Expert Systems With Applications 46 (2016) 129–138

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Recovering the sight to blind people in indoor environments with smart technologies

Mohamed L. Mekhalfi[a], Farid Melgani[a,*], Abdallah Zeggada[a], Francesco G.B. De Natale[a], Mohammed A.-M. Salem[b], Alaa Khamis[c]

[a] Department of Information Engineering and Computer Science, University of Trento, Via Sommarive, 9, I-38123 Trento, Italy
[b] Department of Scientific Computing, Ain Shams University, Egypt
[c] Vestec Inc. Canada and Department of Engineering Science, Suez University, Egypt

ARTICLE INFO

Keywords:
Assistive technologies
Blind people
Computer vision
Indoor navigation
Machine learning
Multiobject recognition
Real time processing

ABSTRACT

Assistive technologies for blind people are showing a fast growth, providing useful tools to support daily activities and to improve social inclusion. Most of these technologies are mainly focused on helping blind people to navigate and avoid obstacles. Other works emphasize on providing them assistance to recognize their surrounding objects. Very few of them however couple both aspects (i.e., navigation and recognition). With the aim to address the aforesaid needs, we describe in this paper an innovative prototype, which offers the capabilities to (i) move autonomously and to (ii) recognize multiple objects in public indoor environments. It incorporates lightweight hardware components (camera, IMU, and laser sensors), all mounted on a reasonably-sized integrated device to be placed on the chest. It requires the indoor environment to be 'blind-friendly', i.e., prior information about it should be prepared and loaded in the system beforehand. Its algorithms are mainly based on advanced computer vision and machine learning approaches. The interaction between the user and the system is performed through speech recognition and synthesis modules. The prototype offers to the user the possibility to (i) walk across the site to reach the desired destination, avoiding static and mobile obstacles, and (ii) ask the system through vocal interaction to list the prominent objects in the user's field of view. We illustrate the performances of the proposed prototype through experiments conducted in a blind-friendly indoor space equipped at our Department premises.

1. Introduction

Visual disability is one of the most serious troubles that may afflict an individual. Despite the fact that 80% of all visual impairment is claimed to be preventable and even curable, still blindness/partial-sight represent a serious problem worldwide (WHO). Besides the great efforts spent in medicine, neuro-science and biotechnologies to find an ultimate solution to such problems, technologies can provide tools to support those people by providing basic functionalities such as the ability to navigate and recognize their entourage independently, to improve their quality of life and allow better integration into the society. This objective is ambitious but not out-of-reach, thanks to the recent technological advances.

In pursuit of satisfying the needs of visually disabled people and promote better conditions for them, several designs have been put forth in the last years. From an overall perspective, they can be framed into two mainstreams. The former addresses the guidance/navigation concern, while affording the possibility to avoid potential obstacles. The latter is focused on recognizing the nature of nearby moving/static obstacles. Considering both aspects, various contributions have been suggested, often referred to as electronic travel aids (ETAs) (Capi & Toda, 2011; Chen, Li, Dong, & Wang, 2010; Ganz, Gandhi, Wilson, & Mullett, 2010; Loomis, Golledge, & Klatzky, 1998; Simpson et al., 2005; Tian, Yi, & Arditi, 2010). In Nanayakkara, Shilkrot, and Maes (2012), an autonomous device, called EyeRing, has been presented. It comprises a finger-worn ring equipped with a VGA mini camera and an on/off switch, as well as an android mobile application. The user is required to turn the switch on, then the camera captures the scene and carries it on to the mobile phone via Bluetooth for further computer vision-based processing. Depending on the chosen mode (e.g., object, color, or currency), which is verbally selectable by the user, a vocal statement is output by the mobile application through a TTS (Text-To-Speech) module. The notable features of the designed instrument are the ease-of-use and lightweight. Ulrich and Borenstein (2001) proposed a guide-cane consisting of a round housing, wheelbase and a handle. The housing is surrounded by ten

* Corresponding author. Tel.: +39 0461281573.
E-mail addresses: mohamed.mekhalfi@unitn.it (M.L. Mekhalfi), farid.melgani@unitn.it, melgani@disi.unitn.it (F. Melgani), abdallah.zeggada@unitn.it (A. Zeggada), francesco.denatale@unitn.it (F.G.B. De Natale), salem@cis.asu.edu.eg (M.A.-M. Salem), akhamis@pami.uwaterloo.ca (A. Khamis).

http://dx.doi.org/10.1016/j.eswa.2015.09.054
0957-4174/

ultrasonic sensors, eight of which are placed on the frontal side and spaced by 15° so that to cover a wide sensed area of 120°, while the remaining two are located on the edgewise for side-objects detection (doors, walls, etc.). The user can use a mini joystick to control the preferred direction and push the cane through in order to inspect the area. When an obstacle is detected by the sensors, an embedded obstacle avoidance algorithm is launched to estimate an alternative obstacle-free path. The feedback to the user is given by steering the cane through, which results in a force felt by the user on the handle. A somehow similar concept, called NavBelt, was also presented in Shoval, Borenstein, and Koren (1998). In this work, the ultrasonic sensors are integrated on a worn belt and spaced by 15°. The information about the context in front of the user is carried within the reflected signal and is processed within a portable computer. The outcome of the analysis is relayed to the user by means of earphones. The distance to objects is represented by the pitch and volume of the generated sound (i.e., the shorter the distance, the higher the pitch and volume). As an attempt to facilitate the use and bring more comfort, a wearable smart clothing prototype has been designed in (Bahadir, Koncar, and Kalaoglu, 2012). The model is equipped with a microcontroller, ultrasonic sensors, as well as indicating vibrators. The sensors explore the area of concern, whilst a neuro-fuzzy-based controller detects the obstacle position (left, right, and front), and provides navigation tips such as “turn left”, or “turn right”. A similar approach is also proposed in Shin and Lim (2007). Another study (Bousbia-Salah, Bettayeb, & Larbi, 2011), provides an ultrasonic-based navigation aid for the blind, permitting him/her to explore the route within 6 m ahead via ultrasonic sensors placed on the shoulders as well as on a guide cane. The underlying idea is that the sensors emit a pulse, which in case of an obstacle is reflected back: the time between emission and reception (time of flight) allows estimating the distance of the obstacle. The indication is carried to the user by means of two vibrators (also mounted on his/her shoulders), and verbally for guiding the cane. The control of all the process is attributed to a microcontroller. In Lee, Kang, and Lee (2008), the authors propose a different approach including many tasks. In particular, the proposed system contains the following modules: object detection, pedestrian recognition, ultrasonic-based object distance sensing, and a positioning system through the GPS (global positioning system). Object detection module generates a disparity image, which is processed in the object recognition module using support vector machines (SVM). The classifier is trained on vertical silhouettes and takes charge of face detection. Text recognition is also included and is achieved using a commercial engine. The main drawback is that the above modules are run sequentially. Another design was considered in Scalise et al. (2012). In that paper, a new electromagnetic concept for obstacle detection was introduced, based on the idea of scanning the frontal area through a wideband antenna, emitting an electromagnetic wave. The presence of an obstacle generates a reflection of the signal, which is amplified and analyzed to assess the distance of the reflecting object. The presence of objects has been achieved at a signal-to-noise ratio (SNR) within 10–23 dB. Robot-based assistance for visually impaired has also received attention (Kim & Yi, 2009; Kulyukin, Gharpure, Nicholson, & Osborne, 2006). Robots can perform many assistive tasks for people with vision disability such as navigation guidance, handling various household duties, providing medical care, entertainment and rehabilitation. In order to effectively assist blind and low vision people in an interactive and team-based way, and to improve the acceptance and usability of these technologies, assistive robots must be able to recognize the current activity that the human is engaged in, the task and goal context of the current activity, as well as be able to estimate how the human is performing and whether assistance is required and appropriate in the current setting. Social assistive robots, which may be particularly expensive, must also ensure the physical safety of the human users with whom they share their workspace. Banknote recognition for the blind has also been addressed in Hasanuzzaman, Yang, and Tian (2012), where the speeded-up robust features (SURF) have been employed. A supermarket shopping scenario has been treated in López-de-Ipiña, Lorido, and López (2011). In this work, radio-frequency identification (RFID) has been used as a means for localization and navigation, while product recognition has been performed by reading QR codes through a portable camera. Product barcodes detection and reading has also been suggested in Tekin and Coughlan (2009). In another work (Pan, Yi, & Tian, 2013), a portable camera-based design for bus line-number detection was considered as a travel assistant. Staircase detection in indoor environments was proposed in Tang, Lui, and Li (2012). In Chen and Yuille (2004), the authors suggest assistive text reading in natural scenes. RFID technology was also exploited for bus detection at public bus stations as to further ease blind people mobility (Al Kalbani, Suwailam, Al Yafai, Al Abri, & Awadalla, 2015). Another worth-noting contribution was proposed in Kulkarni and Bhurchandi (2015). It consists of a device designed for facilitating e-book reading for blind individuals via a built-in Braille script. The device was claimed to be handy and at an affordable cost. Another work, which considers clothes color as well as pattern recognition as a means of facilitating recognition capabilities of blind people, was put forth in Thilagavathi (2015). It combines three kinds of features, which are further fed into a SVM classifier as a decision making paradigm. In Neto and Fonseca (2014), assistive text reading was propounded. Its underlying idea is to acquire the text zones by means of a camera, and afterwards exploit optical character recognition capabilities to recognize the text and forward it to a text-to-speech engine as to deliver a vocal feedback.

Overall, although the current literature proposes several interesting technologies to address specific guidance or recognition problems for the blind, there is still a remarkable lack of integrated solutions able to provide a usable “sight substitute”. In this context, we propose in this paper a new design that incorporates guidance and recognition capabilities into a single prototype. These two needs, as observed throughout the literature, have very scarcely (just one previous work to the best of our knowledge) been coupled together. The tool is designed for indoor use, and is fully based on computer-vision technologies. The components of the system include a portable camera attached to a wearable jacket, a processing unit, and a headset for commands/feedback. The navigation system is launched as soon as the prototype is powered on, and keeps instructing the blind person whenever he/she moves across the indoor environment. In order to avoid information flooding, the recognition system is activated upon request of the user. The prototype was implemented and tested in an indoor environment, showing good performance in terms of both navigation and recognition accuracy.

The rest of this paper is structured as follows. Section 2 provides an overview of the prototype architecture. In Section 3, the functioning of the guidance system and its modules are reported in detail. Section 4 describes how the recognition task is performed. Section 5 illustrates the operational use of the prototype in a real indoor environment. Finally, conclusions are drawn in Section 6.

2. Proposed prototype

The proposed prototype accommodates two complementary units: (i) a guidance system, and (ii) a recognition system. The former works online and takes charge of guiding the blind person through the indoor environment from his/her current location and leading him/her to the desired destination, while allowing avoiding static as well as moving obstacles. By contrast, the latter works on demand. The whole prototype is based on computer vision and machine learning techniques. The inputs of the prototype include:

- A speech recognition module, which acquires verbal instructions from the headset and serves for determining the command of the