Answers to the comments for Professor Zoltán Gál

1- General impression

The subject and title of the dissertation are relevant and correspond to the development trends of assisted technologies for helping people with visual impairment (PVI) in their everyday life. Research motivation is based on the obvious requirement of implementing user-friendly end efficient mobile assistive technology in indoor environment to the over 280 million people affected worldwide. Main target of the research work is finding the best usable technology for PVI and build navigation system to avoid physical obstacles improved with convolutional neural network solutions. Results of the research work are usable in practice to enhance mobile assisted technologies and services for PVI. The overall format of both documents makes easy readable each of them. The list of references includes relevant studies related to the topic of the research work. The quantity and quality of them are sufficient and satisfactory.

- First, I would like to thank you for your review and useful feedback. They have provided me with valuable insight to improve the thesis in terms of its content and analysis. Therefore, I have done my best to address all the issues raised.

2- Thesis booklet

This 14 pages document includes introduction of research subject of the PVI navigation in indoor environment, research objectives of the candidate, summary of the theses, application of the scientific results and list of own publications. The target, the structure and the specification of the document contains all the necessary elements.

Chapters 1 and 2 serve as problem description and objectives of the scientific research in satisfactory level of elaboration.

Chapter 3 is listing the theses summary. Radical modifications has been made in this final version of the document offering right presentation form to the scientific assumptions.

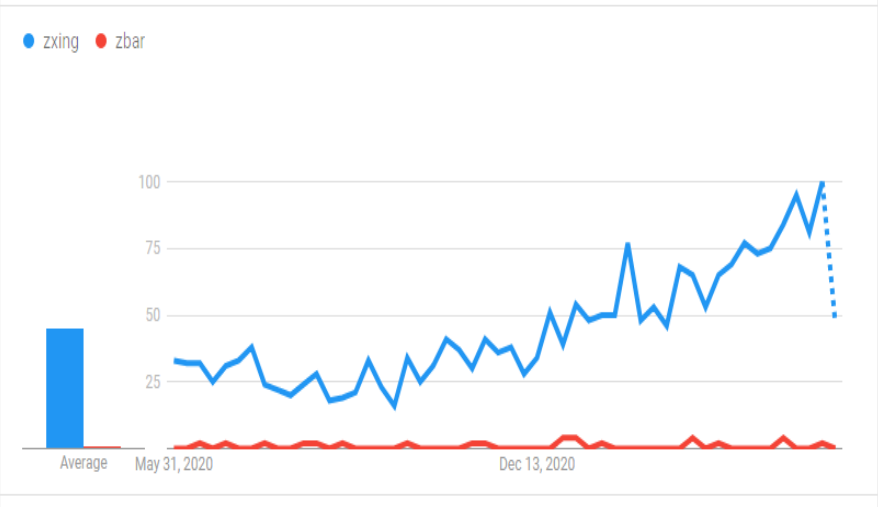
- Thank you for your helpful comment. Based on your previous comments, I modified the thesis booklet to offer the right presentation form to the scientific assumptions.

3- Question to the candidate

Q1- In section 4.1.1 proof of higher performance of Aruco markers than of QR codes is based on Aruco and open-source library (Zxing), respectively. Which considerations guarantee that Zxing method is the best and representative for QR code recognition? What other methods exist for QR code recognition in the literature? What are their performance compared to Zxing method?

- To select Zxing, I found that there are several methods: Zbar. Cppcheck, Google Mobile Vision API. I have selected Zxing because, it returns not only the result of the QR code but also the coordinates of three position detection patterns. The coordinates can be used to calculate the distance to it, while others such as Zbar can't achieve this. This library has been written in java so, it can be easily integrated with android applications. Finally, it was the most popular used

one. The following figure showed a comparison between the usage of Zxing and Zbar where the blue curve is for the used library.



Q2- To reduce the uncertainty and movement time easy help is required to PVI. How the routing algorithm described in subsection 4.1 manages situation when the PVI forgets the source marker ID or gets lost? Would nearest marker be detected and announced by audio feedback "This marker is the closest" or similar with the same meaning during the 360 ° rotation of the uncertain PVI?

- The navigation system stores the first position and destination to be used if PVI is lost. To deal with wrong directions such as going right instead of going left, PVI will find another marker most of the time while walking in this wrong direction as markers cover most places inside the building. In this case, if PVI finds another marker which is in the list of markers to the destination point, the system continues giving navigation commands from this marker to the destination point. If the found marker is not on the list, the system starts to find the shortest path from that marker to the destination point.
- However, failing to detect any marker for some time such as 20 seconds, the system gives feedback to PVI that they are walking in the wrong direction.
- For future work, integrating orientation sensors is needed to warn PVI on time that they turned to the wrong direction.

Q3- It is stated that having eight identical markers at each point of interest, the detection of signals is more accurate. Does the physical size of merged marker matrix or the number of identical markers improves the efficiency? Why?

- Sometimes PVI cannot detect markers because they move their hands a lot and markers move out of the smartphone's camera view or be captured with occlusion. I improved this by installing eight markers with the same id at each interest point instead of adding only one. This implementation makes detection easier as if one of these markers is hidden there is a chance that the other markers can be detected which solved the problem of occlusion and

minimize the chance for the markers to be out of the camera view. It also helps PVI of different heights detect markers easily.

Q4- Convolutional neural network type was applied in the deep learning model during the study. Could be used any other type or neural network to enhance the identification efficiency of the markers? Why?

- Marker identification is considered as a detection problem. Deep learning algorithms for detection can be divided into two categories: two-stage and one-stage. Region-based Convolutional Neural Networks (R-CNNs), Fast R-CNNs, and Faster R-CNNs are two-stage algorithms that use a region proposal network to generate regions of interest in the first stage and propagate the region proposals down the pipeline for object classification and bounding-box regression.
- On the other hand, single-stage detectors such as You Only Look Once (YOLO) and Single Shot Detector (SSD) were proposed to improve the detection efficiency to be suitable for real-time applications by treating object detection as a simple regression problem. They take an input image and learn the class probabilities and bounding box coordinates. Such models reach lower accuracy rates but are much faster than two-stage object detectors. YOLO is a CNN specifically designed for making object detection fast, accurate, and suitable for real-time usage. It uses a single convolutional neural network to predict object categories and find their locations. Several versions of the YOLO model were proposed to improve the accuracy without notable effects on speed.
- We choose YOLOv3 because it can balance the performance on accuracy and processing time well; however, the execution time needs to be improved when using it for real-time applications, especially on smartphones. Tiny-YOLOv3 simplified the original YOLOv3 model by reducing the number of the convolutional layers to be suitable for real-time applications without losing much accuracy

Finally, I want to thank you for your work, valuable comments, and efforts.

Sincerely,

Veszprem, 8 November 2021

Mostafa Abdallah Abbas Atwa Elgendy
PhD student, Faculty of Information Technology,
University of Pannonia, Veszprem, 8200, Hungary