Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Computer Science 14 (2012) 94 – 101
Procedia Computer Science

Proceedings of the 4th International Conference on Software Development for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2012)

Blind navigation support system based on Microsoft Kinect

Vítor Filipe[*a], Filipe Fernandes[b], Hugo Fernandes[c], António Sousa[d], Hugo Paredes[e], João Barroso[f]

[a]*CIDESD-UTAD, vfilipe@utad.pt, Portugal*
[b]*UTAD, al28983@utad.eu, Portugal*
[c]*UTAD, hugof@utad.pt, Portugal*
[d]*CITAB-UTAD, amrs@utad.pt, Portuga*
[e] *INESC TEC (formerly INESC Porto) / UTAD – University of Trás-os-Montes e Alto Douro, 5001-801 Vila Real, Portugal*
[f]*UTAD-GECAD, jbarroso@utad.pt, Portugal*

Abstract

This paper presents a system which extends the use of the traditional white cane by the blind for navigation purposes in indoor environments. Depth data of the scene in front of the user is acquired using the Microsoft Kinect sensor which is then mapped into a pattern representation. Using neural networks, the proposed system uses this information to extract relevant features from the scene, enabling the detection of possible obstacles along the way. The results show that the neural network is able to correctly classify the type of pattern presented as input.

Keywords: blind, navigation systems, image processing, pattern recognition, neural networks;

1. Introduction

People with special needs always resorted to support tools while performing their daily tasks. The evolution of technology enables new ways of support other than the traditional, like the cane or the guide dog. Obtaining information from the surrounding environment using artificial sensors, and acting accordingly, is becoming more common nowadays. The possibility of creating technology which simplifies the daily life of a person with special needs is easier today. Technologies that are able to analyze, in real time, the surrounding environment and produce useful and interactive information are definitely an added value. The World Health Organization estimates that 285 million people are visually impaired worldwide: 39 million are blind and 246 have low vision [1]. People with vision disability have great difficulty in perceive and understand the physical reality in an unknown environment [2][3]. Their motion difficulties in new and non-familiar spaces are increased not

[*]Corresponding author. Tel.: +351-259-350-000; fax: +351-259-350-480.
E-mail address: vfilipe@utad.pt

 doi:10.1016/j.procs.2012.10.011

only due to the specific disability, but also to the lack of useful and contextual information in those kinds of scenario. The system presented in this paper aims to counteract that situation, proposing a solution that uses artificial vision sensors to assist blind people in their navigation, delivering information about the surrounding environment in real time.

The overall organization of the paper is as follows: referring to related work in the context of this paper, section 2 describes how related research projects address the problems related with the creation of navigation systems specifically designed to help and enhance the mobility of the visually impaired; section 3 explains the proposed system, describing how depth data is acquired and processed to be classified, in a further step, by the neural network; section 4 presents some results obtained with an implementation of the proposed system; finally, in section 5, the conclusions about the work done so far are presented, as well as some features that can be developed as future work.

2. Related work

In the last decades, several guidance systems for blind and visually impaired pedestrians were proposed [4]. One of the most important features of these devices is the obstacle avoidance module, which provides information about obstacles along way.

Bousbia-Salah suggests a system where obstacles on the ground are detected by an ultrasonic sensor integrated into the cane and the surrounding obstacles are detected using sonar sensors coupled on the user shoulders [5]. *Shoval et al.* proposes a system which consists of a belt, filled with ultrasonic sensors called *Navbelt* [6]. One limitation of this kind of system is that it is exceedingly difficult for the user to understand the guidance signals in time, which should allow walking fast.

Other authors, like *Castells et al.*, use vision sensors in their system setups. In this case, part of a vision system is proposed to detect possible obstacles as a complement to normal navigation with the cane. Using computer vision, images are analyzed to detect sidewalk borders and two obstacle detection methods are applied inside a predefined window [7]. Another system using a vision sensor is presented by *Sainarayanan et al.* to capture the environment in front of the user. The image is processed by a real time image-processing scheme using fuzzy clustering algorithms. The processed image is mapped onto a specially structured stereo acoustic pattern and transferred to stereo earphones. as seen in the system description in [8].

Some authors use stereovision to obtain 3D information of the surrounding environment. *Sang-Woong Lee* proposes a walking guidance system which uses stereovision to obtain 3D range information and an area correlation method for approximate depth information. It includes a pedestrian detection model trained with a dataset and polynomial functions as kernel functions [9]. Genetic algorithm methods are used by *Anderson et al.* to perform stereovision correlation to generate dense disparity maps. These disparity maps, in turn, provide rough distance estimates to the user, allowing them to navigate through the environment [10]. In [11] the overall idea is the detection of changes in a 3-D space based on fusing range data and image data captured by the cameras and creating the 3-D representation of the surrounding space. This 3-D representation of the space and its changes are mapped onto a 2-D vibration array placed on the chest of the blind user. The degree of vibration offers a way of sensing the 3-D space and its changes to the user. In [12] *A. Penedo et al.* also proposed a real time stereo vision system that uses one relative view (right camera) and a depth map (from the stereo vision equipment) to feed a fuzzy-based clustering module which segments the scenario into object clusters. Knowing the clusters' locations, the system is able to detect near and far obstacles and feed this information to the user.

3. Proposed system

In this work the stereoscopic vision is replaced by the Microsoft Kinect sensor, which is affordable and

widely available. It also supports a large feature set and has the ability to work in low light environments. There are some related works that also use this sensor [13][14]. In *Shrewsbury et al.* the sensor is used to calculate the distance from the user to objects within its field of view. The depth image obtained is mapped and fed via wireless to a haptic glove. *Zöllner et al.* uses the Kinect sensor to identify optical imprints and use them to guide a blind person [14].

The system proposed in this paper enables the recognition of pre-defined features/patterns on the surrounding environment using a neural network to analyze depth images obtained from the Microsoft Kinect sensor (Fig. 1).

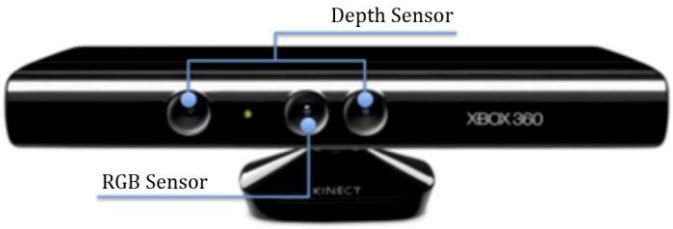

Fig. 1. Microsoft Kinect sensor

Neural networks enable data processing to be performed in a similar way to the human brain. A neural network is a distributed processor composed by simple processing units, which singularly have the natural tendency to store experimental knowledge and make it available for posterior use. The network obtains the knowledge from its environment; then, through a learning process, the strength of the connection between neurons, known as synaptic weights, is used to store and represent the obtained knowledge. The procedure used to perform the learning process is designated as 'learning algorithm'. Its function is to modify the network's synaptic weights in order to achieve a desired (known) output from a known input. Neural networks have been applied to many fields ranging from modeling and temporal series analysis to pattern recognition, signal processing and control. In the system proposed, a neural network is used to classify features/patterns taking advantage of its distributed parallel structure as well as of its learning ability and, therefore, generalize. This means that the neural network has the ability to produce appropriate outputs from the inputs presented during the training (learning) process. These two abilities make it possible for neural networks to solve problems with high degree of complexity [16].

3.1 Experimental setup

Depth images are acquired with the Kinect sensor, which includes a depth sensor and an *RGB* camera (Fig. 1). The depth sensor is composed by an infrared laser source that projects non-visible light with a coded pattern combined with a monochromatic *CMOS* image sensor [15] that captures the reflected light. The pattern received by the *RGB* sensor is a deformed version of the original pattern, projected by the laser source and deformed by the objects on the scene. The algorithm that deciphers the light coding generates a depth image representing the scene.

Depth data is acquired by mounting the Kinect sensor on the user's chest with an approximate inclination of *21 degrees* (Fig 2). This way, considering the height of the device to the ground to be of about *1600 mm*, the vertical field of view is about *3660 mm*, starting about *614 mm* in front of the user.

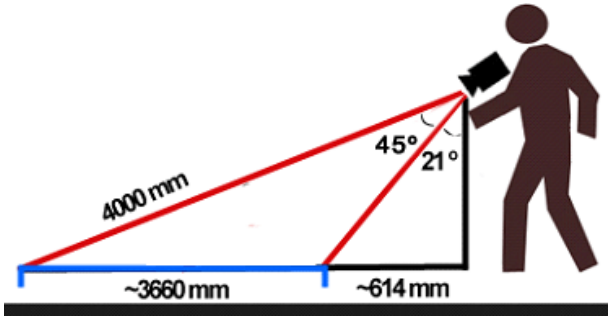

Fig. 2. Image depth acquisition

The depth image obtained from the camera represents distances to objects in the range of *800 mm* to *4000 mm*. Based on gray level representation of depth values, longer distances are mapped with lower intensity gray levels and shorter distances are mapped into higher intensity gray levels.

3.2 Depth image processing

The depth images were acquired with a resolution of *640x320 pixels* at a frame rate of *30 fps*. For each depth image, six vertical lines (line profiles) are extracted at pre-defined locations. Figure 3(a) shows the distribution of the six vertical lines over a depth image containing upstairs, also visible in the respective *RGB* image (Fig 3(b)).

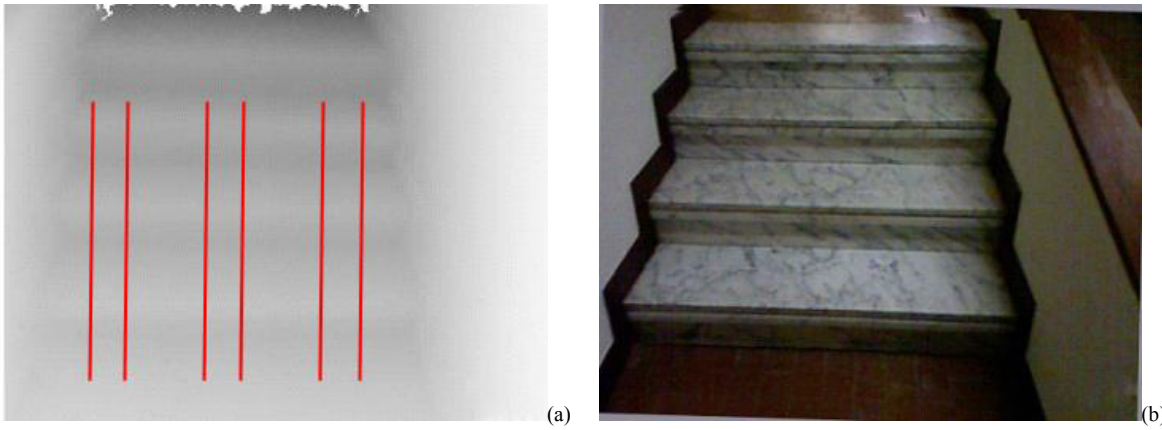

Fig. 3. (a) Depth image with highlights of the six vertical lines to be analyzed; (b) *RGB* image.

Figure 4 illustrates the signature of one line profile extracted from the image in Figure 3(a), where a distinctive pattern of the stair steps can be observed.

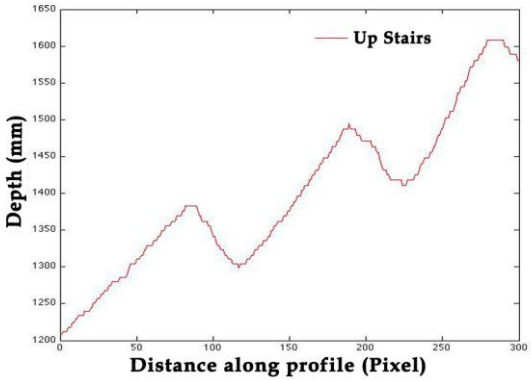

Fig. 4. Line profile signature

One practical example can be observed by looking at the signatures of the two line profiles presented in Figure 6: one concerning a scene without obstacles (free path), as seen in Figures 5(a) and Figure 5(c), and another representing a scene with an obstacle (a wall), as seen in Figures 5(b) and Figure 5(d). The difference between the two signatures is very clear (Figure 6). While the signature of the profile from Figure 5(c) is nearly linear (a distinctive pattern of free path) the signature of the profile from Figure 5(d) denotes a sudden variation in distance values due to the presence of the wall.

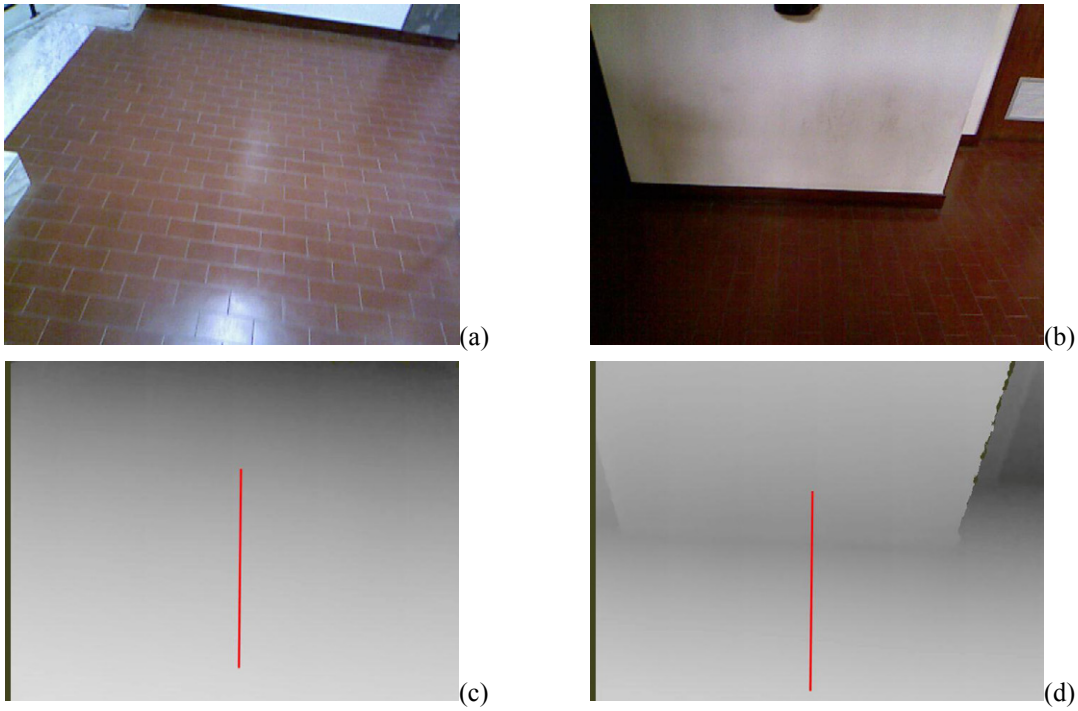

Fig. 5. RGB and depth images: in a scene with no obstacle (a) (c); in a scene with an obstacle (b) (d).

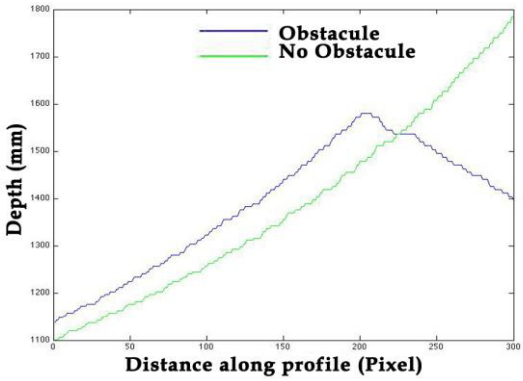

Fig. 6. Signature of line profiles corresponding to the images in Figure 5 (c), with no obstacle, and Figure 5(d), with a wall as obstacle.

The scene analysis is performed by using a neural network to classify six line profiles extracted from the depth image (Fig. 3(a)). The role of the neural network is to classify each line profile as fitting one of the four classes associated with the following situations: no obstacle, obstacle, upstairs and downstairs.

Combining the output from the neural network to each line profile input, the system is able to provide information about the location of the obstacles found, in terms of right, left and center, according to user's perspective.

3.3 Line profile classification

Using a neural network, several line profiles extracted from the depth image are classified according to certain features that are expected to be present, occasionally, in the user's field of view. In this work, these features are classified into four different classes: (1) no obstacles in the way (free path), (2) obstacle ahead (wall, for instance), (3) upstairs ahead and (4) downstairs ahead. The proposed system will be able to deliver warnings based on obstacle locations, assisting blind people with a usable guide system that increases their mobility, security and autonomy. In this work, the system proposed used a neural network to process the six signatures extracted from the depth image. In order to train the neural network, a set of input/output samples was used, where the inputs were the depth values from the line profiles and the outputs corresponded to one of the four predefined classes: obstacle, no obstacle, upstairs and downstairs. A feedforward neural network with three layers was trained using the backpropagation learning algorithm (supervised learning). The network was structured as follows: *300* neurons in the input layer, *10* neurons in the hidden layer and *4* neurons in the output layer. For each input sample the neural network only triggered one of the output neurons. Although the use of an artificial neural network to identify just 4 simple types of depth line profiles may seem excessive, this choice was made with the objective of a higher number of different obstacles identifications in the future. This way the system will be able to detect a much higher number of features without loosing its ability to perform in real-time, which is one of the requirements of this project.

4. Results

In order to test the system, a real time computer application was developed using the *C#* programming language. Image acquisition control and neural network training were implemented using two different tools. Microsoft's SDK for the Kinect sensor [17] was used to acquire both the depth and color images. The neural network was implemented using the open source library FANN (Fast Artificial Neural Network) [18].

The dataset used was composed of *1200* samples (input/output pairs) obtained from a large amount of images, each representing one of the four predefined classes. The dataset was split into three subsets: training, validation and test subsets. Table 1 presents the confusion matrix retrieved from tests performed on another dataset, used for system evaluation, consisting of *714* input samples. None of the image samples used in the training process was present in this final dataset.

Table 1. Confusion matrix of dataset used in system evaluation.

Network outputs	Targets (correct outputs)				
	No obstacle	Obstacle	Upstairs	Downstairs	Total
Obstacle	205 (28.7%)	0 (0%)	1 (0.1%)	0 (0%)	**99.3%**
No obstacle	0 (0%)	152 (21.3%)	2 (0.3%)	2 (0.3%)	**98.1%**
Upstairs	2 (0.3%)	0 (0%)	222 (31.9%)	0 (0%)	**99.1%**
Downstairs	0 (0%)	0 (0%)	0 (0%)	128 (17.9%)	**100%**
Total	**98.7%**	**100%**	**98.7%**	**98.5%**	**99%**

It is clear that the neural network correctly classifies approximately *99%* of the samples. In the specific case of the samples from the class "obstacle", all samples were correctly identified (*100%* accuracy).

Two samples from "upstairs" class and two samples from "downstairs" class were misclassified as "no obstacle". This may present a potential danger situation to the blind user since he is not informed about a potential risk on his way.

The tests performed with the real time application showed that, typically, the system provides information about the presence of obstacles and stairs approximately *2* meters before the blind reaches them. This distance may be considered as appropriate for a timely response to the blind user. The system presents encouraging results since, from an overall perspective, the network was able to differentiate situations where obstacles and stairs were present from situations with no obstacles at all.

5. Conclusions and future work

This paper proposes a system to assist blind users in their navigation. The proposed system is able to provide information about the surrounding environment, in real time, based on depth data acquired by Microsoft Kinect sensor. The system is able to detect different patterns in the scene like no obstacles (free path), obstacle ahead (wall), and stairs (up/down). The neural network proved to be efficient in the classification of lines profiles extracted from depth images.

However, the practical use of the Microsoft Kinect sensor for data acquisition is still a partial solution. Concretely, in terms of portability, its dimensions and the need to be plugged in to a computer doesn't allow it to be conveniently nor comfortably carried by the user. Another restriction is posed by the difficulty in obtaining depth information on surfaces exposed to sunlight or covered by water. This is a limitation to the use of the system in outdoor environments.

In the future, a solution using a smartphone will be implemented to improve the overall portability of the system. It is also intended to deliver real time information to the user through haptic devices or through sound. Another set of tests will be made in order to enhance the accuracy, avoiding the existence of false positives in the neural network classification.

Acknowledgements

This research was supported by the Portuguese Foundation for Science and Technology (FCT), through the project RIPD/ADA/109690/2009– "Blavigator: a cheap and reliable navigation aid for the blind".

References

1. WHO: Visual impairment and blindness, Fact sheet nº 282, October 2011 (2011).

2. Z. H. Tee, L. M. Ang, K. P. Seng, J. H. Kong, R. Lo, and M. Y. Khor. SmartGuide system to assist visually impaired people in a university environment. In Proceedings of the 3rd International Convention on Rehabilitation Engineering & Assistive Technology (i-CREATe '09). ACM, New York, NY, USA, Article 2 , 4 pages. DOI=10.1145/1592700.1592703 (2009).

3. Z. H. Tee, L. M. Ang and K. Seng. Smart Guide System to Assist Visually Impaired People in an Indoor Environment. IETE Technical Review, 27(6), 455. doi:10.4103/0256-4602.68522 (2010).

4. S. Shoval, I. Ulrich, J. Borenstein, et al. Computerized obstacle avoidance systems for the blind and visually impaired. Rehabilitation. Citeseer. doi=10.1.1.23.8523 (2000).

5. M. Bousbia-Salah, M. Bettayeb and A. Larbi. A Navigation Aid for Blind People. Journal of Intelligent & Robotic Systems, 64(3-4), 387-400. doi:10.1007/s10846-011-9555-7 (2011).

6. S. Shoval, J. Borenstein and Y. Koren. Auditory Guidance with the Navbelt - A computerized Travel Aid for the Blind. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 28(3), 459–467 (1998).

7. D. Castells, J. M. F. Rodrigues and J. M. H. du Buf. Obstacle detection and avoidance on sidewalks. Proc. Int. Conf. on Computer Vision-Theory and Applications (VISAPP2010), Angers, France (Vol. 2, pp. 235–240) (2010).

8. G. Sainarayanan, R. Nagarajanand S. Yaacob. Fuzzy image processing scheme for autonomous navigation of human blind. Applied Soft Computing, 7(1), 257-264. doi:10.1016/j.asoc.2005.06.005 (2007).

9. S. Lee and S. Kang. A Walking Guidance System for the Visually Impaired. Int. J. Patt. Recog. Art. Intel., 22(6), 1171-1186 (2008).

10. J. Anderson and D. Lee. Embedded stereo vision system providing visual guidance to the visually impaired. Life Science Systems and Applications Workshop, LISA 2007. 229-232 (2007).

11. N. Bourbakis. Sensing surrounding 3-D space for navigation of the blind. Engineering in Medicine and Biology Magazine. 49-55 (2008).

12. A. Penedo, P. Costa, H. Fenandes, A. Pereira, J. Barroso, "Image segmentation in systems of stereo vision for visually impaired people", DSAI2009 - Proceedings of 2nd International Conference on Software Development for Enhancing Accessibility and Fighting Info-exclusion, pp. 149-156, Lisbon, Portugal, UTAD (2009)

13. B. T. Shrewsbury. Providing haptic feedback using the kinect. The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility (pp. 321–322). ACM (2011).

14. M. Zöllner, S. Huber, Hans-Christian Jetter, and H. Reiterer. NAVI: a proof-of-concept of a mobile navigational aid for visually impaired based on the Microsoft Kinect. In Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part IV (INTERACT'11), P. Campos, N. Nunes, N. Graham, J. Jorge, and P. Palanque (Eds.), Vol. Part IV. Springer-Verlag, Berlin, Heidelberg, 584-587 (2011).

15. Kinect Sensor, from: http://msdn.microsoft.com/en-us/library/hh438998.aspx (2011).

16. S. Haykin. Neural Networks: an comprehensive foundation, 2nd Edition, Prentice Hall (1999).

17. Microsoft Kinect SDK, from: http://www.microsoft.com/en-us/Kinectforwindows/ (2011).

18. FANN, from: http://leenissen.dk/fann/wp/ (2012).