NARRATIVE SYSTEM FOR VISUALLY IMPAIRED

Project Reference No.: 42S_BE_1348

College : K.L.E. Institute of Technology, Hubballi  
Branch : Department of Electronics and Communication Engineering  
Guide : Mr. Arun A. Badiger  
Students :  
Ms. Afifa Anjum Doddamani  
Ms. Afreen M Shaikh  
Ms. Dhanashree Pai  
Ms. Azimunissa M  

Keywords:  

Introduction:  
There are about 285 billion visually impaired people in the world; 39 billion people in that are blind and 285 billion people are low vision. This system aims to provide this missing experience for them. The system uses state of art deep learning methods from Microsoft Cognitive Services for image classification and tagging. The experience is powered by Bone Conduction plates and Microphone. The narration generated by converting the scenes in front of them to text which describes the objects in the scene.

Objectives of the project:  
1. Capture the image present in front of visually impaired person upon receiving voice commands from the user and identify the objects from the captured image.  
2. Convert the captured images to text.  
3. Convert text to voice to narrate the objects present in the image.  
4. The narration is heard with the help of bone conduction plates.

Methodology:  
The block diagram in Figure 1 (a) & (b) shows a visually impaired person using narrative system with camera fitted on cap, ear phone with bone conduction plates, USB Microphone and Raspberry Pi. The user issues voice commands and receives the voice response from Raspberry Pi.
State-of-the-art components and tools used to implement this project work is as shown below:

1. **CAMERA**: In the intelligent system 8 MP camera is used to capture the real time image of scene around the blind person, which is placed on the cap.

2. **Raspberry Pi**: The main module is of Raspberry Pi which is on its own a mini computer is used for interfacing camera, Mic and Speakers. Python programming is used for integrating Microsoft Cognitive Service and Alexa Voice Service.

3. **Microsoft cognitive services**: Microsoft Cognitive Services are used for emotion and video detection, speech and language understanding. The image recognition process is done using deep learning techniques and the objects in the images are described in the form of text.

4. **Amazon Web Service (AWS) DynamoDB**: After recognition of image, the image is converted into text and uploaded into AWS DynamoDB which is one type of the software used in a fully managed NoSQL data base service that provides fast and predictable performance with seamless scalability.

5. **Amazon Web Service (AWS) Lambda Function**: Lambda is the Serverless environment offered by AWS. The code uploaded can be triggered with events from mobile applications, HTTP endpoints and other AWS services. It is used to trigger the access to Dynamo DB to fetch the text based on voice commands.

6. **Alexa Skill Kit**: The Alexa skill kit triggers Amazon Lambda function to bring the data from the data base DynamoDB. The use of Lambda functions as triggers for Amazon DynamoDB table. The various utterances like “Describe the scene”, “Describe What is there in front of me” and “What is their around me” are added to Alexa in skill kits.

7. **Bone Conduction for hearing**: The Bone Conduction plates helps to hear the clear instructions given by the device. It is a substitute for earphones so that the person can hear outside world sound and the narration at the same time.

8. **Audio amplifier and piezoelectric plates**: PAM 8403 is used as an audio amplifier and piezoelectric plates are used for bone conduction.

   The working of bone conduction plates is shown in Fig. 2 (a) and (b). Fig 2(a) shows the actual ear working using air conduction and also bone conduction. Fig 2(b) shows the audio amplifier with piezoelectric plates connected to temporal bones.
Figure 2(a): Ear working using air conduction and bone conduction, (b): Audio amplifier with piezo electric plates.

**System Execution:**

- User gives voice commands
- Capture Image from RPI
- Upload Image to MS Cognitive Services
- Get response as text from MS Cognitive Services
- Aggregates text to form sentence
- Upload text to Dynamo DB
- Alexa fetches text from DynamoDB and narrates it to the user

Figure 3: Flow diagram of system execution

**Results and Conclusion:**

**Results:**

The results are discussed in Table 1, the various input images and respective outputs are shown below. Figure 5(a), 6(a), 7(a) are the input images given for narrative system. On voice request “Describe the scene”, the narrative system has given correct voice responses which are shown in figure 5(b), 6(b), 7(b).
Table 1: Results and Discussion

<table>
<thead>
<tr>
<th>Figure 5(a): Image of a Computer</th>
<th>Figure 5(b): Computer is identified by the Narrative system</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Computer Image" /></td>
<td>Trying to print I think it is a desk with a laptop computer sitting on top of a table. And the keywords are indoor, computer, desk, table, laptop</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Figure 6(a): Image of a Chair</th>
<th>Figure 6(b): Chair is identified by the Narrative system</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image2" alt="Chair Image" /></td>
<td>I think it is a blue chair. And the keywords are indoor, chair, blue, table, sitting</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Figure 7(a): Image of a Bag</th>
<th>Figure 8(b): Bag is identified by the Narrative system</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image3" alt="Bag Image" /></td>
<td>I think it is a piece of luggage sitting on top of a suitcase. And the keywords are indoor, luggage, sitting, black, suitcase</td>
</tr>
</tbody>
</table>

**Conclusion:**

The proposed project “Narrative System for Visually Impaired” is very helpful for visually impaired people, it provides narration of world around us, hence it can be a very effective product in improving the lifestyle of visually impaired people.

**Scope for Future work:**

1. Reading the newspaper out loud: Using the text to speech converter we convert the information in text format to an audible format which help the blind people to read.
2. Keeps track of the location of the person continuously: It provides security. It helps the family members to keep a track of the blind person.
3. Helps in Home automation: It helps to automatically on and off the electronic appliances.