Computer Vision-based System for Impaired Human Vision Compensation

Povilas Daniušis*, Andrius Budrionis*†, Audrius Indriulionis*‡ and Darius Plikynas*
*Vilnius Gediminas Technical University,
Vilnius, Lithuania.
†Norwegian Centre for E-health Research, University Hospital of North Norway,
Tromsø, Norway.
‡Nature Research Center,
Vilnius, Lithuania.

***Abstract*—This paper presents a high level architecture for computer vision-based system for partial compensation of lost or impaired human vision. It combines standard smartphone device, external deep learning-based image processing infrastructure and audio/tactile user interface. The proposed architecture is based on input from user-centered design process, involving end users into system development. The paper discusses user needs and expectations towards electronic travelling aids for the blind and highlights limitations of the existing solutions. The suggested architecture may be used as a basis for developing computer vision-based tools for visually impaired individuals.**

Keywords–Computer vision; Deep learning; Mobile application; Aid for blind and visually impaired; Audio feedback; Tactile feedback; Impaired human vision compensation; User-centered design.

I. Introduction

More than 250 million people worldwide have moderate to severe vision impairment, while ($\approx$ 36 million are blind) [1]. During the past decades significant effort was devoted to develop computer vision and other sensor based aids (e.g. [2], [3], [4], [5]) for helping the blind and visually impaired users to perceive the surrounding world better. However, computer vision is a rapidly evolving field; systems based on the aforementioned approaches become outdated in a relatively short time and often lack accuracy and reliability in real-world conditions in comparison to the state-of-the-art technologies.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 [5], Microsoft Kinnect sensors [6], helmet-mounted photo sensors and cameras [7]. 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 utilizing the latest achievements in the field of computer vision.

II. Method

Following the principles of the participatory design [8], [9] 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 proposed product.

Design process included testing of the existing mobile phone app (Seeing AI) for assisting visually impaired people. Semi-structured tests were performed by the visually impaired participants accompanied by a researcher from the 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 (UI) 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.

UI consists of a wearable forehead belt, integrating RGB and depth cameras, inertial measurement unit (IMU), 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 and the system at the same time, which is essential for the visually impaired people. Some types of guidance information may be more efficiently perceived through tactile feedback device than audio channel. For instance, indicating the direction of a detected object with a vibromotor on the corresponding area on the forehead belt, instead of providing audio description. More research is needed to clarity how various types of guidance information should be represented to the user in the most convenient and efficient way. USB and Wifi data links can

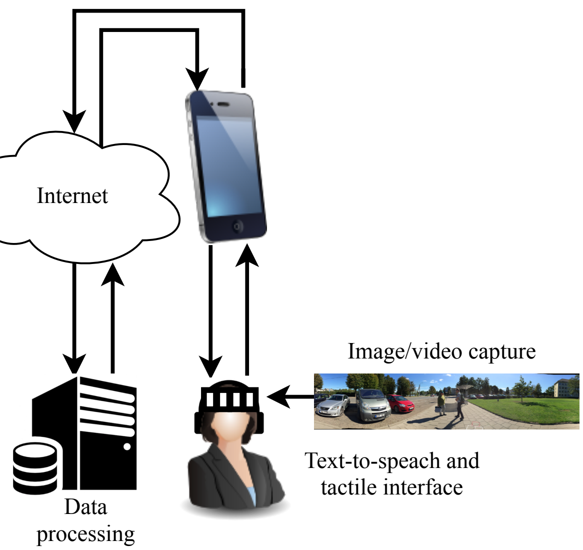

Figure 1. Schematics of vision compensation device.

be used for bidirectional data transmission between UI and smartphone component.

Although modern high-end smartphones are powerful computational devices, many computer vision algorithms require hardware of significantly higher processing power (e.g. a server equipped with a high-end graphics processing unit (GPU)). Therefore, an external server is used for live vision data processing. In our design we envision a computational server hosting a set of computer vision algorithms implemented and exposed via web services: faster regions with convolutional neural network features (faster R-CNN) object detector [10], trained to detect important objects, convolutional and recurrent neural network-based (CNN and RNN) scene descriptor [11], which provides textual annotation of a given RGB image, place recognition [12], face detection and recognition [13], obstacle detection [14], action recognition [15], among others. After video analysis is completed on the server side, results are transmitted back to the smartphone component and represented to the user through the aforementioned UI.

The user interacts with the system through gestures, voice commands or specifically designed user interface on the smartphone. User input is required for selecting operation mode of the system, i.e. switching between optical character recognition (OCR), currency recognition, object detection, navigation aid and other available functions.

Social networking functionality was highlighted as a potentially important feature allowing other users/volunteers contribute to up-to-date and high quality guidance (for instance, temporary road works) with exact location and description. Navigating around such obstacles can be challenging for computer vision-based tools and may decrease user's trust in the tool. Integration of additional information from other users, governmental organizations and volunteers could be of high value for the visually impaired people.

Discussions on the key features of the proposed solution revealed a gap between the functionality 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 an easy-to-use and minimalistic tool, which can be trusted. Reliable object detection, direction and approximate distance to an obstacle were highlighted as the main 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 in many real-world problems allow to achieve accuracies, similar to that of human decision (e.g. face recognition [13], object detection [10], among others). These models are now well supported by software libraries (e.g. [16]) 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 partial or in certain cases even full vision 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 [5], [4], however, existing knowledge on user preferences and evaluation of various options of tactile feedback (types of actuators, 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 no 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 nearby obstacles, 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

This paper outlined a high level architecture of a computer vision-based system, which may help to partially compensate impaired or lost human vision. The main advantages of therein suggested system are: ability to use efficient (but still computationally intensive) modern computer vision algorithms via the Internet connection and present the output of the video processing to the user through a combined audio/tactile interface.

Similar functionality has been previously addressed by research communities [2], [3], [4], [5] and industry [17], [18]. The main difference between the aforementioned commercial solutions and the proposed system is the ability to utilize external resources for computationally intensive image or video processing tasks. While the availability of the computational power could be seen as the main advantage, it comes with high cost - dependency on a well-functioning mobile broadband. The development of the high-speed mobile networks (4G and 5G in the nearest future) is likely to make the major limitation of the proposed architecture obsolete, especially in densely populated areas containing many hazards for the blind. On the other side, increasing computational capacity of smartphones also may allow to conduct more advanced image or video processing locally. Combined audio/tactile interface may also be more convenient for the users, allowing to provide feedback in a more efficient way than using tactile or audio interfaces separately.

Both open source (e.g. [16]) and state of the art commercial computer vision software libraries (e.g. [19]) may be applied implementing the suggested architecture. Moreover, detalization and specification of the suggested system's hardware components following open hardware approach - by providing open repository with the detailed list of electronics components, schematics, and 3D models of mechanical parts, could provide a solid basis for semi-standard reference platform for the researchers in the field.

It is important to emphasize that 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 enhance and 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 highlighted by the end users - direction and distance estimation to a selected object. The proposed hardware architecture can be used as 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.

Acknowledgment

This research is/was funded by the European Regional Development Fund according to the supported activity Research Projects Implemented by World-class Researcher Groups under Measure No. 01.2.2-LMT-K-718.

References

[1] 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.

[2] Caraiman, S., Morar, A., Owczarek, M.,Burlacu, A., Rzeszotarski, D., Botezatu, N., Herghelegiu, 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] Carroll, J. M., Rosson, M. B. Participatory design in community informatics. Des. Stud., vol. 28, no. 3, pp. 243-261, May 2007.

[10] 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.

[11] 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.

[12] Ohn-Bar, E., Kitani, K., Asakawa, Ch. Personalized Dynamics Models for Adaptive Assistive Navigation Interfaces. arXiv:1804.04118, [cs.LG], 2018.

[13] 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.

[14] 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.

[15] Varol, G., Laptev, I., Schmid, C. Long-Term Temporal Convolutions for Action Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 40, issue 6, pp. 1510 - 1517.

[16] Abadi M., Agarwal A., Barham P., Brevdo E., Chen Z., Citro C., Corrado G., Davis A., Dean J., Devin M., et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2016. arXiv:1603.04467

[17] Orcam system, https://www.orcam.com. [2018-10-16].

[18] Horus system, http://www.horus.tech. [2018-10-16].

[19] Neurotechnology, VeriLook SDK. https://www.neurotechnology.com/verilook.html. [2018-10-16].