CONESTOGA
Connect Life and Learning

Formal Project Report

Group: 21 | Vision: A Virtual Assistant For The Visually Impaired

Deep Shah (8750086) | Jeevan Dsouza (8716171)

Engineering Capstone Project

Table of Contents

Abstract:

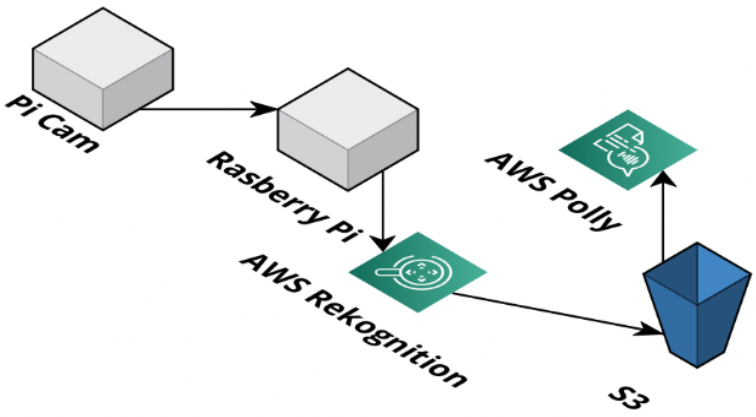
Today, there are nearly 285 million people in the world that are visually impaired. Although technology has grown leaps and bounds, the accessibility, especially that of the internet for differently-abled people, is still far-fetched. The major challenge in developing stable software is to include as few keystrokes as possible and to provide an end-to-end experience with the help of voice alone. The inclusion of multiple languages and setting the right pace of the speech when played back to the user are important factors to consider. We would like to incorporate our skills in AWS and Data Structures by working on a virtual assistant for visually impaired individuals. In our first phase of the prototype, we would like to design the architecture with the help of many services offered by AWS and would like to first work on text detection from the live video feed of Rasberry Pi Cam. Then, we can give live feedback on the text by converting it to a speech by Amazon Polly. After this, we can also integrate AWS Alexa on the Pi and write custom Alexa skills. We can store all the logs in Amazon S3 buckets and do a log analysis with the help of Athena. Once the initial architecture is implemented, the opportunities on this project are endless as we can leverage an ample amount of services on AWS and the computing power of Rasberry Pi.

Problem Statement:

The driving force behind this project is to work on a Virtual Assistant by leveraging the power of cloud computing. The first and primary objective is to start with text detection (It can be a billboard, a book's cover letter, etc.). After successful detection, the feedback on the detected text can be given by voice commands. The second phase will make this assistant more advanced by installing Alexa on the pi. We want to create custom Alexa skills to make our assistant more interactive. We could ask Alexa what the user saw at a particular time and date... This can be achieved using S3 and Athena. As all the data will be stored in S3 buckets, we could do a log analysis using simple SQL queries with the help of AWS Athena.

Proposed Solution:

Our assistant will take images through a cam and convert them into text files.
This file will then be converted to an audio file. This will be feedback for the visually impaired individuals.



Here we are leveraging AWS Rekognition, which is trained with the vast amount of data on the internet. To perform and interconnect all the services, we will be using Lambda for serverless computing.

Technology/Key Components Used:

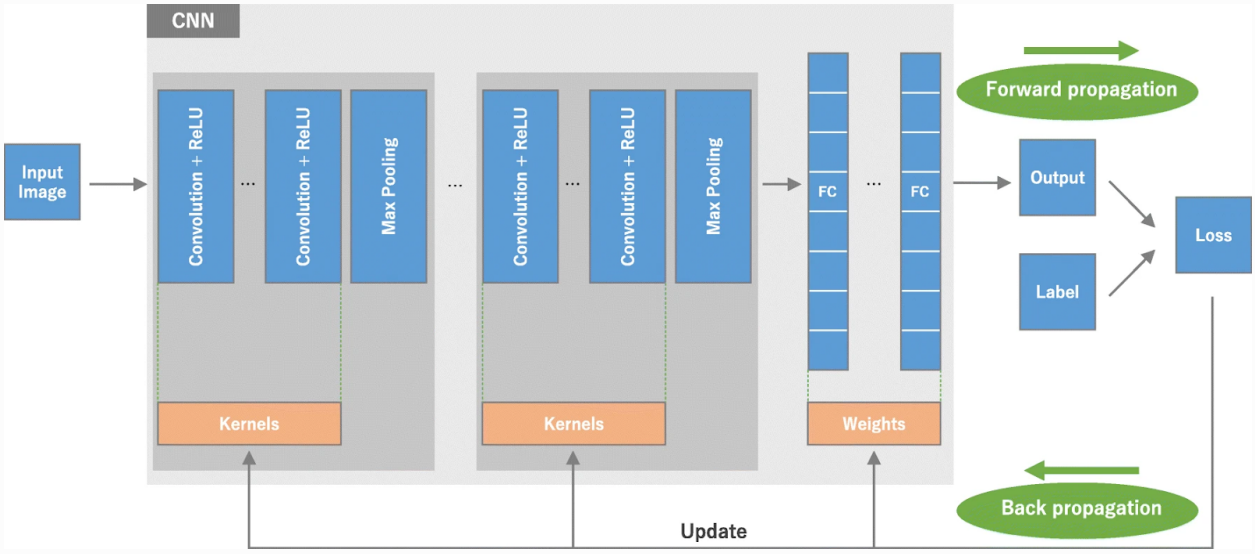
We are building our solution on the AWS ecosystem. We have used multiple AWS services, and the hardware requirement of this project is Rasberry Pi and a camera. This assistant will heavily leverage cloud computing, and we will be using AWS for the hosting.

The AWS services we would like to leverage but are not limited to

- AWS Rekognition: To perform a visual analysis of the live video feed
- AWS Kinesis Video Streams: To transfer live video feed to AWS via Rasberry Pi
- AWS Lambda: To do serverless computing and interconnect all the services
- AWS S3: To store all the data
- AWS Alexa: To make the assistant more interactive
- AWS Polly: To perform the text-to-audio conversion

Background Theory needed to understand the report:

AWS Rekognition: The addition of image and video analysis to your applications is simple with Amazon Rekognition. The Amazon Rekognition API only needs a picture or video to recognize objects, people, text, scenes, and actions. Any inappropriate content can also be found using it. Additionally, very accurate face comparison, face search, and facial analysis capabilities are offered by Amazon Rekognition. For a wide range of use cases, such as user authentication, cataloging, people counting, and public safety, you may identify, analyze, and compare faces.



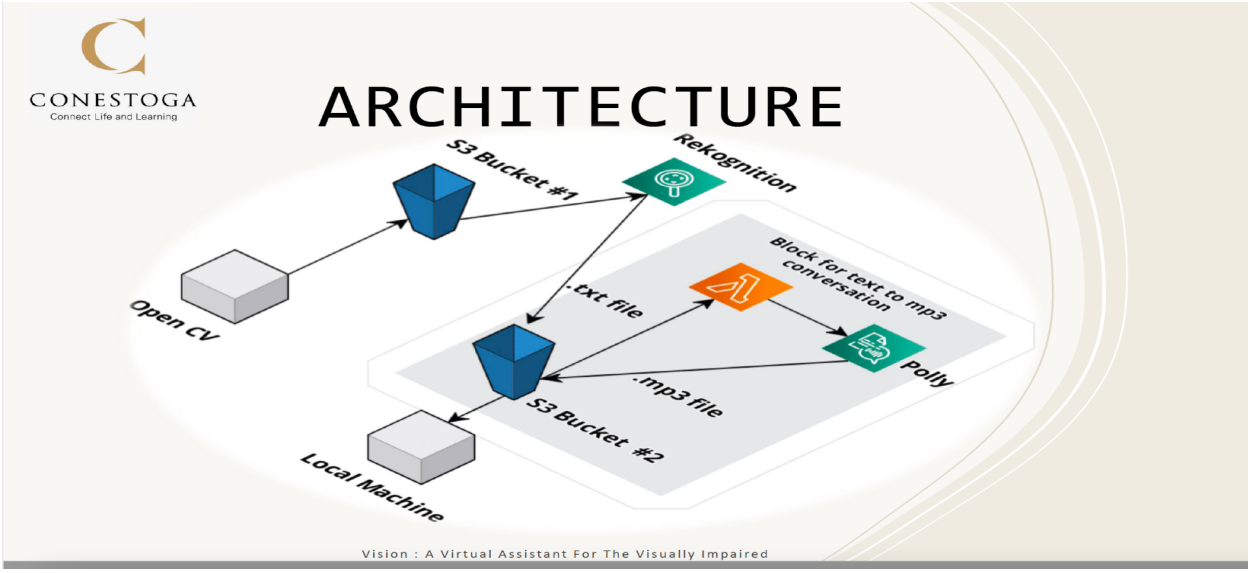
AWS Lambda: With the compute service Lambda, you may run code without setting up or maintaining servers. Lambda executes your code on a high-availability compute infrastructure while handling all compute resource management tasks, such as server and operating system upkeep, capacity provisioning and automatic scaling, and logging. You may execute code for practically any kind of application or backend service with Lambda. Simply provide your code in one of the languages supported by Lambda.

AWS S3: It provides performance, security, and scalability that are unmatched in the market. Any quantity of data can be stored and protected by customers of all sizes and sectors for practically any use case, including data lakes, cloud-native applications, and mobile apps. You may reduce expenses, organize data, and set up precise access controls to satisfy unique business, organizational, and compliance requirements using cost-effective storage classes and simple administration tools.

AWS Polly: A service called Polly converts text into realistic speech, enabling you to design entirely new categories of speech-enabled products and talkative applications. With the use of cutting-edge deep learning technology, Polly's Text-to-Speech (TTS) service creates human speech that sounds natural. You may create speech-enabled applications that function in numerous nations by using dozens of authentic voices in a wide range of languages.

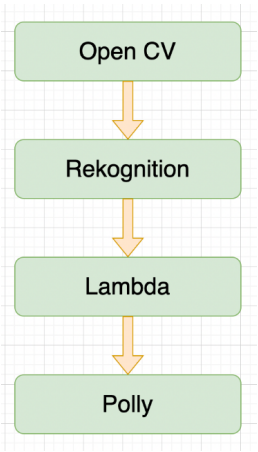
AWS Boto3: Boto is an SDK (software development kit) created to enhance the use of Python on Amazon Web Services. The Boto project began as a customer-contributed library that translated application programming interface (API) replies from AWS into Python classes to assist developers in creating Python-based cloud applications.

Methodology/Steps taken in the development process:



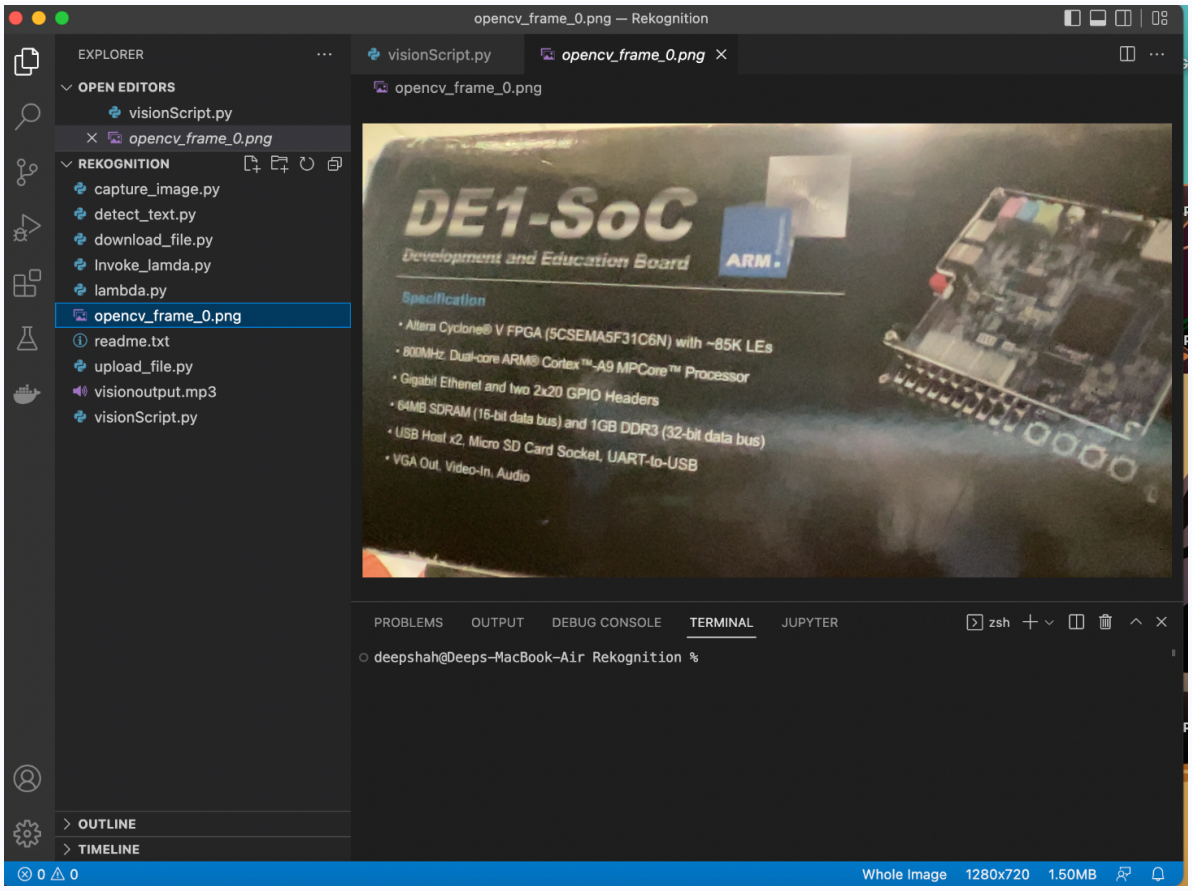
So first, we are capturing a live feed using the Open CV to S3 bucket One. Rekognition will fetch this image and perform text detection on the image. Then we will leverage lambda to perform serverless computing on the feed, and it will feed to AWS Polly. Now AWS Polly will convert the .txt into an mp3 file and store it into S3 bucket two. Finally, the mp3 file will be downloaded on the local machine.

We are performing each and every operation using AWS boto3, which is the python library to deploy AWS resources. Initially, when we started the project, we started to do each and every task like uploading to S3, Rekognition Script and downloading to local machine script individually. Then we created a script where everything is happening in a logical order and in only one script.

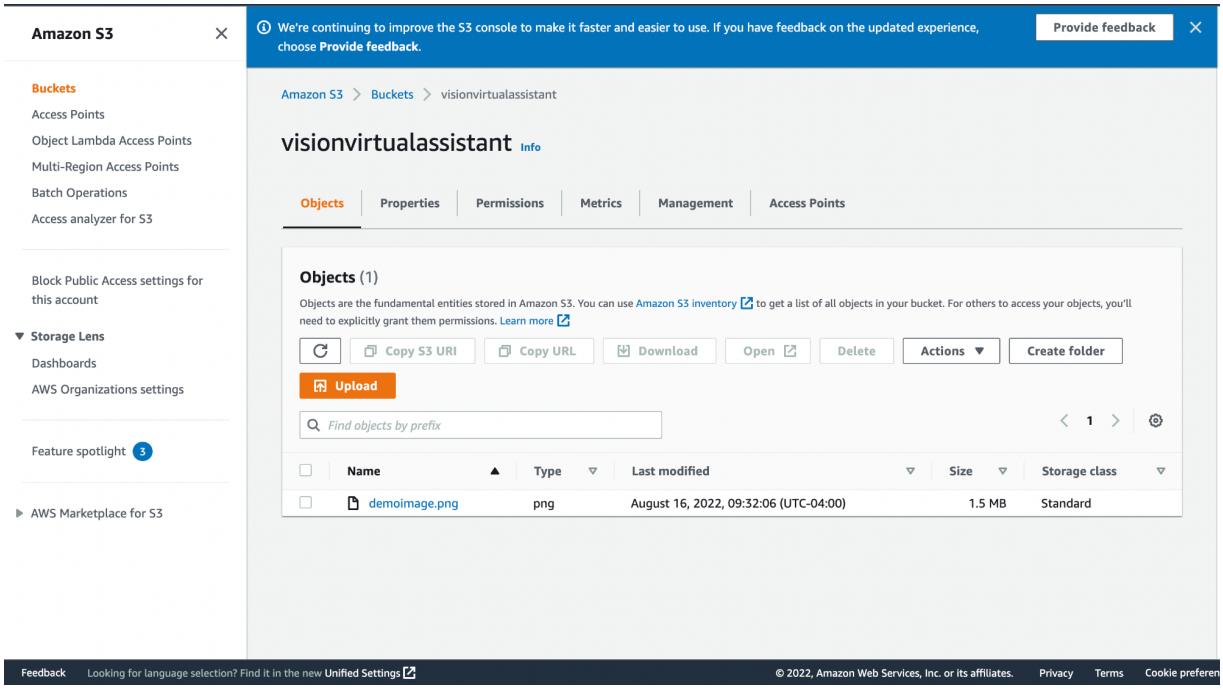


Results/Analysis:

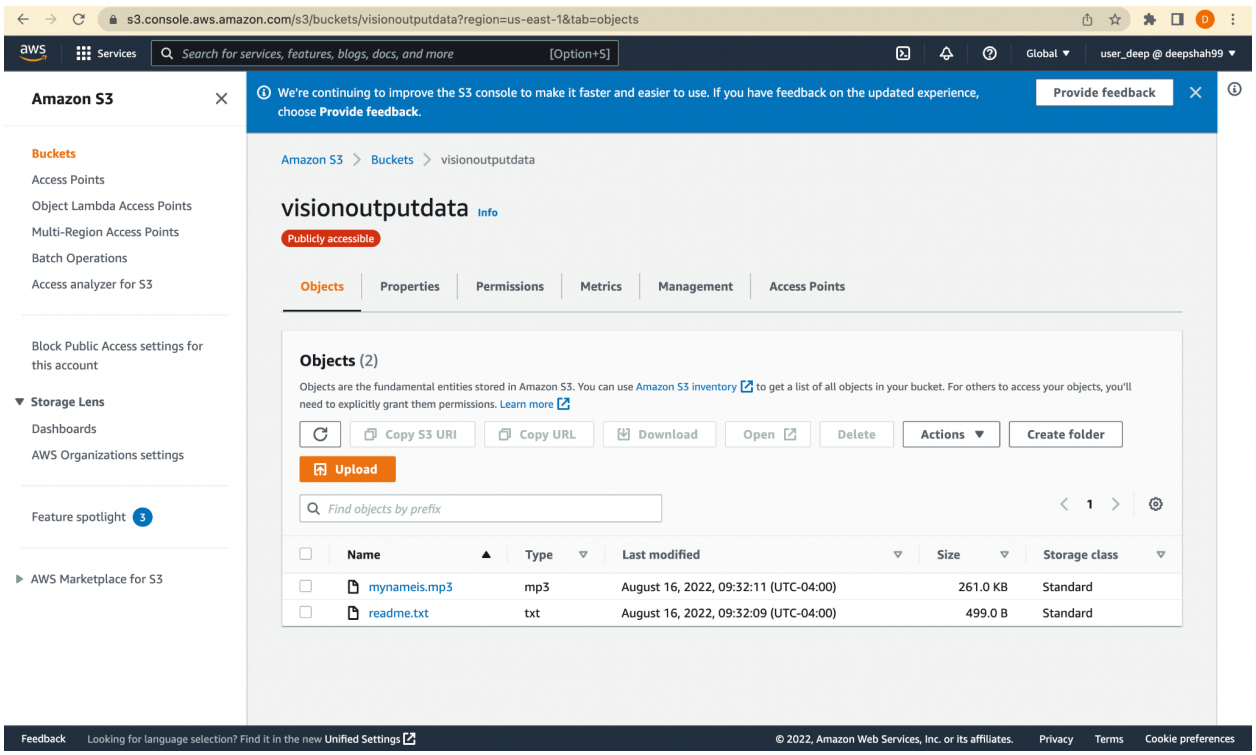
For this project, we are heavily leveraging boto3 library of python offered by AWS. It makes automation of AWS resources much easier and quicker.
Here, we are capturing the image via Webcam and Open CV from our local machine.



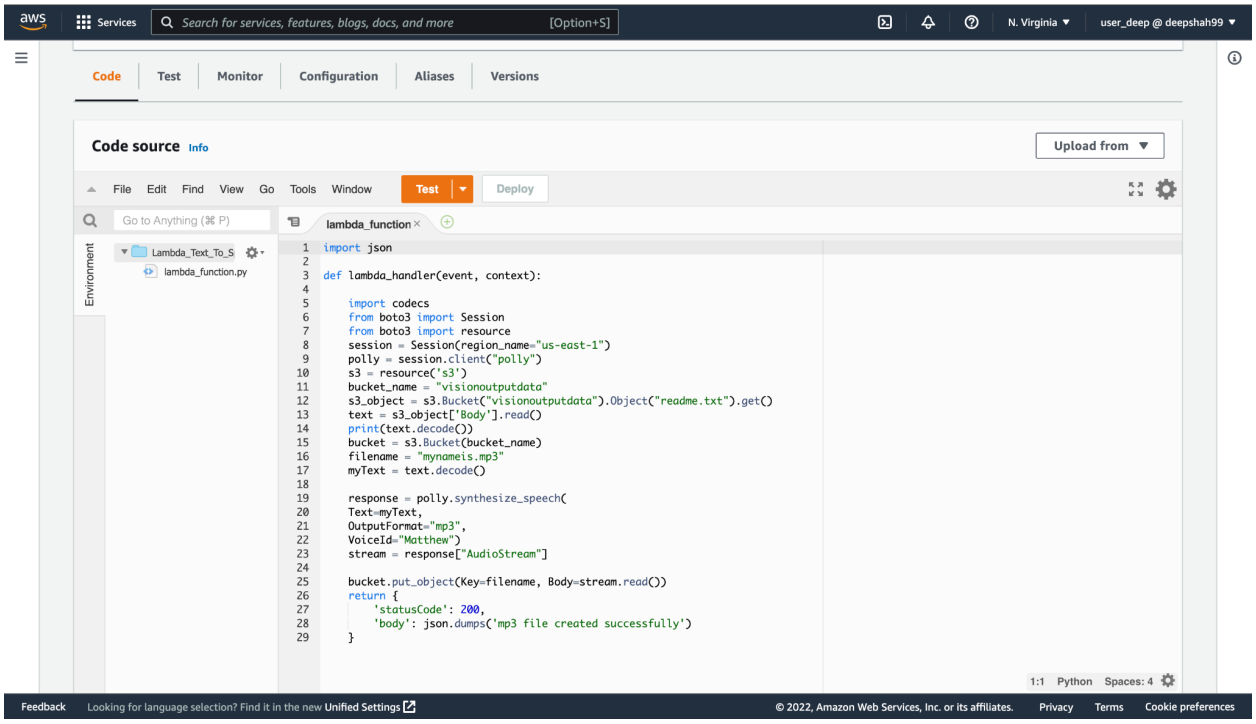
Now we are sending that png file to the S3 bucket.



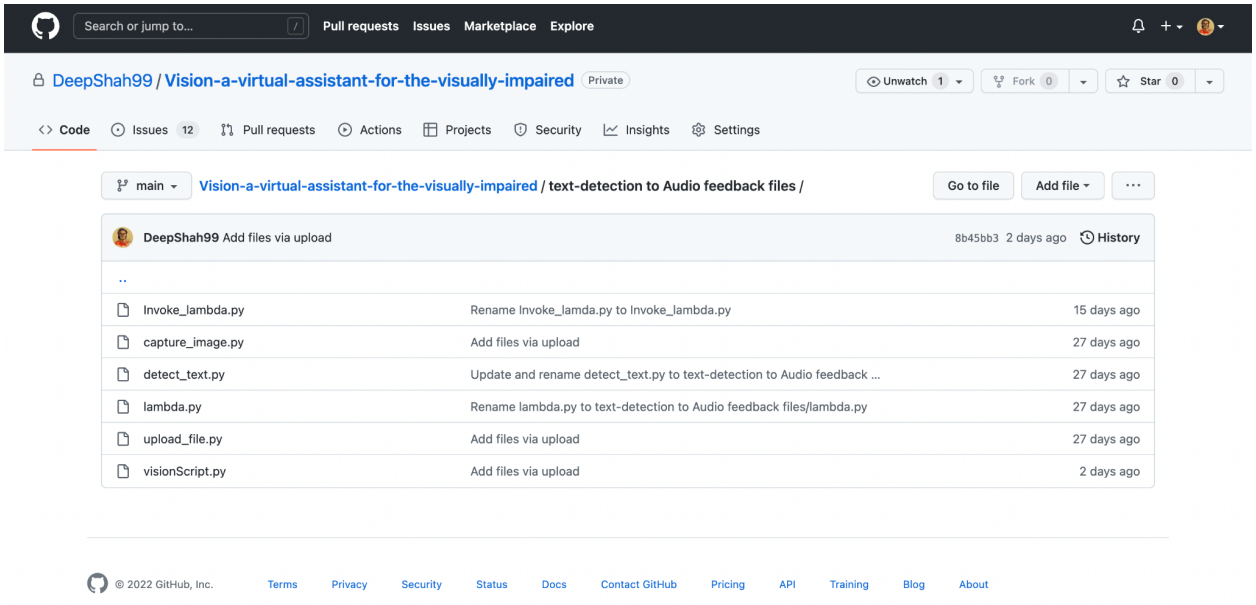
After getting this file, we are converting the image into a txt file using AWS Rekognition. Now AWS lambda will pull this txt file and convert it into an mp3 file.



Instead of invoking the lambda function through the console, we are invoking it via the script of boto3.



All of our code is uploaded to the GitHub Repo.



Issues faced with Raspberry Pi:

RPi Cam is a plain circuit board that is highly sensitive to electrical interference, static electricity, or physical damage. Even if we touch the cable, it will cause video interference or can potentially damage the board. The RPi Cam comes with a very short cable. If we buy third-party cables, then the reliability of the Pi cam gets reduced. Also, we may need to install extra device drivers, which will have to be stored on the RAM of pi, thereby reducing the speed. The aim of this project is to provide live feedback, and hence any delay will further delay the algorithm. Since our project is heavily dependent on the quality of the image. If the quality of the image gets reduced, then it may so happen that even after performing image cleaning and segmentation, the image captured will be blurred, and the CNN algorithm which we have used for text detection in amazon rekognition will not be able to give the corresponding text file, and as a result, the text and audio both will fail. Built-in web camera circuitry is already concealed and has been already been interfaced with the system OS.

Future Scope of the Project:

Compatibility: We can install the camera module on the glasses, and this module will share the images with the mobile phone. Then the mobile will share the data to and from the AWS cloud. Basically, the mobile device will be a master link between glasses and the cloud.

Log Analysis: As we are storing the .txt files in S3 buckets, we can perform query the txt files by using AWS Athena. Athena is a query tool that has a similar syntactical structure to SQL.

Custom Alexa Skills: We integrate and create our own custom Alexa skills for visually impaired individuals. They will help them in navigation. For example, “Hey Alexa, which stores I walked by when I was on XYZ street?” As all the data is stored in S3, Alexa can easily pull the data with the help of Athena.

Cost Analysis:

The following table represents the Cost breakdown for the hardware as well as the various AWS services that we used to do our project.

Sr. No.	Hardware	Cost
1.	Raspberry Pi 4	90
2.	Pi - Cam	70
3.	Micro SD	28
4.	Delivery Charges & Tax	13

AWS Services Cost

Sr. No.	AWS SERVICE USED	Cost
1.	AWS S3	Currently in Free Tier
2.	AWS Lambda	Currently in Free Tier
3.	AWS Rekognition	Currently in Free Tier
4.	AWS Polly	Currently in Free Tier

Recommendations/References:

https://aws.amazon.com/lambda/

https://aws.amazon.com/sdk-for-python/

https://aws.amazon.com/s3/

https://aws.amazon.com/rekognition/

Appendixes:

AWS Services

AWS Service	Description	URL
AWS S3	Amazon Simple Storage Service (Amazon S3) is an object storage service offering industry-leading scalability, data availability, security, and performance	https://aws.amazon.com/s3/
AWS Lambda	AWS Lambda is an event-driven, serverless computing platform provided by amazon as a part amazon web services. It is a computing service that runs code in response to events and automatically manages the computing resources required by that code	https://aws.amazon.com/lambda/
AWS Rekognition	Amazon Rekognition offers pre-trained and customizable computer vision capabilities to extract information and insights from your image and videos	https://aws.amazon.com/rekognition/
AWS Polly	Amazon Polly is a service that turns text into lifelike speech, allowing you to create applications that talk, and build entirely new categories of speech-enabled products	https://aws.amazon.com/polly/

Hardware Components

Component	Description	URL
Raspberry Pi	Rasberry Pi a quad-core 64-bit processor, 4GB of RAM, wireless networking, 40 pin GPIO header and dual display output.	https://www.raspberrypi.com/products/raspberry-pi-4-model-b/
Pi -Cam	Sony Exmor IMX219 Sensor Capable of 4K30 1080P60 720P180 8MP still	https://www.amazon.ca/Raspberry-Pi-Camera-Module-Megapixel/dp/B01ER2SKFS

Acknowledgments:

We would like to thank our mentor Amrinder for guiding us throughout the project and for sharing his valuable knowledge with us. He was very cooperative during the entire capstone journey and never hesitated to help along the way. We would also like to thank professor Allan for developing our technical skills from the start of the entire course. Also, we would thank the entire Embedded System program professors for helping us during the course and for motivating us, thereby bringing our best capabilities.