IntelliNavi: Navigation for Blind Based on Kinect and Machine Learning

Alexy Bhowmick[1], Saurabh Prakash[1], Rukmani Bhagat[1], Vijay Prasad[1], and Shyamanta M. Hazarika[2]

[1] School of Technology, Assam Don Bosco University
Guwahati, Assam, India
{alexy.bhowmick,saurabhp,rukmanib,vijay.prasad}@dbuniversity.ac.in
[2] School of Engineering, Tezpur University
Tezpur, Assam, India
smh@tezu.ernet.in

Abstract. This paper presents a wearable navigation assistive system for the blind and the visually impaired built with off-the-shelf technology. Microsoft Kinect's on board depth sensor is used to extract Red, Green, Blue and Depth (RGB-D) data of the indoor environment. Speeded-Up Robust Features (SURF) and Bag-of-Visual-Words (BOVW) model is used to extract features and reduce generic indoor object detection into a machine learning problem. A Support Vector Machine classifier is used to classify scene objects and obstacles to issue critical real-time information to the user through an external aid (earphone) for safe navigation. We performed a user-study with blind-fold users to measure the efficiency of the overall framework.

Keywords: Kinect, RGB-D, Blind, Navigation systems, Object recognition, Machine Learning.

1 Introduction

The blind and visually impaired (VI) face many challenges in their everyday life. People with vision disabilities are handicapped in perceiving and understanding the physical reality of the environment around. Searching, walking, crossing streets, recognizing objects, places and people becomes difficult or impossible without vision. Hence support tools become absolutely essential while performing activities of daily living. Recent statistics from the World Health Organization estimate that there are 285 million people worldwide who are visually impaired: 39 million are *blind* and 246 million have *low vision*; 90% of world's visually impaired live in developing countries [3]. Herein lies the motivation for innovation of affordable tools to aid the affected community. Research in the area indicates that Computer Vision embodies a powerful tool for the development of assistive technologies for the blind[10],[18],[11]. Reliable and robust vision systems for the blind and visually impaired involving cutting-edge technology, provided at an affordable cost can have a very relevant social impact.

M.N. Murty et al. (Eds.): MIWAI 2014, LNAI 8875, pp. 172–183, 2014.

One major challenge for the visually handicapped everyday is safe navigation by detecting and avoiding objects or hazards along their walking path in an indoor or outdoor environment. This paper presents a wearable computerized navigation system that uses Microsoft Kinect[2] and state-of-the-art techniques from Computer Vision and Machine Learning to detect indoor objects or obstacles and alert the blind user through audio instructions for safe navigation in an indoor environment.

Till today, the blind and visually impaired people rely heavily on their canes, guide dogs, or an assistant for navigating in an unfamiliar environment. In case of familiar environments the blind mostly depend on their sense of orientation and memory[6]. The traditional *white cane* helps the blind user to familiarize oneself with the immediate surroundings. However, this process requires memorizing the locations of doors, exits or obstacles and can be arduous, time taking, and mentally taxing. Moreover, any change in a familiar environment's configuration demands the familiarization process to be repeated again. The advent of the very affordable Microsoft Kinect sensor opens up new possibilities of creating technology which provides a degree of situational awareness to a blind person and simplifies the daily activities. Kinect[2] is a widely deployed off-the-shelf technology that has evolved to be a very suitable sensor to detect humans and objects and navigate along a path [5][15][11][16][12]. Released in the context of video games and entertainment, it is a recent device that has been applied in several areas such as – Computer Vision, Human-Computer Interaction, Robot navigation, etc. The goal of our system is to facilitate *micro-navigation*, i.e. sensing of immediate environment for obstacle and hazards. The system is affordable, and provides features such as obstacle detection, auditory assistance and navigation instructions. These features can be extremely helpful to the blind in performing their daily tasks such as walking by a corridor, recognizing an object, detecting staircases, avoiding obstacles, etc. The proposed navigation system is meant to complement the traditional navigational tools and not necessarily replace them.

The paper is organized as the follows. Section 2 discusses various methods on visual object detection in vision-based navigation systems involving Kinect. In Sect. 3 we present the system architecture of the navigation system and the proposed framework for obstacle detection and safe indoor navigation. In Sect. 4 we report on implementation and present results of a user study with few blind-folded users. The evaluation of the system showed above 82% overall object detection accuracy in a controlled environment. Section 5 presents the conclusion and possible future works.

2 Related Work

Visual object recognition of everyday objects in real world scenes or database images has been an ongoing research problem. The release of Kinect and the wide availability of affordable RGB-D sensors (color + depth) has changed the landscape of object detection technology. We review current research work in these fields - methods for addressing vision problems and development of Kinect

based navigation systems specifically developed to empower the blind and VI community. Vision-based navigation is one popular approach for providing navigation to blind and visually impaired humans or autonomous robots [15].

NAVI [18] - is a mobile navigational aid that uses the Microsoft Kinect and optical marker tracking to help visually impaired (VI) people navigate inside buildings. It provides continuous vibro-tactile feedback to the persons waist, instead of acoustic feedback as used in [5] and [13], to give an impression of the environment to the user and warn about obstacles. The system is a proof-of-concept and a lacks a user study. Mann et. al. [14] present a novel head-mounted navigational aid based on Kinect and vibro-tactile elements built onto a helmet. The Kinect draws out 3D depth information of the environment observed by the user. This depth information is converted into haptic feedback to enable blind and visually impaired users to perceive depth within a range and avoid collisions in an indoor environment. Depth information has been extensively used in a variety of navigation applications. Brock and Kristensson [6] present a system that perceives the environment in front of the user with a depth camera. The system identifies nearby structures from the depth map and uses sonification to convey obstacle information to the blind user. Khan, Moideen, Lopez, Khoo, and Zhu [13] also aimed at developing a navigation system utilizing the depth information that a Kinect sensor provides to detect humans and obstacles in real time for a blind or visually impaired user. Two obstacle avoidance approaches *one-step approach* and *direction update approach* were experimented with and a user-study was performed. However, in real time, the authors reported slow execution.

Other authors have used Kinect to create a 3D representation of the surrounding environment. Bernabei, Ganovelli, Benedetto, Dellepiane and Scopigno [5] present a low-cost system for unassisted mobility of the blind that takes as input the depth maps and data from the accelerometer produced by the Kinect device to provide an accurate 3D representation of the scene in front of the user. A framework for time-critical computation was developed to analyze the scene, classify the obstacles on ground, and provide the user with a reliable feedback. KinSpace [11] is another Kinect-based passive obstacle detection for home that monitors the open space for the presence of obstacles and notifies the resident if an obstacle is left in the open space.

Some authors have employed Kinect to provide autonomy to mobile robots, such as Sales, Correa, Osrio, and Wolf [15], who present an autonomous navigation system based on a finite state machine learned by an artificial neural network (ANN) in an indoor environment. The experiments were performed with a Pioneer P3-AT robot equipped with a Kinect sensor. The authors claimed to achieve excellent results, with high accuracy level for the ANN individually, and 100% accuracy on navigation task.

The use of machine learning (ML) presents a promising new approach to process RGB-D data stream acquired by Kinect. In recent years increasing research efforts in the machine learning community has resulted in a few significant assistive systems. Filipe, Fernandes, Fernandes, Sousa, Paredes, and Barroso [10]

present an ANN-based navigation system for the blind that perceives the environment in front of the user with the help of the Kinect depth camera. The system is a partial solution but is able to detect different patterns in the scene like - no obstacles (free path), obstacle ahead (wall), and stairs (up/down). The authors in [16] and [9] have both experimented with a framework based on Support Vector Machines (SVMs) combined with local descriptors to report on the performance of the classifier in categorization of varied objects such as cups, tomatoes, horses, cars, stairs, crossings, etc. Over the past decade increasing efforts and innovations have focused on generic object detection in images or real world scenes. Visual object recognition for assistive systems and vision-based navigation are fertile fields attracting many researchers with a growing interest in developing Computer Vision algorithms and mechanisms to address actual problems.

3 Proposed System - *IntelliNavi*

In this section we describe the proposed framework for recognizing indoor object from RGB-D sensor data and issuing audio feedback to the user. Feature extraction from RGB-D images is described next and finally the classification of indoor objects encountered and the navigation mechanism is presented.

Figure 1 shows the proposed Computer Vision-based framework for the detection of common indoor objects. The Kinect sensor is used to extract the RGB-Depth information of the environment being observed by the user. First, detection and description of *interest points* from training data is performed by the Speeded Up Robust Features (SURF) local descriptors. These descriptors of image patches are clustered using the Bag-of-Visual-Words (BOVW) model [8] - a novel approach to generic visual categorization to construct feature vectors. The feature vectors are of fixed dimension and serve as input to a multi-class SVM classifier to distinguish between objects.

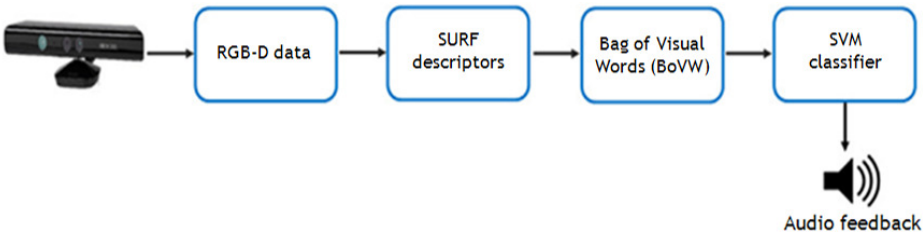

Fig. 1. Framework of the proposed navigation assistive system for blind using Kinect and Machine Learning

Thus a novel mechanism involving Kinect, RGB-D data, SURF, Bag-of-Visual-Words and SVM classifier, (which were not used together in literature before) is

developed. The audio feedback program notifies the user of obstacles ahead and suggests the blind user with an instruction to move or turn to a specific route to reach the endpoint. Figure 2 shows the user description of the navigation system. The navigation system integrates a Microsoft Kinect[2], a Laptop with earphone connected, a battery pack, and a backpack construction for the laptop. The Kinect device is fixed at the height of the pelvis. The backpack construction carries the laptop as the program modules analyze the scene and process objects.

Fig. 2. User description

3.1 System Architecture

Instead of a statically placed Kinect that tracks moving objects (as in [11]), we track the static environment with a moving waist-mounted Kinect (Fig. 2). To power the mobile Kinect we use a 12V - 7Ah battery pack that provided sufficient hours of power supply during our tests. The setup may be seem to be a cumbersome computer vision system because of the weight and size of the components, but our objective was to test the technological concept, hoping that further advancements could handle the portability issues. The system has two modes: Firstly, after being trained, detect foreseen indoor objects and classify them using a state-of-the-art supervised classifier. Secondly, communicate the environment information back to the user through an audio feedback.

Data Acquisition. Kinect provides a depth sensor, an RGB camera, an accelerometer, a motor and a multi-array microphone. Kinect leads to low-cost solutions in object recognition and works well in low light as well. The RGB camera (capable of capturing data at 640x480 pixels in 30 Hz) is associated to an infrared transmitter and receiver allowing depth estimation of the environment elements. The angular ranges, covered by the depth map, are about 60

degrees (horizontal) and 40 degrees (vertical) in front of the device. We adapted the Kinect to feed its images into a PC where vision processing modules get activated by the depth sensor at approximately 1-1.5 meters distance from any object. The system determines if there are objects or obstacles in the path, how close they are, and which way should one move to avoid them.

Fig. 3. a) Depth image of a chair acquired by Kinect b) Depth image of a RGB laptop on a table. Colors indicate distance ranging from *safe* to *very close* c) SURF interest points detected in a sample RGB chair image d) SURF interest points detected in a sample RGB laptop image.

Feature Extraction from RGB-Depth Data. SURF[1] is a fast, performant scale and rotation-invariant detector and descriptor. The SURF detector provides an almost real-time computation without loss in performance making it ideal for Computer Vision applications such as ours. The SURF descriptor is based on sums of Haar wavelet components [4]. We chose to extract SURF descriptors (from the training set images) as interest points since they are robust to image distortions and transformations including rotation, scale, illumination, and changes in viewpoint. These interest points contain rich local information of the image and local image descriptors have shown very good performance in object recognition tasks [9]. SURF constructs a descriptor vector of length 64 to

[1] `http://www.vision.ee.ethz.ch/~surf`

store the local features of the RGB-D images. The SURF implementation used in this paper is provided by the Accord.NET framework [2].

Most supervised classifiers (including SVMs) require feature vectors of fixed dimensions as input [17]. To transform the raw interest points provided by SURF into an image feature with a fixed dimension we opt for the BOVW [8] representation. Bag-of-Visual-Words assigns the SURF descriptors from training images to N clusters of visual words using the K-means algorithm. This BOVW model representation of these descriptors results in fixed dimensional feature vectors perfectly suited for a supervised classifier, which is next in our framework.

Training Support Vector Machines. SVMs have faster training times and generalize very well on smaller training sets [7], hence we use SVMs to train SURF descriptors and visual words. We built a supervised classifier based on the visual word features and applied them to predict objects in the indoor scenario. In order to extend the original binary SVM to a multiple class problem like ours, we built a *one-vs-one* scheme where multiple SVMs specialize to recognize each of the available classes. A multi class problem then reduces to n*(n-1)/2 smaller binary problems. Given an object encountered in test scenario, it is then assigned to the class with the maximum votes (i.e. the class selected by the most classifiers). An alternative strategy to handle the multi class case is to train N *one-vs-rest* classifiers as described in [7],[9].

Audio Feedback - Output. Although many navigation applications for visually impaired [5][13][14], have used audio as the primary interface for providing feedback, there is a strong argument against this solution: In micro-navigation scenarios, acoustic feedback can be annoying or even distracting since we are depriving the blind person of his/her sense of hearing [5][18]. The blind and visually impaired (VI) have limited sensory channels which must not be overwhelmed by information. Hence, we tried with very brief audio instructions that are triggered by the SVM classifier. Simple audio messages (e.g. "Wait","Obstacle ahead", "Take left", "Take right", etc) were provided via the earphone with the sole goal of leading a blind user safely from an initial point to an end-point.

4 Implementation and Experimental Results

In order to test the system, we developed a real-time application of the framework in C# using the Microsoft Kinect Software Development Kit (SDK) v1.8, Microsoft Visual Studio 2012, and Accord.NET framework v2.12. The proposed framework was tested on a Windows 8 platform running atop a 2.66 GHz processor with a 4GB RAM.

[2] `www.accord-framework.net` - a C# framework for building Machine Learning, Computer Vision, etc applications.

4.1 RGB-D Object Dataset

To experiment with the proposed framework, we developed a personal dataset consisting of images of common indoor objects e.g. chair, paper bin, laptop, table, staircase and door. We captured indoor workplace settings for our dataset, besides using subsets of the Caltech 101 dataset [1]. The training dataset is composed of 876 images with approximately 145 images representing each of the predefined objects (classes). The adopted learning algorithm is supervised, so the Support Vector Machine (SVM) was provided with labeled instances of the objects in training.

Fig. 4. Sample images of common indoor objects from our dataset

4.2 System Integration and Testing

We assembled the components mentioned in Fig. 2 for testing purposes. The Kinect was fixed at waist height with support from the neck. We performed a user-study with two blind-folded users to measure the efficiency and robustness of the algorithms and the framework. In a series of repeated experiments, the blind-folded users were instructed to walk along a path, in a controlled environment consisting of chairs, tables, laptops, staircases, etc. Figure 5 presents a top view of two obstacle courses where a user is required to move from a starting point to an end point along a path. The users had to walk slowly along a path

since the feature extraction module was not fast enough. We found the BOVW computation module to be computationally expensive which required the user to be still for a while. The users were guided safely by the brief audio messages triggered by the SVM classifier immediately on classification of an object ahead. We observed that '*Paper Bins*' and '*Staircases(Upper)*' were detected and classified with 100% accuracy in test scenarios A and B. In a few cases, mis-classifications in case of the other objects left the user stranded. The matrices in Table 1 and 2 show that the classifier did get confused between '*door*', '*laptop*', and '*chair*'. We observed a deterioration of performance in detection and classification with a reduction in lighting in the environment. We also noted that the users do not walk straight from the front towards an obstacle, they may approach from varying angles. Hence the angle at which Kinect RGB-D camera captures the images is crucial for a good field of view. We conclude from the confusion matrices in Table 1 and 2 that the feature extraction and classification algorithms worked satisfactorily well.

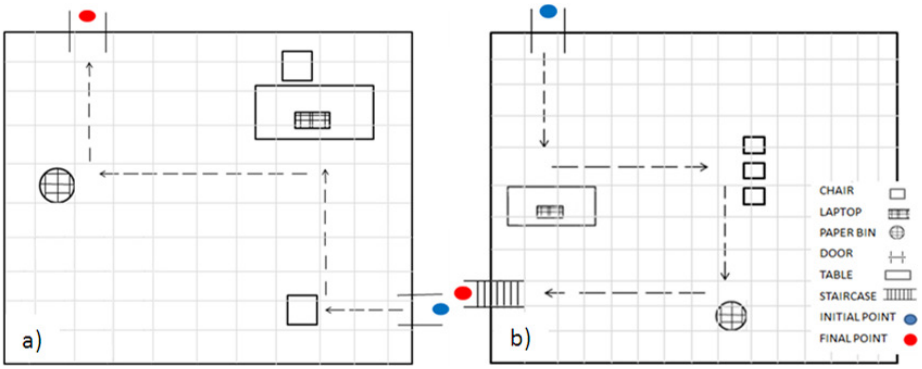

Fig. 5. Top view of two obstacle courses a) Test scenario A b) Test scenario B

As stated earlier, our environment was a controlled one (with fixed and limited indoor objects) and the navigation instructions for each route had to be coded into the classifier and was dependent on correct classification. Khan et. al. [13] have improvised on this and employed novel obstacle avoidance approaches to guide users. The multi-class SVM employed a *one-vs-one* strategy to classify multiple objects successfully. The audio feedback module was programmed to provide information about obstacles or staircases approximately 1-1.5 meters before the user reaches them, providing crucial time for a timely response. The different objects used in the test scenario are denoted in the legend in Fig. 5. The dashed arrow represents the direction along which the system prompted the blind-folded user to proceed.

Table 1. Confusion matrix of indoor objects seen in Test scenario A. Error cells in the bottom left indicate *false negatives* (2.5 %), while error cells in the top right indicate *false positives* (15.0%).

		Predicted Class				
		Chair	Laptop	Paper Bin	Door	Table
Actual Class	Chair	**7 (17.5%)**	2 (5.0%)	0 (0%)	0 (0%)	0 (0%)
	Laptop	1 (2.5%)	**6 (15.0%)**	0 (0%)	2 (5.0%)	1 (2.5%)
	Paper Bin	0 (0%)	0 (0%)	**8 (20.0%)**	0 (0%)	0 (0%)
	Door	0 (0%)	0 (0%)	0 (0%)	**6 (15.0%)**	1 (2.5%)
	Table	0 (0%)	0 (0%)	0 (0%)	0 (0%)	**6 (15.0%)**
	Total: 40, Errors: 7, *Accuracy*= 82.5%					

Table 2. Confusion matrix of objects seen in Test scenario B. Error cells in the bottom left indicate *false negatives* (5.4%), while error cells in the top right indicate *false positives* (5.35%).

		Predicted Class				
		Table	Laptop	Chair	Paper Bin	Staircase (Upper)
Actual Class	Table	**7 (12.5%)**	0 (0%)	0 (0%)	0 (0%)	0 (0%)
	Laptop	1 (1.8%)	**6 (10.71%)**	3 (5.35%)	0 (0%)	0 (0%)
	Chair	0 (0%)	2 (3.6%)	**21 (37.5%)**	0 (0%)	0 (0%)
	Paper Bin	0 (0%)	0 (0%)	0 (0%)	**8 (14.3%)**	0 (0%)
	Staircase (Upper)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	**8 (14.3%)**
	Total: 56, Errors: 6, *Accuracy*= 89.31%					

5 Conclusion and Future Works

The traditional *white cane*, which is the standard navigation tool for the blind will be hard to replace because - it is cheaper, lighter, foldable and requires no power-source when compared to its modern competitors *i.e.* navigation assistive systems. *IntelliNavi* is a navigation assistive system that is meant to help the user gain improved understanding of the environment and thus complement the white cane and not necessarily replace it. We present this system which uses the Kinect sensor to sense the environment and learn indoor objects and deliver crucial information of the surrounding environment to the wearer through an external aid (earphone). An accurate labeling of the objects is done most of the time by the multi-class SVM trained with SURF descriptors and visual words. The experimental results show that our system can detect objects or obstacles with more than 82% accuracy consistently (which is comparable to [15] and [11]) and navigate a user successfully through predefined indoor scenarios in the presence of many real-world issues.

In spite of the promising results in experiments, several limitations of the system and considerations will need to be addressed. We observed that our application is not yet fully optimized. The BOVW module in the framework is found to be computational expensive, thus slowing up detection. This can be sped up with optimization. The multi-class SVM classifies multiple objects but detects only one object at a time. We plan to address these limitations and iteratively improve this application in future. The use of Computer Vision to recognize complex objects in cluttered scenes, detect people, localize natural or artificial landmarks, and thus assist in blind way-finding are promising research directions. Few researchers have worked on providing navigation to blind users in an outdoor environment [16]. The main challenge is to formulate a novel algorithm on navigation for indoor and outdoor environments.

The concept of this project was to test a technological concept that could, in the future, evolve into something more advanced. Obstacle detection and avoidance and blind navigation using Computer Vision and Machine Learning techniques is a very fertile field of research. Despite the prospect of increased independence enabled by assistive systems, we believe much advancement still needs to be made in terms of user-friendliness, portability, and practicality before such systems gain acceptance by the Visually Impaired community.

References

1. Caltech101 dataset, http://www.vision.caltech.edu/Image_Datasets/Caltech101/
2. Kinect for Windows, http://www.microsoft.com/en-us/kinectforwindows/
3. Visually Impairment and Blindness. WHO Fact sheet–282 (2014), http://www.who.int/mediacentre/factsheets/fs282/en/
4. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding (CVIU) 110, 346–359 (2008)
5. Bernabei, D., Ganovelli, F., Benedetto, M.D., Dellepiane, M., Scopigno, R.: A Low-Cost Time-Critical Obstacle Avoidance System for the Visually Impaired. In: International Conference on Indoor Positioning and Indoor Navigation, Portugal, pp. 21–23 (September 2011)
6. Brock, M.: Supporting Blind Navigation using Depth Sensing and Sonification. In: ACM Conference on Pervasive and Ubiquitous Computing - UbiComp 2013, pp. 255–258. ACM, New York (2013)
7. Burges, C.J.C.: A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery 2(2), 121–167 (1998)
8. Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C., Maupertuis, D.: Visual Categorization with Bags of Keypoints. In: Workshop on Statistical Leanring in Computer Vision, pp. 1–22 (2004)
9. Eichhorn, J., Chapelle, O.: Object categorization with SVM: Kernels for Local Features. Tech. Rep. 137, Max Planck Institute for Biological Cybernetics, Tubingen, Germany (2004)
10. Filipe, V., Fernandes, F., Fernandes, H., Sousa, A., Paredes, H., Barroso, J.A.: Blind Navigation Support System based on Microsoft Kinect. Procedia Computer Science 14, 94–101 (2012)

11. Greenwood, C., Nirjon, S., Stankovic, J., Yoon, H.J., Ra, H.-K., Son, S., Park, T.: KinSpace: Passive Obstacle Detection via Kinect. In: Krishnamachari, B., Murphy, A.L., Trigoni, N. (eds.) EWSN 2014. LNCS, vol. 8354, pp. 182–197. Springer, Heidelberg (2014)
12. Hicks, S.L., Wilson, I., Muhammed, L., Worsfold, J., Downes, S.M., Kennard, C.: A depth-based head-mounted visual display to aid navigation in partially sighted individuals. PloS One 8(7), e67695 (2013)
13. Khan, A., Moideen, F., Lopez, J., Khoo, W.L., Zhu, Z.: KinDectect: Kinect Detecting Objects. In: Miesenberger, K., Karshmer, A., Penaz, P., Zagler, W. (eds.) ICCHP 2012, Part II. LNCS, vol. 7383, pp. 588–595. Springer, Heidelberg (2012)
14. Mann, S., Huang, J., Janzen, R.: Blind Navigation with a Wearable Range Camera and Vibrotactile Helmet. In: Proceedings of 19th ACM International Conference on Multimedia, pp. 1325–1328. ACM, New York (2011)
15. Sales, D., Correa, D., Osório, F.S., Wolf, D.F.: 3D Vision-Based Autonomous Navigation System Using ANN and Kinect Sensor. In: Jayne, C., Yue, S., Iliadis, L. (eds.) EANN 2012. CCIS, vol. 311, pp. 305–314. Springer, Heidelberg (2012)
16. Wang, S., Pan, H., Zhang, C., Tian, Y.: RGB-D Image-Based Detection of Stairs, Pedestrian Crosswalks and Traffic Signs. Journal of Visual Communication and Image Representation 25(2), 263–272 (2014)
17. Yang, J., Hauptmann, A.G.: Evaluating Bag-of-Visual-Words Representations in Scene Classification. In: International Workshop on Multimedia Information Retrieval (MIR 2007), pp. 197–206. ACM, New York (2007)
18. Zöllner, M., Huber, S., Jetter, H.-C., Reiterer, H.: NAVI – A Proof-of-Concept of a Mobile Navigational Aid for Visually Impaired Based on the Microsoft Kinect. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 584–587. Springer, Heidelberg (2011)