Real-Time Object Detection Application for Visually Impaired People: Third Eye

Selman TOSUN
Computer Engineering Department
Mugla Sitki Kocman University
Mugla
selmantsn@gmail.com

Enis KARAARSLAN
Computer Engineering Department
Mugla Sitki Kocman University
Mugla
enis.karaarslan@gmail.com

Abstract — Visually impaired people can't move safely outdoors because they can not perceive the outside obstacles as normal people. The prototype application in this study aims to make the visually impaired people's lives easier with the mobile devices. The mobile application with the designed chest apparatus will be like a virtual third eye that is not too costly and easily accessible. All the design and codes is shared with GPL free software licence. The application is developed for the Android platform. Image processing and machine learning technologies are used. Methods, design, and findings are discussed.

Keywords - *Mobile Computing, Image Processing, Machine learning, Visually Impaired*

1. INTRODUCTION

According to the Social Security Administration data the world over 40 million, there are 220 thousand visually impaired people in Turkey.

Nowadays, visually impaired people have difficulties in the transportation especially in the modern crowded cities. The tools such as wand and cane used by these people are not enough compared to the facilities of today's technology.

This study aims to produce an application prototype to make the life of visually impaired people a little bit easier by using intelligent mobile devices.

The opinions of visually impaired people are very important as they will use this application. The opinions of the president of the Association of Visually Impaired in Education is taken into consideration during this study .

The application should be free and easily reachable. It is available with a free software licence. Application should use advanced techniques like machine learning and image processing.

In the next section, fundamentals will be given. In section 3, previous studies similar to this project will be given. In section 4, the experiment and the implementation is explained. In the last section, the potential benefits and promises of this study will be given.

2. FUNDAMENTALS

This study is based on mainly two main techniques; Machine learning and image processing.

A. Machine Learning

Machine learning is an algorithm category that allows software applications to be more accurate than estimating results without being explicitly programmed. The main basis of machine learning is the use of statistical analysis to generate input data and to generate algorithms for updating output when new data are given. Machine learning processing cycle is shown in Figure 1. Machine learning algorithms can be categorized as supervised or unsupervised [9].

Supervised algorithms require a data scientist or data analyst with machine learning skills who train the algorithm by providing data to enhance the accuracy of the estimations. Data scientists determine the variables or features the model will use for this process. The algorithm will use this learning to the new data afterwards [9].

Unsupervised algorithms do not need desired data for training. Rather, they utilize a iterative way which called deep learning to get results on the given data. These algorithms work fine if they have big data. They use neural networks which scan this massive amount of data and detect fine correlations automatically. The learned associations will be used to interpret new data. Unsupervised learning algorithms are used for complex processing tasks like image recognition, text-to-speech, and natural language processing [9].

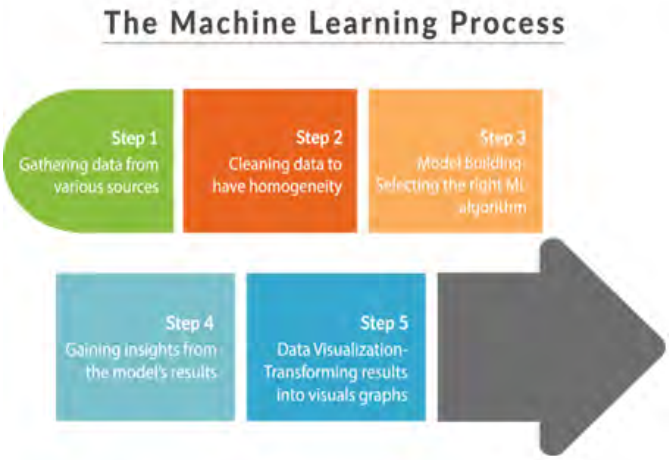

Figure 1 : ML Processing Chart [10].

B. Image Processing

Image processing is the process of transforming a view into a digital form and performing some operations to obtain an enhanced image or to get useful information from it. An image, such as a video frame or photo and output, may be an image signal or an image or feature associated with the image. Generally, the Image Processing system involves processing images as two-dimensional signals while applying pre-set signal processing methods. Image processing usually formed of the following three steps [11]:

- Importing the image to the system,
- Analyzing and manipulating the image by data compression, image enhancement and spotting patterns,
- Output as an altered image or image analysis report,

The purpose of image processing may be the following [11]:

- Visualization to determine the objects that are not visible,
- Image sharpening and restoration to get a better image,
- Image retrieval to reach the image of interest,
- Pattern measurement to determine the measures of the various objects,
- Image Detection to mark the objects as different.

3. Related Works

There is not much work being developed using image processing for visual impaired people. Some of the significant related works will be given in this section;

In a recent study [1], a mobile alerting system is developed for visually impaired people, where the current location infrastructure and construction jobs are detected by using web facilities on the internet, mobile devices and marked coordinates are determined using the GPS feature. The aim of the device is to develop a mobile program that moves people with visual impairments according to their distances to these coordinates.

Camfind application is not directly blind, but the method used in this study is similar. When the scanned image is scanned and the detected objects are written on the screen, the user is transferred through the voice command system. And the web searches about the detected object.

LCW Sense application is an application that will facilitate the use needs of visually handicapped people. LCW Sense works with barcode reading system. Barcodes are located on the carton label on the product and on the inner label of the product. Basic information about the product is transmit to the user such as size, color, pattern, product content, washing instructions and cost.

Microsoft Seeing AI, designed for the blind and low vision community, uses the power of AI to describe people, texts, currencies, colors and objects. AI viewing is a project that brings together the power of the cloud and AI to deliver an intelligent application designed to helps to navigate the day.

4. Experimental

The study is implemented for the Android platform and tested on mobile device.

Android Studio is preferred as a development platform as it is the official integrated development environment (IDE) designed specifically for Android development. Python programming language is preferred as it is suitable

for Rapid Application Development, also can be used as a scripting or glue language to put the existing components together.

OpenCv library is used for image processing. It has a wide range of libraries and is designed for computational efficiency and with a real-time application focus.

TensorFlow is preferred for the machine learning process. It provides high performance numerical computation. It has a flexible architecture which makes easy deployment of computation across a variety of platforms possible.

The steps of the implementation can be summarized as follows:

1. The objects that are needed to be defined at the application are determined.

2. Positive and negative images of these objects are downloaded from the internet, resized and unnecessary images are cleaned with the written python script.

3. Camera of the smartphone is used and the objects in the data set were identified on Android Studio. Profiles were created according to the data set and the voice commands for the detected objects are defined.

4. A suitable chest strap for the experiment is designed to hold the phone is designed and printed with a 3D printer.

5. The created demo is used and tested by the targeted users.

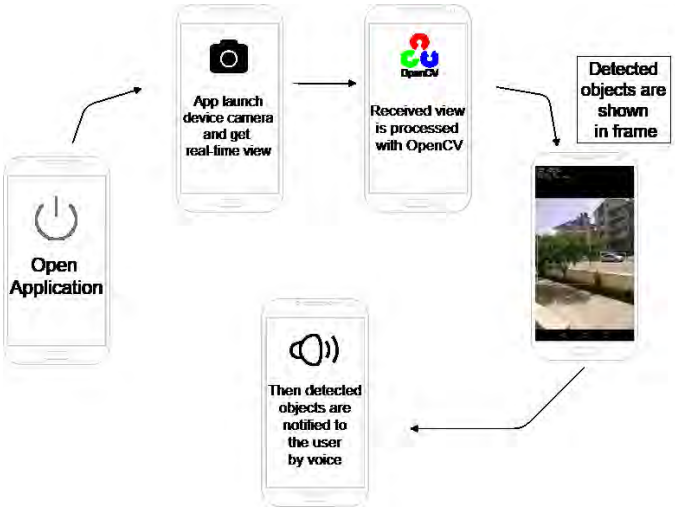

Figure 2: Flowchart of the project

A. GUI and Chest Apparatus Design

The use of the chest apparatus in the study designed with a 3D printer to make appropriate positioning. The users can download the design files and can make their own by using 3D printers.

Figure 3: SolidWorks Designs for chest strap [8].

As it can be seen in Figure 4, the application has a very simple interface for visually impaired people as its main purpose is to inform the person in sound. The objects recognized at the specified distance is shown on the mobile device screen by showing a rectangular marker.

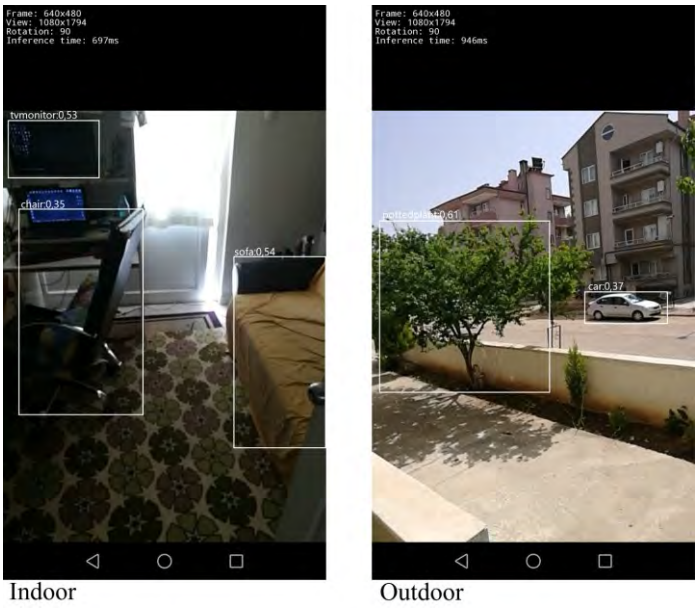

Figure 4: Application screenshots (Indoor & Outdoor)

B. Dataset

In this project, data set is prepared by training it to recognize the images of that object. . YOLOv2 is a dataset trained to recognize specific objects [13]. YOLOv2 dataset

with the GPU so it is faster and accurate than other algorithms while training our dataset.

The values of mAP and FPS of YOLOv2 and other datasets are given;

Detection Models	Train	mAP	FPS
Fast R-CNN [2]	2007+2012	70.0	0.5
Faster R-CNN VGG-16 [3]	2007+2012	73.2	7
Faster R-CNN ResNet [4]	2007+2012	76.4	5
YOLO [5]	2007+2012	63.4	45
SSD300	2007+2012	74.3	46
SSD500	2007+2012	76.8	19
YOLOv2 288 × 288 [12]	2007+2012	69.0	91
YOLOv2 352 × 352 [12]	2007+2012	73.7	81
YOLOv2 416 × 416 [12]	2007+2012	76.8	67
YOLOv2 480 × 480 [12]	2007+2012	77.8	59
YOLOv2 544 × 544 [12]	2007+2012	78.6	40

Table 1: Detection frameworks on PASCAL VOC 2007 [12].

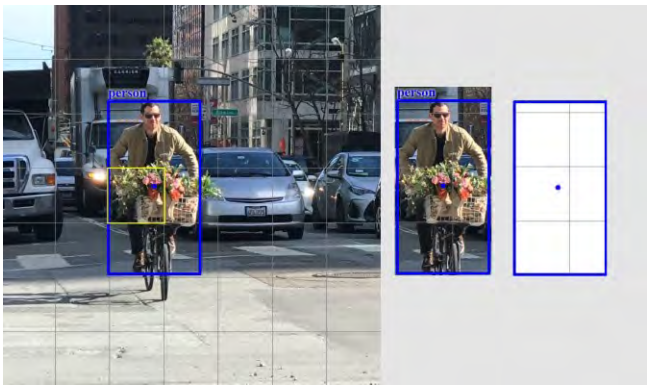

Figure 5: Working logic of YOLOv2 [5].

Each grid cell of an image has a fixed number of boundary boxes. In Figure 6, the yellow grid cell generates two boundary box estimates (blue box) to find where the person is. Each grid cell detects only one object. In this study, high box confidence scores (greater than 0.25) is kept as our final predictions

Each boundary box contains 5 items: (x, y, w, h) and a box confidence score. The confidence score reflects an object of the box (objectness) and possibly how accurate the bounding box is. In operation, the bounding box width w and height h are normalized by the image width and height. x and y offset to the corresponding cell. Therefore x, y, w and h are all between 0 and 1. Each cell has 20 contingent class possibilities. The probability of a conditional class is the probability that the perceived object belongs to a particular class (one probability per category for each cell). So, YOLO's prediction has a shape of (S, S, B×5 + C) = (7, 7, 2×5 + 20) = (7, 7, 30)[12].

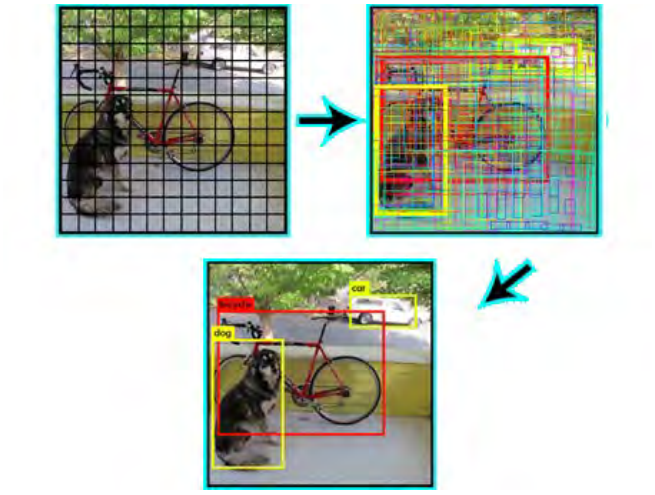

Figure 6: class confidence score = box confidence score x conditional class probability [5].

YOLOv2 is faster and more accurate than prior detection methods. YOLOv2 dataset had 9,000 objects, it worked with low FPS even on the GPU processor. Image processing to be a process that requires high performance. So in order to run the dataset with the mobile phone processor, we had to work with a dataset with fewer objects. So in the study used the Tiny-YOLO dataset the next stage.

C. Implementation

The application can also work as a background process when it is started. The audio notification wii be still on when the mobile device has activated the screen lock. There is no account creation and no sign-in process. It is necessary to adjust the positioning of the mobile device so that it can see the full course direction in order to be able to carry out the identification well. The designed chest strap is recommended to be used.

In this implementation, GPS location data is not used. It performs the operation using the camera and audio output modules of the mobile device. It will recognize and display the defined objects in the application interface. There is also a certain perception distance to recognize objects. Objects recognized at the specified distance will first be reported to the user on the mobile device screen by showing a rectangular marker. The name of the object recognized on this rectangle will be displayed to the user. Immediately after these two steps, the name of the object identified to the user, the distance in meters of the object to the device, and

the direction in which the user will be notified will be reported via the voice unit of the mobile device.

Since the time between this object recognition and the notification of the user as a voice command is a millisecond event, the user will be informed directly without any delay. The intended use of the unit as a sound unit is to reduce the external noise, so that the user can hear the sound more clearly.

D. Results & Discussion

The dataset (Tiny YOLO) which is used in the tests, has worked very efficiently on Android. As you can see the Table 2, Tiny YOLO has also slightly lowered the mAP, but a much better FPS value is obtained. This value is sufficient for detection on the mobile device.

Tiny YOLO dataset values is shown in Table 2.

Detection Models	Train	mAP	FPS
Tiny YOLO	COCO trainval	23.7	244

Table 2: Performance on the Tiny YOLO dataset.

Currently there are totally 20 classes in the dataset. Indoor classes are bottle, cat, chair, dog, person, potted plant, sofa, tv monitor, dining table. Outdoor classes are car, bus, bicycle, motorbike, train, boat, aeroplane, person, cat, dog, bird, cow, horse, sheep. In the study, it is intended to further increase the classes in this dataset at the next stage.

5. Conclusion and Future Works

This study is designed to make visually impaired people more comfortable and aware of their daily life without any help from anyone. The visually impaired people will be able to notice the threats that may arise during transportation with voice feedback and this will help preventing possible accidents. The mobile devices can be carried easily and the camera of the device can be used as a third eye to the visually impaired people. This work is open to development with new conditions and different needs.

The application has two profiles (Indoor and outdoor) which is used to reduce the active data set. The user must select these profiles manually. It is planned to make profile transitions automatically by using sensors in the future. Also the application is going to notify the distance and the direction of the detected objects from the user. And custom profile creation is planned be added as a future works.

Project source codes are available with free software licence (GPL) in Github https://github.com/selmantsn/Third-Eye-master

References

[1] *Ünal, E., & Yğce, H. (2017). Development Of Mobile Warning System For The Visually Impaired People. Marmara Fen Bilimleri Journal, 3: 102-110, DOI: 10.7240/marufbd.298380*

[2] Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448).

[3] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems* (pp. 91-99).

[4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[5] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

[6] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp. 21-37). Springer, Cham.

[7] Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. *arXiv preprint*.

[8] *SolidWorks Designs*. From https://github.com/selmantsn/Third-Eye-master

[9] *Rouse M., "Machine learning (ML)" 2018, [Online]. Available:* https://searchenterpriseai.techtarget.com/definition/machine-learning-ML, (accessed on August 30, 2018)

[10] "15 Algorithms Machine Learning Engineers Must Need to Know", *2017, [Online]. Available:* https://www.favouriteblog.com/15-algorithms-machine-learning-engineers/, (accessed on August 30, 2018)

[11] "Introduction to Image Processing", *2018, [Online]. Available:* https://www.engineersgarage.com/articles/image-processing-tutorial-applications, (accessed on August 30, 2018)

[12] Hui J., “Real-time Object Detection with YOLO, YOLOv2 and now YOLOv3”, *2018, [Online]. Available:* https://medium.com/@jonathan_hui/real-time-object-detection-with-yolo-yolov2-28b1b93e2088, (accessed on August 30, 2018)