

Computer Vision-based System for Impaired Human Vision Compensation

Povilas Daniušis

Audrius Indriulionis

Andrius Budrionis

Vilnius Gediminas Technical University
Vilnius, Lithuania

Vilnius Gediminas Technical University
Vilnius, Lithuania

Vilnius Gediminas Technical University
Vilnius, Lithuania
Norwegian Centre for E-health Research
University Hospital of North Norway
Troms, Norway

Abstract—In this short article we describe and discuss architecture of computer vision-based system for partial compensation of lost or impaired human vision.

Keywords—Computer vision; Mobile application; Aid for blind and visually impaired; Tactile feedback; Impaired human vision compensation.

I. INTRODUCTION

More than 250 million people worldwide have moderate to severe vision impairment, while (≈ 36 million are blind) [2]. During the past decades significant effort was devoted to develop computer vision and other sensor based aids (e.g. [3], [5], [12], [14]) for helping the blind and visually impaired users to better perceive the surrounding world. However, computer vision is a rapidly evolving field; systems based on the aforementioned approaches become outdated in a relatively short time and often lack accuracy and reliability in real-world conditions in comparison to the state-of-the-art technologies. In this article we describe a system, which employs modern computer vision techniques for compensation of lost or impaired vision function in humans. Many of the previously proposed electronic aids for the blind count on highly specialized hardware, for instance smart glasses [14], Microsoft Kinect sensors [11], helmet-mounted photo sensors and cameras [6]. Such systems often lack convenience and bring additional complexity into daily tasks, which visually impaired people currently manage by the help of a white cane. This project puts major emphasis on developing a tool, which seamlessly integrates with the hardware visually impaired persons already have on-hand and is used to (smartphone). This paper describes a high level architecture of the technology aid for visually impaired people utilising the latest achievements in the field of computer vision.

II. METHOD

Following the principles of the participatory design [7], [4] we included the end-users into the design process of the system. Two visually impaired persons participated in discussion and focus group meetings together with the project team to identify the key properties of the solution. During the six discussion sessions major attention was paid to:

- 1) Functionality and key features of the solution addressing real-life scenarios of visually impaired persons;
- 2) Technical feasibility, connecting user needs to the latest development in computer vision and assisted technologies;

- 3) Interface between the visually impaired person and the assisted technology;
- 4) Appearance, usability and potential costs of the final product.

Design process involved testing of the existing mobile phone app (Seeing AI) for assisting visually impaired persons. Semi-structured tests were performed by the visually impaired participants accompanied by a researcher from a project group. Strengths and weaknesses of the existing solutions are reflected in the design of the proposed system.

III. RESULTS

User-centered design sessions with visually impaired persons provided substantial input for the design of the proposed system (Fig. 1). The process made it clear that understanding of the user needs was limited in the project team. Based on the results of the user-centered design process, we present the outline of the system for impaired human vision compensation.

The key part of the proposed system is a smartphone device, connecting user interface and data processing components into a well-functioning ecosystem. Smartphone was selected due to its high availability and capability to perform required computations or forward visual information for processing to the external server via the Internet connection.

User interface consists of a wearable forehead belt, integrating RGB and depth cameras, tactile feedback device and bone conduction headphones. Although more expensive than standard ones, bone conduction headphones do not block the ears and allow the user to perceive audio information from the surrounding environment, which is essential for the visually impaired people.

External server processes visual data employing a set of computer vision algorithms: faster RCNN object detector [13], trained to detect important objects, CNN-RNN-based scene description [9], place recognition [10], face recognition [1], obstacle detection [8], among others. After completing the analysis results are transmitted back to the smartphone and represented to the user through the aforementioned interface.

Discussions on functionality and key features of the solution revealed a gap between the aims of the existing tools, which are often focused on advanced navigation and scene description features, and actual user needs. For example, the end users highlighted that a simplistic solution for navigating is lacking. Major focus was put on a easy to use and minimalistic tool, which can be trusted. Reliable object detection, direction

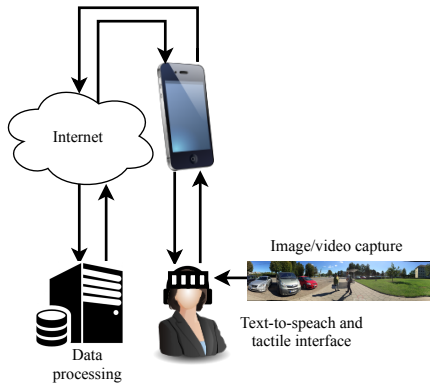


Figure 1. Schematics of vision compensation device.

and approximate distance to an obstacle were highlighted as major requirements. To our surprise, only 5-8 distinctive object classes (for instance, buss stop, pedestrian crossing, doors, stairs, etc) were of major importance while navigating. Additional information, such as type of a passing car, blossoming flowers in the park or a person riding a bicycle was considered as overwhelming and distractive.

Computer vision technology is under rapid development and during the last years made a major breakthrough in performance and efficiency. Deep learning neural networks, which are de facto standard in modern image/video processing, are now well supported by software libraries and dedicated processing hardware is built-in even in relatively low-power computational devices, such as smartphones. These devices are also equipped with 4G Internet connectivity and sufficient computational power to perform initial data processing locally. The maturity of the aforementioned technologies suggest that our proposed computational aid for visually impaired persons may be highly feasible.

The cornerstone of this project is the interface between the visually impaired person and the assistive technology. We are proposing a combination of audio and tactile feedback, which may improve the interaction between the assistive technology and the user. Similar interfaces were suggested by research communities earlier [14], [12], however, existing knowledge on user preferences and evaluation of various options of tactile feedback (placement on the body, frequencies, strength, etc.) is limited.

Appearance, usability and cost of the end product are of major importance to the users. The proposed assistive technology should supplement the tool that visually impaired persons used for decades - the white cane. Using a white cane requires little to none training, has very low costs and is relatively reliable. It provides information on the surrounding objects 1-2 meters in front of the user. While the cane can be used to detect a nearby obstacle, electronic travel aid could be beneficial for longer range (2-10m) route planning and object detection. However, it should integrate seamlessly into the existing navigation practices of visually impaired people.

IV. DISCUSSION

In this article we outlined an idea of complex computer vision-based sensor architecture, which may help to partially

compensate impaired or lost human sight. Main advantages of therein suggested system are: ability to use efficient (but still computationally intensive) modern computer vision algorithms via the Internet connection, present the output of video processing to the user through a combined audio/tactile interface. The suggested system does not aim to replace the main travelling aid of visually impaired people - the white cane. Instead, this computer vision tool aims to push the perception of the surrounding environment boundary from 1-2 meters (achieved by using the white cane) to 2-10 meters. It may improve route planning and identification of objects of interest (for instance, doors, stairs, elevators, bus stops, etc.). This functionality corresponds with the main requirement for electronic travelling aids identified by the end users - direction and distance estimation to a selected object. The proposed hardware architecture can be used as a basis for various computer vision-based software modules, aiming to assist visually impaired users in daily activities (e.g. outdoor/indoor navigation, object/face recognition, obstacle detection, etc.).

The proposed architecture is based on several assumptions, which may be characterized as limitations of the system. Technical feasibility of the solution is dependent on the availability of high bandwidth Internet connection ensuring access to high-power data processing components. Insufficient bandwidth may result in latency, which may not be tolerated by the users.

Economical feasibility of the proposed solution may also be questioned. Advanced technology (high-end smartphone, depth camera, tactile feedback device, bone conduction headphones) is needed to ensure reliable functioning of the system. A combination of such components may be perceived as costly by the end users. More research is needed to demonstrate the benefits of the system in real-world scenarios.

REFERENCES

- [1] Amos, B., Ludwiczuk, B., and Satyanarayanan, M. Openface: A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science, 2016.
- [2] Bourne RRA, Flaxman SR, Braithwaite T, Cicinelli MV, Das A, Jonas JB, et al.; Vision Loss Expert Group. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *Lancet Glob Health*. 2017 Sep;5(9):e888-97.
- [3] Caraiman, S., Morar, A., Owczarek, M., Burlacu, A., Rzeszutarski, D., Botezatu, N., Hergehelegiu, P., Moldoveanu, F., Strumillo, P., Moldoveanu, A. Computer Vision for the Visually Impaired: the Sound of Vision System. *IEEE International Conference on Computer Vision Workshops*, pp. 1480-1489, 2017.
- [4] Carroll, J. M., Rosson, M. B. Participatory design in community informatics. *Des. Stud.*, vol. 28, no. 3, pp. 243-261, May 2007.
- [5] Csapó, A., Wersényi, G., Nagy, H., Stockman, T. A survey of assistive technologies and applications for blind users on mobile platforms: a review and foundation for research. *Journal on Multimodal User Interfaces*. Vol. 9, issue 4, pp. 275-286, 2015.
- [6] Dunai, L. D., Lengua, I. L., Tortajada, I., Brusola Simon, F. Obstacle detectors for visually impaired people. *2014 International Conference on Optimization of Electrical and Electronic Equipment (OPTIM)*, 2014.
- [7] Kensing, F., Simonsen, J., and Bødker, K. MUST: A Method for Participatory Design. *Hum-Comput Interact*, vol. 13, no. 2, pp. 167-198, Jun. 1998.
- [8] Laina, I., Rupprecht, C., Belagiannis, V., Tombari, F., Navab, N. Deeper depth prediction with fully convolutional residual networks. *Fourth International Conference on 3D Vision (3DV)*, pp. 239-248, 2016.
- [9] Liu, Ch., Mao, J., Sha, F., Yuille, A. Attention Correctness in Neural Image Captioning. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pp. 4176-4182, 2017.

- [10] Ohn-Bar, E., Kitani, K., Asakawa, Ch. Personalized Dynamics Models for Adaptive Assistive Navigation Interfaces. arXiv:1804.04118, [cs.LG], 2018.
- [11] Owayjan, M., Hayek, A., Nassrallah, H., Eldor, M. Smart Assistive Navigation System for Blind and Visually Impaired Individuals. 2015 International Conference on Advances in Biomedical Engineering (ICABME), 2015.
- [12] Poggi, M., Mattoccia, S. A wearable mobility aid for the visually impaired based on embedded 3d vision and deep learning. Proceeding of IEEE Symposium on Compututers and Communication, pp. 208-213, 2016.
- [13] Ren, Sh., He, K., Girshick, R., and Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pp. 91-99, 2015.
- [14] Zientara, P.,A., Lee, S., Smith, G., H., Brenner, R., Itti, L., Rosson M., B., Carroll, J., M., Irick K., M., Narayanan, V. Third Eye: A shopping assistant for the visually impaired. Computer Vol. 50, Issue 2, pp. 16-24, 2017.