Do you know the way?

Leave a vision behind.

ABSTRACT

The world is not a safe place for people with sensory impairments. Since they make a minority of the community, advancements in the standard of living tend to leave them out and move faster than accessibility technologies can keep up. This leads to them being ostracized out of the amenities of modern life. This app aims to close the gap between the visually impaired and the rest of the populace by empowering them with a tool that puts the best of AI and machine learning into their palms via, now commonplace, smartphones. This app aims to make smartphones truly the digital companions of people with sight disabilities.

There is a plan and a purpose, a value to every life, no matter what its location, age, gender or disability.

TECH-STACK

- ➔ Python, Bash scripting, Machine learning , Deep learning, Voice recognition

MOTIVE TO CREATE THE APP

An estimated 253 million people live with vision impairment. While they make only a minority out of more than 7 billion people, they too have the right to all the amenities and facilities afforded to the people who do not have these impairments. Seeing this disparity in the standard of life inspired us to do something and put our technological know-how to their welfare.

Sight is by far our primary means of learning our environment. Though the blind have come to learn of their frequented environments and are habituated to them, new environments and surroundings pose a constant danger to them.
These new surroundings are unpredictable, uncertain and in general tend to disturb the regularity of their life. This app reduces this by periodically informing the user about the items in their vicinity and alarming them if any object enters their personal space, informing them about the direction and speed of said object. Another problem that we wish to solve is that the visually impaired find it generally difficult to 'look' for an object that they have misplaced or are seeking. This app has a **search mode** that actively seeks a specified object in their field of vision and notifies the user when it is found. Needless to say, the app is completely speech activated.

WORKING OF OUR AP

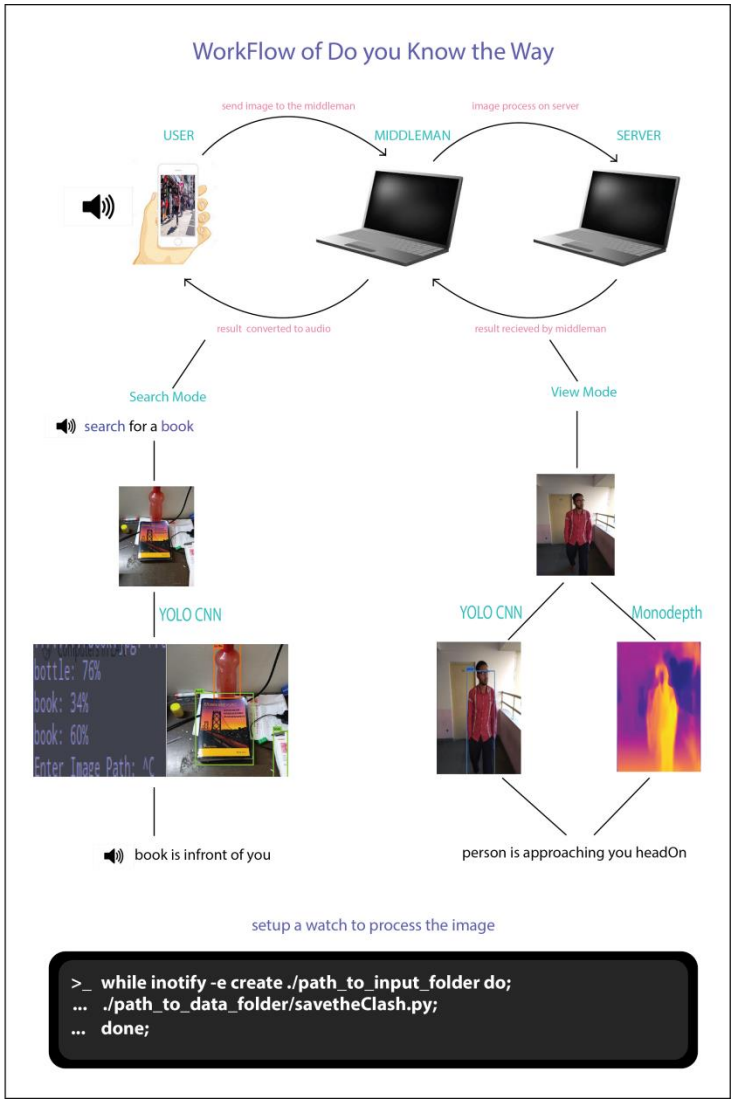

Image is transferred from the device to the middleman through IPWEBCAM. And middleman to server via SCP an IP.

Search: Image processed using yolo net and the result is compared with a voice generated txt file. If the search object is matched, an audio is generated to inform the user.

View: Image is processes simultaneously on Monodepth and YOLO via subprocesses. Monodepth is the RDS of our network and on our gpu provides a speed of 1frame/2sec. The movement of the center of the images from the current and previous transfer to estimate the motion of the object.

Objects lying in a specific hue range (selected through experimentation on various images of varied object distance) are classified as a potential threat and are notified via audio.

Pass: for cases where the user wants to switch off the View mode (eg talking to a known person)

Although many apps exist that claim to provide a similar functionality, none of them come even close to processing vision at the speed facilitated by our approach. Most systems rely on sending requests to a remote server via HTTP where it is processed and the results are delivered back, taking a long time. Since time is very critical in situations like navigation where one misstep could be the difference between life and death, our app eliminates this time lag by processing the data locally, where the processing happens on a server connected to the device via IP and SCP technologies. Also the plan is to move this processing into the ML chipset of the phone as this technology becomes more and more mainstream. Also no approach exists that processes both object detection and depth simultaneously, the coupling of which enables new possibilities for the technology to assist the blind.

LIMITATIONS AND FUTURE IMPLEMENTATIONS:

1.) Presently our app is dependent on a middleman hardware which may be resolved by introducing a GPU for computations in a device, say the Goggles of the user.
2.) We use hue range of the Stereo image (produced by the Monodepth CNN) which sometimes tend to give false results by ignoring to be concerned objects. We wish to solve this by extracting ground truth data from the stereo image so that we may be more accurate.

1.) Solving the above mentioned and using a better hardware to improve the speed of our network, currently working at 1frame/sec to maybe actual real-time.
2.) Adding more features like OCR , introducing user to this form of mass media.
3.) Extracting ground truth depth so that we can approximate the time or number of steps before collision(if any) with the incoming objects.