An AR Mobile Application for Visually Challenged People

Aayush Sharma:-	1222229268
Danish Pasricha:-	1223573559
Prakhar Praveen:-	1211591729
Tanuj Singh:-	1219525476
Meghana Sukhavasi:-	1220187449

Significance :

Visually Challenged people can use this application to perform their everyday activities conveniently.

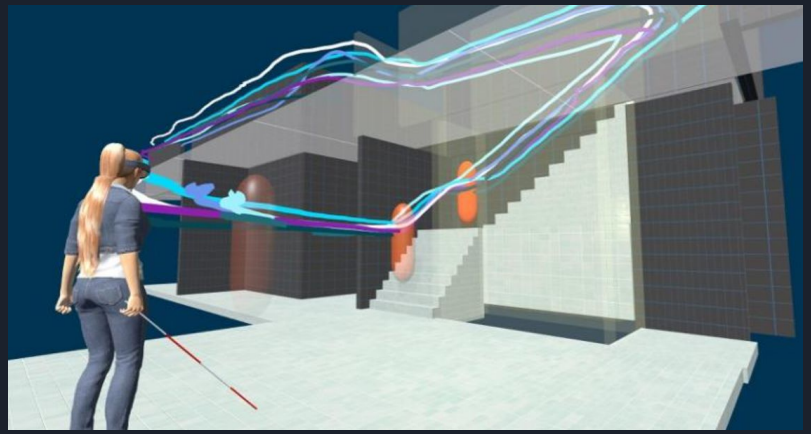
An example of this kind of application is Google Maps Walkthrough feature which can detect buildings and roads in your surroundings using Camera and Deep Learning plus Geolocation and walk you to your chosen destination as if it were your personal guide . This app throws voice and vibration notifications when the user is in close proximity to an obstacle. This could prove very beneficial in case of blind users .

On top of that, it can translate texts in the local language of the foreign country to their native language, and make suggestions on what they can do with that information and what they can say to ask people for help ,also the Alexa-like Voice Assistant can read out the English expressions for another person to hear so they can help the users complete a task or find their way, i.e. directions using public transit, taxi, walking, driving or a mix of some or all of the above.

Problem Definition:

There is no app like this one that can identify nearby items and surroundings in real time while the user is interacting with them.

For visually disabled people, an application becomes the replacement of a human helper and facilitates their daily activities. For a better user experience, we can use AR sensors that are connected to our mobile application which will give him voice commands when an obstacle is in close proximity.



Solution Approach:

We are going to use data like location , time stamp and technologies like Computer Vision based Surrounding Detection that effortlessly identifies objects, buildings using a fast data exchange algorithm for communication with Software services . It combines these services with our lightweight Machine Learning algorithms using Convolution Neural Network models stored locally on your device locally that run in the background. This can be improved using volunteer computing, which is cloudlets wherein resources are shared across mobile phones connected to the application’s user network.

Novelty:

Even though our idea draws inspiration from existing AR-enabled apps we have used, the likes of which are Live View feature in Google Maps and the immersive walking directions in Apple Maps (iOS 15 onward) with rich 3D views of cities like London, Los Angeles, New York or AR implementations like Pokemon Go's world or batman's detective vision; we are making this application for a different and novel reason of increasing accessibility of people that are specially-abled, such as blind and can later be extended to deaf, dumb, or paralyzed.

This application helps user to use AR enabled sensors connected with the application to make the user perceive the environment. An example is to use proximity sensors inbuilt in the Mobile to show any bumping chances with surrounding objects or people.

Why is it Mobile Computing:

We are creating an app that uses smart features found in smart phones, using artificial intelligence to detect user's physical surroundings, both indoors and outdoors and guides users to complete the actions they want to do around their surroundings. It requires a technology that is allowing its users to transmit data from one device to another without physical cables or links and using concept of volunteer computing and cloudlets.

Why is the problem novel?

There is no such app that can detect objects, surroundings near by in real time as the user is interacting with the surroundings. This problem can be solved by our application. This app helps user to interact with surroundings realtime. This application works in place of human assistant for visually disabled and makes their daily actions more convenient.

Now a days people walking on the footpath bump into each other while texting on their phones. This can be avoided by using this application which will let you know if you are about to crash into someone. For this mobile need to have proximity sensors which are not available right now and can be added into, in near future.[1]

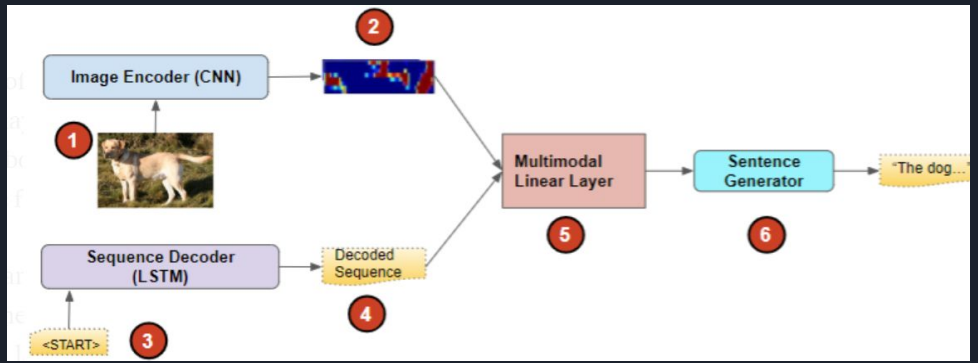
Why is the solution novel?

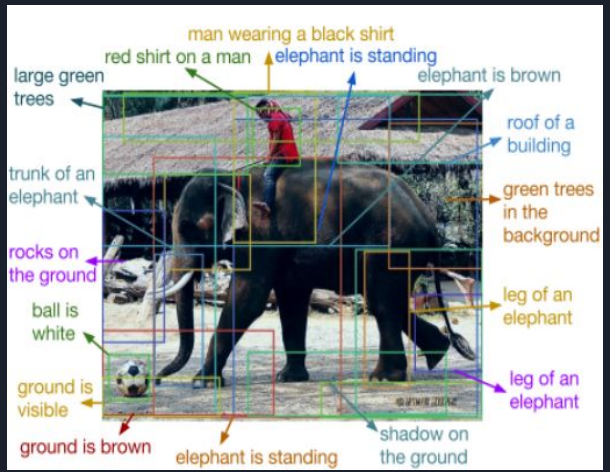
Our app will be an extension of currently available Digital Assistants. It will be an AR-based Surrounding detection Guide app that uses your phone camera and your current location. The smart Surrounding detection will use Computer Vision and other fast Machine Learning algorithms to recreate your “Current View through your Eyes” on your mobile internally and it will help visually challenged users to access and interact with surrounding objects which would not be possible for visually impaired in other cases. To conclude, the solution tries to compensate for the visual impairment of those individuals empowering them and making them equal. More importantly this solution removes the barriers between separate apps. The integrated functionality of our app combines Augmented Reality, Computer Vision, and Natural Language Processing. Our app helps the blind to communicate with other people in their surroundings and do everyday life things around them seamlessly as if the Assistant were a personal caregiver.

Solution : Algorithm, Methods

Solution 1:-

In many machine learning applications, you typically have to deal with the location data. The context that location data often adds to the data in your application is very rich. Deep learning has made great strides recently in the fascinating field of **image captioning**. It's even more intriguing because it combines both computer vision and natural language processing. It accepts an image as input and outputs a brief textual summary that summarizes the information in the image. In our Idea, with the help of **peer-to-peer computing and cloudlets**,we can make use of this method. By using the computation power from our ad-hoc mobile network,we will be able to get the images of the surrounding s and by using **CNN and LSTM**, we can caption the images and thus will be able to get the knowledge of our surroundings and detect the objects





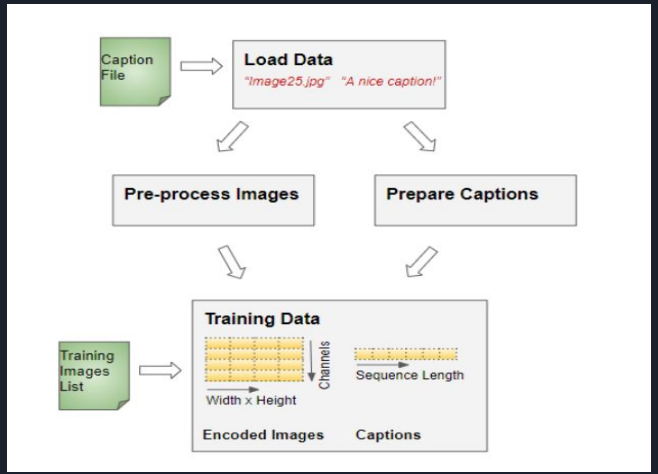
Solution : continued

Solution 2:-

One another possible solution could be the use of **posenet** and **K-NN** Algorithms. We could fetch the visual data from our ad-hoc mobile networks which will act as a source data for our mobile application. From here frames of the visual data will be extracted and then json data files could be generated which can help in detection of the category of the obstacle seen in the visual data. Posenet will be used in **Frame extraction** and K-NN algorithms like **Decision tree** could be used for categorization of the data. For the implementation of this method, we first have to train our data. Visualization of the above method could be represented like this:-

1. Load the image and caption data
2. Pre-process images
3. Pre-process captions
4. Prepare the training data using the pre-processed images and captions

We employ transfer learning to pre-process the untrained CNN-based network's raw pictures. The encoded picture vectors that capture the key aspects of the image are created using the input images. We then feed our Picture Caption model these encoded image attributes rather than the original raw images. Additionally, we send the target captions for each encoded image. The model learns to predict captions that match the target captions by decoding the visual features.



Feasibility:

Even though this has been tried to be implemented from long time and a lot progress has already been made but still it will take 5 years easily to complete this. We can draw an analogy of this to autonomous driving, even after so much progress, we still cannot completely shift to autopilots as the mainstream driving on roads in traffic. Similarly, as part of this project, we need input from all the devices around us, need to train the model to categorise descriptions of visuals and then further predict them into suggestion, warnings respectively. If we try to look at it from a broader perspective, we can clearly conclude that this is exactly what our brain and our senses perform before making a decision. We need to collect lot of sensory data belonging to all different situations and this needs to be collected across a very long period of time so that the prediction of the model becomes more accurate

Feasibility:

Lightweight machine learning model is a very big hiccup as best of the best models right now are compute expensive as well as storage expensive. If these models are not lightweight, they cannot be installed in the devices and kills whole purpose. Similarly if they are not fast and the predictions are not at right time, it will be of no use.

Volunteer computing at such a large scale is also quite difficult. Google maps have become now relevant because of such a big customer base feeding data from a long period of time. Similarly for the processing and data collection for the ML models in volunteer computing, a huge number of devices need to be considered to feed in data for a long amount of time

References:-

[1] https://slate.com/technology/2018/05/google-maps-how-augmented-reality-is-going-to-change-walking-directions.html

https://towardsdatascience.com/image-captions-with-deep-learning-state-of-the-art-architectures-3290573712db

https://towardsdatascience.com/image-captions-with-attention-in-tensorflow-step-by-step-927dad3569fa