

# Obstacle detection and warning system for visually impaired people based on electrode matrix and mobile Kinect

Van-Nam Hoang<sup>1</sup> · Thanh-Huong Nguyen<sup>1</sup> · Thi-Lan Le<sup>1</sup> · Thanh-Hai Tran<sup>1</sup> ·  
Tan-Phu Vuong<sup>2</sup> · Nicolas Vuillerme<sup>3,4</sup>

Received: 7 December 2015 / Accepted: 18 July 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** Obstacle detection and warning can improve the mobility as well as the safety of visually impaired people specially in unfamiliar environments. For this, firstly, obstacles are detected and localized and then the information of the obstacles will be sent to the visually impaired people by using different modalities such as voice, tactile, vibration. In this paper, we present an assistive system for visually impaired people based on the matrix of electrode and a mobile Kinect. This system consists of two main components: environment information acquisition and analysis and information representation. The first component aims at capturing the environment by using a mobile Kinect and analyzing it in order to detect the predefined obstacles for visually impaired people, while the second component tries to represent obstacle's information under the form of electrode matrix.

**Keywords** Mobile kinect · Obstacle detection · Point cloud · Assistive system for visually impaired

## 1 Introduction

Travel activity, even a simple one, consists a long list of travel subtasks. There are two main categories of the subtasks in travel activity that are mobility and environmental access [5]. Mobility itself can be divided in obstacle avoidance and orientation/navigation, while environment access consists of hazard minimization and information/sign. Most of subtasks in travel activity are based on the vision information. For this, sighted people mainly rely on their sense of sight. Visually impaired are only able to use their sense of sight to a limited extent or possibly not at all. Therefore, visually impaired people require support from assistive technology to carry out different travel activity subtasks. In our work, we focus on developing assistive technology for obstacle avoidance for visually impaired people, because it has always been considered a primary requirement for aided mobility. Obstacle avoidance technology needs to address two issues: obstacle detection and obstacle warning. The obstacle detection means the perception of potentially hazardous objects in the environment ahead of time, while the latter one concerns the manner to convey obstacle information to the visually impaired people. White cane can be considered as the first obstacle avoidance assistive tool. However, this tool is generally not used to detect obstacles above knee height. Recently, the advance in sensor technology makes a number of obstacle avoidance technologies available for visually impaired people [19]. However, most researches focus on obstacle detection, obstacle warning is not well studied.

In our previous work, we have proposed an obstacle detection and warning system based on a low-cost device (Kinect)

---

✉ Van-Nam Hoang  
Van-Nam.Hoang@mica.edu.vn

Thanh-Huong Nguyen  
Thanh-Huong.Nguyen@mica.edu.vn

Thi-Lan Le  
Thi-Lan.Le@mica.edu.vn

Thanh-Hai Tran  
Thanh-Hai.Tran@mica.edu.vn

<sup>1</sup> International Research Institute MICA, HUST-CNRS/UMI  
2954-Grenoble INP, Hanoi University of Science and  
Technology, Ha Noi, Vietnam

<sup>2</sup> IMEP-LAHC, Grenoble Institute of Technology (GINP),  
Grenoble, France

<sup>3</sup> Institut Universitaire de France, LAI Jean-Raoul Scherrer,  
University of Geneva, Geneva, Switzerland

<sup>4</sup> University Grenoble Alpes, Grenoble, France

and electrode matrix [6]. We extend our previous work with three main contributions. Firstly, we improve obstacle detection method in order to decrease the detection miss by using plane segmentation on organized point cloud and eliminating the assumption that obstacles are on the ground. Secondly, instead of using stimulation signal for obstacle warning based on visual substitution as described in [6], we input the obstacle warning by the output of obstacle detection. Finally, we introduce the new patterns on electrode array for mapping information of obstacles and perform different experiments to evaluate the proposed mapping solution.

## 2 Related works

In the literature, different technologies such as WiFi, RFID, laser, ultrasound, or camera have been used for aiding blind people avoiding obstacles in the environment. In this section, we present only vision-based methods that are relatively close to our work in this paper. Methods for obstacle detection and warning could be categorized depending on how the obstacles are detected and how their information is sent to the user.

### 2.1 Vision-based obstacle detection

Obstacle detection is a key problem in computer vision for navigation. Existing methods could be categorized into two main approaches. The first approach learns object model then verifies if a pixel or an image patch satisfies the learnt model. In [18], a camera captures grayscale images, then pixels are classified into background or objects based on neural network technique. Then, the pixels belonging to obstacle are enhanced and the background pixels are removed. Joachim et al. [11] detects obstacles utilizing a model of human color vision. Then lens position of the auto-focus stereo camera was used to measure distance of the object center. In [23], a method was proposed for appearance-based obstacle detection. Firstly, color image is filtered, then converted to HSI color space. Then the color histogram on the candidate area is computed and compared with reference histogram.

The second approach is based upon a definition of objectness and detects regions with the highest objectness measures. In [17], authors developed a method for obstacle avoidance based on stereo vision and a simplistic ground plane detection. The obstacle detection relies on the creation of a virtual polar cumulative grid, which represents the area of interest ahead of the visually impaired user.

Approaches using conventional RGB camera draw some inherent limitations such as shadow, occlusion, illumination sensitivity. The use of stereo camera is expensive and requires highly precise calibration. Recently, low-cost RGB-D sensors (e.g., Microsoft Kinect) have been widely used

to complement RGB data with depth, helping to improve significantly performance of object detection. In [1], a system reads data from Kinect and expresses it as 3D point cloud then the floor plane and the occupancy of the volume in front of the user are detected. The occupancy represents an obstacle. In [9], the authors proposed a method combining depth and color. First, the depth map is denoised using dilation and erosion morphological operations. Then, least squares method is applied to approximate ground curves and to determine the ground height. The obstacles are decided based on the dramatic change in the depth value. Finally, object labeling is carried out with region-growing technique. Color information is used for edge detection and staircase identification. In [24], Vlamincx et al. presented a method for static obstacle detection consisting of four steps: point cloud registration, plane segmentation, ground and wall detection and obstacle detection. For plane segmentation, the authors employ RANSAC in order to estimate plane. They achieved a state-of-the-art result in obstacle detection using RGB-D data. However, their system is time consuming because of normal estimation and plane segmentation using RANSAC on 3D point cloud takes a lot of time to process. Moreover, the authors assume that the obstacles are on the ground; that assumption is not always satisfied.

### 2.2 Obstacle warning

Once detected, information of obstacles must be conveyed to the blind. In general, the user could be informed through auditory and tactile sense.

*Audio feedback* In [11], obstacle information is sent to the user using text-to-speech engine and the loudspeaker. In [25], the vOICE system translates live images into sounds for the blind person to hear through a stereo headphone. The position of visual pattern corresponds to the high pitch, while the brightness is represented by the loudness. In [18], segmented image is divided into left and right parts, transformed to (stereo) sound that is sent to the user through the headphones. In [17], acoustic feedback is in charge of informing the visually impaired users about the potential obstacles in their way. However, to avoid blocking the ears, the authors use audio bone conducting technology which is easy to wear and ears-free.

*Tactile feedback* Another approach is to transform obstacle information into a vibrotactile or electrotactile stimulations on different parts of the body. Visually impaired users are then trained to interpret the information. This approach allows the hearing sense to be free for the task of precautions or warning dangers. Johnson and Higgins [12] created a wearable device consisted of vibrator motors, each motor is assigned to detected regional obstacles. The value of the closest object in each region is transformed to vibration applied on the skin of abdomen. In [14], obstacle information is trans-

formed to electrical pulses that stimulate the nerves in the skin via electrodes in the data gloves.

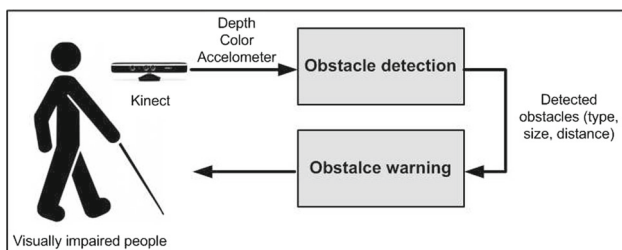
Among all the areas on the skin, tongue is very sensitive and mobile since it has the most dense number of receptors. A number of methods conveying electrotactile stimulate on the tongue have been conducted. The first tongue display unit (TDU) [27] translates the optical images captured by a head-mounted camera into electrotactile stimuli that are carried to the tongue by an array of  $12 \times 12$  electrodes via a ribbon cable. This prototype was then commercialized and called Brainport [22]. Tang and Beebe [20] created a two-way touch system to provide directional guidance for blind traveler. It consists of an electrotactile display of 49 electrodes to provide directional cues to the blind users. Recently, [26] has fabricated a matrix of 36 electrodes which sends the electrical impulses to the tongue in order to detect and correct the posture and stability for balance-impaired people.

From these studies, we find that the assistive systems for blind people are various and different from obstacle definition, detection and warning. Kinect sensor has great advantages than conventional RGB camera. This motivates us to use Kinect sensor for obstacle detection. Instead of combining RGB and depth data, we will explore accelerometer information for ground plane detection and remove wall and door planes as possible; thus, false alarms will be reduced. Concerning obstacle warning, we believe that conveying electrotactile pulses on the tongue is an efficient way. We continue our research direction on tongue display unit [14, 15] and build a complete system from obstacle detection to obstacle warning.

### 3 Prototype of obstacle detection and warning for visually impaired people

#### 3.1 Overview

The proposed system is composed of two modules: obstacle detection and obstacle warning (see Fig. 1). The main aim of obstacle detection is to determine the presence of interested obstacles in the scene in front of the users, while the obstacle warning represents and sends this information to the users.



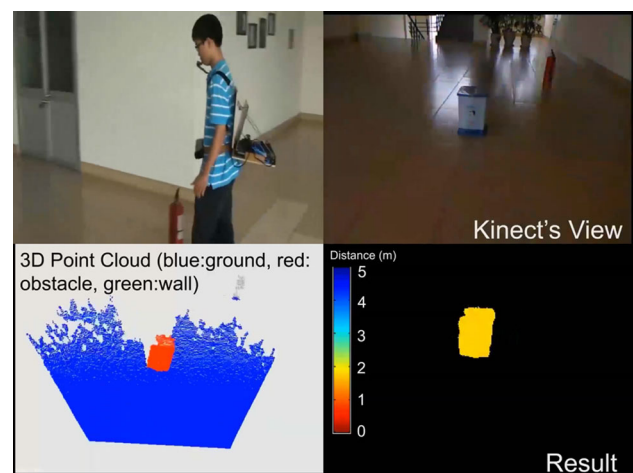
**Fig. 1** System flow chart

The obstacle detection module takes scene information from a mobile Kinect. In our prototype, the obstacle detection is running on a laptop mounted on a backpack of the visually impaired people and mobile Kinect is the Kinect with battery so that it can be mounted easily on the human body for collecting data and transferring data to the laptop. The scene information, in our case, is the color image, depth image, and accelerometer information provided by Kinect.

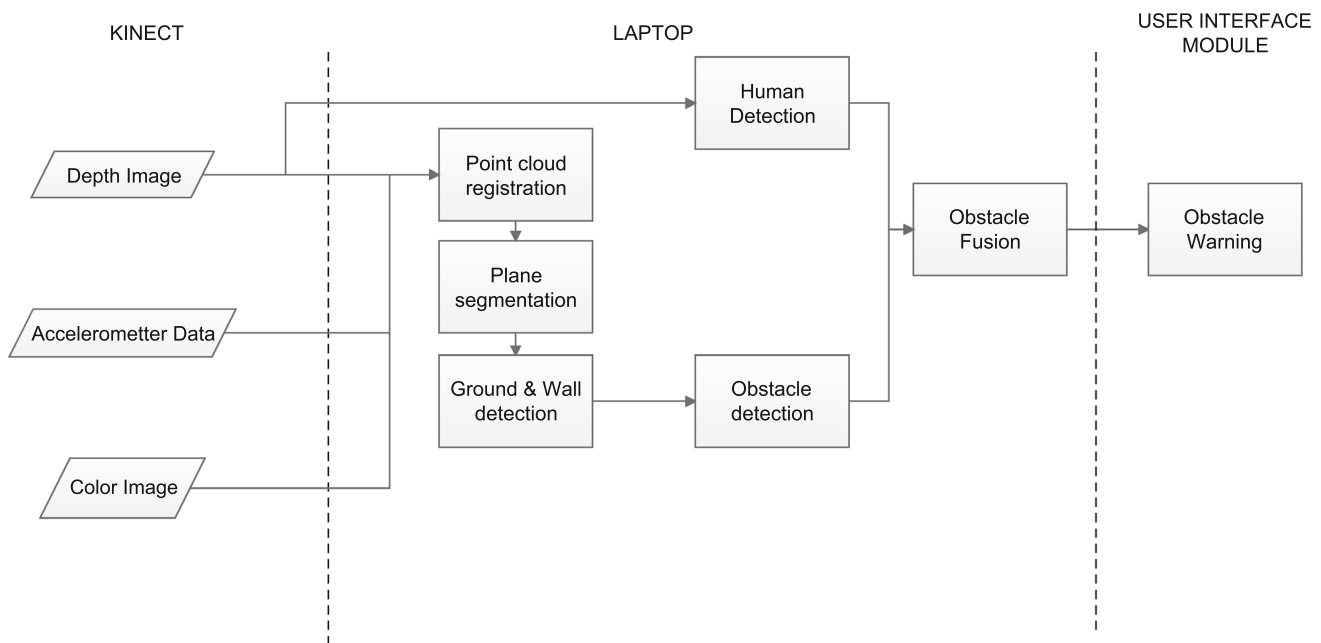
Concerning obstacle warning module, we reuse our tactile–visual substitution system which uses the tongue as the human–machine interface, gives a warning to the visually impaired people user to avoid the obstacles on the corridor way. This system is an embedded system that is equipped with an electrode matrix, a microprocessor unit (MCU), a communication module using RF wave [15]. For this module, we have to encode the obstacle information into the electrode matrix.

The prototype of our system is shown in Fig. 2. All the system can be mounted on the human body by backpack which hold the laptop, RF transmitter, and belt to anchor the Kinect. Although the current system is quite bulky and heavy and everything must be mounted on the user body, in the future, where all those things can be miniaturized and integrated into a small, wearable device like Google Glass, this problem can be solved. Especially, with the depth sensor, Microsoft have successful fabricated a device which is similar to the Kinect's depth sensor and can be attached to a normal mobile phone.

In our work, we consider indoor environment where obstacles are defined as objects in front, obstructing or endangering while visually impaired people moving. Specifically, we focus on detecting moving objects (e.g., people) and static objects (e.g., trash, plant pots, fire extinguisher). Staircase



**Fig. 2** Prototype system mounted on body (top left). Color image of the scene captured by Kinect (top right). Obstacle detection result in point cloud (bottom left). Estimated distance of the detected obstacle (bottom right)



**Fig. 3** Static and moving obstacle detection flowchart

has different characteristics and require another approach for detection.

In the following, we will describe in detail the obstacle detection and warning.

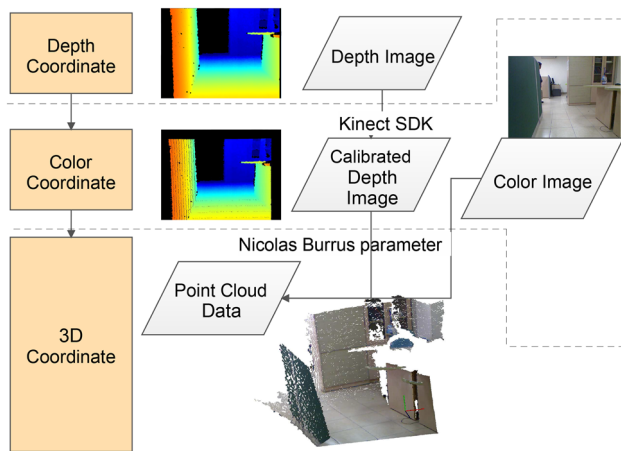
### 3.2 Obstacle detection

With obstacle detection module, we extended the works of Vlaminck in [24] while the objective and all other assumptions are still remained: visually impaired user moving along the hallway in the indoor environment with mobile Kinect and the system will detect an obstacle and give a warning message to the user. For data acquisition, we use mobile Kinect with a laptop as mentioned in Sect. 3.1. Kinect was chosen as the receiver sensor because it can provide many kinds of information such as color data, depth data, audio, etc. Moreover, depth data is the big advantage of Kinect because it is robust under lighting condition and can be used to calculate the distance from the user to obstacle to giving a warning message. The flowchart of static and moving obstacle detection is shown in Fig. 3. Concerning moving obstacle detection, we employ the human detection module provided by Kinect SDK. This module takes depth image as an input and provides a list of detected persons. Static obstacle detection consists of four steps: point cloud registration, plane segmentation, ground and wall detection and obstacle detection. As analyzed in Sect. 2, for static obstacle detection, we improve the work of Vlaminck presented in [24] in-plane segmentation step and ground and wall detec-

tion. First, for plane segmentation step, we use organized point cloud with the segmentation algorithm proposed in [7] instead of using RANSAC as in the work of Vlaminck. This allows us to perform the plane segmentation faster. Second, in [24], the authors base on an assumption that the obstacles are on the ground; therefore, if the ground plane is not detected, the obstacle detection process will terminate. Our work tries to detect ground and wall planes in order to remove that from the point cloud. The obstacle module still works even no ground plane is detected. In the following section, we present in detail the static obstacle detection.

#### 3.2.1 Point cloud registration

Point cloud registration step aims at taking information (color, depth and accelerometer data) from Kinect to build a point cloud. With Kinect, the color and depth image are captured by two different sensors, so they are not aligned. That means that given a pixel in the color image, we cannot get corresponding pixel in depth image directly as well as 3D coordinate. To make a 3D Point Cloud from Kinect data, with each pixel in both color and depth image, we must know exactly the location of this pixel in the 3D coordinate to create an RGB-XYZ point in Point Cloud. To solve that problem, a lot of work has focused on developing a good calibration method in order to transform between color coordinate, depth coordinate and real world coordinate such as Microsoft Kinect SDK, Burrus [8], Tang [21].



**Fig. 4** Coordinate transformation process

In our work, we used Microsoft Kinect SDK to convert depth coordinate to color coordinate, then use parameter from [8] to convert to 3D coordinates. Given a depth and color image. For each pixel in the depth image, we can find it is 3D coordinate in meter using the following formula:

$$P3D.x = (x_c - cx_c) * depth(x_c, y_c) / fx_c$$

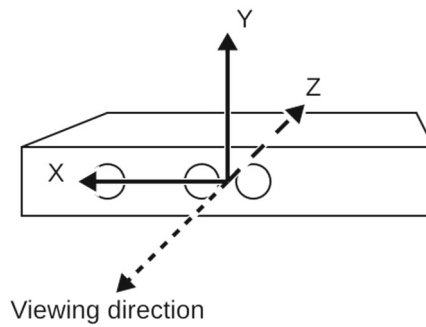
$$P3D.y = (y_c - cy_c) * depth(x_c, y_c) / fy_c$$

$$P3D.z = depth(x_c, y_c)$$

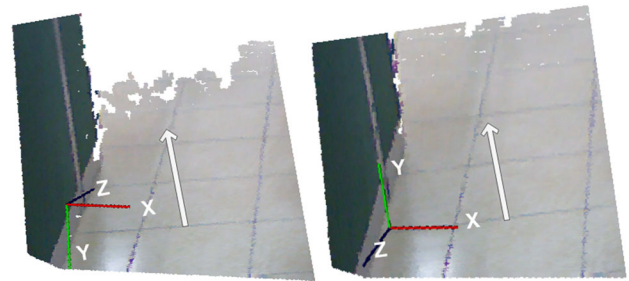
where  $x_c$  and  $y_c$  is the pixel coordinate in color image,  $cx_c, cy_c, fx_c, fy_c$  is taken from color intrinsic matrix,  $depth(x_c, y_c)$  is the depth value of pixel. This process is illustrated by Fig. 4. Because there are a lot of points in point cloud (about 300.000 points with VGA resolution), so the system becomes time-consuming and cannot run in the real-time. To reduce the execution time, point cloud will be down-sampled using  $2 \times 2$  block. So that the number of points in the cloud will be reduced by four times.

As mentioned in Sect. 3, our system uses mobile Kinect, which means Kinect mounted on the body. Therefore, while the visually impaired are people moving, Kinect may be shocked and shaking so that the point cloud will be rotated due the changing of Kinect direction. In our project, we used accelerometer data provided by Kinect SDK to rotate point cloud in order to align the ground plane with the xz-plane in reference system.

The accelerometer data is actually a 3-D vector pointing in the direction of gravity with coordinate system centered on the sensor shown in Fig. 5. With the default Kinect configuration (horizontal) represented by the  $(x, y, z, w)$ , the vector value is  $(0, -1.0, 0, 0)$ . We use this vector to build rotation matrix and then apply it into point cloud data in order to rotate point cloud. Figure 6 shows the output of this stage.



**Fig. 5** Kinect coordinate system [3]



**Fig. 6** Point cloud rotation using normal vector of ground plane (white arrow): left before rotating, right after rotating

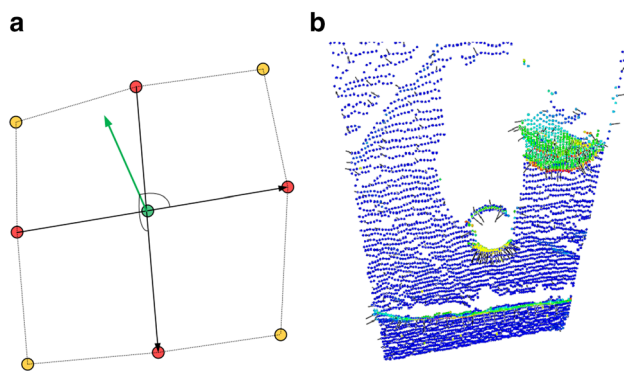
### 3.2.2 Plane segmentation

The plane segmentation step is to determine dominant planes from point cloud. For this step, we propose to use the plane segmentation method proposed in [7] that allows to segment point cloud data into multiple planes in real time. The main advantage of this algorithm is that plane segmentation can be done very fast using both information in image structure and point cloud data. For this, the normal vector estimation is performed by using an integral image. The normal vector of a single point is calculated by a cross product of two vectors of four neighbor points: bottom-top and left-right (see Fig. 7a). Based on the normal vector of each single point, first, two maps of tangential vectors, one for x- and the other for y-dimension, are computed. Then, planes are detected by segmentation in normal space (see Fig. 7b). An example of plane segmentation result of the scene illustrated in Fig. 8a is shown in Fig. 8b.

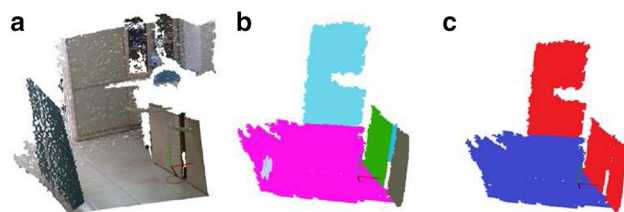
### 3.2.3 Ground and wall detection

After planes have been segmented, ground and wall planes can be detected easily using some constraints. Because our point cloud has been rotated to align with ground plane in the previous step using gravity vector, so the ground plane must satisfy the following conditions:





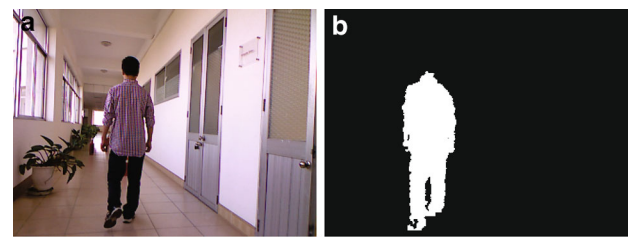
**Fig. 7** Normal vector estimation: **a** normal vector of the center point is calculated by a cross product of two vectors of four neighbor points (in red); **b** normal vector estimation of a scene



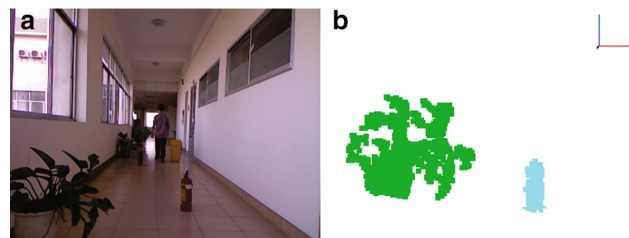
**Fig. 8** Plane segmentation and ground and wall detection results: **a** point cloud; **b** segmented planes; **c** detected ground (in blue) and wall planes (in red)

- The angle between gravity vector and ground plane's normal vector is almost 0 degree;
- Ground plane must be large enough. In our case, we checked the number of points inside a ground plane, if the number of points is larger than 10,000, then we consider it is a ground plane candidate;
- Since Kinect is mounted on the human body, distance between ground plane and Kinect (y-axis coordinates) must be in a range of 0.8–1.2 m.

Wall is considered as perpendicular plane to the ground plane. So, in order to detect wall planes, we use similar constraints with ground plane except that the angle between gravity vector and wall's normal vector is almost  $90^\circ$  and we do not need to check distance between wall plane and the Kinect, because wall plane can appear anywhere in our scene. After ground and wall have been detected, all remaining points will be checked again if they belong to those planes by using distance to detected plane, this step aims to remove the missing points in the plane due to the noise in its normal vector. Then, all the points belonging to ground and wall planes will be removed. Figure 8c shows an example of the ground and wall plane detection for the scene Fig. 8a.



**Fig. 9** Example of human detection: **a** color image; **b** human mask



**Fig. 10** Example of detected obstacles: **a** color image of the scene; **b** detected obstacles represented by different colors

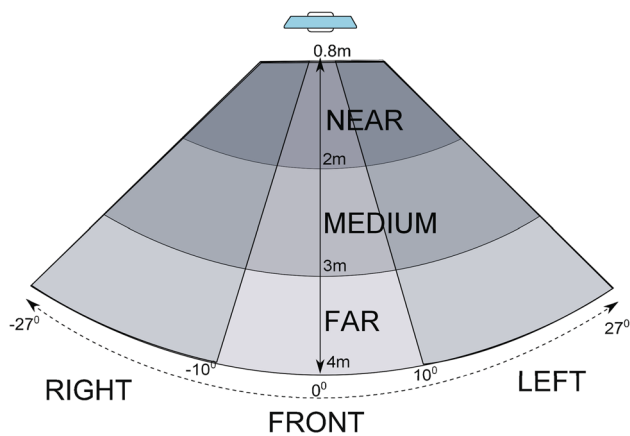
### 3.2.4 Obstacle detection

In this step, we will detect obstacles from the remaining point cloud. There are two kind of obstacle: human and static object. With human detection, Microsoft Kinect SDK also provided human segmentation data. Kinect can track up to six person in a camera field-of-view. This data is encoded as 3 lowest bit for each pixel in depth image and represented index of the person that Kinect has been tracked. Figure 9 shows an example of detected person.

After checking human data in the frame, we remove all points belonging to the detected human and do clustering to find remaining obstacles in the scene. This algorithm is based on the Euclidean distance between neighbor points. From the initial point (seed), the distance between this point and its neighbor will be calculated. Then the points whose distance is smaller than a threshold are kept. This procedure is repeated until all points are checked in the point cloud. And using organized point cloud's structure, the neighbor points will be chosen directly based on 2D coordinate in the depth image. This allows to save a lot of time in comparison with neighbors finding based on the distance between them. Figure 10 illustrates an example of detected obstacle. For obstacles lying on the ground, we calculate the distance to the user to give a warning message.

### 3.2.5 Obstacle fusion and representation

At this step, all detected obstacles will be checked to give a final warning message. These obstacles include wall, human and static objects. Because there may be more than one obstacle in a frame, so we need to know which obstacle has to



**Fig. 11** Obstacle position quantization for sending warning message to visually impaired people

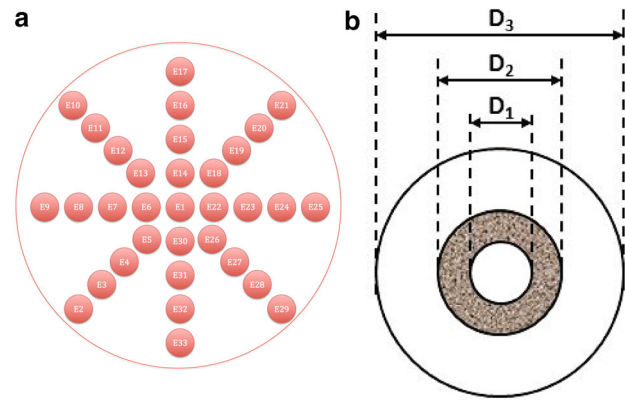
be informed to visually impaired people. For this, among detected obstacles, we keep the nearest one whose size is larger than a predefined threshold. Then we quantize the 3D position into three levels of distance (near, medium and far range) and three directions (left, front and right) (see Fig. 11). The encoded information is written in an output file and sent to warning module.

### 3.3 Obstacle warning

As presented previously, once obstacles have been detected, the second task is to send this information on obstacles to the blind. In our system, the Tongue Display Unit is used for conveying the instructions to the visually impaired users; hence, they will know how to react accordingly. Several methods have been used in literature as the means of transferring the needed information to the users, especially warning signals [2, 4, 10]. However, the tongue has been investigated by Paul Bach-y-Rita in the context of sensory substitution in which stimulus properties of one sense (e.g., vision) can be converted to stimulation of another sense (e.g., vibrotactile or electrotactile matrix in contact with different parts of human body). We proposed to use the tongue since it is the most sensitive organ of the body with the discrimination threshold of one or two millimeters (the tongue has approximately a million nerve fibers) [27]. Based on this idea, the proposed design of the electrotactile matrix and the representation of obstacle warning will be described in this section.

#### 3.3.1 Design the matrix of electrode

Most of the electrode arrays have the square or rectangular shape in which all the pins are arranged into perpendicular rows and columns. However, the matrix can only be placed on the inner superior part of the tongue in order for all the pins to get in contact with the surface. In our design, we propose



**Fig. 12** Design of electrode matrix (a) and typical dimension of an electrode pin:  $D_1 = 0.2$  mm,  $D_2 = 0.4$  mm;  $D_3 = 2$  mm (b)

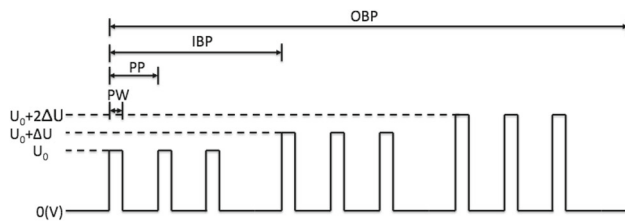
a round matrix of tactile arrays which better conforms to the shape of the tongue. Normally, it is easier for humans to perceive according to directions; therefore, we made use of this feature to arrange electrode pins into 45-degree-difference diameters as shown in Fig. 12a. This arrangement is composed of 2-mm disc-shaped electrode pins with a via of 0.2 mm for connecting to the ground. The distance between two electrodes is 2.7 mm. Figure 12b shows the dimension of an electrode pin.

#### 3.3.2 Information representation

In our TVSS system, the electrotactile stimulation is responsible for informing the visually impaired users about the potential obstacles in their way. Based on the signal in the form of tingles on the tongue, they will obtain information and warning of environment obstacles and react accordingly. The electrotactile stimulation is used to generate tactile sensation on the skin site, specifically the tongue surface. A local electric current is passed through the tongue receptor to stimulate cutaneous afferent nerve fiber. This interface is a good site for electrotactile display, because it does not block the ears of visually impaired users.

After receiving the data of obstacle, we will define the kinds of obstacles into different representation on the electrode matrix. Then, according to the depth information, we will define the degree of warning by changing the level of electrical signal. Actually, the local current is delivered through electrical pulse. A control module is included in the TVSS system to produce these pulses.

For electrotactile stimulus, positive rectangular pulses are chosen to deliver in series to the TDU [13]. According to [16], the pulse period is approximately 100ms and the duty cycle of each pulse should be 20% for rather good perception. Since the purpose of informing is in the form of warning, we chose the method of increasing regularly the intensity of electrical stimulation. By doing this, when users come closer



**Fig. 13** The stimulation waveform is composed of three levels of pulse groups (bursts) to obtain warning goal. Each burst contains three pulse with period of 100ms and 20ms of “on-time”

**Table 1** Electrotactile stimulation parameters

Symbol	Meaning	Range	Unit
OBP	Outer burst period	1200–1400	ms
IBP	Inner burst period	400–600	ms
PP	Pulse period	100	ms
PW	Pulse width	20–50	ms
$U_0$	Lowest voltage level	5–10	V
$\Delta U$	Voltage difference	0.5–3	V

Parameter values are controllable in real time by the control module program

to obstacles, the alert signal becomes stronger and makes them respond and take action to avoid objects. In our scenario, three stimulating voltages were defined: the lowest level, the higher level and highest level. At the lowest level, users can feel clearly the signal. The higher level start to create an uncomfortable feeling and the highest level can cause a strong sensation. Figure 13 and Table 1 display the waveform with three consecutive bursts of pulses. The magnitude of voltage increases steadily.

### 3.3.3 Obstacle warning representation

To prove the capability of the system to give warning message to the visually impaired individuals, we have to decide what information needs to be conveyed. Not all the objects are defined as an obstacle and after the detection step, the object types or classes and the position of objects need to be distinguished. As a consequence, the electrical stimulation can correspond to the warning of object classes. Besides, the intensity of each stimulation can be leveraged to give the suitable warning message to instruct the users' reaction. In the indoor environment of the experimental part, the object classes will be divided into two, the stationary (e.g., flower pot, fire extinguisher or dustbin) and the moving one (e.g., human or opening door). The object position in front of the users consists of three positions—left, front, right and the warning intensity increases to three levels—near, medium and far. Table 2 demonstrates the division of warning representation.

**Table 2** Classification of warning representation

Feature	Type		
Object classes	Stationary	Moving	
Warning level	High	Medium	Low
Position	Left	Center	Right

According to Table 2, a complete feasibility study was performed so as to evaluate the sensitivity of the tongue towards the intensity and electrode position on the tongue as well as the efficiency of this biofeedback device in warning the obstacles on the mobility path of the test subjects.

## 4 Experimental results

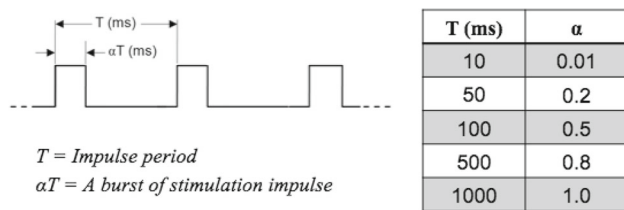
### 4.1 Material setup

Our prototype device is constructed upon off-the-shelf hardware components including a Kinect sensor which captures the color and depth data, a laptop computer for image processing, a control module and a matrix of electrodes which is arranged on a round substrate. The Kinect sensor is operated by a 12-V source of  $8 \times 1.5$  V AA batteries (we removed the original adapter and replaced it by the battery source); the control module and the electrode matrix attached to it are powered by a 3-V battery. The Kinect sensor is mounted on the user's belt to record the environment and the matrix of electrodes is placed inside the mouth and attached to the control module through the cable. Figure 14 shows the real prototype of the obstacle detection and warning system.



**Fig. 14** Illustration of the warning device. **a** Kinect sensor on user and **b** control module and electrode matrix on user. **a** Kinect sensor mounted on the belt worn by a blind student. Video processing is conducted by a laptop placed on a backpack. **b** The tongue electrotactile device worn on a blind user. The matrix of electrodes is place totally inside the mouth in contact with the dorsal part of the tongue and is controlled by the module through cables





**Fig. 15** Testing waveform parameters

The experiments were conducted with 20 young adults who voluntarily participated. Subjects were recruited at Grenoble University and Hanoi University of Science and Technology. Each volunteer was eager to participate and all provided informed consent to participate. Three main evaluations were implemented: waveform evaluation, intensity evaluation and efficiency evaluation. In each evaluation, all the subjects must be trained for a couple of minutes and then give feedback by their recognition or take part in a real mobility in an indoor environment on one floor.

#### 4.2 Electrical stimulation waveform calibration

In order to have an effective stimulation on the tongue, the waveform was calibrated. As a result, different values of electrical pulse parameters were tested with participants. Five healthy subjects performed this assessment. Their task was to test with one electrode at the front part of the tongue. Different values of impulse period and the duty cycle (the activation duration of the electrode in one impulse) were applied at 3 V and two trials were done with each couple of period and duty cycle. Figure 15 shows the waveform and its testing parameters.

The impulse values were first changed in order for several times and told to the participants. Then the values were generated randomly and each subject was asked about his/her perception. The results are shown in Fig. 16. It seems to give good perception and good speed of recognition at period  $T = 100$  ms and duty cycle = 0.2. In other cases, if the period is too high, it is too slow for recognition and if the period is low, it is too fast to distinguish. In case of high duty cycle, the electrical stimulation is so strong that it caused pain while in the case of low duty cycle, it is not a clear signal. Giving this timing parameter, the participants were then required to take part in the intensity evaluation.

#### 4.3 Electrical stimulation intensity calibration

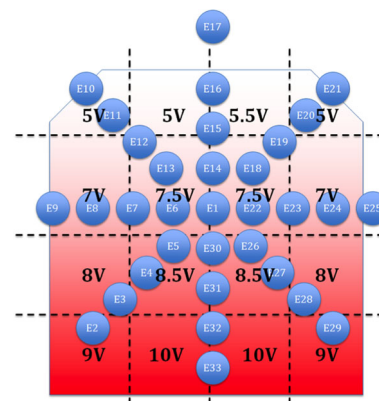
The TDU is very variable and may be used with any kind of electrodes, we have designed a particular geometry which is appropriate for the tongue application. The round shape can proliferate the convenience and comfort because it follows the contour of the tongue. This matrix is fabricated on

$T(ms) \backslash \alpha$	<0.01	0.2	0.5	0.8
10	Too fast, Not clear	Too fast, Not clear	Too fast, Strong	Too fast, Strong
50	Fast, Not clear	Fast, Not clear	Fast, Strong	Fast, Strong
100	Good speed, Not clear	Good speed, Good perception	Good speed, Strong	Slow, Strong
500	Slow, Not clear	Slow, Good perception	Slow, Strong	Slow, Too strong
1000	Too slow, Not clear	Too slow, Good perception	Too slow, Too strong	Too slow, Too strong

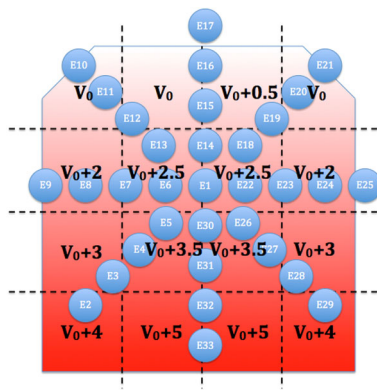
**Fig. 16** Waveform parameters perception

FR4 substrate which is very common for commercial circuit vendor. Each of the electrode has the diameter of over 2 mm and the center-center spacing is 2.34 mm. The overall dimension is 25 mm × 25 mm which fits easily on the tongue. The exposed surface of the electrode is gold-plated to reduce the harm to user's health. Although the tongue electrotactile display has been experimented in many applications, the perception on the electrical stimulation intensity has not yet been studied in detailed. Due to the limited size of the tongue, the electrode diameters must be small and reduce resistance. Aside from this, the region on the tongue determines the intensity. We performed a real test on five different users aging from 25 to 40. The preliminary results show that the contour of the tongue requires much low power than the center and rear part is less perceptive than the front part. A voltage generator produces voltages from 5 to 15 V and the average value is depicted in Fig. 17.

Because the intensity is an important factor for obstacle warning, this result is considered as the average voltage level that users can afford. From the obtained average voltages, the voltage values of different tongue regions are designated based on the lowest average voltage which is defined as V0 in Fig. 18. They are then written in the control program to adjust the voltage level automatically for the next tests. The value of V0 depends on the perception of each participant and is determined prior to the obstacle warning test.



**Fig. 17** Average voltage results measured on different regions of the tongue



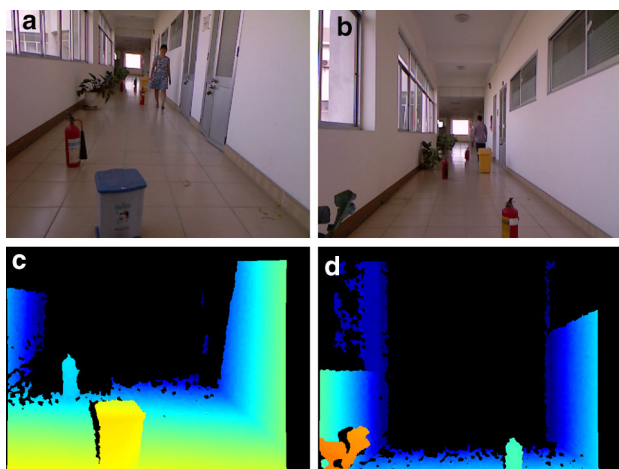
**Fig. 18** Voltage-level calculation

#### 4.4 Validation of obstacle detection module

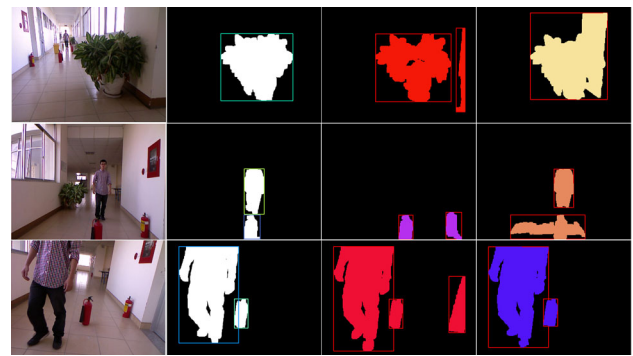
We evaluate the static obstacle detection method with 200 images captured at two different times with visually impaired people in MICA building. We named them dataset 1 and dataset 2. Each dataset contains 100 frames including color image, depth image and accelerometer data. With dataset 1, the ground plane in depth image has a large area; whereas the dataset 2 ground only takes a small area, as can be seen in Fig. 19. We compared our method with the method of Vlaminck et al. [24].

With each dataset, we made two different evaluations: pixel level and object level. Concerning pixel level, for the ground-truth, we apply Watershed algorithm on depth image in order to separate objects from background. The obstacle detection result in point cloud is back projected into 2D image.

For object level, we define manually obstacles of the scene. Each obstacle is determined by a rectangle. A detection result is a true detection if the ratio between the intersection of the



**Fig. 19** Example images: **a** and **c** are color and depth images in dataset 1; **b** and **d** are color and depth images in dataset 2



**Fig. 20** Obstacle detection result. From *left to right* color image, ground truth, detected obstacles of our method and the method in [24]

detected and the ground-truth rectangles and the union of these rectangles is larger than 0.5.

We employ three evaluation measures that are precision, recall and F-measure. These measures are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$F = 2 \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

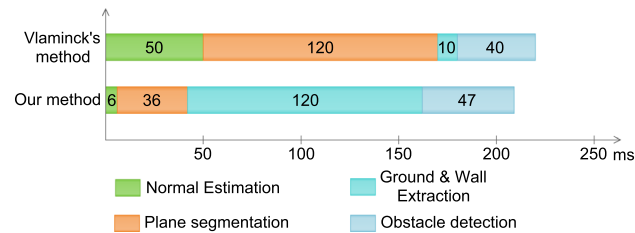
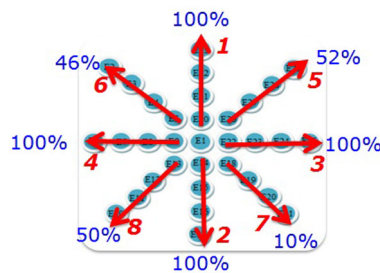
Figure 20 illustrates some examples of detection while Table 3 shows the quantitative evaluation. Our algorithm has a slightly higher F-score than method in [24], its has lower precision score but higher recall score, especially in the dataset 2, which has small ground region, the recall is significantly different between two methods (5.6% higher in pixel level and 12.4% higher in object level). In overall, our method produces less false alarms with a acceptable rate of true detection. This is because in Vlaminck's method [24] using RANSAC algorithm to segment plane and ground plane must be well identified in order to rotate the point cloud based on normal vector of detected ground plane then detect obstacle. So when the ground plane is wrongly detected or missed, it tends to consider the whole ground plane as a obstacle. That is why the precision with pixel level of method [24] is significantly higher than recall.

Concerning computational time, Fig. 21 shows the detection time of two methods. We tested both of them in same configuration of PC (an Intel Core i7-2720QM processor and 12 GB memory inside) and down-sample rate ( $2 \times 2$  block, which produces 76,800 points in point cloud). Both methods operate with average speed of 4–5 Hz (200 ms/frame). In our method, due to plane refinement by calculating distance from all points to detected plane, it occupied most of time while in [24] method, the most time-consuming part is plane segmentation using RANSAC. In general, this processing time is enough to be used in practice.

**Table 3** Obstacle detection results comparison with the method in [24]

		Pixel level			Object level		
		P	R	F	P	R	F
Overall	Our	76.9	80	78.4	63.5	73.4	68.1
	[24]	81.9	73.7	77.6	61.9	66.9	64.3
Dataset1	Our	68.3	73.6	70.9	51.7	56.2	53.8
	[24]	69.9	66.9	68.3	46.8	54.9	50.6
Dataset2	Our	85.6	<b>85.9</b>	85.8	75	<b>92.5</b>	82.8
	[24]	94.9	<b>80.3</b>	87	81.8	<b>80.1</b>	81

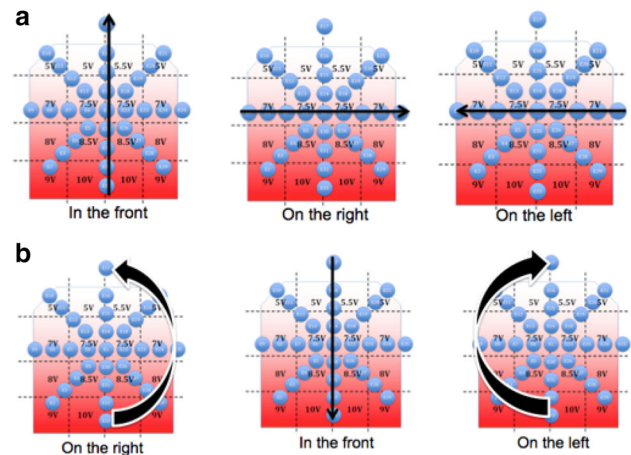
*P* Precision (%), *R* Recall (%), *F* F-Measure (%)

**Fig. 21** Detection time of each step of our method and the method in [24]**Fig. 22** Average accuracy of eight direction on the tongue

#### 4.5 User perception validation

In order to evaluate the performance of the proposed prototype system, a perception experiment was conducted for users. Based on the design of the electrode matrix and the idea of stimulation pulses, we used a sequence of electrodes to represent eight directions. Each direction corresponds to one radius line and the order of stimulating electrodes is from center to the edge of the tongue. Five participants took part in a training session to adapt to the device then they were asked for randomized directions. Figure 22 shows the average accuracy of perception calculated on five participants. The electrical intensity is generated based on the perception evaluation in Fig. 18.

According to the feedback of users, the edge regions of the tongue often gives good perception. Besides, the left and right-front parts of the tongue achieve higher accuracy than the rear parts. As a result, the obstacle warning representation is suitable for users.

**Fig. 23** Electrotactile representation of stationary and moving obstacle warning. **a** stationary object and **b** moving object

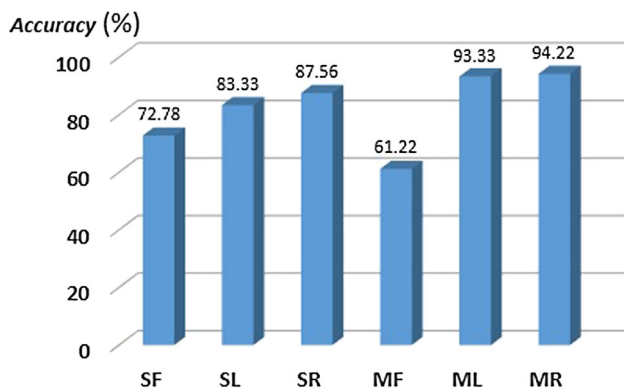
The resulted perception for main directions (left, right, forward and backward) are very promising to be used not only for supporting navigation in terms of directivity, but also can further improve the safety by giving detailed information through different representations on electrodes. Several research groups used tongue electrotactile feedback for different purpose for blind people and unbalanced people.

In existing researches [20,26,27], the systems normally have their basic forms of square or rectangular. Our prototype is destined to consume less energy and to be able to change voltage level. It is very important as the warning task requires informing the danger before the user gets very near the obstacle. The experiment and results on warning representation will be described in the next sections. Firstly, we will test with the direction when on the path, there is no obstacle. Then the experiment on obstacle warning will be detailed and discussed.

#### 4.6 Obstacle warning evaluation

The obstacle detection and warning is the major function that we aim at in our research. Based on the output information, the warning signals were generated and the tongue electrotactile system was again used to test this function. Due to the above results on the directions of stimulation impulses on the tongue, we choose the most precise directions: forward, left, right. In addition, the experiment on part 4.5, the edge of the tongue is more sensitive than the interior of the tongue. Figure 23 depicts the representation for stationary and moving obstacle warning for our system.

In Fig. 23, the arrangement of electrodes was made so as to bring the good perception to the users. As a consequence, we made use of the more sensitive regions on the tongue such as the edge of the tongue and the high percentage correction regions on the tongue. The stationary obstacle

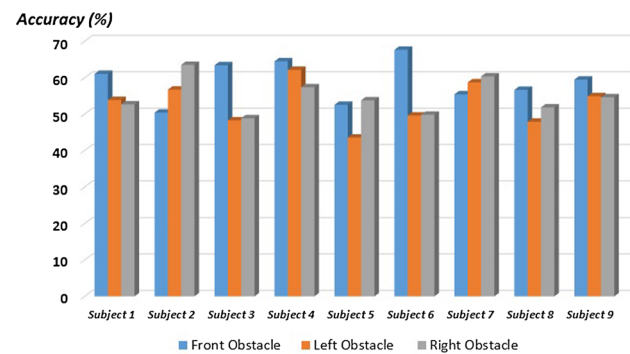


**Fig. 24** Distinction accuracy for obstacle warning: *S* stationary object, *M* moving object, *F* on the front, *L* on the left, *R* on the right

was warned by utilizing nine electrodes to indicate its position, while the moving one was alerted by employing the edge electrodes and backward direction. Firstly, the sensitivity test was implemented with nine blindfolded subjects with one voltage level to evaluate their perception capacity towards the position and the kind of object. Each participant performed two stages: the training stage and the perceiving stage.

In training stage, after the V0 value was decided for each participant, they will be trained for adaptation without moving to associate the electrical stimulations with the corresponding command. In perceiving stage, subjects were asked to say the command without knowing in advance. Figure 24 displays the accuracy of distinction of command for indicating position and status of objects. Among six stimulations, the sensitivity results for using the edge of the tongue are higher than using the interior of the tongue. In addition, using nine electrodes can sometimes cause confusion to users about two opposite directions because their stimulating signals use the same electrodes. If the two chains of impulses were struck too close in time, such as two SF impulses, user easily confuses SF for MF. This is also what the test subjects mentioned after the experiment. The same situation happens with the case of SL and SR. That is why the accuracies for SF, SL, SR and MF are below 90%.

In order to encode the warning signal to tactile representation, electrical stimulating intensity was varied according to the distance to the obstacles. Nine subjects were asked to take part in the obstacle avoidance experiment based on a pseudo-warning signals corresponding to moving and stationary obstacles at different positions while completing a trajectory in a building corridor. Some stationary obstacles such as fire extinguishers, flower pots and dustbins were placed arbitrarily along the way. Each participant must be trained for adaptation with the electrode array during 30 min before conducting the experiment. When the subject got nearer to the obstacle, the intensity of the results are shown in Fig. 25.



**Fig. 25** Obstacle warning result based on the position

Actually due to the hearing sense and the environment perception of the test subjects, the results here could not be totally accounted for the tongue electrotactile system. However, nearly all the subjects obtained higher than 50% accuracy when they travel in reality. For the case of front obstacle, the capacity of avoidance is really high because the representation on the electrode matrix for the front objects lies in only one region of the tongue, while left and right object can reach from 45% to around 62% of avoidance capacity. Not all subjects travelled at normal or low speed to have better perception and they were often curious about the tongue system and did not follow strictly the training stage. That is also why the results were not totally as expected to have higher rate of recognition. However, the accuracy rate can be promisingly improved if more subjects should be required to participate and asked to follow carefully the training stage.

## 5 Conclusion

In this paper, we proposed a system which is an integration between mobile Kinect with electrode matrix to help visually impaired people from obstacle while moving. Our system is designed to act as a mobility aid and perform the obstacle detection and warning task. Keeping in mind that users are visually impaired people, the information representation is simple, portable, hands and ears-free by using human tongue as the interface. The results indicate that under certain constraints, the imaging technique has so far been able to provide guidance cues, detect both stationary and moving obstacle, calculate rather precisely the depth information in order to give warning information at the right time. Although using tongue as the representation interface requires intensive study on the perception, the preliminary perception results show that it is totally possible to express the alert signal in this form and the electrical stimulation intensity can be adjusted attentively for the users.



The results of our experiment demonstrated that subjects were able to correctly interpret the directional signal provided by the wireless TDU. Interestingly, our results further showed that the tongue behavior is very flexible. Different regions on the tongue adapt to different voltages and recognition also based on the stimulation impulse. Moreover, different users have different levels of stimulation intensity. The outer and front part of the tongue have good perception and low voltage level, while the inner and rear part needs higher voltage activation. It is proved that people can be trained to adapt to a new sense to recover lost information due to impaired sensory modality.

Indeed, not all users can totally get used to this kind of device and the mobility still depends mainly on their natural feeling and instinct. Some visually impaired are not totally blind and they can follow the instruction by light cue. However, our results show that subjects can move independently with the instruction from the TDU but with care. This observation could be relevant for conducting future studies.

**Acknowledgments** This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant Number FWO.102.2013.08.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Bernabei, D., Ganovelli, F., Di Benedetto, M., Dellepiane, M., Scapigno, R.: A low-cost time-critical obstacle avoidance system for the visually impaired. In: International conference on indoor positioning and indoor navigation (IPIN) (2011)
- Calder, D.J.: Assistive technology interfaces for the blind. In: 3rd IEEE international conference on digital ecosystems and technologies, pp. 318–323, June (2009)
- Multi-kinect camera calibration. <http://doc-ok.org/?p=295>. Accessed 25 July 2016
- Chen, G., Can, Z., Jun, P.: An intelligent blind rod and navigation platform based on zigbee technology. In: 2011 International conference on E-Business and E-Government (ICEE), pp. 1–4, May (2011)
- Hersh, M., Johnson, M.A.: Assistive Technology for Visually Impaired and Blind People, 1st edn. Springer, London (2008)
- Hoang, V.N., Nguyen, T.H., Le, T.L., Tran, T.T.H., Vuong, T.P., Vuillerme, N.: Obstacle detection and warning for visually impaired people based on electrode matrix and mobile kinect. In: 2nd National foundation for science and technology development conference on information and computer science (NICS), pp. 54–59, Sept (2015)
- Holz, D., Holzer, S., Rusu, R.B., Behnke, S.: Real-time plane segmentation using rgb-d cameras. In: Röfer, T., Mayer, N.M., Savage, J., Saranlı, U. (eds.) RoboCup 2011: robot soccer world cup XV, pp. 306–317. Springer, Berlin (2012)
- Nicolas Burrus HomePage. <http://nicolas.burrus.name>. Accessed 25 July 2016
- Huang, H.C., Hsieh, C.T., Cheng-Hsiang, Y.: An indoor obstacle detection system using depth information and region growth. *Sensors* **15**, 27116–27141 (2015)
- Jameson, B., Manduchi, R.: Watch your head: a wearable collision warning system for the blind. In: 2010 IEEE sensors, pp. 1922–1927, Nov (2010)
- Joachim, A., Ertl, H., Thomas, D.: Design and Development of an indoor navigation and object identification system for the blind. In: Proc. ACM SIGACCESS accessibility, computing, pp. 147–152 (2004)
- Johnson, L.A., Higgins, C.M.: A navigation aid for the blind using tactile-visual sensory substitution. In: 28th Annual international conference of the IEEE engineering in medicine and biology society, pp. 6289–6292 (2006)
- Kaczmarek, K.A., Webster, J.G., Bach-y Rita, P., Tompkins, W.J.: Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Trans. Biomed. Eng.* **38**(1), 1–16 (1991)
- Nguyen, T.H., Le, T.L., Tran, T.T.H., Vuillerme, N., Vuong, T.P.: Antenna design for tongue electrotactile assistive device for the blind and visually impaired. In: 7th European conference on antennas and propagation (2013)
- Nguyen, T.H., Nguyen, T.H., Le, T.L., Tran, T.T.H., Vuillerme, N., Vuong, T.P.: A wearable assistive device for the blind using tongue-placed electrotactile display: design and verification. In: International conference on control, automation and information sciences (ICCAIS), pp. 42–47 (2013)
- Nguyen, T.H., Nguyen, T.H., Le, T.L., Tran, T.T.H., Vuillerme, N., Vuong, T.P.: A wireless assistive device for visually-impaired persons using tongue electrotactile system. In: Advanced technologies for communications (ATC), 2013 international conference on, pp. 586–591, Oct (2013)
- Rodriguez, S.A., Yebes, J.J., Alcantarilla, P.F., Bergasa, L.M., Almazan, J., Cela, A.: Assisting the visually impaired: obstacle detection and warning system by acoustic feedback. *Sensors* **12**, 17476–17496 (2012)
- Sainarayanan, G., Nagarajan, R., Yaacob, S.: Fuzzy image processing scheme for autonomous navigation of human blind. *Appl. Softw. Comput.* **7**(1), 257–264 (2007)
- Solomon, N., Bhandari, P.: Patent landscape report on assistive devices and technologies for visually and hearing impaired persons. Technical report, Patent landscape report project (2015)
- Tang, H., Beebe, D.J.: An oral tactile interface for blind navigation. *IEEE Trans. Neural Syst. Rehabil. Eng.* **14**(1), 116–123 (2006)
- Tang, T.J.J., Lui, W.L.D., Li, W.H.: Plane-based detection of staircases using inverse depth. In: Australasian conference on robotics and automation (ACRA) (2012)
- Brainport Technology. <http://www.wicab.com>. Accessed 25 July 2016
- Ulrich, I., Nourbakhsh, I.: Appearance-based obstacle detection with monocular color vision. *AAAI* (2000)
- Vlaminck, M., Jovanov, L., Van Hese, P., Goossens, B., Wilfried, P., Aleksandra, P.: Obstacle detection for pedestrians with a visual impairment based on 3d imaging. In: 2013 International conference on 3D imaging (IC3D), pp. 1–7. IEEE (2013)
- The VOICE. <http://www.seeingwithsound.com>. Accessed 25 July 2016
- Vuillerme, N., Pinsault, N., Chenu, O., Fleury, A., Payan, Y., Demongeot, J.: A wireless embedded tongue tactile biofeedback system for balance control. *Pervasive Mob. Comput.* **5**, 268–275 (2009)
- Bach y Rita, P., Kaczmarek, K.A., Tyler, M.E., Garcia-Lara, J.: Form perception with a 49-point electrotactile stimulus array on the tongue: a technical note. *J. Rehabil. Res. Dev.* **35**(4), 427–430 (1998)