Internet of Things (IoT) based Object Recognition Technologies

Dr. K. Srinivasan[1], V.R. Azhaguramyaa[2]

[1]Associate Professor, Department of Electronics and Communication Engineering,
Sri Krishna college of Technology,
Coimbatore, Tamil Nadu, India.
k.srinivasan@skct.edu.in
[2]Assistant Professor, Department of Computer Science and Engineering,
Sri Krishna college of Engineering and Technology,
Coimbatore, Tamil Nadu, India.
azhaguramyaa@skcet.ac.in

ABSTRACT

Over the past years, the Object recognition technologies have matured to a great extent, where it has enabled the development of exciting solutions for visually impaired. A variety of solutions has been proposed using computer vision and object recognition to help the visually impaired with their day-to-day activities. This paper aims at the development of a solution that can be adopted by the visually impaired for identifying and locating household objects in their daily life. The solution includes a wearable device that listens for the user's voice, understands the user's command and locates the object in the surrounding environment. Once the target object is located, it gives user the information about the object and the maximum possible distance of the object from the user. The performance of the device has been increased and the battery consumption has been decreased by adopting an efficient algorithm that makes the device usable in day-to-day life.

Keywords: voice, vision, visually impaired, IoT, object recognition

INTRODUCTION

Vision is one of the most essential human senses and it plays an important role in perceiving the surrounding environment. According to the World Health Organization (WHO), about 285 million people are visually impaired worldwide out of which 39 million are blind and 246 million have low vision [12]. These people often find it difficult to perform their day-to-day activities such as locating household objects, reading and writing, navigating on the roads, etc. Since 1970, several solutions have been proposed for visual substitution[1]. The advancement in the areas of machine learning and computer vision have definitely contributed to the development of efficient solutions for visual substitution [8]. But, most of these solutions focus on navigation in unknown localities and obstacle detection while only few focus on object detection to identify the different components in the environment [10]. For this reason, we propose a solution employing robust and fast computer vision algorithms that can be practically employed in day-to-day life to locate house-hold objects [9]. In this paper, we first present an overview of the existing related vision substitution systems and then introduce our proposed system and compare their performance[5].

OVERVIEW OF EXISTING VISUAL SUBSTITUTION SYSTEMS

The traditional solutions for visual substitution such as dogs and white canes are very useful in navigation in unknown places but they aren't much helpful in performing the day-to-day activities such as identifying and locating house hold objects. So, modern solutions involving computer vision are employed for helping the visually impaired with their day-to-day activities [1].

A visual substitution system can be evaluated in terms of its input sensor which receives the input stimuli from the environment, the time taken for processing, the output after processing and the accuracy of the output. A visual substitution system usually takes the live video feed as the input and gives auditory information as the output. The quality of the input video feed is usually enhanced by using computer vision algorithms for accurate processing. Since the processing is done in real time, the processing time must be as low as possible while preserving the output accuracy.An overview of various vision substitution systems of our interest are given below:

1-The vOICe

The vOICe was developed by Leslie Kay from the University of Birmingham in 1974. The visuo-auditory device uses the idea of echolocation of bats. The system has a camera that processes the live video feed and transforms the video feed into sound. The input video feed is processed from left to right and corresponding sound is produced whose frequency depends upon the brightness and location of objects in the video feed. The visually impaired person hears the sound and tries to understand the environment. This system however does not give exact information about objects in the environment. It produces sound depending upon the brightness and location of objects. So the user might have to guess the objects and its location with the output sound from the system.

2- The Vibe

The Vibe uses a similar technique as that of the vOIce. The subjects had to wear a wide angle camera on their head that captures the live feed. Later the video is processed from left to right, converted into grey scale and transformed into sound. The pitch of the sound gives information about the top and bottom areas of the feed whereas intensity of the sound gives information about the brightness. Therefore, a sound of less intensity implies that the object is dark and vice versa. Similar to the vOIce, the user might not have an accurate understanding of their environment because the device does not give information about the shape or size of the object [4].

3- Martinez et al's approach

Martinez et al developed a system which scans the image and produces the sound output which is similar to the vOIce but it differs in the way the image is scanned. The scanning starts at the center and expands to the left and right simultaneously. The sound patterns are generated and those patterns obtained while scanning from center to left are provided to the left earphone and those obtained while scanning from center to right are provided to the right earphone [9].

Jung Uk Kim et. al introduced Object detection in a road scene has received a paramount attention from research fields of developing autonomous conveyance and automatic road monitoring systems. The proposed object detection network is robust in occlusions [2] but required adequate power.

Liyan Yu et. al [1] Came with the method by which the computer identifies the main objects and understands the relationship between the main objects and the other objects, but it performed well without occlusion. Based on the occlusion in the inputs, accuracy is getting varied.

PROPOSED SYSTEM

Sensory substitution systems are very encouraging with the evolution of machine learning and computer vision technologies. These sensory substitution systems are aimed at solving the major problems of the visually impaired people. Multiple solutions have been proposed to recognize the objects in the surrounding environment and report them to the visually impaired through audio. However there are limitations for such devices such as high power consumption, low speed and accuracy resulting in low performance, network connectivity requirement, etc. In this paper, we propose a completely offline solution that involves a device with low power consumption that detects objects in the surrounding environment and reports them to the user. Since the device requires no network connectivity, it can also be used outside the local environment for object detection. The overall setup illustrated in Figure 1.

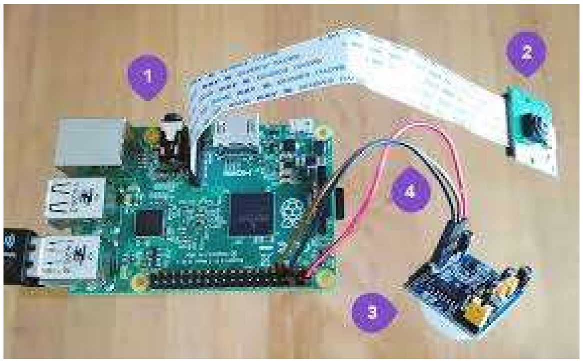
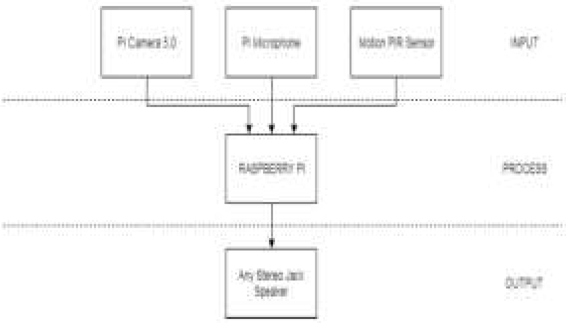


Fig.1. Setup for Object Recognition.

SYSTEM ARCHITECTURE:

The input sensors include Pi Camera for recording the live video feed, Pi Microphone for recording the user's commands and motion sensor for detecting the motion of the raspberry pi. The processing is done by the pre-trained model which resides in the raspberry pi. The raspberry pi then evaluates the input and after processing the video feed, the output is given out through any stereo jack speaker.

Fig. 2. System Architecture



METHODOLOGY

We have trained the Tensorflowssdlitemobilenet model to recognize the common objects in the household environment. The entire process explained in Figure 2.

The training involved 4 basic steps:

1- Collection of 850 images of 30 commonly used household objects.
2- Creating an XML file that contains information about objects in every image.
3- Converting the XML files into TFRecord files.
4- Training the ssdlitemobilenet model with the collected images as the input tensorand the generated TFRecord file as the configuration file.

We have used Sopare for speech to text conversion. Sopare is a Sound Pattern Recognition tool which scans for patterns in the input sound and gives the text corresponding to the matching pattern as output. The patterns are recorded and stored during the training phase. Once Sopare is trained, it can recognize patterns in the input sound and convert it into corresponding text. The object label can be extracted from the text. We trained Sopare with 12 different voices and 30 object labels. The training involved 2 basic steps:

1- Recording 12 different voices of persons reading the same object labels.
2- Training Sopare with the recorded voices and the label text as the input.

CMU Flite is small and fast run-time open source text to speech synthesis engine. Flite works completely offline. So we use Flite to provide information to the visually impaired about the found object and its distance from him. The flite takes a piece of string as input and provides the converted speech as output.

ALGORITHM

1- When the device is switched on, it continuously looks for patterns in the surrounding audio.
2- When a pattern is found in the audio, the device extracts information from the pattern which has the label of the target object to be found and stores it along with the time out limit.
3- If the motion detector has detected a motion after the last search,objectsNotFoundarray(contains a list of objects which are not in the current frame)is cleared andwe continue to step 4. If the motion detector has not detected a motion after the last search, then the objectsNotFound array is checked and if it already has the current object label, then the user is notified that the object is not found in the current frame and step 3 is repeated until the time out limit exceeds. If the current object label is not in the objectsNotFound array, then we proceed to step 4.
4- Now the current frame is recorded and given as input to the trained ssdlitemobilenet model and the last search time is updated.
5- If the model detects the target object in the current frame, we proceed to step 6. If the model does not detect the target object in the current frame, then the object label is pushed into the objectsNotFound array and continue to step 3.

6- The user is notified that the object is found using the text-to-speech API.
7- The distance of the object from the camera is found using an IR sensor and notified to the user. This step continues until the time out limit is reached or the object goes out of sight. Then we proceed to step 1.

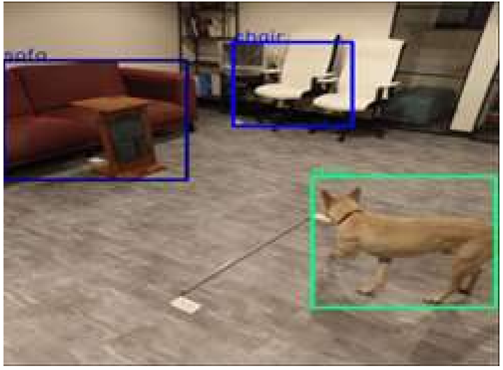


Fig 3. Results

EXPERIMENTAL RESULTS AND ANALYSIS

The performance of the proposed solution is of great importance when it's needed to be employed in daily life. The device was tested with different target objects illustrated in figure 3 in different lighting scenarios and obtained the following accuracy given in Table 1.

TARGET OBJECT	ACCURACY	AVERAGE TIME TO DETECT OBJECT IN THE FRAME
Bottle	94%	0.96 s
Chair	85%	1.2 s
Television	83%	1.02 s
Walking stick	92%	1.14 s
Spectacles	87%	1.12 s

Table 1.Target Object Recognition Accuracy.

The algorithm increased the battery life of the device to approximately 32% because it can work in offline so it can be employed in the day to day life of the visually impaired to detect common household objects.

CONCLUSION

In the recent years, several solutions have been proposed for visually impaired object detection. However, only a minority is used daily because of its limited performance and high battery usage. So this paper proposes a solution that includes a device with low battery usage and works completely offline. With the help of an efficient algorithm using the motion detection technique and storing the objects that were not found in the current frame, a lot of processing power has been saved [6][11-13].

FUTURE WORK

Our future work involves improving the performance of the device by using feature extraction algorithms such as SIFT, SURF, etc. With improved performance, these devices can definitely help the visually impaired to a great extent in their daily lives [7].

References:

[1] Liyan Yu, Xianqiao Chen andSansan Zhou, "Research of Image Main Objects Detection Algorithm Based on Deep Learning", *IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, June 2018.

[2] J. U. Kim ,Jungsu Kwon , Hak Gu Kim , Haesung Lee and YongMan Ro , "Object Bounding Box-Critic Networks for Occlusion-Robust Object Detection in Road Scene" , *25th IEEE International Conference on Image Processing (ICIP)* October 7-10, 2018.

[3] Zeeshan Saquib, Vishakha Murari and Suhas N Bhargav,"BlinDar: An invisible eye for the blind people making life easy for the blind with Internet of Things (IoT)", *2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)* May 2017.

[4] Rasika Phadnis, Jaya Mishra and Shruti Bendale, "Objects Talk - Object Detection and Pattern Tracking Using TensorFlow", *Second International Conference on Inventive Communication and Computational Technologies (ICICCT)* April 2018.

[5] J. Redmon, S. Divvala, R Girshick, et al. "You only look once:Unified, real-time object detection", *Computer Vision and Pattern Recognition*, June 27, 2016, pp. 779-788.

[6] N. Bodla, B. Singh, R. Chellappa, et al. "Soft-NMS□Improving Object Detection with One Line of Code", *IEEE International Conference on Computer Vision. IEEE*, 2017, pp. 5562-5570.

[7] A.Karpathy, F F. Li, "Deep visual-semantic alignments for generating image descriptions", *Computer Vision and Pattern Recognition. IEEE*, 2015, pp. 3128-3137.

[8] Huynh, Tri, Jay Pillai, Eunyoung Kim, Kristen Aw, Jack Sim, Ken Goldman, and Rui Min. "Bringing Vision to the Blind: From Coarse to Fine, One Dollar at a Time." In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 481-490. IEEE, 2019.

[9] Guo, Shi, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. "Toward convolutional blind denoising of real photographs." In Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1712-1722. 2019.

[10] Giva Andriana Mutiara, Gita Indah Hapsari, Ramanta Rijalul, "Smart Guide Extension for Blind Cane", *IEEE Int. Conf. Information and Communication Technologies*, 2016.

[11] Z. Liu, W. Dai, M. Z. Win, "Mercury: An infrastructure-free system for network localization and navigation", *IEEE Trans. Mobile Comput.*, vol. 17, no. 5, pp. 1119-1133, May 2018.

[12] "Vision impairment and blindness: Fact sheet", *World Health Org.*, 2017.

[13] Kumar, R. Praveen, and S. Smys. "A novel report on architecture, protocols and applications in Internet of Things (IoT)." In *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, pp. 1156-1161. IEEE, 2018.