# Reducing the minimum range of RGB-D sensor to aid navigation in visually impaired individuals

**KAILUN YANG, KAIWEI WANG,* HAO CHEN AND JIAN BAI**

*State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou 310027, China*

**Corresponding author: wangkaiwei@zju.edu.cn*

Received XX Month XXXX; revised XX Month, XXXX; accepted XX Month XXXX; posted XX Month XXXX (Doc. ID XXXXX); published XX Month XXXX

**The introduction of RGB-Depth (RGB-D) sensors harbors a revolutionary power in the field of navigational assistance for the visually impaired. However, RGB-D sensors are limited by a minimum detectable distance of about 800mm. This paper proposes an effective approach to decrease the minimum range for navigational assistance based on a RGB-D sensor of RealSense R200. A large-scale stereo matching between two IR images and a cross-modal stereo matching between one IR image and RGB image are incorporated for short-range depth acquisition. The minimum range reduction is critical not only for avoiding obstacles up close, but also in the enhancement of traversability awareness. Overall, the minimum detectable distance of RealSense is reduced from 650mm to 60mm with qualified accuracy. A traversable line is created to feedback visually impaired individuals through stereo sound. The approach is proved to be with usefulness and reliability by a comprehensive set of experiments and field tests in real-world scenarios involving real visually impaired participants.**

***OCIS codes:*** *(100.6890) Three-dimensional image processing; (130.6100) Sensors; (150.6044) Smart cameras.*

http://dx.doi.org/10.1364/AO.99.099999

## 1. Introduction

According to World Health Organization, 253 million people are estimated to be visually impaired and 36 million are blind in the world [1]. Short-range obstacle avoidance and traversability awareness are two fundamental topics in the research area of navigational assistance for the visually impaired [2-10], as well as mobile robotics [11-14], unmanned driving [15-18], autonomous agriculture [19-20] and augmented reality [21-22].

RGB-Depth (RGB-D) sensors, which deliver RGB streams together with depth information, are becoming increasingly popular for these tasks. The main reason is that ranging technique with RGB-D sensors provides good portability, functional diversity and cost effectiveness. However, typical RGB-D sensors, including light-coding sensors and stereo cameras all have a minimum range to output valid depth value.

Light-coding sensors, such as PrimeSense, Microsoft Kinect, Asus Xtion Pro as well as Mantis Vision MV4D, consist of an IR laser projector which emits structured near-IR patterns of speckles into the scene to encode objects and then an IR image sensor captures the speckles [23]. The distortions of speckles are deciphered and the depth map is generated through triangulating algorithms. However, short-range speckles are hard to identify due to over-exposure in IR images, which means the reflected structured light pattern is sufficiently bright to saturate the image sensor. As a result, these light-coding sensors leave out short-range speckles which restrict the minimum range of detection, i.e. about 800mm in the case of Microsoft Kinect and Asus Xtion [24-25].

Stereo cameras, such as PointGrey Bumblebee, ZED and DUO, estimate depth map through stereo matching of images from two or more lenses. Points on one image are correlated to another image and depth is calculated via disparity, which is the shift between a point on one image and another image. The minimum range of stereo camera is determined by the overlapping field view of both cameras and the search range of disparity in corresponding algorithms [26]. For example, the minimum depth range of ZED stereo camera is 1000mm [9].

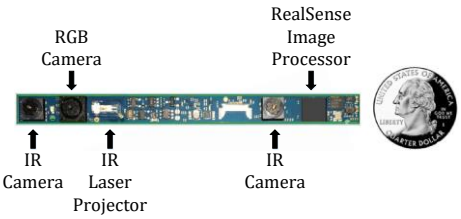


Fig. 1. The RGB-D sensor of RealSense R200, which is a good option for navigational assistance from miniaturization perspective.

RealSense R200 [27] is a RGB-D sensor that exploits a combination of active speckles projecting and passive stereo triangulating to derive disparity maps, and subsequently depth images, which is referred to as active stereo [28]. As shown in Fig. 1, RealSense R200 consists of an IR laser projector, a RGB camera, two IR cameras and an image processor. IR laser projector emits static non-visible near-IR speckles into the scene which is then attained by the left and right IR cameras. A depth map is

generated by the image processor through embedded stereo matching algorithm. The combination endows RealSense R200 the ability to work under both indoor and outdoor circumstances. In texture-less indoor environment, the projected patterns enrich textures on simple scenarios such as white wall. In sunny outdoor environment, although projected patterns are submerged by natural light, near-IR component of sunlight shines on the scene to form well-textured IR images. With the contribution of abundant textures to robust stereo matching, RealSense R200 is quite suitable for assisting the visually impaired thanks to its small size and environmental adaptability.

However, the combination brings about range restriction originated from both active speckles projecting and passive stereo matching. At close range, the overexposed regions in IR image lack in sufficient textures for stereo matching. Additionally, the narrow overlapping field of view leads to the obvious close blind area given the baseline distance of about 70mm. According to the technique overview [27], RealSense uses a Census cost function and performs a limited disparity search to compare left and right images. All depth points generated by the hardware correlation engines are high-quality photometric matches between the left-right stereo pairs. This allows the algorithm to scale well to noisy images. However, due to the fixed disparity search range, the overexpose and the narrow overlapping field of view, the embedded algorithm fails to acquire close-range depth information and RealSense R200 only outputs depth value greater than the threshold of 650mm.

In indoor environment, the normal operating distance ranges from 650mm to 2100mm in VGA resolution [29] while the sensor is able to deliver larger depth values outdoors, depending on illumination and textured conditions. If an object is within the detection range of 650mm, there is a black hole in the depth image as shown in Fig. 2(a)(b), and pixels in the black hole have no valid depth. In the case of navigational assistance, obstacles in close blind area cannot be perceived well and detected easily. As a result, sight impaired individuals are frequently left vulnerable in dynamic environment suffering from finding walkable directions as shown in Fig. 2(c)(d). Thereby, minimum range reduction for obstacle avoidance and traversability awareness is clearly desirable.

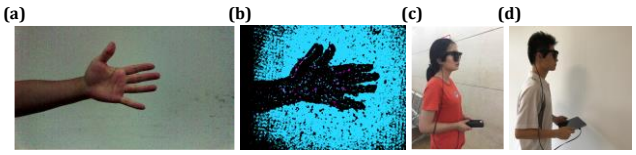


Fig. 2. (a)(b) The color image and depth image acquired with RealSense R200, the hand is within minimum range of detection so as to form black holes and mismatching pixels in depth image; (c)(d) Users who encounter obstacles in close blind area when wearing an assisting prototype.

## 2. Related work

In order to expand the range of RGB-D sensors, researchers have investigated the possibilities of many approaches for short-range depth acquisition. These approaches fall into several dominating categories of techniques: optics modification, deployment of multiple RGB-D sensors, information fusion as well as 3D simultaneous localization and mapping (SLAM) [30].

As for the modification of optics, Nyko Zoom is a commercial wide-angle optical adaptor for Microsoft Kinect. It reduces both the minimum and maximum range of Kinect, but pronounced distortion in depth image is introduced. Draelos et al. [12, 25] compensated the lens-introduced distortion through a depth calibration procedure and decreased the minimum range of Kinect by approximately 30%. Tomari et al. [31] addressed the distortion issue with a two-staged strategy by correcting pixel locations with an inverse radial distortion model and rectifying depth values with a neural network filter based on laser-assisted training.

Multiple RGB-D sensors are integrated in some systems to obtain a wilder field of view and decrease the minimum detection range. However, multiple light-coding sensors with overlapping views produce interference effects from overlapping speckles. Alhwarin et al. [24] used two IR images of two light-coding sensors as a stereo pair to generate a depth map. By combining active stereo depth and original depth, this method sidesteps the interference problem and decreased the minimum distance of Asus Xtion from 800mm to 500mm at a baseline of 45mm. Vorapatratorn et al. [6] performed a similar combination by utilizing two light-coding sensors for segmenting obstacles from the background. Nevertheless, the working distances were not reported. Shröder et al. [32] used a spinning shutter to block the IR emitter on each Kinect in turn to mitigate the interference. The framerate decreases as laser speckles from each light-coding sensor cannot access the scene all the time. While inspiring, appending a spinning shutter is unsuitable for a wearable assistance system with the additional weight. Maimone and Fuchs [33] applied a small vibration with a simple motor to a subset of light-coding sensors to alleviate the interference, which contributes the burring in RGB images as the color camera works in rolling shutter mode and poses challenges for detection algorithms to accurately locate obstacles with the continuous movement of sensors.

Information fusion against the range limitation is adopted by some researchers through combining RGB image with depth image or IR image. In order to provide the visually impaired with obstacle-free paths, Aladren et al. [5] firstly detects ground with RANdom Sample Consensus (RANSAC) [34], then extends the depth based ground segmentation with RGB image. This method is quite suitable to expand detection result to longer range, but not robust enough to acquire short-range information and the algorithm with unpromising framerate fails to provide upper-level assistance at normal walking speed. There are researchers complemented the depth image of RGB-D sensor through cross-modal stereo matching between RGB and IR image [35-37]. The minimum range is reduced, since a wide overlapping field could be obtained due to the short baseline of IR camera and RGB camera in a RGB-D sensor. Although related, these methods focus more on the depth restoration of transparent and specular surfaces, and attach less importance to depth accuracy at close range.

In terms of 3D simultaneous localization and mapping (SLAM), the based solution could build a vicinity map. In this fashion, instead of original depth image, short-range information is acquired through the vicinity map. Lee and Medioni [2, 38] adopted a metric-topological SLAM approach to provide the visually impaired with 3D traversability on the map. This method achieves real-time processing speed and improves the mobility performance, but still suffers from the loss of short-range depth when spinning too fast or working in crowded real-world environments with many independently moving objects.

Although a number of related work have addressed the problem, they do not decrease the minimum range to a large extent or cause intolerable side-effects in navigational assistance. In this paper, we focus on the short-range depth imaging to enhance the assistance in terms of obstacle avoidance and traversability awareness for visually impaired individuals. A commercial off-the-shelf RGB-D sensor in RealSense R200 is used without modifying hardware nor requiring any additional cameras. The minimum range reduction scheme is the combination of a large-scale stereo matching algorithm between two IR images as well as a cross-modal stereo matching algorithm between RGB image and IR image. To feedback visually impaired individuals through depth-sound mapping, a traversable line is generated after minimum range reduction. The approach has been integrated in a wearable prototype and tested in real-world scenarios with real visually impaired volunteers.

We have already presented some preliminary studies related to blind assistance in [7-10]. Detection algorithms of local ground plane and long-range traversable area were developed in [7-9] while [9] also addressed the avoidance of water hazards. In [10], we have made the first attempt to decrease the minimum range of RGB-D sensor, namely from 650mm to 165mm. In this paper, we considerably extend previously established proof-of-concepts, where the scheme to largely reduce the minimum range is explained. In addition, we include novel contributions and results tested with real visually impaired people to validate the effectiveness of our solution:

(1) A minimum range reduction approach for the RGB-D sensor in RealSense R200, namely from 650mm to 60mm with qualified accuracy.

(2) A depth perception scheme with the combination of large-scale stereo matching algorithm and cross-modal stereo matching algorithm.

(3) A navigational assistance framework on a wearable prototype for visually impaired individuals.

This remainder of this paper is organized as follows. In Section 3, the proposed approach is presented in detail. Section 4 describes the experiments in terms of accuracy test, detection results and field tests. Section 5 are the conclusions and outlooks to future work.

## 3. Approach

In this section, the approach to reduce the minimum range of navigation assistance is elaborated in detail. As the pipeline of the approach shown in Fig. 3, color image, original depth image and IR image pair are acquired with the RGB-D sensor, while we implement our algorithm to obtain a minimum range decreased depth image and use the depth image to provide audio feedback to visually impaired people through the created traversable line. The minimum range reduction serves as a crucial step to triangulate the large-scale stereo pair and cross-modal stereo pair. In this contribution, the minimum depth range of RGB-D sensor in RealSense R200 is firstly decreased from 650mm to 160mm by large scale stereo matching, and the cross-modal stereo matching is the key enabler to further decrease the minimum detectable range to 60mm, which are described in the subsections.

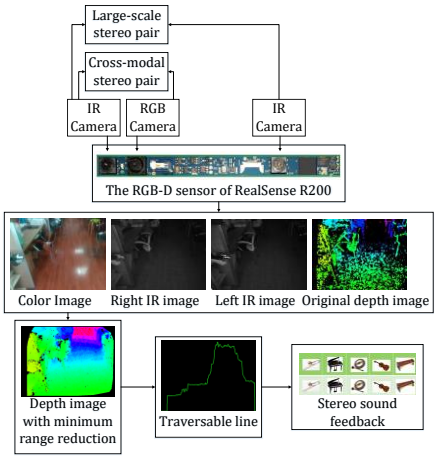


Fig. 3. The pipeline of the presented approach.

### A. Large-scale stereo matching

The minimum range of original depth image is around 650mm and there are many holes, noises and mismatched pixels as shown in Fig. 4(a)(b). However, the original depth image is delivered by the stereo matching algorithm fixed in the RealSense processor which is unable to be altered. Still, the embedded algorithm is based on local correspondences, which prevents the sensor from delivering dense original depth map in texture-less scenarios, especially short-range regions as shown in Fig. 4(a)(b). In addition, parameters are preset with the algorithm, such as the matching score and texture threshold.

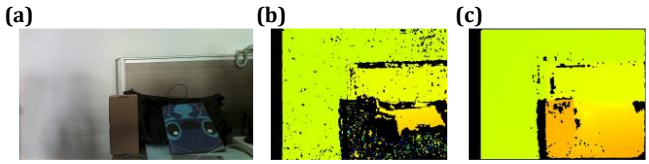


Fig. 4. (a) Color image; (b) Original depth image; (c) Large-scale matched depth image.

Comparatively, IR images from the RealSense are large-scale matched in our work. To yield a minimum range decreased depth map with calibrated IR images, not only the disparity search range is expanded in the algorithm, namely from 64 to 255, original efficient depth pixels are also included in the design of a variant version of efficient large-scale stereo matching algorithm [39]. Here, supported pixels are denoted as pixels which can be robustly matched due to their textures and uniqueness, and these pixels are obtained using Sobel masks with the fixed size of 3 by 3 pixels and a large-scale disparity search range to perform stereo matching. In our implementation, instead of uniformly selecting support points, we perform several simple and effective steps to determine the support points. Beyond Sobel filters responses, which are insufficient for stereo matching, original depth image pixels are added to the support pixels. In addition, a multi-block-matching principle [40] is employed to obtain more robust and sufficient support matches from real-world textures even with short-range overexposed regions in the IR pair as shown in Fig. 5. Given the resolution of IR images 628 by 468, the appropriate block sizes found are 41 by 1, 1 by 41, 9 by 9 and 3 by 3. After that, following [39], the approach estimates the depth map by forming triangulation on a set of support pixels and interpolating disparities. As shown in Fig. 4(b)(c) and Fig. 5(d)(e), the large-scale matched depth image is much denser than the original depth map in close range.


Fig. 5. (a)(b) Left and right IR image pair; (c) Color image; (d) Original depth image; (e) Large-scale matched depth image.

### B. Cross-modal stereo matching

In order to further decrease the minimum range to a great extent, we use the RGB camera and the IR camera close to the RGB camera to establish a stereo camera system whose baseline distance is short enough (11.7mm in the case of RealSense R200) to ensure the detection of objects in close range.

As a prerequisite step of cross-modal matching, the RGB image is rectified using parameters from RealSense calibration software [41]. Subsequently, instead of searching for the optimal weights to convert the RGB image into a fake IR image for IR stereo matching as proposed in [35-36], the idea from vector quantified imaging coding [42] is introduced to reconstruct the fake IR image. For obstacle avoidance, note that this reconstruction assumes that particularly close objects cover quite a large area in the images. Following the assumption, we adequately considered the real-time circumstance by reconstructing the fake IR image based on K-means clustering algorithm to improve the suitability in different environments. The flow chart of the reconstruction is fully depicted in Fig. 6 where we aim to cluster the color image in RGB space. For pixels of color image in each cluster, the grayscale in the reconstructed fake IR image is assigned with the same

value as the grayscale of the pixel corresponding to the initial clustering center in IR image.

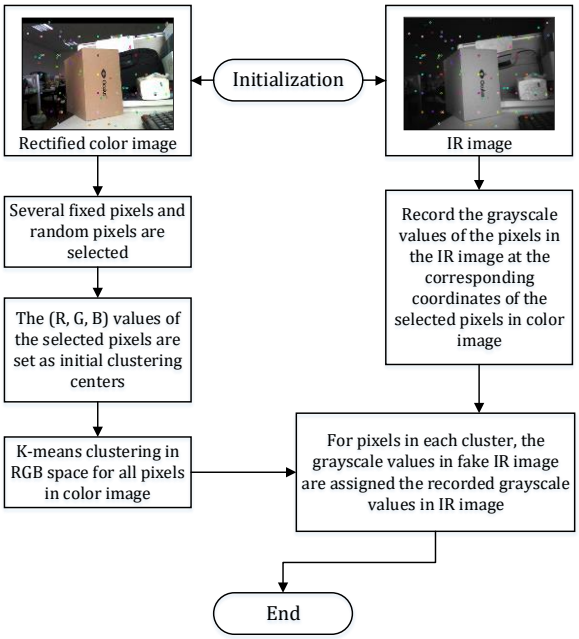


Fig. 6. The flow chart of the reconstruction of color image into a fake IR image based on K-means clustering algorithm.

The initial clustering center pixels contain two types: pixels at fixed coordinates and pixels which are randomly elected. Given the resolution of the rectified color image 628 by 468, we collected $M$ equally distributed fixed pixels and randomly elected $N$ pixels at different coordinates. The (R, G, B) values of the selected pixels are set as the initial clustering centers. In this way, we obtain $K$ clustering centers as calculated in Eq. (1) where $M$ is empirically set to 35 to ensure a basic number of grayscale levels for cross-modal stereo matching. Comparably, $N$ could be specifically set and equals to 15 generally in the case of RealSense R200 to leverage a rational trade-off between efficiency and matching capacity.

$$K = M + N. \quad \textbf{(1)}$$

In the IR image with the resolution 628 by 468, grayscale values of the pixels at the corresponding coordinates of the selected clustering center pixels in color image are recorded. Thereupon, K-means clustering is implemented in RGB space for all pixels in the color image. To speed up real-time assistance, only one iteration is executed instead of using the convergence of objective function in K-means as the stopping criterion. After the clustering, each pixel in the color image belongs to a cluster. For pixels in each cluster, the grayscale values in the fake IR image are assigned with the recorded grayscale values in IR image. In this light, the reconstructed fake IR image shares the same data domain of the IR image and has $K$ levels of grayscale as shown in Fig. 7.


Fig. 7. (a) Color image; (b) IR image; (c) The reconstructed fake IR image.

As discussed in Section 1, the over-exposed or over-dense speckles occurring in IR images is one of the main reasons why light-coding sensors and active stereo type of RGB-D sensor RealSense could not calculate the depth of short-range objects. In respond to this issue, we proposed to exploit the over-exposed regions of IR speckles in the IR image and the fake IR image as a stereo pair to generate short-range depth. As these regions tend to be edge-less with over-dense speckles, they could be easily extracted with a typical Canny edge detector. After that, over-dense regions correspond to short-range objects. These regions in the IR image and the fake IR image reconstructed from the rectified color image are adopted as an IR stereo pair for disparity calculation through a block matching stereo algorithm. This algorithm is previously presented in [10], which is based on local correspondences to acquire an edge depth image. In this regard, it allows to generate early warning of extremely close obstacles in real time by using this edge depth image. As shown in Fig. 8(b), the large-scale depth image is combined with the cross-modal depth image to form the synthetic depth image, which comprises the edge depth regions acquired with cross-modal stereo matching. The synthetic depth image is generated by replacing depth value of invalid pixels in large-scale depth image with the corresponding one from the cross-modal depth image as calculated in Eq. (2). As for each pixel, the depth equals large-scale depth, if the $Bool$ of the pixel equals to 1, which means the depth of the pixel in the large-scale depth image is valid. Otherwise, the depth equals cross-modal depth.

$$d^{synthetic} = d^{large\text{-}scale} \times Bool + d^{cross\text{-}modal} \times (1 - Bool). \quad \textbf{(2)}$$

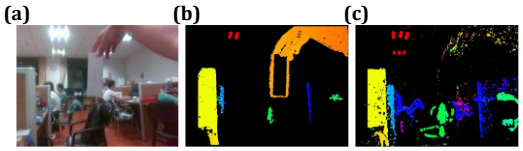


Fig. 8. (a) Color image; (b) Synthetic depth image; (c) Original depth image without close-range depth information.

### C. Traversability awareness through stereo sound feedback

After large-scale stereo matching and cross-modal stereo matching to reduce the minimum range of the RGB-D sensor, a traversable line to represent the traversable distances in different directions is proposed. For producing the stereo sound and yielding a complete line, a guided filter introduced in a previous work [7] is used in advance to refine the depth map in terms of hole-filling and density-enhancing. Subsequently, we first separate feasible ground area from hazardous obstacles with the method proposed in our previous work [5], then locate the closest valid pixel with the minimum depth value $Z_m$ in each direction. In this manner, the traversable line constituted by the minimum depth values is obtained as shown in Fig. 9. Additionally, we implemented an obstacle detection method to warn against ultra-close hazards based on Mean-Shift algorithm [43] to cluster depth pixels which tend to be congregated together. As a result, the audio interface generates a prompt to help visually impaired people to be aware of the close hazards after the obstacle segmentation.

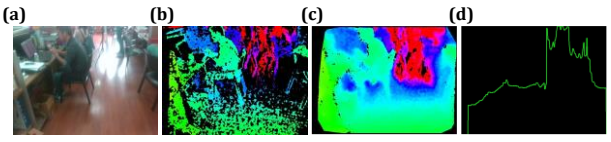


Fig. 9. (a) Color image; (b) Original depth image; (c) Depth image with minimum range reduction and guided enhancement; (d) Traversable line.

In addition, this paper uses a variant of non-semantic audio interface with respect to previous work [9]. To feedback to visually impaired people, we perform depth-sound mapping by using the traversable line which signifies the traversability in different directions. As shown in Fig. 10, the directions of traversable distances are differentiated not only by sound source locations in virtual 3D space and the directions of stereo sound, but also by the musical instruments, whose timbre differs from each other. The generation of the stereo sound follows rules below to guide and attract visually impaired individuals to take the prioritized direction to navigate the traversable path and detour around hazardous obstacles:

(1) Divide the traversable line into five sections which correspond to the five different musical instruments.

(2) The horizontal field view of the RGB-D sensor is 57.1°, so each musical instrument corresponds to the traversable line with a range of 11.42°.

(3) Each direction of traversable distance is represented by a musical instrument in virtual 3D space.

(4) For each musical instrument, the bigger the sum of height in the corresponding section of the traversable line, the greater the sound from the instrument.

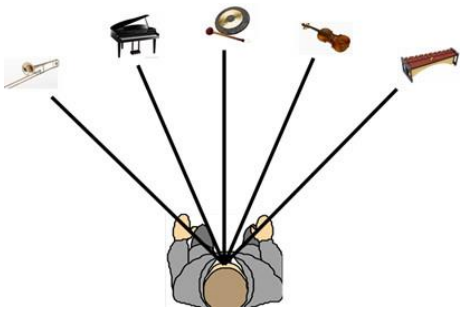

Fig. 10. Stereo sound feedback to signify the traversability in different directions to visually impaired individuals via the sonification of five instruments including trumpet, piano, gong, violin and xylophone in 3D space.

## 4. Experiments

The presented approach has been evaluated with several experiments including ranging accuracy test, obstacle detection, traversability awareness as well as field tests. The accuracy test is performed to analyze the ranging accuracy of large-scale and cross-modal stereo matching and study whether the requirement for accuracy of navigational assistance has been met. Obstacle detection and traversability awareness are performed to study the effectiveness of detecting various obstacles and determining traversable directions as well as the running time of the algorithm. Field tests are designed to check whether the presented approach effectively assists navigation.

### A. Accuracy test

The accuracy test is performed separately in terms of original depth ranging, large-scale depth ranging and cross-modal depth ranging and the results are shown in Fig. 11. The relative accuracy is calculated in comparison with the result of the laser ranging, which is set as the true value and the accuracy of the laser ranger is 0.001m. In terms of the original depth ranging whose minimum detection distance is around 650mm, the relative accuracy is less than 1.2%, and the slope coefficient of the linear fit equals to 1.0003 with the fitting goodness $R^2$ equaling to 0.9999. As for the large-scale depth ranging whose minimum distance has been decreased to around 160mm, the relative accuracy is less than 1.1%, while the slope coefficient equals to 1.0006 with $R^2$ close to 1. It could be seen that the large-scale depth ranges not only closer but also slightly more accurate than the original depth image. Meanwhile, the cross-modal depth ranges from 60mm to 200mm, which is very close, and the slope coefficient equals to 0.9992 with $R^2$ equaling to 0.9939, and the relative accuracy is less than 4.8%.

Generally, the ranging deviation increases as distance increases in stereo systems. In the ranging formula Eq. (3), depth $d$ is calculated where $T$ is the baseline distance and $\Delta$ is the disparity value generated with stereo matching. From Eq. (3), we deduce Eq. (4)(5). In a passive stereo system, the errors $|\partial\Delta|$ in disparity space are usually constant which stem from imaging properties and the quality of the matching algorithm. For active stereo, the condition holds until imaging noise overwhelms projector intensity. As the metric we use in the accuracy test is to record depth at which a perpendicular white wall returns greater than 95% of its measurements in the center of the field of view, so the condition of disparity error constancy is satisfied. Simple manipulations of Eq. (5) gives Eq. (6), which means the relative accuracy $\varepsilon_{\%}$ is proportional to the depth $d$ .

$$d = \frac{f \times T}{\Delta}. \tag{3}$$

$$\frac{\partial d}{\partial \Delta} = -\frac{f \times T}{\Delta^2}. \tag{4}$$

$$|\partial d| = \frac{f \times T}{\Delta^2}|\partial \Delta|. \tag{5}$$

$$\varepsilon_{\%} = \frac{|\partial d|}{d} = |\partial \Delta|\frac{d}{f \times T}. \tag{6}$$

However, the relative accuracy of large-scale stereo matching is better than cross-modal stereo matching which has a shorter minimum detectable distance. This is because the results are acquired through two stereo systems with different baselines. The large-scale stereo matching is carried out between two original IR images and the cross-modal stereo matching is carried out between one IR image and the fake IR image reconstructed from the rectified RGB image. For this reason, the shorter baseline gives rise to the relative error. In addition, textures which benefit stereo matching are poorer in closer range due to the overexposure in active stereo type of RGB-D sensor, thus the errors $|\partial\Delta|$ in disparity space could not be assumed equivalent in the two stereo systems. To briefly summarize, the ranging error of the range within 3m is lower than 1.0cm and lower than 3.2cm within 5m. Apparently, the accuracy satisfies the requirement of navigational assistance for the visually impaired in terms of obstacle avoidance and traversability awareness.

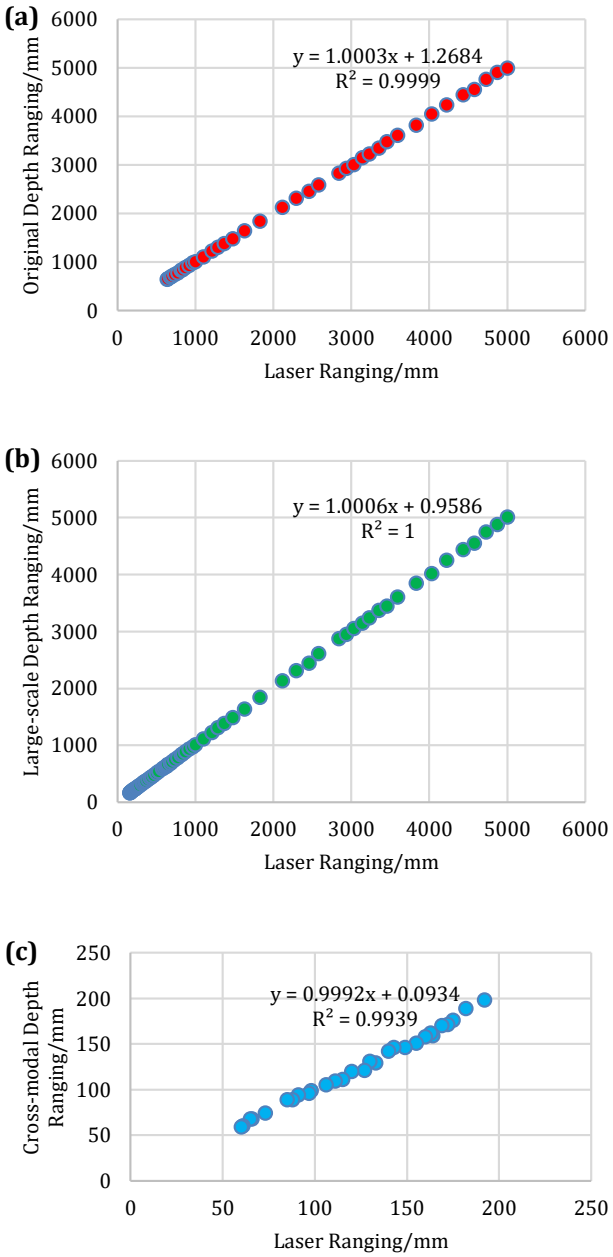


Fig. 11. (a) Original depth ranging (minimum detectable distance 650mm) in comparison with laser ranging; (b) Large-scale depth ranging (minimum detectable distance 160mm) in comparison with laser ranging; (c) Cross-modal depth ranging (minimum detectable distance 60mm) in comparison with laser ranging.

### B. Obstacle detection and traversability determination

The effectivity of the approach to reduce minimum detectable distance is evaluated for detecting obstacles. Fig. 12 shows examples of depth maps and detection results using the Mean-Shift segmentation algorithm, where the second column and the third column show the original depth images and the depth images with minimum range reduction respectively. Based on RealSense depth estimation, we observe that nearly all short-range objects are either missed or filled with mismatched pixels and noises in the original depth image. In comparison, the depth images with minimum range reduction appear to deliver dense depth information of the close-range objects. Thereby, the Mean-Shift based segmentation algorithm is able to detect close-range obstacles as shown in the fourth column of Fig. 12. As a result, the minimum range reduction is quite effective to enhance the obstacle avoidance for the visually impaired in their daily life.

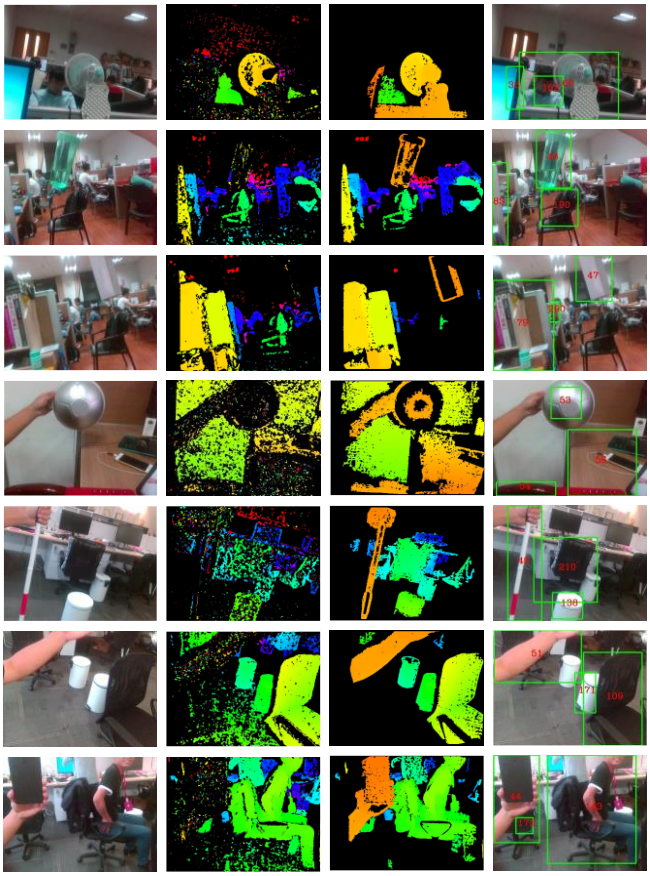

Fig. 12. Obstacle detection results using depth images with minimum range reduction. First column: color images; second column: original depth images; third column: depth images with minimum range reduction; fourth column: obstacle segmentation results with the average depth marked in scale of centimeter.

As the minimum range of the depth perception with the RGB-D sensor of RealSense has been decreased to around 60mm, different obstacles at different distances ranging from 60mm to more than 5000mm can be detected. Fig. 12 and Fig. 13 exhibits a wide set of qualitative examples, where the detection of obstacles of different materials, textures and distances are addressed. Here, we produce a bounding window to present the segmentation results of close-range objects in color images. Additionally, the average depth of the 2D bounding segmented object is marked in scale of centimeter. Based on our minimum range reduction scheme, common objects within the reach of visually impaired people are correctly detected thanks to the range expansion, including electric fan, semitransparent water bottle, texture-less boxes, metallic surface, matte surface, white cane, human face, hand, arm, mobile phone and loudspeaker. Moreover, obstacles of different sizes and locations on the ground which would impede the navigation of visually impaired people are also correctly detected, including chairs, garbage cans, human body, cart, cabinet, air conditioner, football, traffic cone and umbrella.

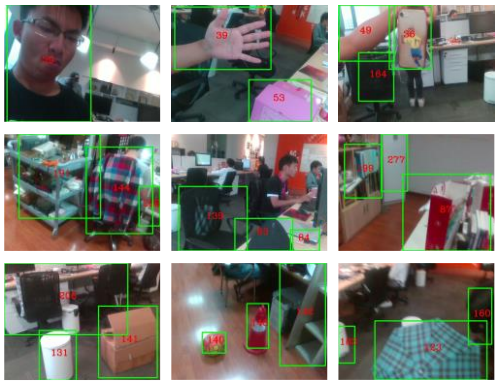

Fig. 13. Obstacle detection results.

As far as the traversability awareness is concerned, we found out that the depth map with minimum range reduction not only benefits the early warning of close obstacles but also enables a complete and smooth traversable line compared with the result generating from the original depth map as shown in Fig. 14, which is full of noises and black holes within close range. As a result, the stereo sound feedback is comparably more stable and would not confuse visually impaired people. The total computing time of a single frame on a portable processor with a 2.4GHz CPU is 194ms, which leads to a feasible 5 FPS framerate for obstacle avoidance and traversable direction determination.

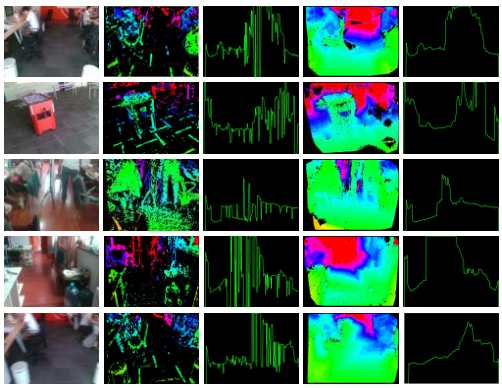

Fig. 14. Traversability awareness and the stereo sound interface generated with/without minimum range reduction. First column: color images; second column: original depth images; third column: traversable lines generated with original depth image; fourth column: depth images with minimum range reduction and guided enhancement; fifth column: traversable lines generated with minimum range decreased depth images.

### C. Field test

Two closed-loop field tests were conducted in an office and a corridor respectively. In these field tests, the presented approach has been integrated in an assisting system. As shown in Fig. 15, the wearable prototype is composed of a RGB-D sensor of RealSense R200 to capture three dimensional information of the environment, a bone conduction headset to transfer stereo sound feedback to visually impaired individuals, a pair of environmentally friendly resin lenses, a USB3.0 high frequency communication cable to transmit data and a portable processor. The assisting prototype is not only wearable but also ears-free and hands-free, because the bone conducting stereo sound feedback will not prevent the ears of visually impaired people from hearing environmental sounds, and the processor could be carried in pockets (Fig. 15(c)) instead of being held in hand (Fig. 15(b)).

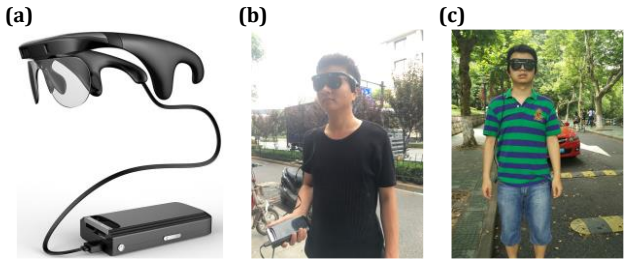


Fig. 15. (a) The wearable assisting prototype including the RGB-D sensor of RealSense, a bone conduction headset and a portable processor; (b) The user wears the prototype and holds the processor; (c) The user wears the prototype with the processor in the pocket.

Twenty-one visually impaired volunteers participated in the field test in an office as shown in Fig. 16. In this field test, the navigation assistance performance was compared with/without the minimum range reduction of the RGB-D sensor. As a comparison task, each one of them first completed with short-range depth information complemented. After that, they were asked to finish the task without close-range depth information, which means the obstacle avoidance and traversability awareness were operated with original depth output from RealSense.

During this assistance study, participants would learn the stereo sound feedback in the first place. The working pattern of the system and signals from the bone conduction headset were introduced. Each participant had five minutes to learn, adapt to the audio interface, and wander around casually. After that, participants were asked to traverse through obstacles without collisions and walk around the office, finally return to the start region where the Fig. 16 was taken. All visually impaired participants completed the test. The number of collisions and time to complete the test were recorded. Collisions include collisions with obstacles such as desks, foosball table, chairs, walls and so on. The timer started when a participant was sent to the start region and stopped when the participant completed a single test.

Fig. 16. Field test scenario in an office.

As shown in Fig. 17, participant needed an average time of 130.81s to finish a single test to walk around the office hearing the sound to detour around close obstacles and find the traversable directions with minimum range reduction. Each one collided into obstacles 0.95 times on average. In comparison, when finishing the task without minimum range reduction, the mean collision times and traversing time were 1.86 and 136.67s respectively. It is convinced that the minimum range reduction is extremely important for safety-critical blind assistance as the collision times were nearly 2 times of the number in the condition without close-range information. In addition, average traversing time were slightly lower with short-range information owing to the consistent feedback enabled by the denser depth image. It can be ruled out the possibility that decrease of the number of collisions is due to variation of familiarity of the sound feedback or the system. Because the test was performed with minimum range reduction first, it would help

improve rather than weaken the performance of the navigation without short-range information, if they were more familiar or better trained with the device afterwards.

Intriguingly, most of the collisions occurred when the user had already bypassed the obstacle but still scratched the sides. It is worth mentioning that the participant who collided with obstacles most times misunderstood the directions of the stereo sound, which guides the user to take prioritized direction to walk instead of notifying the directions of obstacles. Thereby, if his data is pruned, each one collided into obstacles 0.8 times on average with the full depth information. Altogether, eleven participants, more than half of all volunteers, never collided into obstacles when performed the navigation task using our assistive approach.

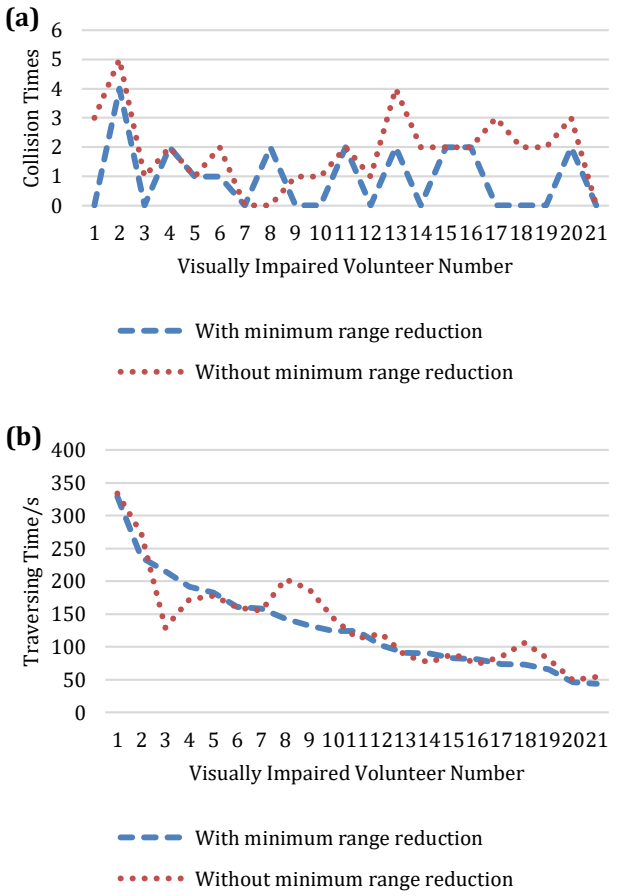


Fig. 17. Experiment data sorted in descending orders of the traversing time. (a) The collision times of visually impaired volunteers; (b) The traversing time of visually impaired volunteers.

To analyze the major concern of the navigation performance for long-range obstacle avoidance task, another field test was conducted in a corridor about 2.5m wide and 38m deep with many umbrellas popped as shown in Fig. 18. An example of the detection of umbrella is evaluated in Fig. 13. This test involved ten visually impaired individuals who have also experienced the learning stage to get adapted to the stereo sound feedback in the beginning. Afterwards, they were asked to walk collision-free within the environment. As shown in Fig. 19, all participants finished the test of traversing from one end of the narrow corridor to the other even with many low obstacles on the ground. Each visually impaired individual spent an average time of 163.6s to finish the route and collided into obstacles 1.2 times on average. The number of collisions were few when the participants walked through the narrow passage with many disorganized umbrellas on the floor. It can be proved convincingly that the traversability awareness is endowed with usefulness and improved reliability with minimum range reduction, which helps visually impaired individuals to avoid close-range obstacles and determine traversable directions. In other word, the safety and robustness are ensured on navigation.

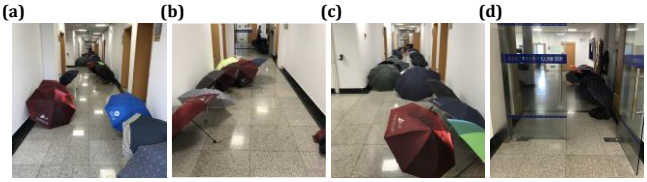


Fig. 18. Field test scenario in a corridor.

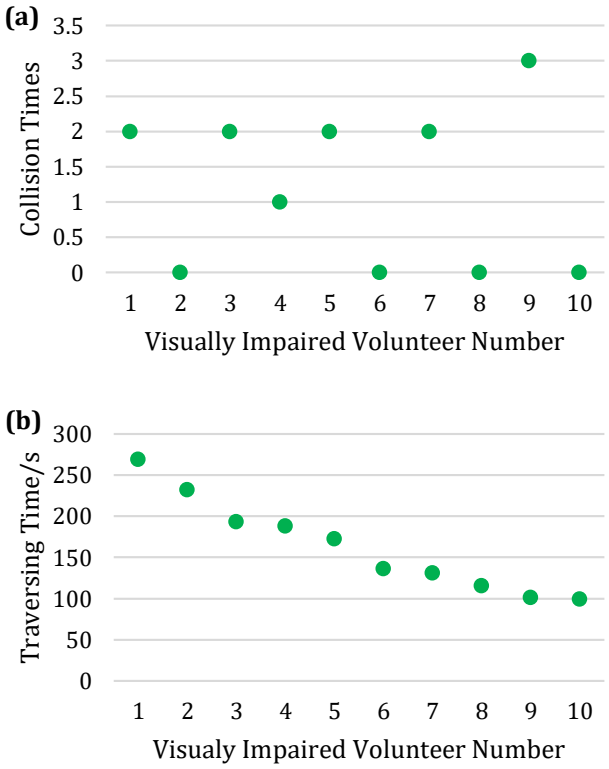


Fig. 19. Experiment data sorted in descending orders of the traversing time. (a) The collision times of visually impaired volunteers; (b) The traversing time of visually impaired volunteers.

## 4. Conclusion

RGB-D sensors are widely applied in navigational assistance for visually impaired people. However, most solutions, such as traversable area detection and obstacle avoidance, suffer from the limitations imposed by RGB-D ranging in terms of active speckles projecting or passive stereo matching.

In this paper, an effective approach to decrease the minimum range of a RGB-D sensor of RealSense R200, which is a hybrid type of RGB-D sensor to enable environmental compatibility and maintain miniaturization advantage. Overall, the minimum depth range of RealSense has been decreased from 650mm to 60mm with qualified accuracy. A traversable line is created to feedback visually impaired individuals through stereo sound to detour around close obstacles and determine traversable directions.

Accuracy test, obstacle detection, traversability awareness and field tests are described in detail which prove the approach to be effective and reliable.

In the future, we aim to incessantly enhance our navigational assistance approach for the visually impaired. We are looking forward to not only decreasing the minimum detection range, but also expanding the maximum range and amplifying the field of view of the RGB-D sensor. Additionally, a pRGB-D-SS framework which incorporates polarization imaging and real-time semantic segmentation would be interesting and useful to acquire complementary information and cover the perception needs of navigational assistance in a unified way.

**Funding Information.** This work has been partially funded by the Zhejiang Provincial Public Fund through the project of visual assistance technology for the blind based on 3D terrain sensor (No. 2016C33136) and cofunded by State Key Laboratory of Modern Optical Instrumentation.

**Acknowledgment**. We acknowledge Prof I. C. Khoo for the heuristic discussions on liquid crystals photonics during the research. We also acknowledge Hangzhou KR-VISION Technology for the organization of the user study.

## References

1. R. R. Bourne, S. R. Flaxman, T. Braithwaite, M. V. Cicinelli, A. Das, J. B. Jonas, J. Keeffe, J. H. Kempen, J. Leasher, H. Limburg, K. Naidoo, K. Pesudovs, S. Resnikoff, A. Silvester, G. A. Stevens, N. Tahhan, T. Y. Wong and H. R. Taylor, "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis," The Lancet Global Health, **5**(9), e888-e897 (2017).
2. Y. H. Lee and G. Medioni, "Rgb-d camera based navigation for the visually impaired," in Proceedings of RSS 2011 RGB-D: Advanced Reasoning with Depth Camera Workshop, (RSS, 2011), pp. 1-6.
3. A. Rodríguez, J. J. Yebes, P. F. Alcantarilla, L. M. Bergasa, J. Almazán and A. Cela, "Assisting the visually impaired: obstacle detection and warning system by acoustic feedback," Sensors, **12**(12), 17476-17496 (2012).
4. A. Rodríguez, L. M. Bergasa, P. F. Alcantarilla, J. Yebes and A. Cela, "Obstacle avoidance system for assisting visually impaired people", in Proceedings of IEEE Intelligent Vehicles Symposium Workshops (IEEE, 2012), **3**, p. 16.
5. A. Aladren, G. López-Nicolás, L. Puig and J. J. Guerrero, "Navigation assistance for the visually impaired using RGB-D sensor with range expansion," IEEE Systems Journal, **10**(3), 922-932 (2016).
6. S. Vorapatratorn, A. Suchato and P. Punyabukkana, "Real-time Obstacle Detection in Outdoor Environment for Visually Impaired using RGB-D and Disparity Map," in Proceedings of International Convention on Rehabilitation Engineering & Assistive Technology (START Center, 2016), p. 8.
7. K. Yang, K. Wang, W. Hu and J. Bai, "Expanding the Detection of Traversable Area with RealSense for the Visually Impaired," Sensors, **16**(11), 1954 (2016).
8. K. Yang, K. Wang, R. Cheng and X. Zhu, "A new approach of point cloud processing and scene segmentation for guiding the visually impaired," in Proceedings of Iet International Conference on Biomedical Image and Signal Processing (IET, 2016), pp. 1-6.
9. K. Yang, K. Wang, R. Cheng, W. Hu, X. Huang and J. Bai, "Detecting Traversable Area and Water Hazards for the Visually Impaired with a pRGB-D Sensor," Sensors, **17**(8), 1890 (2017).
10. K. Yang, K. Wang, X. Zhao, R. Cheng, J. Bai, Y. Yang and D. Liu, "IR Stereo RealSense: Decreasing minimum range of navigational assistance for visually impaired individuals," Journal of Ambient Intelligence and Smart Environments, **9**(6), 743-755 (2017).
11. D. Kim, J. Sun, S. M. Oh, J. M. Rehg and A. F. Bobick, "Traversability classification using unsupervised on-line visual learning for outdoor robot navigation," in Proceedings of IEEE International Conference on Robotics and Automation (IEEE, 2006), pp. 518-526.
12. M. Draelos, "The Kinect Up Close: Modifications for Short-Range Depth Imaging," M.S. thesis, (2012).
13. M. Bellone, A. Messina and G. Reina, "A new approach for terrain analysis in mobile robot applications," in Proceedings of IEEE International Conference on Mechatronics (IEEE, 2013), pp. 225-230.
14. G. Reina, M. Bellone, L. Spedicato and N. I. Giannoccaro, "3D traversability awareness for rough terrain mobile robots," Sensor Review, **34**(2), 220-232 (2014).
15. R. Hadsell, P. Sermanet, J. Ben, A. Erkan, M. Scoffier, K. Kavukcuoglu, U. Muller and Y. LeCun, "Learning long‐range vision for autonomous off‐road driving," Journal of Field Robotics, **26**(2), 120-144 (2009).
16. A. Robledo, S. Cossell and J. Guivant, "Outdoor ride: Data fusion of a 3d kinect camera installed in a bicycle," in Proceedings of Australasian Conference on Robotics and Automation, (2011).
17. M. Bellonea, L. Pascalib and G. Reina, "A kinect-based parking assistance system," Advances in Robotics Research, An International Journal, **1**(2), 127-140 (2014).
18. G. Reina, A. Milella and R. Rouveure, "Traversability analysis for off-road vehicles using stereo and radar data," in Proceedings of IEEE International Conference on Industrial Technology (IEEE, 2015), pp. 540-546.
19. G. Reina and A. Milella, "Towards autonomous agriculture: Automatic ground detection using trinocular stereovision," Sensors, **12**(9), 12405-12423 (2012).
20. G. Reina, A. Milella, R. Rouveure, M. Nielsen, R. Worst and M. R. Blas, "Ambient awareness for agricultural robotic vehicles," Biosystems Engineering, **146**, 114- 132 (2016).
21. F. Katz, S. Kammoun, G. Parseihian, O. Guitierrez, A. Brilhault, M. Auvray, P. Truillet, M. Denis, S. Thorpe and C. Jouffrais, "NAVIG: augmented reality guidance system for the visually impaired," Virtual Reality, **16**(4), 253-269 (2012).
22. Miksik, V. Vineet, M. Lidegaard, R. Prasaath, M. Nießner, S. Golodetz, S. L. Hicks, P. Pérez, S. Izadi and P. H. S. Torr, "The semantic paintbrush: Interactive 3d mapping and recognition in large outdoor spaces," in Proceedings of ACM Conference on Human Factors in Computing Systems (ACM, 2015), pp. 3317-3326.
23. K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," Sensors, **12**(2), 1437-1454 (2012).
24. F. Alhwarin, A. Ferein and I. Scholl, "IR stereo kinect: improving depth images by combining structured light with IR stereo," in Proceedings of Pacific Rim International Conference on Artificial Intelligence (Springer, Cham, 2014), pp. 409-421.
25. M. Draelos, N. Deshpande and E. Grant, "The Kinect up close: Adaptations for short-range imaging," in Proceedings of IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems (IEEE, 2012), pp. 251-256.
26. M. Kytö, M. Nuutinen and P. Oittinen, "Method for measuring stereo camera depth accuracy based on stereoscopic vision," Proc. SPIE **7864**(13), 78640I-1-9 (2011).
27. L. Keselman, J. I. Woodfill, A. Grunnet-Jepsen and A. Bhowmik, "Intel RealSense Stereoscopic Depth Cameras," arXiv preprint arXiv:1705.05548 (2017).
28. K. Konolige, "Projected texture stereo," in Proceedings of IEEE International Conference on Robotics and Automation (IEEE, 2010), pp. 148-155.
29. T. Propst and T. Morrison, "Getting started with the Depth Data provided by Intel RealSense Technology," (Intel, 2015), http://software.intel.com/en-us/articles/realsense-depth-data.

30. K. Berns and E. von Puttkamer, “Simultaneous localization and mapping (SLAM),” In Autonomous Land Vehicles. Vieweg+ Teubner, 146-172.
31. R. Tomari, Y. Kobayashi and Y. Kuno, “Wide field of view kinect undistortion for social navigation implementation,” Advances in Visual Computing, 526-535 (2012).
32. Y. Schröder, A. Scholz, K. Berger, K. Ruhl, S. Guthe and M. Magnor, “Multiple kinect studies,” Computer Graphics, **2**(4), 6 (2011).
33. A. Maimone and H. Fuchs, “Reducing interference between multiple structured light depth sensors using motion,” in Proceedings of IEEE Conference on Virtual Reality (IEEE, 2012), pp. 51-54.
34. M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” Communications of the ACM, **24**(6), 381-395 (1981).
35. W. W. Chiu, U. Blanke and M. Fritz, “Improving the Kinect by Cross-Modal Stereo,” BMVC, **1**(2), 3 (2011).
36. W. C. Chiu, U. Blanke and M. Fritz, “I spy with my little eye: Learning optimal filters for cross-modal stereo under projected patterns,” in Proceedings of IEEE International Conference on Computer Vision Workshops (IEEE, 2011), pp. 1209-1214.
37. G. Saygili, L. van der Maaten and E. A. Hendriks, “Hybrid kinect depth map refinement for transparent objects,” in Proceedings of International Conference on Pattern Recognition (IEEE, 2014), pp. 2751-2756.
38. Y. H. Lee and G. Medioni, “RGB-D camera based wearable navigation system for the visually impaired,” Computer Vision and Image Understanding, **149**, 3-20 (2016).
39. A. Geiger, M. Roser and R. Urtasun, “Efficient large-scale stereo matching,” in Proceedings of Asian conference on computer vision (Springer, Berlin, Heidelberg, 2011), pp. 25-38.
40. N. Einecke and J. Eggert, “A multi-block-matching approach for stereo,” in Proceedings of IEEE Intelligent Vehicles Symposium (IEEE, 2015), pp. 585-592.
41. Intel, “Intel RealSense Camera Calibrator for Windows,” (Intel, 2016), **https://downloadcenter.intel.com/download/24958/Intel-RealSense-Camera-Calibrator-for-Windows-**.
42. Y. G. Wu and K. L. Fan, “Fast vector quantization image coding by mean value predictive algorithm,” Journal of Electronic Imaging, **13**(2), 324-333 (2004).
43. D. Comaniciu and P. Meer, “Mean shift: A robust approach towards feature space analysis,” IEEE Transactions on pattern and machine intelligence, **24**(5), 603-619 (2002).