Improving RealSense by Fusing Color Stereo Vision and Infrared Stereo Vision for the Visually Impaired

Hao Chen, Kaiwei Wang*, Kailun Yang
College of Optical Science and Engineering
Zhejiang University
Hangzhou, China
e-mail: wangkaiwei@zju.edu.cn

***Abstract*—The introduction of RGB-D sensor has attracted attention from researchers majored in computer vision. With real-time depth measurement provided by RGB-D sensor, a better navigational assistance than traditional aiding tools can be offered for visually impaired people. However, nowadays RGB-D sensor usually has a limited detecting range, and fails in performing depth measurement on objects with special surfaces, such as absorbing, specular, and transparent surfaces. In this paper, a novel algorithm using two RealSense R200 simultaneously to build a short-baseline color stereo vision system is developed. This algorithm enhances depth estimation by fusing color stereo depth map and original RealSense depth map, which is obtained by infrared stereo vision. Moreover, the minimum range is decreased by up to 84.6%, from 650mm to 100mm. We anticipate out algorithm to provide better assistance for visually impaired individuals.**

Stereo vision; RGB-D sensor; RealSense; sensor fusion; visually impaired (key words)

I. INTRODUCTION

The introduction of RGB-D sensor, such as Intel RealSense (developed by Intel based in Santa Clara, CA, USA) [1], Microsoft Kinect (developed by Microsoft based in Redmond, WA, USA) [2], and ASUS Xtion (developed by ASUS based in Taipei, Taiwan) [3], has provided an efficient and reliable way to obtain color images and perform a depth measurement simultaneously, which has many applications in several areas. While RGB-D sensor is originally designed as gaming interface, it is also a popular choice for assisting visually impaired people. Compared to the traditional assistive tools, such as a white cane or a guide dog, RGB-D sensor can provide much more abundant information for visually impaired people [4-7]. With real-time depth measurement provided by RGB-D sensor, reliable navigational assistance could be offered for visually impaired individuals by using algorithms of computer vision and proper interacting ways, during which the depth information is essential for obstacle avoidance and traversable area detection.

While the RGB-D sensor can provide real-time and high resolution depth maps, it also has many restrictions. Take Intel RealSense R200 for example, it still fails to acquire precise depth information on some special surfaces such as absorbing, specular, transparent surfaces, which are common in everyday household objects such as computer screens, mirrors, and glass bottles. Defect of depth measurement on these objects can be a big flaw of RGB-D sensors when used for assisting visually impaired people. Otherwise, RealSense R200 has a minimum detecting range of about 650mm. Objects can be detected only when they are 650mm or farther away from the sensor. When objects are within 650mm, there will be a black hole in the depth map, which means those pixels have no depth information. This poses challenges to obstacle detection algorithms a lot.

To overcome these disadvantages, we present a novel approach to combine two RealSense R200 to build a short-baseline color stereo vision system. Since the original depth map is poor and invalid in short range, we implement a stereo matching algorithm based on the color stereo image pair to compute a depth map. The color stereo depth map is fused with the original depth map provided by RealSense R200, which is acquired through stereo matching based on the infrared image pair. In this regard, a much denser and more robust depth map is obtained, and the resulting depth map can be used in obstacle avoidance and traversable area detection.

Some related work will be discussed in Section II. Our approach which comprises four stages will be presented in Section III. A set of experiments will be performed in Section IV and we draw the conclusion in Section V.

II. RELATED WORK

In order to improve the performance of RGB-D sensor, many methods have been proposed in the literature.

An already commercialized product called Zoom of Kinect has implemented to decrease the minimum range of Microsoft Kinect, another efficient RGB-D sensor, by adding wide-angle lens to it, which can increase the FOV (field of view) of the RGB sensor and IR sensor on Kinect. While it does solve the problem that it's not convenient to use Kinect in a narrow space, but it introduces a pronounced distortion in the final depth map. Draelos et al. [8] proposed an algorithm that can produce an empirical depth distortion model which can correct such distortion. The researchers claimed that this can help reducing the minimum range of Kinect by 30%. But this algorithm still cannot offer a solution to get the depth of objects within 450mm, and it doesn't provide solution to the challenging surfaces.

Saygili et al. [9] presented a simple but efficient method that they used two Kinect to compose a stereo system. Actually, they used two IR sensors of Kinect to generate a IR

stereo pair to perform depth estimation, and then fuse it with origin depth map offered by Kinect. For the reason that IR projectors on Kinect will emit IR patterns in infrared images, which is beneficial for stereo matching, the algorithm also performs well in texture-less areas. Their research is efficient in filling the holes in the origin depth map, and the author argued that the minimum range can be decreased from 800mm to 500mm, with the baseline of the generated stereo system set to 45mm. But this method also fails to apply a depth measurement on objects within 500mm, which is still not a short distance.

Cross-modal stereo matching was implemented by Chiu et al. [10] to improve RGB-D sensors. That is to say, to get a disparity map from stereo matching between RGB and IR image pair obtained by the RGB sensor and IR sensor of a Kinect. The researchers investigated a fusion scheme that combine RGB channels to mimic the image response of the IR sensor, in another word, to transform the RGB image to make it as similar as possible to the IR image. In their later work [11], they attained a better result by proposing a more general method of learning optimal filters. Their work results in significant improvement of nearly 30%. But this algorithm has a significant disadvantage, which is that the environment has a great impact on the result of cross-modal stereo matching. The weights of each channel for combining RGB information cannot assure being suitable for all illumination conditions.

In order to decrease minimum detecting range of RealSense and provide better navigational assistance to visually impaired individuals, Yang et al. [12] put forward a simple yet effective approach to reduce the minimum range. This paper makes use of over-dense regions of IR speckles in two IR images that act as a stereo pair to generate short-range depth. Then fusion of original depth image and short-range depth image is performed. The paper successfully decreased minimum range of Intel RealSense R200 by approximately 75%, from 650mm to 165mm. Another algorithm proposed by Yang et al. [13], which is to expand the detection of traversable area based on RealSense, enhance the depth image of RealSense with IR image large-scale matching and RGB image-guided filtering. It improves the performance of RealSense significantly.

Compared with related work, this paper proposes to make use of the RGB sensors of the two RealSense R200 to form a short-baseline color stereo pair, and fuse color stereo depth map and infrared stereo depth map. Through the proposed algorithm, enhanced depth estimations can be obtained both indoors and outdoors. And depth measurement performs well on challenging surfaces. Meanwhile the minimum range is decreased by approximately 84.6%, from 650mm to 100mm, which is beneficial for providing better assistance to visually impaired individuals.

III. APPROACH

A. System Overview

As shown in Fig. 1(a), a RealSense R200 contains one RGB sensor, two IR sensors, and one IR laser projector. The

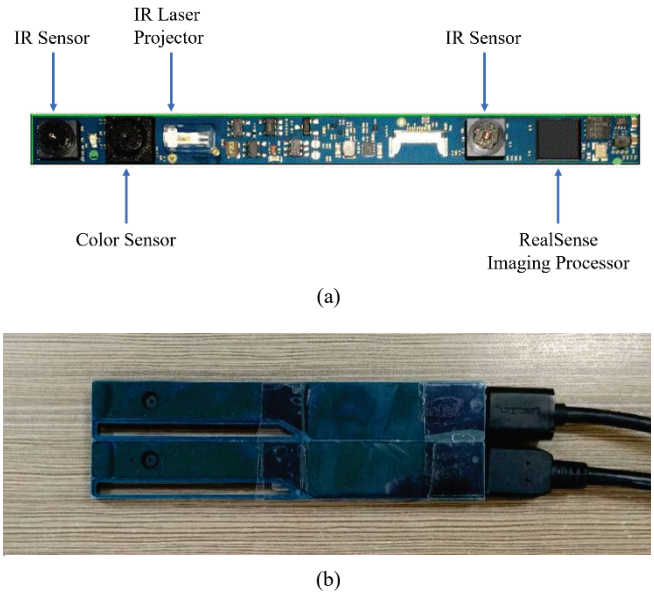


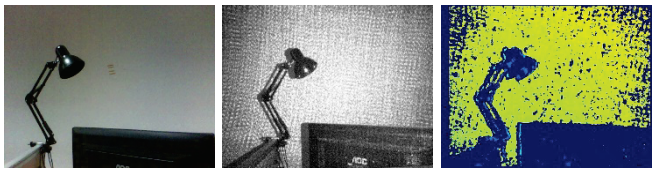
Figure 1. (a) Intel RealSense R200. (b) Dual RealSense

IR laser projector projects non-visible IR speckles, then the left and right sensors acquire maps with those speckles, which provide textures that are beneficial for the embedded stereo matching algorithm to generate a depth map. As shown in Fig. 2, in some texture-less areas such as white walls, there has been projected many IR speckles, and the depth map in such area can be dense enough.

But this kind of algorithm that combines active projecting and passive stereo matching brings about some problems, the main of which can be listed as follows:

- Within short range, speckles are too dense to verify because of the overexposure in IR images, which is harmful for disparity computation and depth measurement. It is the main reason for the short range limit of RealSense R200.
- For reason that the depth measurement of RealSense R200 is carried out by infrared stereo matching, objects can be detected only when they appear simultaneously in the FOV of two IR sensors. But the baseline between the two IR sensors is about 70mm, which may cause the overlapping area of two IR sensors too narrow. This is the second reason for short range limit of RealSense R200.
- The IR speckles will be reflected, transmitted, or absorbed on some special surfaces, which cause the loss of depth measurement on objects with those surfaces, such as the computer screen.
- Compared to RGB maps, infrared maps have inferior quality, because they lose much more details than RGB images and are influenced by infrared light in the environment. It's detrimental to stereo matching algorithm.

Our system architecture is depicted in Fig. 1(b), which combines two RealSense R200 in vertical direction. Our approach's main contributions can be summarized as follows:



(a) (b) (c)

Figure 2. (a) Color image. (b) IR image. (c) Depth image.

- It provides a novel solution for improving original depth map provided by RealSense R200.
- We make use of the RGB sensors of two RealSense R200 to build a color stereo system. By fusing color stereo depth map and infrared stereo depth map (provided by RealSense R200) through an efficient fusion strategy, a much denser and more accurate depth map can be obtained.
- The baseline of our color stereo system is approximately 19mm, which is short enough for disparity computation of objects within short range.
- The stereo matching between two color stereo pairs performs well on challenging surfaces than stereo matching based on IR stereo pairs.
- We apply post-processing to the color stereo disparity map, which makes the final depth map denser and more accurate.
- The approach is computationally efficient, and can be implemented in real time.

B. Camera Calibration

Before performing stereo matching, a calibration process should be carried out to the two RGB sensors of RealSense. We used the camera calibration technique proposed by Zhang [14], which is composed of calibrating a camera and a stereo system. The calibration parameters are listed in Table I, in which 1st RGB sensor stands for the RGB camera of the above RealSense shown in Fig. 1(b), and 2nd RGB sensor refers to the one below. These parameters are used in generating ideal stereo image pairs, transferring disparity value to depth value, and fusing two different depth maps (one is provided by color stereo pair, the other is original depth map obtained by RealSense R200).

TABLE I. CALIBRATION PARAMETERS

Calibration parameters	***Value***
1st RGB sensor's focal length (pixels)	(622.06625, 633.20983)
1st RGB sensor's principle point (pixels)	(319.06693, 258.76695)
2nd RGB sensor's focal length (pixels)	(622.02597, 633.78933)
2nd RGB sensor's principle point (pixels)	(318.22967, 234.59493)
Baseline of two RGB sensors (mm)	19.44680

C. Disparity Computation

Disparity computation comprises two main stages: stereo matching and disparity map post-filtering.

(a) (b) (c)

Figure 3. (a) Color image. (b) Disparity map. (c) Disparity map after post-filtering.

For stereo matching, we apply Semi-Global Matching (SGM) which is implemented by Hirschmuller [15]. SGM is widely used for its high accuracy and fast computational speed.

But stereo matching algorithms tend to make mistakes or fail to compute disparity in uniform texture-less areas, special surfaces and regions near edges. Post-processing must be performed to improve the accuracy of stereo matching and the density of the disparity map. There are several methods of post-processing, one of which is to perform twice different matching to detect potentially inaccurate disparity values and invalidate them. This will make the disparity map sparser than the one before post-processing. Therefore, we take another way widely used, which is the weighted least squares (WLS) smoothing filter. The WLS-filter makes its input image close to the guide image to the greatest extent. Meanwhile the input image after filtering would be as smooth as possible, in addition to the places where the guide image's gradient changes severely (at edges in the guide image). As a result, the disparity map edges are aligned with RGB images, which is the source image of stereo matching. Then the disparity values are propagated from high-confidence to low-confidence regions. After the post-filtering, a denser and more accurate disparity map can be obtained. Fig. 3 [16] shows differences between disparity maps with and without post-filtering.

D. Depth Calculation and Fusion

A disparity map can be acquired by the stereo matching algorithm. Then a transformation method must be applied to convert the disparity map into a depth map. For pixels with non-zero disparity value, the corresponding depth value can be calculated through (1).

$$d = b * f / \triangle \tag{1}$$

Where

b is the baseline length between two RGB sensors,
f refers to the camera's focus length,
d is the depth value,
$\triangle$ is the disparity value.

Accordingly, a stereo depth map can be obtained. Then we fuse it with the original depth map provided by RealSense R200 according to the following fusion scheme.

For each pixel:

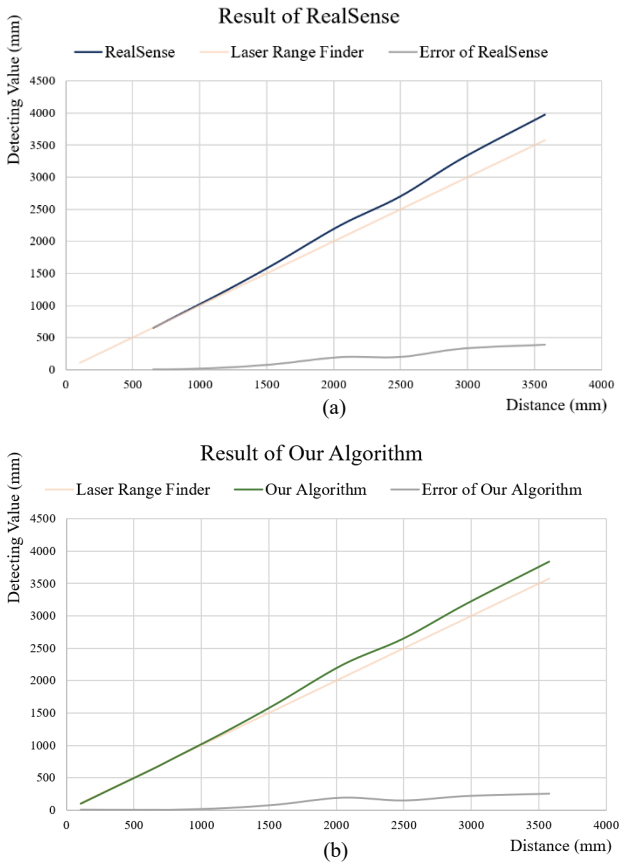


Figure 4. Detecting range and accuracy test.
(a) Detecting range and accuracy of RealSense R200.
(b) Detecting range and accuracy of our algorithm.

- If it has a valid value in only one depth map (either color stereo depth map or RealSense depth map), take the value as the final one.
- If it has valid values in both two depth maps, meanwhile the difference between them is lower than a threshold, the average of the two values is the final depth value.
- If it has two valid values but their difference is larger than the threshold, a voting process will be carried out. A voting window will be set around the pending pixels, and any other pixel in the voting window will vote according to similarity. The one getting the most votes will be the final depth value.
- If it has no valid depth measurement in both depth maps, it is set to be invalid point.

The RealSense depth map and fusion depth map are shown in Fig. 5.

In this work, we bind two RealSense R200 together, and build a short-baseline color stereo vision system using two RGB sensors. By fusing the color stereo depth map and RealSense original depth map, an improving depth map is obtained.

(a) (b) (c)

Figure 5. (a) Color image. (b) Original RealSense depth image.
(c) Depth image obtained by our algorithm.

IV. EXPERIMENTS

In this section, several experiments are conducted to evaluate the performance of our approach. A test on detecting range and accuracy is carried out in the first experiment. In the second one, in order to measure our method's performance on special surfaces, we set up several scenes with objects with challenging surfaces that is non-visible in RealSense depth maps.

A. Detection Range and Accuracy Test

In this experiment, we test our algorithm's performance in terms of detecting range and accuracy. We take the distance value given by a laser range finder as the true value. The test result is shown in Fig. 4. As we can see, the minimum detecting range of RealSense R200 is about 650mm. In the range of 650mm – 1500mm, the distance detecting error using RealSense R200 only is no more than 5%. When the distance comes to 1500mm – 3500mm, the detecting error is no more than 10%. In contrast, our approach decreases the minimum range by 84.6%, from 650mm to 100mm, while maintains the detecting error no more than 5% in 100mm – 1500mm, and less than 10% in 1500mm – 3500mm.

B. Special Surafce Test

In this experiment, we compare the performance of using our approach and using RealSense R200 only in detecting objects within short range or with special surfaces. The RGB images, original RealSense depth maps, and fused depth maps obtained by our algorithm are shown in Fig. 5. The challenging objects in each scene is circled by red circles. As we can see, on specular surfaces (the computer screen), transparent surfaces (the glass bottle), it fails to perform depth measurement using RealSense R200 only. But our algorithm will provide a much denser and more robust depth map. For objects in short distance (the hand), there is a black

hole (it means no depth information) in original depth map given by RealSense, while the hole can be filled appropriately using our approach. Definitely our algorithm performs better than using RealSense R200 only on those challenging situations mentioned.

V. Conclusion and Future Work

RGB-D sensors are a popular choice to navigational assistance for the visually impaired people, which is more efficient than traditional aiding tools. However, they all have problems on special surfaces and close objects. In this paper, we proposed a method that combine two RealSense R200 to build a short-baseline color stereo system. In our work, a better depth map is obtained by fusing RealSense original depth map with our color stereo depth map. As a result, the minimum range is decreased by approximately 84.6%, from 650mm to 100mm. And our algorithm also provides dense depth estimations on transparent objects, reflective objects and absorbing objects. The improved depth maps can be used to provide better assistance to visually impaired people. The approach is tested to realize processing speed of about 6FPS on PC with Intel Core i5-6500 CPU and 8G memory, which satisfies the speed for navigational assistance.

In the future, we are going to build a stereo vision system using two fisheye lenses [17], which can be obtained on Intel RealSense ZR300. Fisheye lenses have a wide FOV, which can definitely provide conditions favorable to navigational assistance for visually impaired people.

VI. Acknowledgement

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 co-funded by State Key Laboratory of Modern Optical Instrumentation.

References

[1] RealSense. Available online: en.wikipedia.org/wiki/Intel_RealSense (accessed on 14 December 2017)

[2] Kinect. Available online: en.wikipedia.org/wiki/Kinect (accessed on 14 December 2017)

[3] Xtion. Available online: www.asus.com/3D-Sensor/Xtion (accessed on 14 December 2017)

[4] Zöllner, M., Huber, S., Jetter, H. C., & Reiterer, H. "NAVI–a proof-of-concept of a mobile navigational aid for visually impaired based on the microsoft kinect," Human-Computer Interaction–INTERACT 2011, 2011, pp. 584-587.

[5] Takizawa, H., Yamaguchi, S., Aoyagi, M., Ezaki, N., & Mizuno, S. "Kinect cane: An assistive system for the visually impaired based on three-dimensional object recognition," 2012 IEEE/SICE International Symposium on System Integration, 2012, pp. 740-745.

[6] Filipe, V., Faria, N., Paredes, H., Fernandes, H., & Barroso, J. "Assisted guidance for the blind using the Kinect device," Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, 2016, pp. 13-19

[7] Hicks, S. L., Wilson, I., Muhammed, L., Worsfold, J., Downes, S. M., & Kennard, C. "A depth-based head-mounted visual display to aid navigation in partially sighted individuals," PloS one, 2013, vol. 8(7), p. e67695.

[8] Draelos, M., Deshpande, N., & Grant, E. "The Kinect up close: Adaptations for short-range imaging," 2012 IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems, 2012, pp. 251-256.

[9] Saygili, G., van der Maaten, L., & Hendriks, E. A. "Hybrid kinect depth map refinement for transparent objects," 2014 22nd International Conference on Pattern Recognition, 2014, pp. 2751-2756.

[10] Chiu, W. C., Blanke, U., & Fritz, M. "I spy with my little eye: Learning optimal filters for cross-modal stereo under projected patterns," 2011 IEEE International Conference on Computer Vision Workshops, 2011, pp. 1209-1214.

[11] Chiu, Walon Wei-Chen, Ulf Blanke, and Mario Fritz. "Improving the Kinect by Cross-Modal Stereo," BMVC, 2011, Vol. 1, No. 2, p. 3.

[12] 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 Environmetns, 2017, 9(6), pp. 743-755.

[13] K. Yang, K. Wang, W. Hu and J. Bai, "Expanding the Detection of Traverasable Area with RealSense for the Visually Impaired," Sensors, 2016, vol. 16(11), p. 1954.

[14] Zhang, Z. (1999). "Flexible camera calibration by viewing a plane from unknown orientations," The Proceedings of 7th IEEE International Conference on Computer Vision, 1999, Vol. 1, pp. 666-673.

[15] Hirschmuller, H. "Stereo processing by semiglobal matching and mutual information," IEEE Transactions on pattern analysis and machine intelligence, 2008, vol. 30(2), pp. 328-341.

[16] Min, D., Choi, S., Lu, J., Ham, B., Sohn, K., & Do, M. N. "Fast global image smoothing based on weighted least squares," IEEE Transactions on Image Processing, 2014, vol. 23(12), pp. 5638-5653.

[17] C. Häne, L. Heng, G. H. Lee, A. Sizov and M. Pollefeys, "Real-time direct dense matching of fisheye images using plane-sweeping stereo," 2014 2nd IEEE International Conference on 3D Vision (3DV), 2014, vol. 1, pp. 57-64.