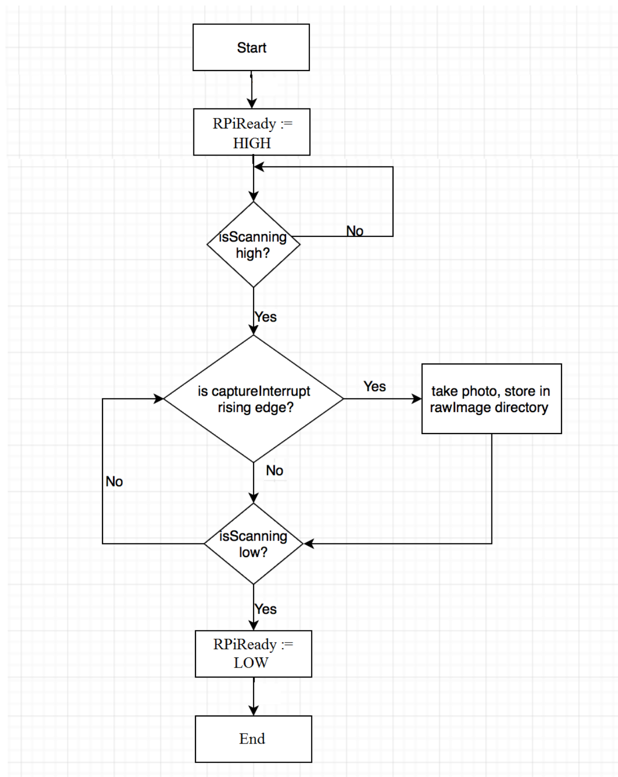

Figure 3: Camera Controller Design

The camera controller depends on feedback from our closed-loop feedback component of the project. A Trinket M0 in combination with a rotary encoder measures the distance travelled by the pen reader while scanning text and outputs a captureInterrupt pulse to the Raspberry Pi. This tells the Raspberry Pi to take a snapshot of what the camera is currently looking at. This is to keep images evenly spaced so future concatenation will result in the entire text being without gaps or repetition (from intersecting images). Since it depends on distance travelled, it also lets the system account for variations in the user's scanning speed. There is a limitation, of course. Too fast of a scanning speed with result in image blurring and a detrimental effect on the output.

The closed-loop feedback waits until the camera controller module is running so it does not waste power while the pen reader is processing text. Then, when the Raspberry Pi is ready for input, it waits until the user holds a button which tells the Raspberry Pi that the user is currently scanning. It then reads the encoder for changes in position and outputs a captureInterrupt pulse for the Raspberry Pi to take a picture. When the button is released, it tells the Raspberry Pi that scanning is finished and the pen reader moves on to concatenate the images and process the input.

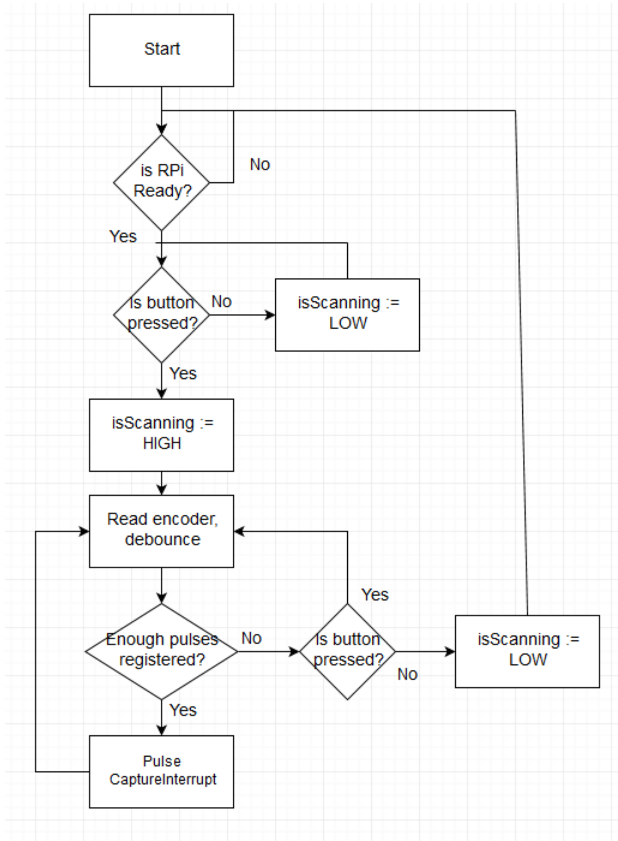

Figure 4: Closed-loop Feedback Design

The image combiner module loads all the images that have been taken by the camera from one scan and combines them into one single image by concatenating them all.

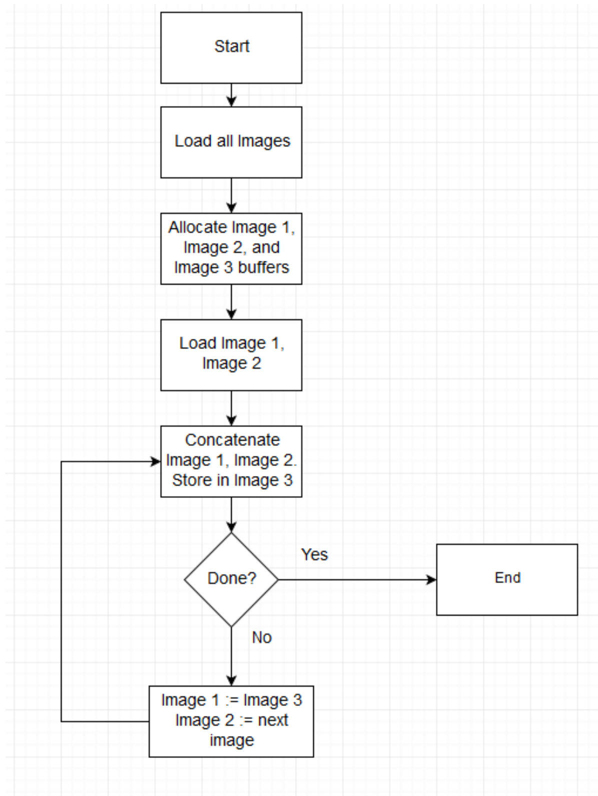

Figure 5: Image Combiner Design

The color filter component was designed starting from edge detection. Usual methods to detect edges include Canny edge detection [4], image derivative (Sobel Filter) [5], and phase stretch transform [6]. The first two are available in software libraries like OpenCV which we also plan to use. The last is not; however, that capability has been open-sourced. We have also explored the possibility of using the cosine similarity method [7] to produce edge detection by putting the color and lightness data of a pixel into a vector and comparing two adjacent pixels. A value close to 1 indicates similarity, and thus no change in color, while a value close to 0 indicates difference, and thus a change in color. This method is not available to implement in OpenCV by default, but is fast and easy to implement. This method has also been used before, albeit not as commonly as the others mentioned, but it was stumbled upon independently by the group and has potential. After experimentation with MATLAB simulations we have decided that Canny Edge Detection works best. Our experimentation with Tesseract showed that edge detection is not enough to reliably read text from images, and that post-processing would need to be performed. This was included in the color filter module. After the differential filtering is performed, the color filter binarizes the image and detects separate letters and their type (specifically, whether the letter has no holes like an L, one hole like an O, or two holes like a B). After determining the type of letter, it fills in the text accordingly.

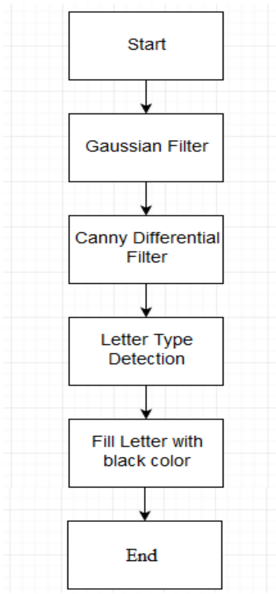

Figure 6: Color Filter Design

After processing the input image, it is fed into the Tesseract OCR engine, which converts the image into text. The output text of Tesseract is then piped into festival which converts it into sound data. This is sent to the speaker for reading aloud.

3.2. Hardware/software Requirements

1. Raspberry Pi 3

Figure 7: Raspberry Pi 3

The Raspberry Pi 3 will serve as the main processing module of our project. It will receive the feedback from microprocessor Trinket M0 and then control the camera module to take photos. The Raspberry Pi 3 will also process the photos taken by camera, filter out the color and send black-and-white image to Tesseract OCR engine and festival text-to-speech software.

2. Microprocessor Trinket M0

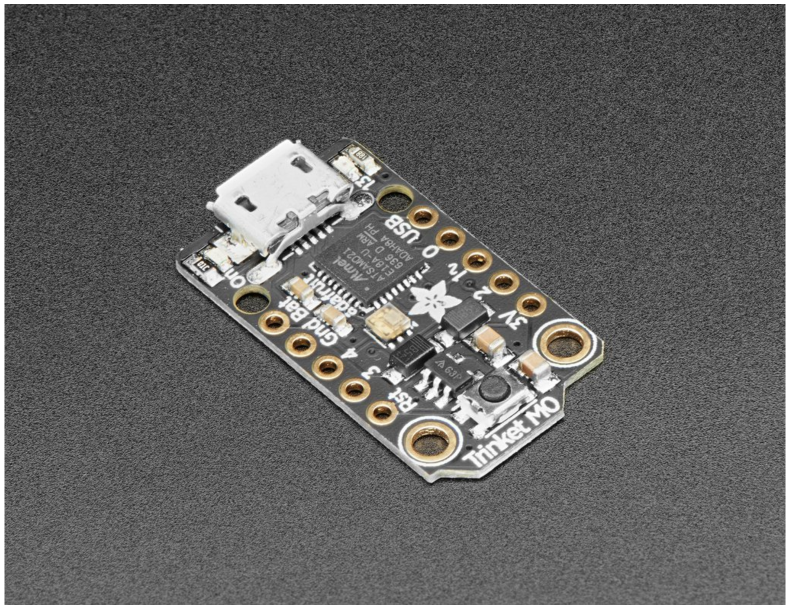

Figure 8: Adafruit Trinket M0

This microprocessor Trinket M0 sends two signals **isScanning** and **captureInterrupt** to the Raspberry Pi and it receives the pulse sent by rotary encoder and signal from a button. If the button is pressed then the Trinket M0 will set the signal isScanning to high and calculate the distance between position of camera module and the next position of camera module based on the length of the received pulse's high level. If the distance reaches a threshold value set in the code, then the Trinket M0 will set the signal captureInterrupt to high.

3. Raspberry Pi Camera Module

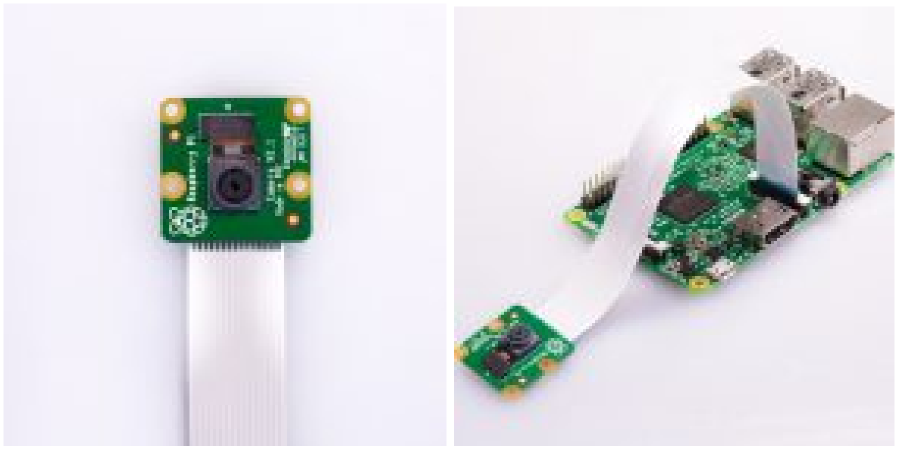

Figure 9: Raspberry Pi Camera Module V2

This camera module V2 uses Sony IMX219 sensor and it’s resolution can reach up to 3280 x 2464 pixels. It will take the pictures of the text to be processed. If processing is slow, the resolution can be reduced.

4. Rotary Encoder

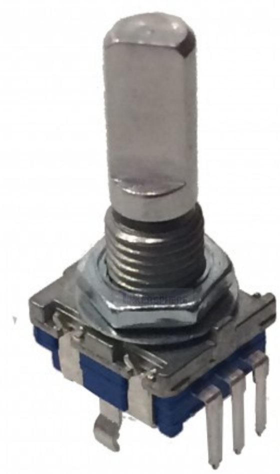

Figure 10: 20p/r Incremental Rotary Encoder

We adhere a wheel to its shaft and use it to measure the distance the pen reader has traveled. When the wheel rolls on the surface of reading materials, the encoder will send out pulse to microprocessor Trinket M0, then the distance will be calculated and sent to the Raspberry Pi.

5. Audio Amplifier

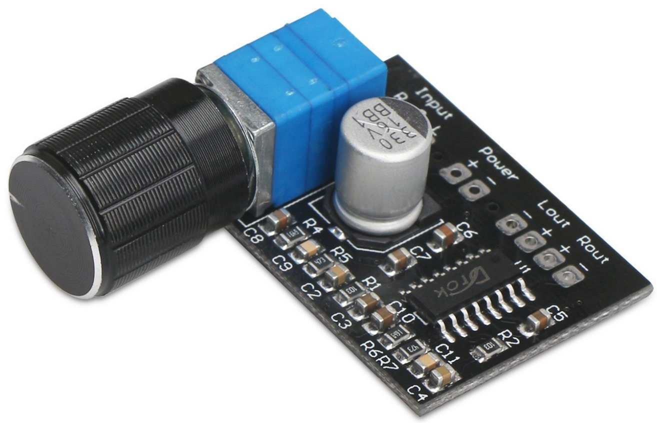

Figure 11: Drok 3 Watt Sound Amplifier

This is connected to the Raspberry Pi's 3.5 mm jack in order to amplify the output of the color pen reader to audible levels.

6. Speaker

Figure 12: Cylewet 4 Ohm 3 Watt Speaker

This is a speaker that outputs the final output of our color pen reader.

7. Battery

Figure 13: Anker Powercore 10000 mAh Li-Ion Battery

This is the battery used to power the color pen reader. With 10 Ah of capacity, it allows the color pen reader to continuously and actively operate for up to 5 hours, with a more realistic estimate being around 8 hours.

4. Commercialization

We have to introduce the rational benefit on our improved reading pen. Our tech makes the reading pen different. When you scan light color characters with darker background, ordinary reading pens can’t recognize any words at all, which means that handicapped people will have a harder time reading, and you still have to struggle on looking up words one by one via cell phone or dictionary. Our project removes the pen reader's color limitation, so you can scan any color of words conveniently as long as there’s a color difference between characters and the background.

Our project will target people with reading difficulties such as dyslexia, which is a real time saver to them. Also, those who suffer from the limitation of original reading pens are willing to see our breakthrough that we remove the color limitation.

We should let customers unconsciously accept the concept of improved reading pen. We will plan and implement BTL marketing activities which will include free improved pen experiencing, comparison between the old one and the improved one, distribution of brochures or banners to introduce feature of our product, scanning QR code to let them feel free to follow our social media account which helps us to push further notifications. Once their feedbacks are received through these activities, it will be more efficient to satisfy their needs and leave a good impression.

We’ll meet a lot of new problems in commercialization. As we all know, there are too many interconnections between the commercialization and the business operations. However, it seems that we are neither experienced enough in business operations nor do we have enough money to commercialize alone. What’s more, our improvement is an integral part of the reading pen. It would be wise if we can cooperate with a reading pen company as we share mutual interest. They own the marketing and have more experience in business operations. Also, we can provide more convenient tech to improve their products, making the cooperation viable. How to maximize the market acceptance is another thing need to take into consideration. We have to admit that it's not an indispensable tool for everyone. It is a specialized product designed for specific people in mind. It cannot be immediately accessed as a platform like a phone app, which means that improved reading pen cannot be spread as easily. In order to speed up the formation of this marketing, we should not only focus on people who have issues on reading but try to create an ambience to people whose English is poor and don't want to look up words one by one on cell phones and dictionaries. We must convince them that it would be a better choice to use our improved reading pen to fix this issue. Through this guidance, they will remember our product and realize how convenient it is, then gradually accept it.

Commercialization of the reading pen does not keep it from being open-sourced; instead, commercialization will help the reading pen become known and will attract more developers to improve it in the future. We have therefore open-sourced the design and the repository containing the software, the proposal, and the hardware connections are available on Github at Comrade96/color-pen-reader.

5. Limitations and further Development:

1. OCR technology’s limitation

Optical character recognition technology which is used in our pen has its limitation in recognizing several parallel character lines. For instance, if the scanner of the pen includes two lines or one line of characters plus portion of other lines in one scanning, then the pen cannot recognize two lines or one specific line. We need to improve this either by restricting the scanning area from outside or improving the algorithm inside the pen. Not only that, but even modern reading pens' success rates range from 81% - 99% [9].

2. Pen’s limitation

What hardware design the existing pen use in scanning port is same as barcode reader. The pen emits light to a surface which is scanned by the user, then the pen receives the reflection of light and process this reflection. We must use a pen that emits only white light to illuminate the text, otherwise the pen cannot scan colorful materials because reflection of colored light on same-colored text would be the same color, however if the background was white, the same color would be reflected too, giving no contrast between text and background. This means if we want to scan colorful materials, the pen needs to emit white color which can reflect any color.

3. Recognition algorithm’s limitation

The precision of character recognition algorithm in existing pen is not very high. When we use the pen in scanning some hand-written characters or noisy text, it cannot recognize them precisely, if at all.

4. Our scanning module’s limitation

If the contrast between text and background is small, the edge detector output will be faint. This will make it hard to fill in the edges by the post-processor in the next stage. We may combine different methods for edge detection and use the output of that, but that may slow the process down. We must find a good balance between filter quality and speed.

5.1. Impact

1. Our reading pen can help people with dyslexics.

It is estimated that 1 in 10 people have dyslexia [1]. Over 40 million American Adults are dyslexic - and only 2 million know it. These people cannot read materials as efficient as people who do not have dyslexia. Our reading pen can help these people solve this problem because our pen can display, can read in high speed and in high accuracy.

2. Our reading pen can help people who are visually impaired.

Our reading pen can read what it scans so that it can help people who are visually impaired to understand reading materials. By listening to reading pen, people who have disability in eyesight can easily understand materials in a normal speed.

3. People who are not familiar with English can also use this pen to help them work.

With our modified algorithm, the reading pen can scan and translate reading materials in a high speed and high accuracy. Then, people who do not understand English can understand English as fast as people who are familiar with English.

When our project is successfully finished, it would make life easier for disabled people, it would improve reading experience and it would make getting jobs easier for disabled people.

5.2. Prior Art

There are many competing technologies like the C-PEN Reader, ScanMarker and Power Pen but they don’t read colored text on colored backgrounds well. We have resolved this limitation.

Open-source software for OCR such as Tesseract (and others) exist and are freely available. Of open-source technologies, Tesseract is the most accurate and will be used for our purposes. Proprietary OCR engines offer higher accuracy in reading text from images, but are not available to us due to our goal to make the pen reader open-source.

Festival is an open-source software tool for converting text to sound data. It is one of the best available in terms of voice clarity. We have also used this to develop our project.

6. Project Management Plan

6.1. Potential Future Improvements

The color pen reader in its current form is not without its limitations. There are notable reliability limitations owing to the color filter's parameters not being totally optimized. As such, it sometimes gives wrong results. There is also the limitation stemming from the fact that it is of a large size that is though handheld and portable, still is quite large and consists of two separate components. Other limitations exist such as the OCR engine not being able to accurately read the text from the image from time to time. There is also the fact that the output is a little quiet due to the speakers not being loud enough, or the amplifier not being as powerful as advertised.

Some of these limitations can be rectified, however. In the future, by finding better parameters for the color filter, the accuracy and reliability of the color pen reader can be improved. The color pen reader can be made smaller by using a Raspberry Pi Zero or a smaller computer to implement the software, as well as use a smaller speaker, a smaller battery, and a more powerful amplifier to facilitate its function. This can potentially reduce the color pen reader's size to a single component, and possibly even pocket-size it, which would be ideal.

6.2. Team and Collaboration

Michael Syrov (Project Manager/Hardware/Software Engineer)- A very experienced engineer who has managed several projects, he is well spoken and very organized. He has also worked with OCR technology and has a clear idea of the direction in which the project should go. He will be responsible for making sure the project is delivered on time, within scope and within budget.

Haoran Wu (Hardware / Software Engineer)- Haoran has worked on different projects which required use of Field – Programmable Gate Arrays (FPGA) and microprocessors. This experience will be useful in our project.

Ellis Zhu (Software engineer)- He has had a lot of internship experience which has prepared him for this project. He is a great programmer, very good with Python and excellent at programming with a Raspberry Pi. He has prior knowledge in modifying existing codes from OpenCV and writing a code from scratch and is willing to do more research.

Oluwadarasimi Adeniyi (Project Manager/ Researcher)- Has previous management experience, led a 5 person team in previous school projects, worked as a general manager intern, and is also a good researcher and finds useful technical detailed information about the project. He will be responsible for gathering useful documents, finding reliable sources and making sure the project is delivered on time.

6.3. Milestones and Timeline

Task	Date
Initial Research	10/07/17 – 12/03/17
AHRC Meeting	10/24/17 – 10/25/17
Project Proposal Presentation	11/14/17 – 11/15/17
Budget Acquirement	11/21/17 – 11/21/17
Pen Reader Acquirement	11/22/17 – 11/25/17
Decision on Main Approach to Solution	12/03/17 – 12/25/17
Materials Acquisition	12/20/17 – 1/17/18
Implement Color Filter	1/17/18 – 3/1/18
Implement Image Combiner	1/25/18 – 2/3/18
Prototype 1	4/3/18
Prototype Tweaking/Prototype 2	4/3/18 – 5/10/18
Final Report	5/9/18

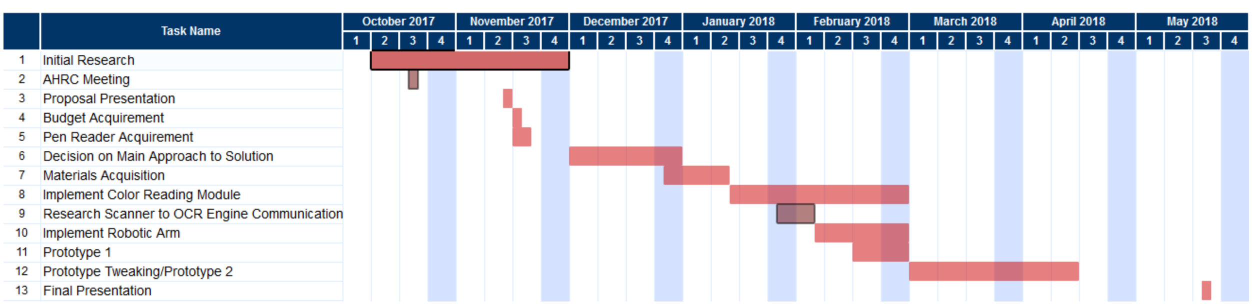

Figure 11: Development Schedule

7. Use of Standards

We have not found any use for standards in this project. The Bluetooth standard would have been used, but the robotic arm part of the project had been cancelled.

8. Resources and Budget

Item	Unit Price	Units	Purpose	Link
C-Pen Reader	$250	1	This is the Pen Reader we plan on expanding.	https://www.amazon.com/C-pen-C-Pen-Reader/dp/B015OL7VNS
Raspberry Pi 3 Kit for develop-ment	$69.99	1	It would greatly expediate development if we have a Raspberry Pi dedicated solely as a development platform so that we wouldn't have to swap Pis in and out of our project to try some new code.	https://www.amazon.com/CanaKit-Raspberry-Complete-Starter-Kit/dp/B01C6Q2GSY
Raspberry Pi 3 for prototype	$35.05	2	One for the Color Reading Module and one for the Robot Arm	https://www.amazon.com/Raspberry-Pi-RASPBERRYPI3-MODB-1GB-Model-Motherboard/dp/B01CD5VC92
Raspberry Pi Camera Module	$28.31	1	This will be used to replace the original scanner sensor of the pen.	https://www.amazon.com/Raspberry-Pi-Camera-Module-Megapixel/dp/B01ER2SKFS/

Lynxmotion AL5D 4DOF Robotic Arm SSC-32U Combo Kit	$350.80	1	This is the Robot Arm hardware to be controlled by one of our Raspberry Pis.	https://www.robotshop.com/en/lynxmotion-al5df-kt-robotic-arm-flowbotics-studio.html
IC3D PLA 3D Printer Filament	$29.99	1	We would like to print an enclosure for the Color Reading Module to defend against accidental dropping and other damage.	https://www.amazon.com/IC3D-White-1-75mm-Printer-Filament/dp/B011RXNIXO
Turnigy 7.4 V 1000 mAH LiPo Battery	$14.99	2	This will power the Raspberry Pi on the Color Reading Module as well as the one on the Robot Arm	https://www.amazon.com/Turnigy-1000mAh-Lipo-HobbyKing-Battery/dp/B0072AEHIC
Iusun Mini-360 DC-DC Buck Converter Step Down Module	$0.69	10	This is used to step down the voltage of the LiPo battery to a safe voltage that the Raspberry Pis can handle.	https://www.amazon.com/Iusun-Mini-360-Converter-Mini360-4-75V-23V/dp/B074SK9SSL

Elegoo HC-SR04 Ultrasonic Module Distance Sensor	$11.49	1	This will provide negative feedback to the Robot Arm while it is positioning itself to pick up an object.	https://www.amazon.com/Elegoo-HC-SR04-Ultrasonic-Distance-MEGA2560/dp/B01COSN7O6
Icstation RFP602 500g to 10kg Resistive Thin Film Pressure Sensor Force Sensing Resistor	$10.48	2	This will provide negative feedback to the Robot Arm while it is closing its gripping arm in order keep the arm from crushing the object. It will also sense when the object encounters resistance to movement, signifying that the place the Robot Arm is trying to place the object in is occupied.	https://www.amazon.com/Icstation-RFP602-Resistive-Film-Sensor/dp/B06ZYJRN8M
Bluetooth 4.0 USB Module	$11.95	2	This will be used so the Color Module and the Robot Arm Control Software can communicate with each other.	https://www.amazon.com/Bluetooth-802-15-1-Modules-Dongle-single/dp/B00HKILG1W
Fluke 113 True-RMS Utility Multimeter	$97.49	1	Multimeter for testing various parts and ensuring the buck converters are set up correctly.	https://www.amazon.com/Fluke-113-True-RMS-Utility-Multimeter/dp/B01EWFE6XK

Total: $989.91

Note: We have $1000 of funding from AHRC already approved.

References

[1] Siegel, Linda S. “Perspectives on Dyslexia.”, Paediatrics & Child Health 11.9 (2006): 581–587. Print.

[2] Bailey R. N., Indian R. W., Zhang X., Geiss L. S., Duenas M.R., Saaddine J. B. “Visual impairment and eye care among older adults—five states,” MMWR 2006; 55(49):1321–1325.

[3] Eikvil, Line. "OCR-Optical Character Recognition.", 1993.

[4] Canny, John. “A Computational Approach to Edge Detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.PAMI-8, No.6, November 1986.

[5] Rebecca Vincent, Folorunso O. “A Descriptive Algorithm for Sobel Edge Detection.” Proceedings of Informing Science & IT Education Conference (InSITE). 2009.

[6] Mohammad H. Asghari, Bahram Jalali. “Edge Detection in Digital Images Using Dispersive Phase Stretch Transform.” International Journal of Biomedical Imaging Volume 2015. 2015.

[7] Shweta Atul Page, Kavita J. Mahajan. “Edge Detection Based Classification of Leukocytes in Blood.” International Journal of Computer Science and Mobile Computing , Vol. 5, Issue. 7,July 2016, pg. 353 – 358. July 2016.

[8] Chentung R. Jih. “Optical Character Recognition Using Baseline Information.”, US4251799 A, Feb 17, 1981.

[9] Holley, Rose. "How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs", *D-Lib Magazine,* April 2009.

Appendix

Given function $f(x, y)$

Gradient vector $\nabla f(x, y) = \begin{bmatrix} \dfrac{\partial f(x, y)}{\partial x} \\ \dfrac{\partial f(x, y)}{\partial y} \end{bmatrix} = \begin{bmatrix} f_x \\ f_y \end{bmatrix}$

Gradient magnitude $|\nabla f(x, y)| = \sqrt{f_x^2 + f_y^2}$

Gradient direction $\theta = \tan^{-1} \dfrac{f_x}{f_y}$

$$\frac{\partial f}{\partial x}[x, y] \approx F[x+1, y] - F[x, y]$$

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|_2 \|\mathbf{B}\|_2} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$