JOURNAL OF MEDICAL INFORMATICS & TECHNOLOGIES Vol. 21/2012, ISSN 1642-6037

computer vision algorithms, blind people support,
threat detection, stereovision

Michał CHORAŚ[1], Rafał KOZIK[1]

THE PRELIMINARY RESULTS OF THE INTELLIGENT COMPUTER VISION SYSTEM TO SUPPORT BLIND PEOPLE

In this paper we present innovative computer vision solutions to support social inclusion of totally blind people. We present the overall framework architecture and we focus on intelligent computer vision system applied to support Instrumental Activities of Daily Living (IADL). Results of the proposed intelligent computer vision system are reported and presented on the basis of sample real-life scenarios.

1. INTRODUCTION

Nowadays, it is extremely difficult for totally blind and visually impaired people to live a normal life in terms of interacting with the society. Most often, such people are excluded and they interact only in the special well-known environment (e.g. special schools). Although various technological solutions have been proposed, none of them clearly addresses assistance for daily activities and social inclusion of totally blind or visually impaired people. Currently available solutions focus mainly on one specific kind of support, e.g., navigation. Partial or full visual impairment has been reported as being a huge factor for social exclusion. For example, the ratio of visually impaired and totally blind people currently employed (and actively working) is significantly low. According to [1][2], ca. 314 million people are visually impaired worldwide, among them 45 million are blind and an average of 3% of Europe's inhabitants experience a loss of sight.

In Europe the average unemployment rate of blind and visually impaired people in the working population is over 75% [3]. However, those blind or visually impaired people who are working are employed in jobs with lower level of qualifications required and lower salaries. There are several factors contributing to this problem such as: inadequate re-habilitation plans; lack of self motivation caused by high unemployment rate and limited government support; lack or limited access to schools adapted for the blind; limited access and high costs of rehabilitation facilities; limited number of facilities adapted to blind person. All those problems yield social exclusion. The problem is even bigger, since in many European countries public facilities (bus or train stations, museums/cinemas, government buildings) are not adapted to blind people. The Eurostat report [4] shows that access for people with visual disabilities to services such as public transport, sport events, workplaces, universities and schools, restaurants and cultural events is significantly more difficult than for people with other types of impairments (e.g. deaf, physically or intellectually disabled people). According to this report, visually disabled people perceive access to buildings, services and events in Europe as fairly difficult to very difficult. Furthermore, traditional means for supporting people suffering from vision impairment like guide dogs or echolocation-based canes are still hardly accessible. For example, in case of guide dogs there is a huge amount of time required for both the animal and human to cooperate and understand each other. Nowadays, only small percentage of blind people have such well-trained dogs (e.g. in Poland according to data obtained from Polish Society of the Blind People only 3% of the totally blind people over 16 years of age have guide dogs).

Therefore, our general goal is to provide a multi sensor-based, distributed service framework for visually impaired and totally blind people inclusion through daily assistance (e.g. travelling aid) and executing everyday duties.

[1] Institute of Telecommunications, University of Technology and Life Sciences, Bydgoszcz,
email:{chorasm, rafal.kozik}@utp.edu.pl.

In this paper we focus on an intelligent computer vision and pattern recognition system. We present the results of the intelligent computer vision system applied to support blind people in Instrumental Activities of Daily Living (IADL) and support their inclusion into society.

The paper is structured as follows: in Section 2 the general architecture of the overall blind people support framework is presented and the role of computer vision system is indicated. In Section 3 the innovations in computer vision applied to support blind people are overviewed. In Section 4 the sample real life scenarios and results of the computer vision system are presented. Conclusions are given thereafter.

2. THE PROPOSED FRAMEWORK ARCHITECTURE

In order to facilitate blind people inclusion we propose the general architecture presented in Figure 1. Obviously, an intelligent computer vision system is only a part of the proposed framework.

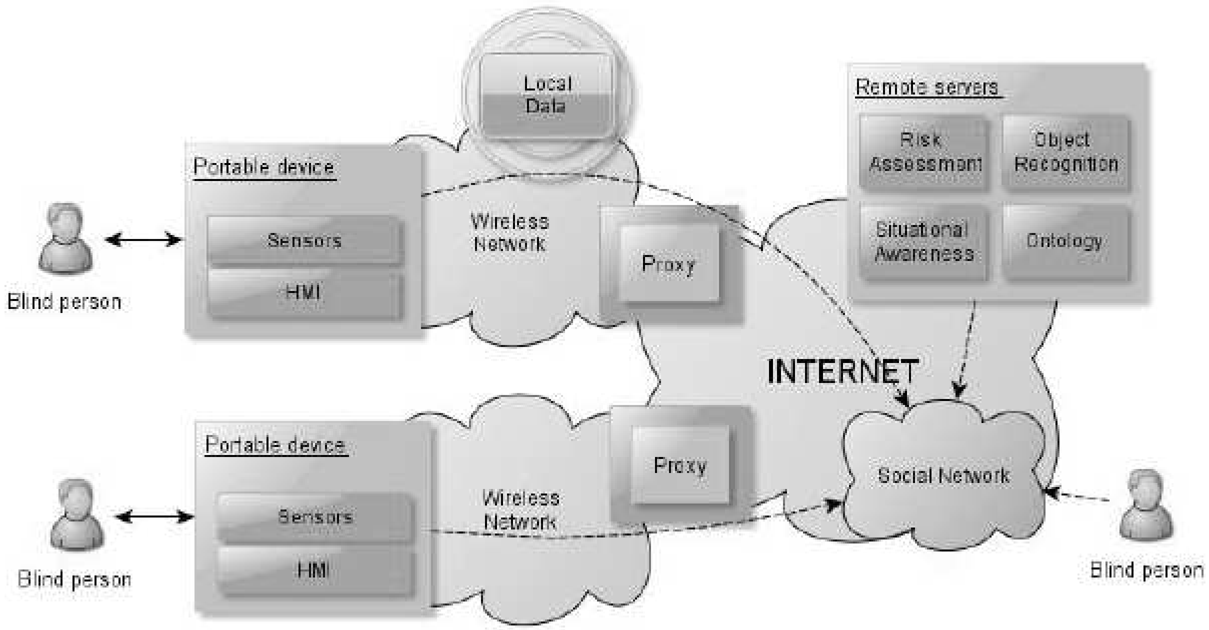

Fig. 1 General architecture of the proposed framework.

The proposed architecture follows the cloud concept allowing the blind users to share experience and information through social network. The network will allow the blind person to post information about particular dangerous or friendly places or past experiences to the community of connected people. Remote servers will be capable of handling such data. Moreover, portable device will also read the "Local Data", which will be broadcasted by remote devices (vending machines, dispensers, etc.) via radio/wireless channels.

We advance the state of the art, for example, using the "guide me to "-service, which guides a blind person from one point to another detecting and helping to avoid obstacles. The blind person will be able to select a destination point (e.g. rehabilitation and training centre) and the system will plan the shortest route from the current location by using GPS localization capabilities. The proposed solution will be capable of avoiding the nearest and colliding objects (detected by computer vision algorithms). Having the information of context (GPS, location, objects recognized by computer vision algorithms) further detectors will be used (like walking path edge tracking, fast movements detectors, etc.) contributing to increasing blind person situational awareness.

During the travel, the proposed solution will find the nearest bus or train station, analyze the timetable, recognize a bus number and find the entrance. In an environment unknown for the blind person, the proposed system will complement guide dog duties, like finding the free seats or exits. Moreover, we propose a self-adaptive, learning approach. The blind user might ask the system for guidance to the close by bakery and the system will propose the optimal route and help avoiding obstacles (e.g., a construction site). Our proposed innovation is, with the use of a social network, to learn that the construction works started recently and to advice the blind user to choose another bakery due to safety reasons (e.g., because of the construction site it would be necessary to cross a road with lots of traffic at this time). We will provide innovative ICT solutions to facilitate social inclusion of the blind on several layers:

- Sensing layer, based on intelligent cameras (computer vision algorithms) and other sensors integrated in the dedicated harness,
- Communication layer, based on resilient wireless personal telecommunication,
- Risk assessment and situational awareness layer, based on semantic notations, correlation and reasoning algorithms,
- Interaction layer for the end user, based on innovative non-visual interfaces. In this paper we focus on the intelligent computer vision system which is a part of the sensing and risk assessment/situational awareness layer.

3. INNOVATIONS IN COMPUTER VISION TO SUPPORT BLIND PEOPLE INCLUSION

Much research has been done on the deployment of computer systems supporting visually impaired people in their interaction with the environment. Most of these solutions are based on computer vision techniques and they can be categorized in two distinct groups: solutions relying on a single video sensor (SVS) and, solutions based on multiple video sensors (MVS). SVS and MVS share similar approaches in data pre-processing: filtering and low level segmentation are usually performed on every image composing the video under analysis. Filtering aims at enhancing images and removing distortion and noise, while low level segmentation is designed to roughly extracts objects from images, classifying image's pixel in two groups: foreground (pixels of interest) and back-ground. Typical low level segmentation algorithms are designed to be fast in order to promptly provide data to users or to other systems for further processing. Based on the specific task, data pre-processing is performed on a frame basis or on a sequence basis. Sequence basis elaboration is more processing and memory demanding, but it is usually more reliable.

After segmentation, post-processing is common for both SVS and MVS approaches. Post-processing usually involves computer vision and machine learning algorithms. Features (e.g. textures or shapes) are extracted from segmented images and correlated with some a-priori knowledge in order to identify, classify or detect particular objects. Typical single-object detectors allow for objects recognition such as door [13], pedestrians [14] or staircase [15]. Their implementation is based on efficient real-time computer vision algorithms (e.g. Haar-like features and rapid AdaBoosted cascades of filters [16][17]) viable for smartphones implementation. The adoption of MVS is motivated by the ability of obtaining additional information on the depth of the observed scene. Indeed, stereo matching and simple triangulation allows for measuring the distance to a given object. This may allow for applications that help people in reaching objects and in avoiding obstacles while moving in unknown environments [18]. Moreover, in multi-sensor imaging the same object is detected from multiple sources and this increases system reliability. However, processing multiple video streams has computational burden that may make this solution unviable for portable devices. To identify or detect a wide range of objects, and to support MVS typical smartphone's capabilities are insufficient. Therefore in such systems a client-server architecture is required. Usually, the mobile is used for video streaming and for simple computer vision tasks with tight real-time constraints, while the remote server performs computationally expensive and delay tolerant tasks. The major limitation of such systems is the data processing model. Indeed, the mobile and the server follow light cooperation patterns: the processing steps are split among them and completed independently. The mobile is in charge to dispatch processing task and to collect and present the results. The server provides reliable on demand services for processing images or image sequences. This solution is effective in most prototypical cases. However, this processing model is too stiff in most practical situations, and especially when serving people with visual impairments [19].

In real scenarios, computer vision applications need to be fast and reliable and in their implementation the developers need to solve a trade-off between these two requirements. The tuning between speed and reliability is difficult and it requires a close match with the expectations of users. In the context we consider, in order to fit the requirements of blind people in a best possible way, computer vision applications need to be modular, distributed, progressive and proactive. These requirements are strictly related one to another. Distributed solutions allow for supporting both delay sensitive and reliability sensitive applications. However, good distributed solutions need a clear modular organization in processing and communications units. In order to be fast and reliable in data processing, the same task can be completed concurrently by two modules, the former providing an approximate solutions for fast feedback, while the latter yielding a more reliable information that can be merged, later, with the previously presented one (progressive processing). Eventually, application responsiveness can benefit from proactive processing, forecasting next users requests and pre-computing results that the user may be willing to know.

The services that are enabled by the proposed computer vision system are listed below:

- Localization (Standard rough GPS position estimated by mobile or smart phone and reinforced by computer vision and prior knowledge),
- Contextual localization (Google maps or other map-based information, Computer vision based recognition and understanding),
- Obstacle/Threat detection (Computer vision based recognition and understanding connected to ontology),
- People counting/People behaviour characterization (Computer vision),
- Street text recognition (Computer vision connected with ontology),
- Risk analysis (Situational awareness, ontology),
- Route planning (GPS, maps, route planning applications, social networking, self-adaptation),
- Finding places and items (Computer vision connected with ontology, situational awareness),
- Empty seats/tables finder (Computer vision),
- Adaptation system-to-user (Historical user experiences and decisions connected with knowledge in ontology),
- Guiding/warning (Non-visual stimulation of user via dedicated actuators). In the next section, some demonstration cases and regarding some of the described scenarios and services are presented.

4. RESULTS

In this section the following scenarios and services are demonstrated and described: free seat detection, door detection, navigation in unknown place and threat detection. We present the results using real-life situations, also in adverse conditions.

4.1. COMPUTER VISION-BASED FREE SEAT DETECTOR

Finding a free seating place in an environment unknown for blind person is a challenging task. In this scenario a blind person enters the room equipped with the proposed system. Two places are suggested by the system. A beeping sound is assigned (and/or special vibrations are used) to these two seat-candidates and the blind person can easily locate the objects. The result of the computer vision system detecting free seats is presented in Figure 2.

Fig. 2. Result of properly detected seats.

4.2. COMPUTER VISION-BASED FREE SEAT DETECTOR (ONE IS OCCUPIED)

This scenario is a modification of the previous one. Now one of the seats is occupied. As the result, only one seat is suggested by the system. The occupied one is not signalized. The armchair standing on the left side is signalized as free. The result of the computer vision system detecting free seat with no false positive (occupied seat) is presented in Figure 3.

Fig. 3. Result of properly detected seat (and no occupied seat detected).

4.3. DOOR DETECTION (ADVERSE CONDITIONS)

Locating appropriate door in a room is also a challenge for a blind person. The task is even more complicated when the blind person is moving across the hall. Without our system blind person is forced to count the doors. However, as it can be noticed (see Fig. 4) there are also notice boards in the hall, which look like doors. Recognizing the notice board without the prior knowledge could be difficult, since door and board are of the same height and width and have characteristic wooden edge.

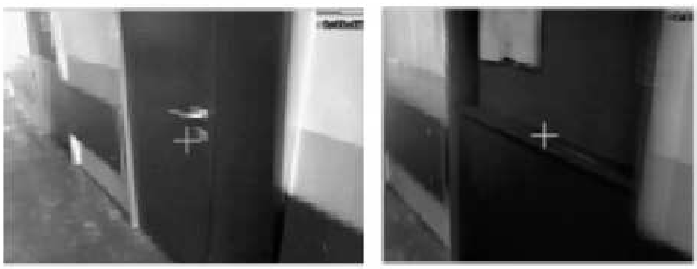

Fig. 4. Deference between notice board and door office is small and hard to distinguish with touch.

This scenario aims at showing that the system can easily recognize the doors with computer vision techniques. Moreover, the system can count the doors itself and perform camera tracking for accurate camera position localization. At first, a door is detected. It can be noticed that door is detected even if the image is slightly blurred due to the camera motion. The notice board is not classified as door (neither from far nor from near). The result of the computer vision system detecting doors is presented in Figure 5, while no errors (no false positives) are presented in Figure 6.

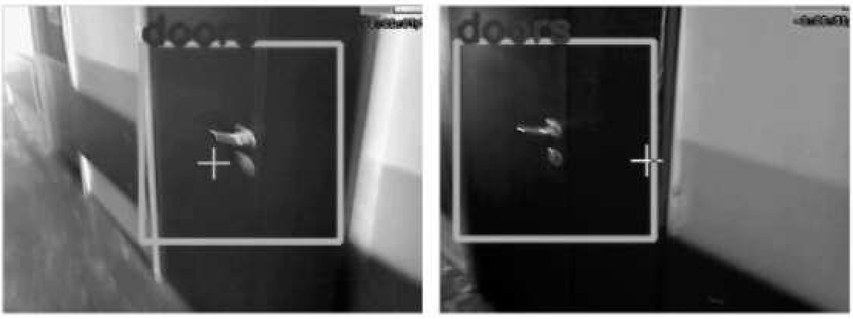

Fig. 5. Examples of properly detected doors.

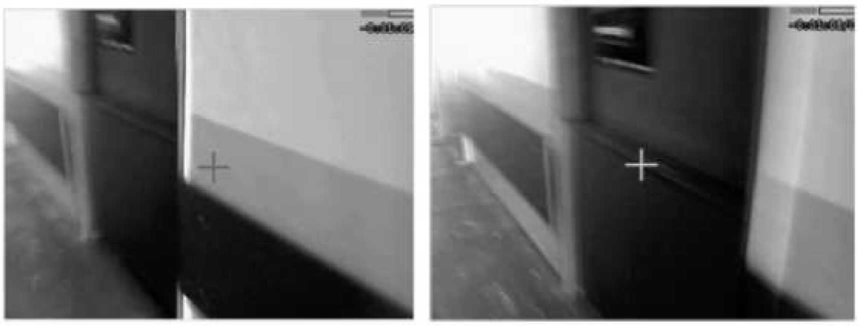

Fig. 6. No doors signalized by the system.

4.4. NAVIGATION IN UNKNOWN PLACE

Unknown places can be very stressful for the blind person. However, many of them such as streets, hotel rooms, training centers, shopping malls etc. can be easily annotated and semantically described. Moreover, taking advantages of social networks both blind persons and their relatives can post an opinion about particular place, threats etc. The following scenario demonstrates how the proposed system can help the blind person to move around in an unknown environment, an example is a hotel room where a blind person spends some time during the training. Commonly there is a guide or trainer, who takes a responsibility of showing to a blind person a localization of particular objects and explaining how to move around. With proposed solution a blind person can explore the apartment on his/her own in a shorter time. Sample result of the computer vision system detecting sink is presented in Figure 7.

Fig. 7. The sink recognized by the system.

4.5. THREAT DETECTION

Threat detection slightly differs from object recognition cases described in the previous demos. First of all, it is more critical and requires a special approach. If a dangerous object is not recognized soon enough, it may expose a blind person to serious injuries. Moreover, threat detector requires more sophisticated communication between server and local machine. Particularly, the system must upload small set of dangerous objects that are likely to appear in the location where the blind person moves. Such detection schema allows to preserve main system functionalities in case when the communication channel (due to some technical reason) is unavailable or slows down. This scenario demonstrates some examples of dangerous object detection capabilities. Scenario concerns indoor office environment where the kettle is an example of dangerous object. In Figure 8 (upper row) the result of kettle detection is shown. It is also classified as dangerous object in the threat ontology. Warning message is played. Using depth map it can be inferred that the object is standing on the corner of a table. This situation may be assigned high severity value since it is more likely that blind person may spill the content moving towards the obstacle. The situation presented in Figure 8 (bottom row) shows kettle detection, but the lower severity value is assigned since the kettle is recognized together with a shelf, where the kettle is supposed to be kept. Hereby, the computer vision system uses the context knowledge formalized in the ontology (which is out of scope of this paper). Therefore adjusting the system threshold for warning generation the blind person might be informed only about severe threats if needed.

Fig. 8. Objects detected and signalized by the system.

5. CONCLUSIONS

In this paper we proposed the general framework architecture and, in particular, we focused on an intelligent computer vision system to support totally blind people. In our consecutive stages we tried to fulfill end-users requirements and expectations and therefore we tested the system in real-life (sometimes adverse) conditions. Therefore, some sample scenarios and results of the proposed intelligent computer vision system in operation are presented.

6. ACKNOWLEDGEMENT

We would like to thank end-users for their help and suggestions to our ideas, especially to Polish Society of the Blind in Bydgoszcz and to Association for the Welfare of the Deafblind in Warszawa.

BIBLIOGRAPHY

[1] World Health Organization, WHO Fact Sheet N282, May 2009, http://www.who.int/mediacentre/factsheets/fs282/en.
[2] European Blind Union (EBU), Key facts and Figures concerning blindness and sight loss, http://www.euroblind.org/resources/information/nr/215.
[3] Action For Blind People, Facts and Figures about issues around sight loss, http://www.actionforblindpeople.org.uk/news/media-centre/facts-and-figures,893,SA.html.
[4] Eurobarometer 54.2, European Commission, 2001.
[5] http://www.mobileaccessibility.info/findphones-results.cfm.
[6] http://www.codefactory.es/en/products.asp?id=326.
[7] http://www.rnib.org.uk/shop/Pages/ProductDetails.aspx category=phones productID=HM3501.
[8] http://www.loadstone-gps.com.
[9] http://www.ubergizmo.com/15/archives/2009/08/braille phone concept.html.
[10] http://www.defense-update.com/products/f/fw-atd.htm.
[11] http://telecom.esa.int/telecom/www/object/index.cfm?fobjectid=12843.
[12] http://casblipdif.webs.upv.es.
[13] CHEN Z., BIRCHFIELD S., Visual Detection of Lintel-Occluded Doors from a Single Image, IEEE Computer Society Workshop on Visual Localization for Mobile Platforms (in association with CVPR), Anchorage, Alaska, June 2008.
[14] GRAY D., TAO H., Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features, ECCV'08, Proceedings of the 10th European Conference on Computer Vision, 2008.
[15] SISONG W., HAN W., 2D staircase detection using real adaboost, Proceedings of the 7th international conference on Information, communications and signal processing, Macau, China, IEEE Press, 2009, pp. 376-380p.
[16] VIOLA P., Jones M., Rapid object detection using a boosted cascade of simple features, In Proceedings IEEE Conf. on Computer Vision and Pattern Recognition, 2001.
[17] CHEN X., YUILLE A.L., A Time-Efficient Cascade for Real Time Object Detection. 1st International Workshop on Computer Vision Applications for the Visually Impaired, In association with CVPR, June, 2005.
[18] STRUMILLO P., PELCZYNSKI P., BUJACZ M., PEC M., Space perception by means of acoustic images: an electronic travel aid for the blind, ACOUSTICS High Tatras 06 - 33rd International Acoustical Conference - EAA Symposium, Strbske Pleso, Slovakia, October 4th - 6th, 2006, pp. 296-299.
[19] SUDOL J., DIALAMEH O., BLANCHARD C., Dorcey T., LookTel - A Comprehensive Platform for Computer-Aided Visual Assistance, IPPLEX, LLC Santa Monica, CA.