**Image-to-Speech**

**Project Synopsis**

Of minor project

Submitted by:

Harshit Kumar (20103011)

Sanil Gupta (20103018)

Abhitash Singh (20103019)

Suyash Rajvanshi (20103053)

**Introduction**

The project that we plan to create is an Image to Speech using OCR. There are a variety of tools that can convert text to speech, however there aren’t any tools that can convert an image into speech directly. This creates problems for blind people, who often have trouble navigating menus on a computer and thus there is a need for a tool that can work seamlessly. The field of our project primarily lies in OCR and image processing, since most OCR tools require images to be relatively noise free to work properly. At the same time, phone cameras can not produce noise-free images, hence our project will deal primarily with cleaning up these images and working on them.

**Rationale**

Often, images of documents are shared across the internet without sharing the original source of the document itself. While there are a variety of speech synthesizers that can work with word documents etc., there are none that can work with images. This presents a problem to blind people who often rely on speech synthesis to work on a computer. Thus our project will aim to bridge this gap, and create a speech synthesizer that can read directly from an image using OCR technology.

**Objectives**

* To be able to de-noise images to the point where existing OCR technologies can work with noisy images.
* To be able to read text from images captured from a smartphone camera.
* To convert text into synthetic speech.

**Plan of work**

Our first task will be to process images to the point where the effects of noise is zero/minimal on the quality of text being read by the OCR. This involves leveraging the fact that text is often a solid colour (most of the times black) and has well defined edges unless the quality of the image is very poor. To reliably simulate and test our techniques, we shall apply varying degrees of noise to non-noisy images, and apply de-noising techniques, such as edge detection, superimposing blurred images etc. We expect these techniques to reliably produce images that are of a fairly good quality.

The next step would be to input these images to the OCR library, and finding out the percentage of text correctly detected. 100% text recognition is ideal, but unrealistic to achieve. Here, we will leverage the human mind’s capability to automatically spell correct words given enough context. We expect that even if the OCR can recognise most of the text correctly, the rest of it can be automatically understood by the reader, hence some inaccuracy can be tolerated here. The bigger challenge here would be to eliminate false positives, i.e., outputting text where there is just noise in the image. However, we expect that our de-noising techniques will be able to take care of this problem.

Finally, converting the text to speech is not a challenging task since it is very well researched. AI assistants in low performance phones can synthesize human voices very well, so we expect this to be a relatively easy task.

All our technology will be enclosed in a WebApp, where the users will be able to upload their images, and receive text, and a voice-over reading that text. We are choosing the platform of Web Apps since browsers have a host of accessibility options that will make it easy for a blind user to navigate and use our project.

**Expected Outcomes**

* We should be able to develop a product that can assist blind people with understanding the text that is written on images.
* The product should have a sophisticated enough de-noising process that it should be able to read text from an image taken with a smartphone camera.
* The product should be able to process one image within a few seconds so that users do not have to wait long.

**Technologies**

* Python – For developing the backend of the WebApp.
* Tesseract – For performing the OCR functions.
* OpenCV – For image processing.
* HTML, CSS, Javascript – For developing the front end of the WebApp.
* gTTS – For text-to-speech conversion.