Concept of Navigation System for Blind People using Image Recognition and Classification

D. Dirvanauskas[1], F. Orujov[2], R. Maskeliūnas
Multimedia Engineering Department, Kaunas University of Technology, Lithuania
[1]*E-mail: darius.dirvanauskas@ktu.edu*
[2]*E-mail: farid.orujov@ktu.edu*

Introduction. Nowadays, nearly 4% of the world's population have problems with eyesight. Statistics from the World Health Organization shows that about 285 million people suffer from visual impairment, of whom 39 million are blind and 246 million have low vision [1]. Approximately 90% of people suffering from visual impairment live in low-income countries. Globally, a major cause of visual impairment is uncorrected refractive error. In the middle and low-income countries, the main cause of blindness is cataract.

Around 65% of people suffering from visual impairment are aged 50 years and older; this age group accounts for 20% of the world's population. With the elderly population growth in many countries, a growing number of people will be at risk of visual impairment because of chronic eye diseases and aging process.

In this article, we consider the concept that should ensure that blind people have an active social life indoors. The concept of our system consists of hardware and software modules that provide the navigation for a blind person. We will analyse the wearable devices and data analysis techniques for photographic images. The convolutional neural network is used to determine the important information using image analysis. It should be noted that this paper focuses on methods of photographic images analysis and does not address to the power of the signal coming from other sensors.

The wearable devices. A range of different wearable devices were developed for different daily tasks. For a better interaction with the users of IT systems one visual interaction is not enough, so large companies such as Google, Microsoft and others are working towards the empowerment of smart devices for full interaction with users. Concepts have been developed, such as the Google Glasses.

Figure 1 show the wearable devices which can be used as sensors for visually impaired people. The Google Glasses is sensor with camera and audio navigation interface [2]. Bracelets, vests and belts serves as a tactile navigation interface for non-sighted person. Smart shoes provide user support when walking, identifies obstacles in the way. Smartphone performs as Control Center for these devices. We will consider photographic image obtained by using Google Glasses or smartphone as a data source.

These photos will be used by algorithms, results of which will the actions that need to be performed by the user of the system to reach their destination.

Since the user is blind, the navigation interface is not visual. Only the audio and tactile commands can be used to transfer information to user.

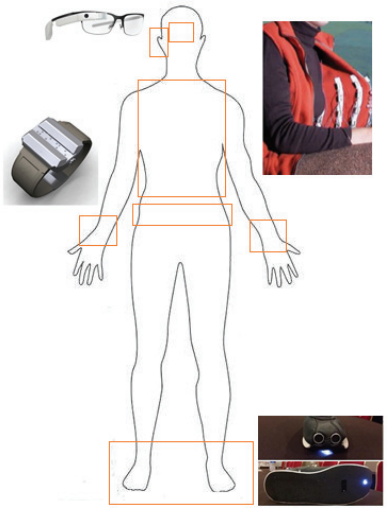

Fig. 1. Overview of wearable devices for the blind persons

It should be noted that recent research [3] have shown that 20-30 minutes of music / speech sound leads to degradation of human response, reduces the person's ability to perform normal tasks and affects posture and balance [4]. This gives the advantage of a tactile interface to the audio interface; it should also be noted that the tactile interface is universal for all users regardless of the language of communication.

The software. In this article, we offer algorithm which will help blind people to find out what is around them. Users using their mobile phone application with camera will capture visual data which will be automatically analysed. Algorithm will process images which were taken and will return answer with important information to user through headphone interface. During image processing, artificial intelligence algorithm will try to recognize important image features. After that detected features will be transferred to classification algorithm, there they will be assigned to most similar objects class. At detected features fusion part, algorithm will analyse objects classes, which there recognized in the image, and locate their place according to the reference point. Main classes which could help people to navigate could be: people, doors, windows, desk, chair, cabinets, stairs and other important objects (Fig. 2).

To do image recognition and classification researchers use variety of machine learning methods: Support vector machine, Naive Bayes, Nearest Neighbour, Decision Tree, Neural Network and others. For our research project, we decided to use Convolutional Neural Network.

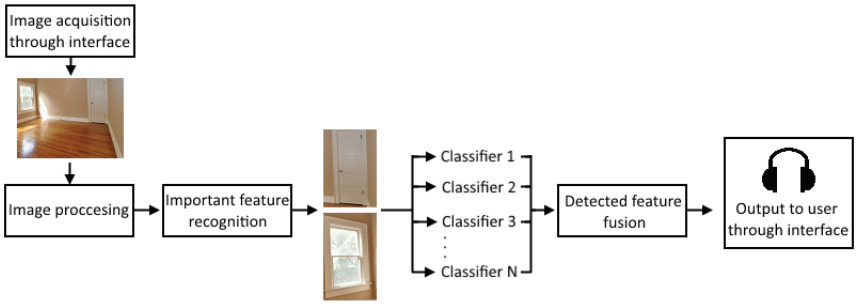

Fig. 2. Algorithm for images analysis

CNN is a biologically inspired class of deep learning models with a single neural network that is trained end to end from raw pixel values to classifier outputs. Neural network consists of layers which have a set of interconnected nodes. CNN combines learned features with input data and uses a 2D convolutional layer that makes this architecture well suited for processing 2D data, such as images.

Convolutional Neural Network model accuracy highly depends on the amount of images used at training stage. To get model which precisely recognize images it is needed to use thousands or even millions of classified training samples. The number of data creates two problems: small number of samples at model training dataset and slow work of CPU processor. The training stage of Convolutional Neural Network lasted for days or even weeks. But from now on these problems are solved. It is possible to reduce network training stage duration from days to hours by transferring calculations from CPU to GPU processor. Now it is possible to get big number of classified data from open source database. ImageNet is one of the biggest databases for researches to get training data for their models. It has about 14 millions of high resolution images which are classified to almost 22 thousands categories. Images are collected from various web pages and classified using Amazon Mechanical Turk application.

Researcher's studies show that it is possible to reach significant results using Convolutional Neural Network for classifying different type and amount of data. ImageNet dataset with 8.9 million images classified to 10.184 classes was used by researchers to investigate CNN accuracy. Half of dataset was used for training and other half for testing. After classification CNN showed 67.4% accuracy [5]. Other research group used two smaller dataset's which consisted of 9.963 and 22.531 images classified to 20 classes. After classification, researchers got 87.26% and 81.16% accuracy results [6]. It shows that it is better to classify images to fewer classes to get better results. And it is not necessary to use big dataset to get good classification results. Convolutional Neural Network usage for video classification was tested by Google research group. They used 1 million YouTube video clips classified to 487 groups,

which consisted of 1000-3000 videos. 80% of database was used for CNN training and 20% for testing. CNN training took a month and after testing it reached 80.2% accuracy [7]. This shows that it is possible to get good results using Convolutional Neural Network not only for image classification but for video classification too.

Conclusions. This paper considers the technology that allows helping people with vision problems. It offers new algorithm with Convolutional Neural Network for images recognition and classification. We concluded that it is necessary to use a hybrid method. The auxiliary telemetry and tactile output devices are better to use for feedback in continuous operation of the system for patient comfort.

Acknowledgement. This research has been supported by European cooperation in science and technology COST Action CA15122, Reducing Old-Age Social Exclusion: Collaborations in Research and Policy (ROSEnet).

References

1. World Health Organization (2014) Visual impairment and blindness – Fact Sheet N°282. [Available online at] http://www.who.int/mediacentre/factsheets/fs282/en/

2. Al-Khalifa Sh., Al-Razgan M., Ebsar: Indoor guidance for the visually impaired, Computers and Electrical Engineering, 2016, pp 26-39.

3. Velázquez R., Wearable Assistive Devices for the Blind. Chapter 17 in A. Lay-Ekuakille & S.C. Mukhopadhyay (Eds.), Wearable and Autonomous Biomedical Devices and Systems for Smart Environment: Issues and Characterization, LNEE 75, Springer, pp 331-349, 2010.

4. Hakkinen J., Vuori T., Paakka M., Postural stability and sickness symptoms after HMD use, Systems, Man and Cybernetics, 2002 IEEE International Conference, vol 1, pp 147 - 152, 2002.

5. Karpathy A., Toderici G., Shetty S., Leung T., Sukthankar R., and Fei-Fei L., "Large-Scale Video Classification with Convolutional Neural Networks," in *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1725–1732.

6. Krizhevsky A., Sutskever I., and Hinton G. E., "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

7. Li Y., Liu L., Shen C., and A. van den Hengel, "Mining Mid-level Visual Patterns with Deep CNN Activations," *Int. J. Comput. Vis.*, pp. 1–21, 2016.

Concept of Navigation System for Blind People using Image Recognition and Classification

D. Dirvanauskas, F. Orujev, R. Maskeliūnas
Department of Multimedia Engineering, Kaunas University of Technology, Lithuania

This paper discusses enabling technologies to help blind people. The work focuses on Image Recognition and Classification methods for determining the important information by analysing photographic images. The article provides system architecture and implementation details of the images analysis system.