Audio Streaming of Classroom Presentations

CSP315

Abdul Hadi Shakir 2010CS50202
Akshay Bhat 2012CS10213
Prateek Jaiswal 2010CS50290
Prince Bhalla 2012PH10863
Shashank Sinha 2010CS10252
Shubham Solanki 2012CS10254

Instructor: Prof. M. Balakrishnan

1 Objective

The objective of this project was to create a system which enables better integration of visually impaired students in modern classrooms, assisting them in understanding classroom slides.
Currently there does not exist any non intrusive system to which enables the visually impaired students to understand classroom slides. The current solution involves handing out braille handouts or manually transferring slides to the visually impaired person's laptop.Current solutions require active efforts made by the presenter. Such a method is intrusive and inconvenient. A bit of research regarding the condition of visually impaired students in India highlights this problem.
The solution which this project intends to provide is to create an all in one embedded system which captures images of the classroom presentations, performs some computation (image processing) on it and then broadcasts the extracted text to visually impaired user's laptop remotely connected to this device, where the user can listen to the audio stream generated from the content. Such embedded device can be mounted at a fixed location and different people can connect to device to receive the stream of text extracted from image.
Such a system can also be used in scenarios outside the classrooms also. It can also be of use to visually impaired people in meetings or any scenario where the speaker is presenting the data visually to the audience.

2 Methodology

The idea is to provide visually impaired students with some control over the information present in the slide. There is a need for a simple solution which doesn't require active efforts made by presenter and yet effectively transfers information to visually impaired students.

2.1 Design Overview

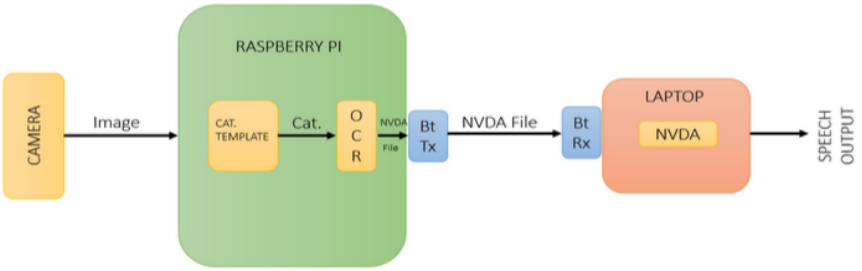
a. The Raspberry camera module is placed at an appropriate location from where it can take quality images of the classroom presentations.

b. The images are transferred to the Raspberry Pi which runs the slide categorization and OCR module.

c. The slides are categorized into respective formats and then OCR module extracts out the text from the slides.

d. The text is transferred to the user's laptop via bluetooth

e. The text received on user's laptop can be converted to an audio stream using NVDA.



So the work and structure of this device can be divided into these broad components.

- Raspberry Setup & Module integration
- Slide Identification & OCR module
- Network connection and transfer of information to user's laptop
- Conversion of text to audio stream

The documentation of each component is provided below

2.2 Raspberry Setup

- The Raspberry Pi needs to be configured with a OS to enable working on it. The latest image pf Raspbian was downloaded, which is a custom built of Debian wheezy.The image was then burned to the SD card.
- To work on Raspberry Pi, insert the SD card, connect the keyboard, display unit and power, the board should start. The boot screen becomes visible on the screen.
- At the configuration screen, select SSH and enable SSH server (assume on by default). Also enable camera for use with the board. Default login credentials are
  username: pi ; password: raspberry.

2.2.1 Capturing images using Raspberry Pi camera

- The camera should be attached to the board properly and secured tight before switching on the Pi board.
- Using terminal, the following command can be executed for capturing image output

```
{raspistill -o cam.jpg}
```

2.2.2 Compiling and running project

OpenCV and Tesseract library has been installed on Raspberry Pi to run scripts pertaining to image analysis and OCR respectively.

- Extract the code and enter directory containing source code
- Run the following command to compile the project

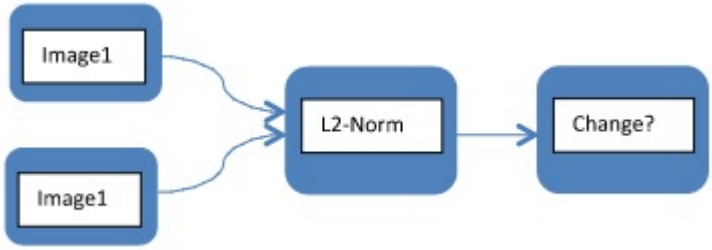
```
g++ -ggdb 'pkg-config --cflags opencv tesseract'
-o 'basename processImage.cpp .cpp' processImage.cpp
'pkg-config --libs opencv tesseract‘
```

- Provide executable rights to shell script
- Invoke script from terminal to start the project

2.3 Slide Identification & OCR Module

2.3.1 Slide change detection

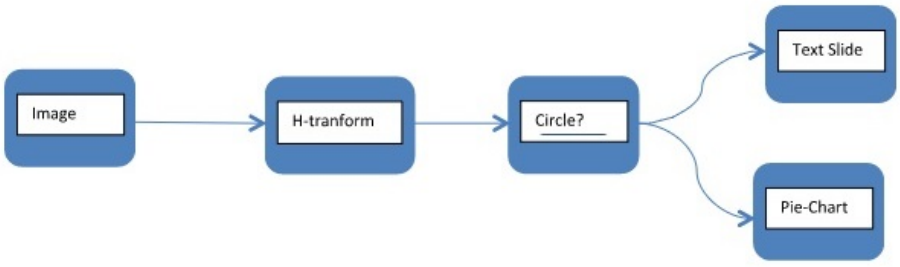
The camera samples images at an interval of 5 second. The first step towards making an efficient system was to identify images of new slide when a slide change happens. Only when the image is of new slide, it is sent to the processing engine.Whenever a new image comes, it is compared against a previous image to check if the it is the image of new slide or of the same slide. For this the two images are first converted to gray-scale and then pixel wise L2-norm difference is taken. If this difference is greater than some threshold, we claim that a slide change has occurred and then it is sent to the processing engine.



2.3.2 Slide type identification

For start, the idea was to support two types of slide: i) Text Slide and ii) Pie-Chart Slide. One of the important features for intelligent streaming of classroom presentation was to identify the type of slide and convey this information to the visually impaired.
In order to achieve this we applied image processing techniques on the captured image. Firstly, the Gaussian filter is applied over the image to reduce noise. Secondly, the fact that images of pie-chart slide will contain circular or elliptical

shapes is taken in consideration. So the circle shape identification can be performed using hough-transform to identify pie charts. Hough transform basically detects circle in images pointing out its center and radius. If the image contains circle greater that have radius greater than some threshold length, we claim that it is a image of a pie-chart slide.



2.3.3 Slide content extraction

The textual content from the slides are extracted by performing Optical Character Recognition (OCR). To recognize characters, the open source OCR engine named Tesseract was utilized. The tesseract engine was incorporated in our code using standard C++ APIs. The tesseract engine takes in an image as input and outputs a textual string. The string is then refined to remove junk and noisy characters.

2.4 Network connection and transfer of information

- The idea was to use bluetooth module to send files from Raspberry Pi to user's laptop
- But certain issues arrised in operating bluetooth module. The raspberry bluetooth module pairs with laptop's bluetooth but there no option to send files to laptop can be found.
- So instead the files are transferred from pi to a laptop connected to operate scripts etc. using bt sync.
- Then files are transferred from this developer laptop to user's laptop using a file sync software named Desychronize. Both laptops have to been connected to a single WiFi network

2.5 Conversion of text to audio stream

- The text is converted to audio on user's laptop using NVDA. Non Visual Desktop Access (NVDA) is a standard software to aid visually impaired people in using laptop. The content generated on Pi by the oCR module has to be arranged to guarantee that it is in a certain layout/template.
- The user's laptop has a script with a GUI application on top of it which with the assistance of NVDA application reads the content out. The objective was to make it as programmed and seemless as could be expected under the circumstances.
- To fulfill this, a script was composed in python with a GUI application on top of it, which would make a custom WIFI network for the Raspberry Pi to connect. This script running parallely recognizes the presence of another new or modified document in a specified folder, the folder will be fixed for receiving files from the broadcast laptop. This document is opened for NVDA to read out emphatically.

3 Status

A successful demo of the project was delivered on Open House Day. The Raspberry Pi was mounted on a raised platform in front of the projector screen. And then then scripts related to OCR and slide identification were run on Pi, and settings pertaining to network connection and audio conversion were configured. The slides were captured and sent to user's laptop in a suitable format which was then converted to audio stream in real time.

Mentioned below are the milestones covered in this project till now. Proposal for improvements are also mentioned with them:

- The slide identification module can now identify two types of slides- bulleted text slides and pie charts. This module can be extended to identify various type of objects in slides like Barcharts, graphs etc.
- The connection module involves sending files using open source software namely Desychronize. Instead Raspberry Pi's WiFi module can be used to send files to user's laptop to make it totally independent.