

RESEARCH

Open Access



A 3D localisation method in indoor environments for virtual reality applications

Wei Song^{1*†} , Liying Liu^{1†}, Yifei Tian^{1†}, Guodong Sun^{2†}, Simon Fong^{3†} and Kyungeun Cho^{4†}

*Correspondence:

sw@ncut.edu.cn

[†]Wei Song, Liying Liu, Yifei Tian, Guodong Sun, Simon Fong, Kyungeun Cho contributed equally to this work

¹ School of Computer

Science, North China University of Technology, Beijing, China

Full list of author information is available at the end of the article

Abstract

Virtual Reality (VR) has recently experienced rapid development for human–computer interactions. Users wearing VR headsets gain an immersive experience when interacting with a 3-dimensional (3D) world. We utilise a light detection and ranging (LiDAR) sensor to detect a 3D point cloud from the real world. To match the scale between a virtual environment and a user's real world, this paper develops a boundary wall detection method using the Hough transform algorithm. A connected-component-labelling (CCL) algorithm is applied to classify the Hough space into several distinguishable blocks that are segmented using a threshold. The four largest peaks among the segmented blocks are extracted as the parameters of the wall plane. The virtual environment is scaled to the size of the real environment. In order to synchronise the position of the user and his/her avatar in the virtual world, a wireless Kinect network is proposed for user localisation. Multiple Kinects are mounted in an indoor environment to sense the user's information from different viewpoints. The proposed method supports the omnidirectional detection of the user's position and gestures. To verify the performance of our proposed system, we developed a VR game using several Kinects and a Samsung Gear VR device.

Keywords: Kinect, LiDAR, Hough transform, Connected-component-labelling, Virtual reality

Introduction

In recent years, head-mounted displays have been widely developed for Virtual Reality (VR) simulations and video games. However, due to the need to wear stereoscopic displays, users cannot view their real environment. Traditionally, the virtual environment's boundary does not match that of a user's real environment. Thus, collisions between the user and the real world always occur in VR applications and cause poor user experiences.

To create an adaptive virtual environment, boundary measurement of the real environment is necessary for warnings. Currently, a light detection and ranging (LiDAR) sensor is utilised to detect the 3D point cloud of the surrounding environment. From the point cloud, large planar regions are recognised as the boundary walls [1]. In order to detect the boundary of an indoor environment, this paper develops a boundary wall detection method based on the Hough transform algorithm [2]. After the Hough transform is implemented on the LiDAR datasets, a connected-component-labelling (CCL) algorithm is applied to classify the segmented intensive regions of the Hough space into

several distinguishable blocks. The corresponding Hough coordinates of the largest four peaks of the blocks are recognised as the wall plane parameters. By scaling the virtual environment to the real environmental range, the user is able to act in the virtual environment without collisions, thus enhancing the user experience.

The tracking of the skeleton of a human body using RGB images and the depth sensors of the Microsoft Kinect has been widely applied for interactions between users and virtual objects in VR applications [3]. When we utilise the Kinect to acquire a user's gesture, the user needs to stand in front of the Kinect within a limited distance and face the Kinect [4]. Otherwise, weak and inaccurate signals are sensed. For omnidirectional detection, this paper proposes a multiple Kinect network using a bivariate Gaussian probability density function (PDF). In the system, multiple Kinect sensors installed in an indoor environment detect a user's gesture information from different viewpoints. The sensed datasets of the distributed clients are sent to a VR management server that selects an adaptive Kinect based on the user's distance and orientation. In our method, only small datasets of the user's position and body joints are delivered from the Kinect clients to the server; this satisfies the real-time transmission requirements [5].

The remainder of this paper is organised as follows. “[Related works](#)” section provides an overview of related works. “[A 3D localisation system](#)” section describes the 3D localisation system, including the environmental boundary walls detection method and wireless Kinect sensor network selection. “[Experiments](#)” section illustrates the experiment results. Finally, “[Conclusions](#)” section concludes this paper.

Related works

To realise a virtual–physical collaboration approach, environmental recognition methods such as plane and feature detection have been researched [6]. Zucchelli et al. [7] detected planes from stereo images using a motion-based segmentation algorithm. The planar parameters were extracted automatically with projective distortions. The traditional Hough transform was usually used to detect straight lines and geometric shapes from the images. Trucco et al. [8] detected the planes from the disparity space using a Hough-like algorithm. Using these methods, matching errors were caused when the outliers overlapped with the plane regions.

To detect continuous planes, Hulik et al. [9] optimised a 3D Hough transform to extract large planes from LiDAR and Kinect RGB-D datasets. Using a Gaussian smoothing function, the noise in the Hough space was removed to preserve the accuracy of the plane detection process. In order to speed up the Hough space updating process, a caching technique was applied for point registration. Compared with the traditional plane detection algorithm Compared Random sample consensus (RANSAC) [10], the 3D Hough transform performed faster and was more stable. During the maxima extraction process from the Hough space, this method applied a sliding window technique with a pre-computed Gaussian kernel. When dense noise exists surrounding a line, more than one peak is extracted in a connected segmented region using this method. In order to maintain stable line estimation, this paper applied a CCL algorithm to preserve only one peak extracted in one distinguishable region [11].

To localise and recognise a user's motion, Kinect is a popular display device in VR development. It is able to report on the user's localisation and gesture information.

However, a single Kinect can only capture the front-side of users facing the sensor. To sense the back-side, Chen et al. [12] utilised multiple Kinects to reconstruct an entire 3D mesh of the segmented foreground human voxels with colour information. To track people in unconstrained environments, Sun et al. [13] proposed a pairwise skeleton matching scheme using the sensing results from multiple Kinects. Using a Kalman filter, their skeleton joints were calibrated and tracked across consecutive frames. Using this method, we found that different Kinects provided different localisation of joints because the sensed surfaces were not the same from different viewpoints.

To acquire accurate datasets from multiple sensors, Chua et al. [14] addressed a sensor selection problem in a smart-house using a naïve Bayes classifier, a decision tree and k-Nearest-Neighbour algorithms. Sevrin et al. [15] proposed a people localisation system with a multiple Kinects trajectory fusion algorithm. The system adaptively selected the best possible choice among the Kinects in order to detect people with a highly accurate rate [16]. Following these sensor selection methods, we developed a wireless and reliable sensor network for VR applications to enable users to walk and interact freely with virtual objects.

A 3D localisation system

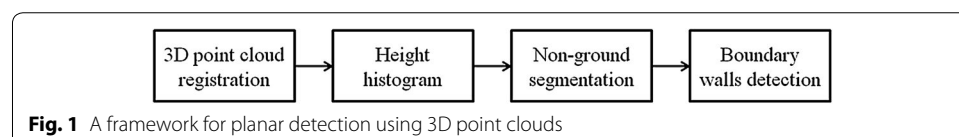
This section describes an indoor 3D localisation system for VR applications. A Hough transform algorithm is applied to detect the indoor boundary walls. A multiple Kinects selection method is proposed to localise a user's position with an omnidirectional orientation.

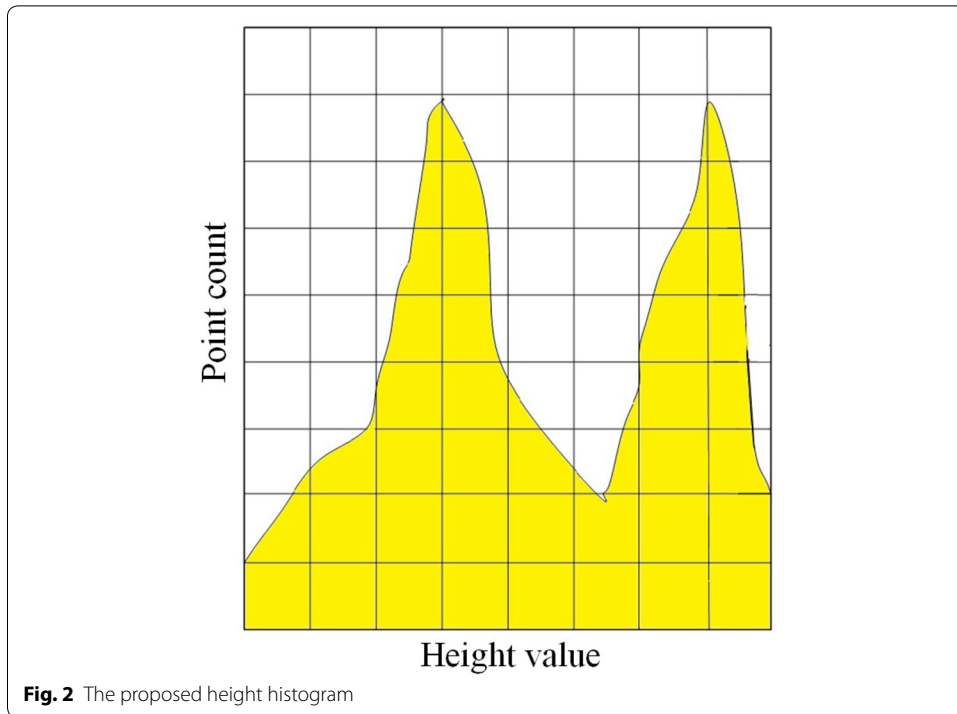
Indoor boundary detection from 3D point clouds

To estimate the localisation of indoor walls, we describe a framework of plane detection in 3D point clouds, as shown in Fig. 1. The framework mainly includes the registration of 3D point clouds, a height histogram of 3D points, non-ground points segmentation and planar surface detection.

An indoor environment always contains six large planes, including four surrounding walls, the floor and the roof. This project aims to segment the non-ground walls from the detected planes to estimate the environmental size. A height histogram, as shown in Fig. 2, is first utilised to estimate the voxel distribution of the height [14]. Since the points located on the floor or roof surfaces always have the same height value, the two peaks of the height histogram are considered to be the floor and roof surfaces. After the peaks are filtered out, the non-ground points are then segmented.

In indoor environments, the planes of boundary walls always form a cuboid shape. Since most LiDAR points are projected onto the walls, the mapped 2D points on the $x-z$ plane from the wall points are combined into four straight lines. The pairwise opposite lines are parallel to each other and the neighbour lines are orthogonal to each other. For indoor boundary detection, a Hough transform algorithm is applied to estimate the

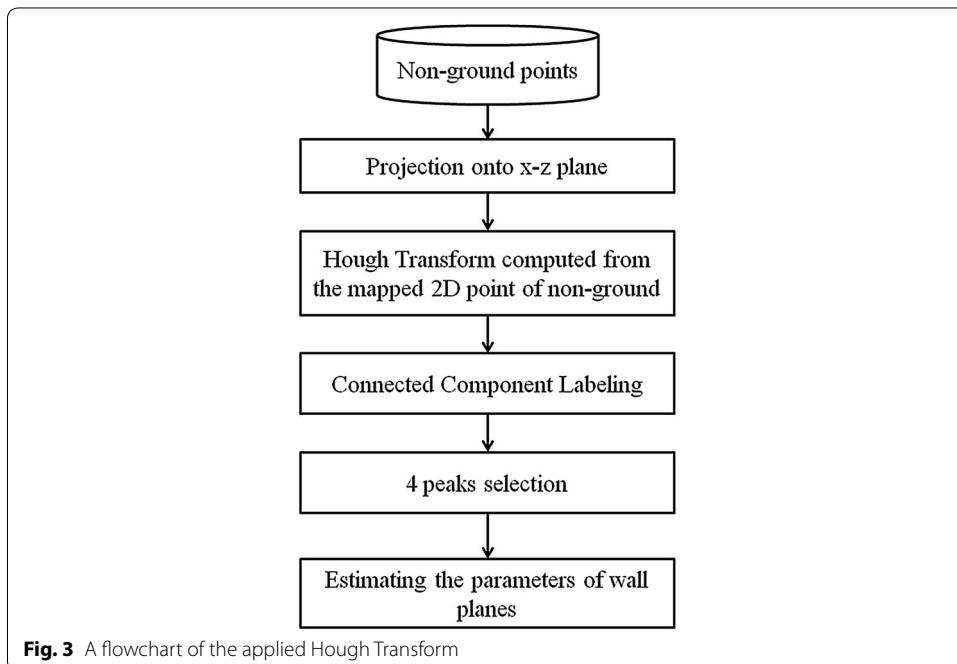




parameters of the mapped lines on x-z plane from the segmented non-ground voxels. A flowchart of the applied Hough Transform is shown in Fig. 3.

We assume that the walls are always orthogonal to the x-z plane. Hence, the wall plane is formulated using the following linear Eq. (1):

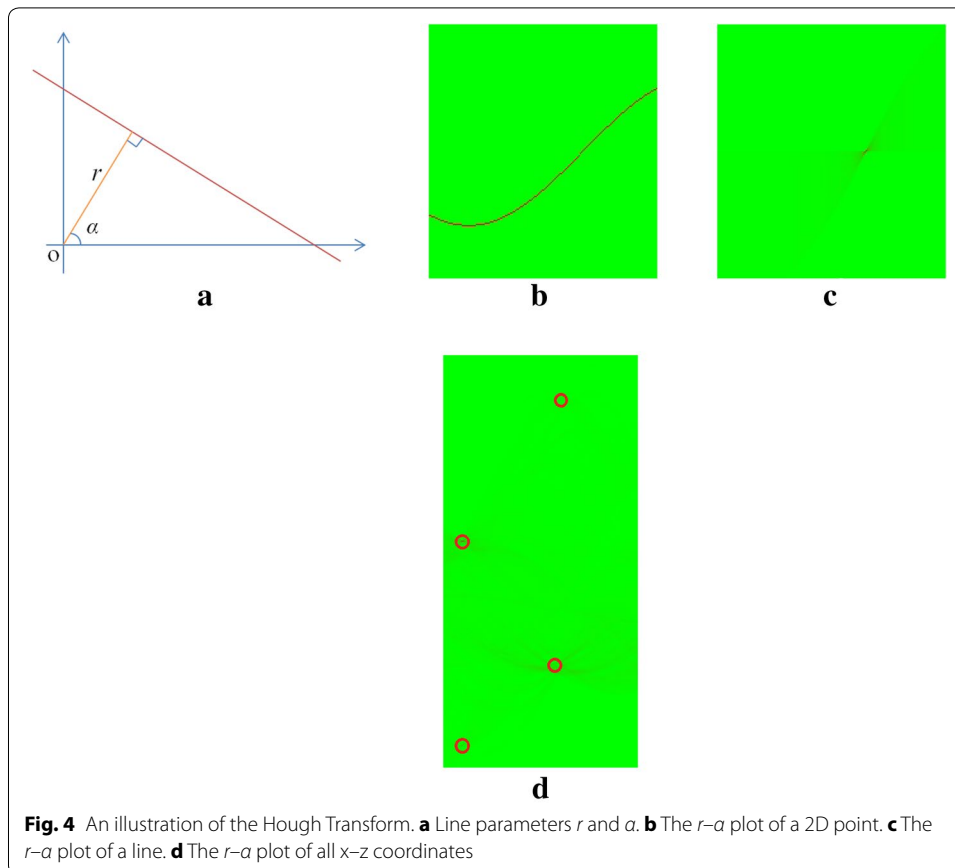
$$r = x \cos \alpha + z \sin \alpha \tag{1}$$

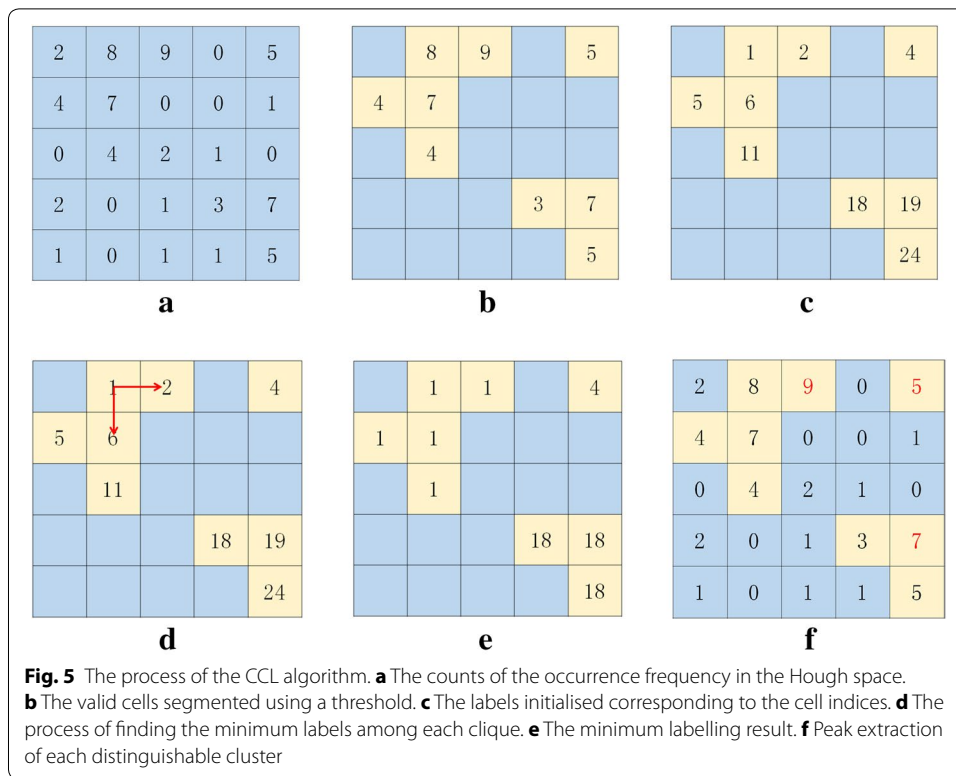


As shown in Fig. 4a, r is the distance from the origin to the straight line and α is the angle between the vertical direction of the line with the x axis. The Hough space is defined as the $(r-\alpha)$ plane calculated from a set of LiDAR points in x and z coordinates. The approximate sinusoidal curve in Fig. 4b represents the Hough space of a 2D point. As shown in Fig. 4c, all sinusoidal curves computed using the Hough transform from the points in a straight line cross at several points. The r and α coordinates of the maxima in the Hough space are the line parameters.

The wall planes contain most of the points that form several straight lines on the $x-z$ plane. Therefore, the four peaks of the Hough space are recognised as the parameters of the boundary wall planes after the (r, α) coordinates are generated from all the sensed indoor points using the Hough Transform. Each (r, α) cell in the Hough space records the count of the mapped LiDAR points; these indicate the occurrence frequency. The four peaks always exist in the intensive areas as shown in Fig. 4d.

Figure 5a presents an instance of the occurrence frequency in the Hough space. To segment the intensive areas, the low frequency cells are filtered out using a threshold based on the occurrence frequency distribution of the cells. The valid cells are segmented as shown in Fig. 5b, and are classified into several distinguishable blocks using the CCL algorithm. In the CCL algorithm, the label of each cell is initialised corresponding to its index, as shown in Fig. 5c. To mark each distinguishable block with a unique label, the minimum label in Fig. 5d is searched for among a clique of each cell that contains the





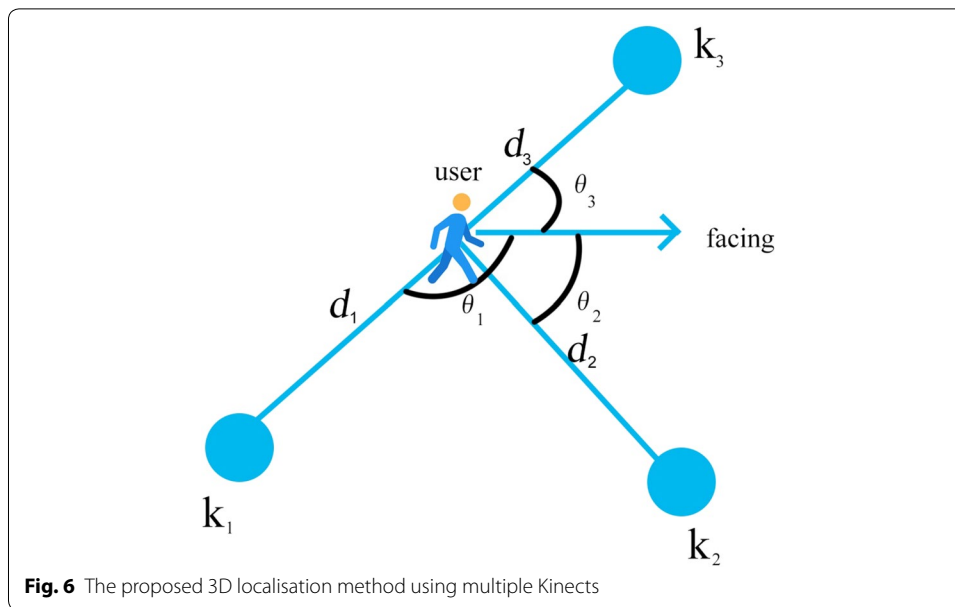
local, right and bottom cells. The clique updates the labels with the minimum label in it. Several seeking iterations of the minimum labels are implemented until all labels remain unchanged. The minimum label in a distinguishable block in Fig. 5e is the indicator of the connected valid cells. Finally, the corresponding (r, α) coordinate of the largest value in each distinguishable block of Fig. 5f is the required straight-line parameter.

Adaptive Kinect selection

We propose a wireless sensor network to localise the VR user using the integration of multiple Kinects. As shown in Fig. 6, the user’s motion and position datasets are detected from multiple views using the Kinects. The distributed Kinects report the sensed datasets to a VR server via a WiFi network. An adaptive Kinect is selected using a bivariate Gaussian PDF.

A Kinect is installed at each client to detect the user’s gesture information from different viewpoints. From several gathered datasets, the effectiveness of each sensor is generated based on the user’s distance d_i and orientation θ_i to the Kinect k_i . If the distance is close and the orientation of the user is facing towards a sensor, the effectiveness of this sensor is then high. To select the best sensor, we apply a bivariate Gaussian PDF for the effectiveness estimation, formulated as follows:

$$f_{k_i}(d_i, \theta_i) = \frac{\exp \left[-\frac{\left(\frac{d_i-d_0}{\sigma_1}\right)^2 - \frac{2\rho(d_i-d_0)(\theta_i-\theta_0)}{\sigma_1\sigma_2} + \left(\frac{\theta_i-\theta_0}{\sigma_2}\right)^2}{2(1-\rho^2)} \right]}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \tag{2}$$



Here, the variables $d \in [0 \sim \infty)$, $\theta \in [-\pi \sim \pi)$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\rho \in [-1, 0]$. The adaptive Kinect is selected using a maximum likelihood function expressed as follows:

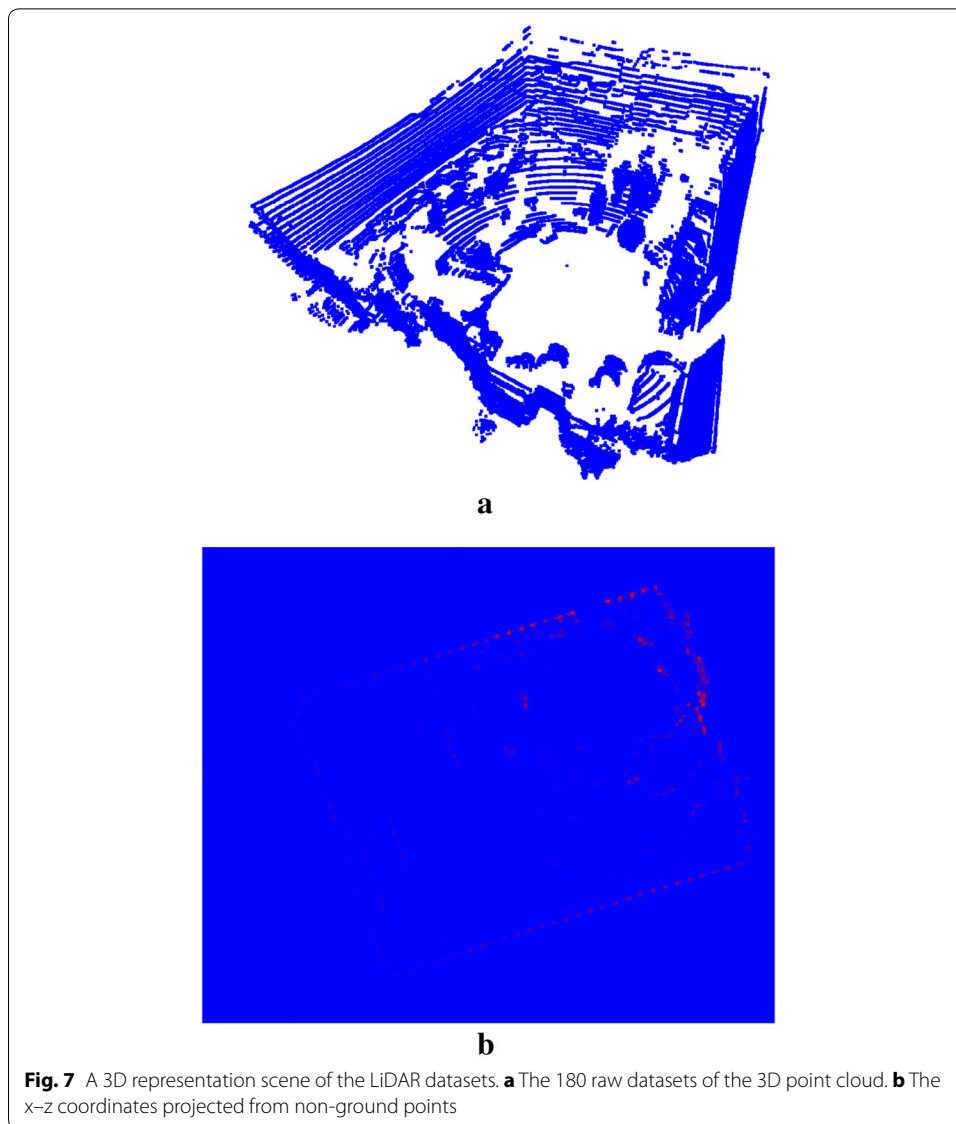
$$k = \arg \max_{k_i} f_{k_i}(d_i, \theta_i). \quad (3)$$

Experiments

In this section, we analyse the performance of the proposed indoor boundary walls detection method from LiDAR points and illustrate a VR application developed using the proposed 3D localisation method. The experiments were implemented using one HDL-32E Velodyne LiDAR and two Microsoft Kinect2 sensors. The wall detection method was executed on a 3.20 GHz Intel® Core™ Quad CPU computer with a GeForce GT 770 graphics card and 4 GB of RAM. The Kinects were utilised to detect a user's gesture on two clients; these were 3.1 GHz Intel® Core™ i7-5557U CPU NUC mini PCs with 16 GB of RAM. The VR client was implemented on a Samsung Gear VR with a Samsung Galaxy Note 4 in it. The Note 4 had a 2.7 GHz Qualcomm Snapdragon Quad CPU, 3 GB of RAM, a 2560 × 1440 pixels resolution and the Android 4.4 operating system.

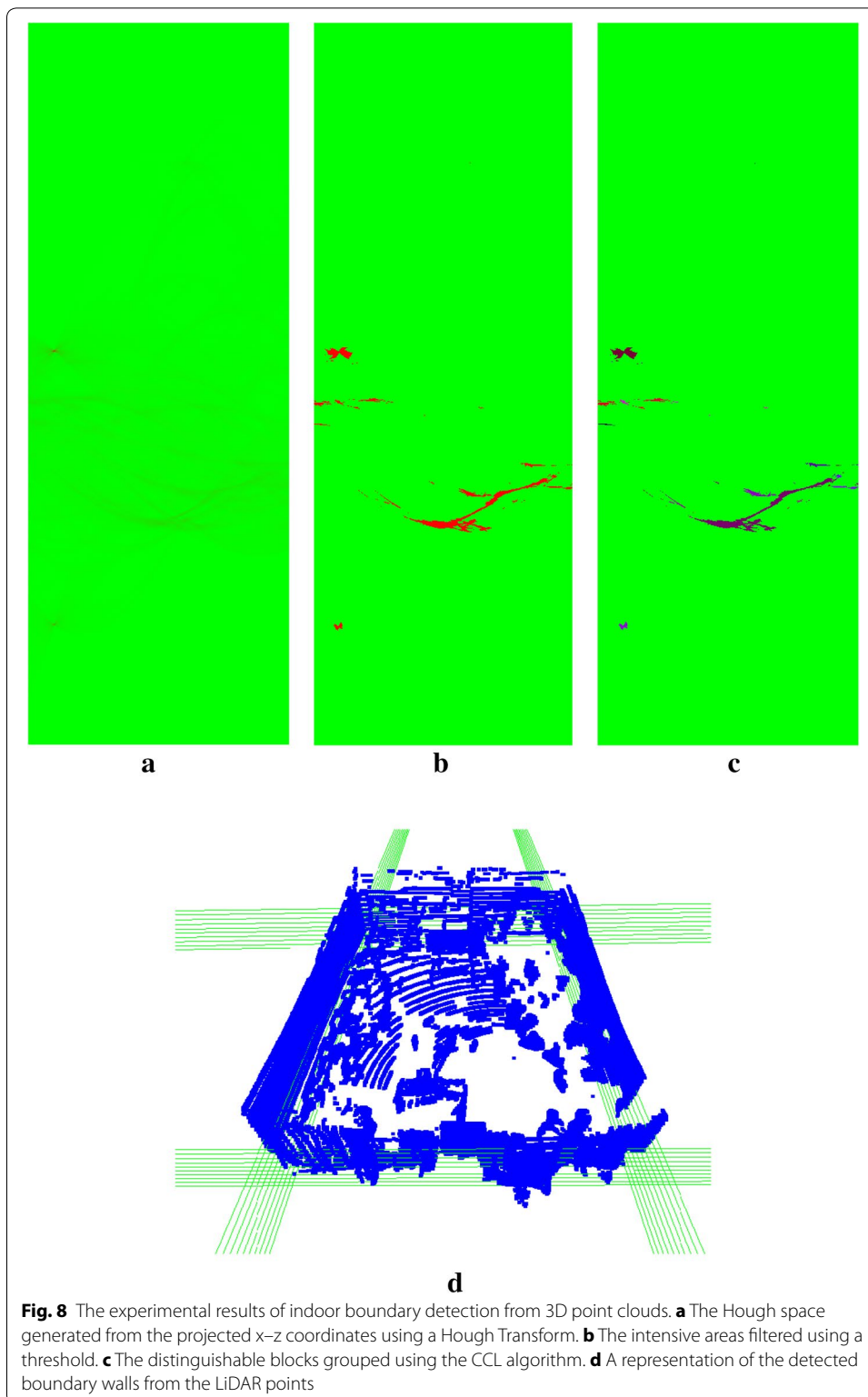
The applied HDL-32E was able to sense 32×12 3D points in a packet per 552.96 μ s. The field of view was 41.34° in the vertical direction and 360° in the horizontal direction with an angular resolution of 1.33°. The valid range was 70 m with an error variance of 2 cm. In our project, the 3D point clouds were reconstructed using DirectX software development kits. Figure 7a presents the raw datasets of $180 \times 32 \times 12$ points sensed by a stationary Velodyne LiDAR in an indoor environment. By projecting the non-ground points onto the x-z plane, a density diagram was generated as shown in Fig. 7b where mapped cells with a high density are represented using red. The intensive regions of line shapes were considered to be the boundary walls.

Using the proposed Hough Transform, the Hough space shown in Fig. 8a was generated from the non-ground points. The brightness of a cell in the Hough space indicates



the occupied frequency of the (r, α) coordinates. In our experiment, the range of the distance r was calculated to be between -10.598 and 5.909 m and the inclination angle α was between 0° and 180° . The system allocated an 825×360 integer buffer for the Hough space cache. Using a threshold computed based on the value distribution of the Hough space, the intensive regions were segmented as shown in Fig. 8b. After the proposed CCL algorithm was implemented using 19 iterations, 55 distinguishable blocks were grouped using different colours as shown in Fig. 9c. By selecting the four largest peaks from the distinguishable blocks, the corresponding coordinates (r, α) were calculated using the parameters of the straight lines. In Fig. 9d, we displayed the detected boundary walls with the LiDAR points. The estimated wall planes are located on the wall voxels, thus proving that our proposed method was accurate.

The range of the indoor environment was estimated to be 9.94 m in length and 7.54 m in width. The virtual environment was resized to correspond to the real environment so



as to achieve virtual–physical synchronisation. The wall detection method was implemented during an initialisation step before the VR application was started. Using the proposed system, we developed a VR boxing game as shown in Fig. 9.



Fig. 9 A VR boxing game developed using the proposed wireless multiple Kinect sensors selection system

In the system, the user's location and orientation were detected by two Kinects. When the player was facing a Kinect with a distance between 2 and 6 m, the motion information was sensed precisely. Through the experiments, we found that $d_0 = 5.0$ and $\theta_0 = 0.0$ is the perfect position for Kinect detection. Through the selection of an effective Kinect, the user was able to make free movements and interact with the virtual boxer from an omnidirectional orientation. Meanwhile, the monitor of the server rendered the game visualisation result synchronously with the VR display. The processing speed of our application including data sensing, transmission and visualisation was greater than 35 fps; this successfully achieved the real-time requirements.

Conclusions

To provide a free movement environment for VR applications, this paper demonstrated a 3D localisation method for virtual–physical synchronisation. For environmental detection, we utilised a HDL-32E Velodyne LiDAR sensor to detect the surrounding 3D point clouds. Using the Hough transform, a plane detection algorithm was proposed to extract indoor walls from point clouds so as to estimate the distance range of the surrounding environment. The virtual environment was then correspondingly resized. To match the user's position between real and virtual worlds, a wireless Kinects network was proposed for omnidirectional detection of the user's localisation. In the sensor selection process, we applied a Bivariate Gaussian PDF and the Maximum Likelihood Estimation method to select an adaptive Kinect. In the future, we will integrate touch sensors to the system for virtual–physical collaboration.

Authors' contributions

WS and LL described the proposed algorithms and wrote the whole manuscript. YT and GS implemented the experiments. SF and KC revised the manuscript. All authors read and approved the final manuscript.

Author details

¹ School of Computer Science, North China University of Technology, Beijing, China. ² Department Digital Media Technology, Beijing University of Technology, Beijing, China. ³ Department Computer and Information Science, University of Macau, Macau, China. ⁴ Department Multimedia Engineering, Dongguk University, Seoul, South Korea.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (61503005), and by NCUT XN024-95. This paper is a revised version of a paper entitled 'A Wireless Kinect Sensor Network System for Virtual Reality Applications' presented in 2016 at Advances in Computer Science and Ubiquitous Computing-CSA-CUTE2016, Bangkok, Thailand [17].

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 26 April 2017 Accepted: 6 October 2017

Published online: 13 October 2017

References

1. Dick A, Torr P, Cipolla R (2004) Automatic 3d modeling of architecture, In: Proc. 11th British Machine Vision Conf. pp 372–381
2. Mukhopadhyay P, Chaudhuri B (2015) A survey of hough transform. *Pattern Recognit* 48(3):993–1010
3. Ales P, Oldrich V, Martin V et al (2015) Use of the image and depth sensors of the Microsoft Kinect for the detection of gait disorders. *Neural Comput Appl* 26(7):1621–1629
4. Mohammed A, Ahmed S (2015) Kinect-based humanoid robotic manipulator for human upper limbs movements tracking. *Intell Control Autom* 6(1):29–37
5. Song W, Sun G, Fong S et al (2016) A real-time infrared LED detection method for input signal positioning of interactive media. *J Converg* 7:1–6
6. Junho A, Richard H (2015) An indoor augmented-reality evacuation system for the Smartphone using personalized Pedometry. *Hum Centric Comput Inf Sci* 2:18
7. Zucchelli M, Santos-Victor J, Christensen HI (2002) Multiple plane segmentation using optical flow. In: Proc. 13th British Machine Vision Conf. pp 313–322
8. Trucco E, Isgro F, Bracchi F (2003) Plane detection in disparity space. In: Proc. IEE Int. Conf. Visual Information Engineering. pp 73–76
9. Hulik R, Spanel M, Smrz P, Materna Z (2014) Continuous plane detection in point-cloud data based on 3D hough transform. *J Vis Commun Image R* 25(1):86–97
10. Schnabel R, Wahl R, Klein R (2007) Efficient RANSAC for point-cloud shape detection. *Comput Graph Forum* 26(2):214–226
11. Song W, Tian Y, Fong S, Cho K, Wang W, Zhang W (2016) GPU-accelerated foreground segmentation and labeling for real-time video surveillance. *Sustainability* 8(10):916–936
12. Chen Y, Dang G, Chen Z et al (2014) Fast capture of personalized avatar using two Kinects. *J Manuf Syst* 33(1):233–240
13. Sun S, Kuo C, Chang P (2016) People tracking in an environment with multiple depth cameras: a skeleton-based pairwise trajectory matching scheme. *J Vis Commun Image R* 35:36–54
14. Chua SL, Foo LK (2015) Sensor selection in smart homes. *Procedia Comput Sci* 69:116–124
15. Sevrin L, Noury N, Abouchi N et al (2015) Preliminary results on algorithms for multi-kinect trajectory fusion in a living lab. *IRBM* 36:361–366
16. Erkan B, Nadia K, Adrian FC (2015) Augmented reality applications for cultural heritage using Kinect. *Hum Centric Comput Inf Sci* 5(20):1–8
17. Li M, Song W, Song L, Huang K, Xi Y, Cho K (2016) A wireless kinect sensor network system for virtual reality applications. *Lect Notes Electr Eng* 421:61–65

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
