

INTEGRATING DEPTH AND IMAGE SEQUENCES FOR PLANETARY ROVER MAPPING USING RGB-D SENSOR

M. Peng^{1,*}, W. Wan¹, Y. Xing², Y. Wang¹, Z. Liu¹, K. Di¹, Q. Zhao³, B. Teng², X. Mao²

¹ State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China - (pengman, wanwh, wangyx716, liuzq, dikc)@radi.ac.cn

² Beijing Institute of Control Engineering, Beijing, China -xingyan_bice@163.com, tengweb@hotmail.com, cicely_mao@163.com

³ Huawei Technologies Co. Ltd, Beijing, China -zhaoqiang20@huawei.com

Commission III, ICWG III/II

KEY WORDS: RGB-D camera, Depth, Planetary Rover Mapping, Bundle Adjustment, Point Cloud

ABSTRACT:

RGB-D camera allows the capture of depth and color information at high data rates, and this makes it possible and beneficial integrate depth and image sequences for planetary rover mapping. The proposed mapping method consists of three steps. First, the strict projection relationship among 3D space, depth data and visual texture data is established based on the imaging principle of RGB-D camera, then, an extended bundle adjustment (BA) based SLAM method with integrated 2D and 3D measurements is applied to the image network for high-precision pose estimation. Next, as the interior and exterior elements of RGB images sequence are available, dense matching is completed with the CMPMVS tool. Finally, according to the registration parameters after ICP, the 3D scene from RGB images can be registered to the 3D scene from depth images well, and the fused point cloud can be obtained. Experiment was performed in an outdoor field to simulate the lunar surface. The experimental results demonstrated the feasibility of the proposed method

1. INTRODUCTION

In planetary exploration rover missions, accurate localization of the rover and mapping of the surrounding terrain are essential for safe and efficient execution of exploration tasks, such as traverse planning, hazard avoidance and target approaching. Simultaneous localization and mapping (SLAM) is the key technology to create a reliable 3D map by estimating the camera pose accurately. Stereo vision based rover localization method, which is called visual odometry (VO), has been successfully applied in the 2003 Mars Exploration Rover (MER) mission to reduce position errors accumulated by dead-reckoning localization. In Chang'e-3 mission, cross-site visual localization is also applied for rover localization. However, rational SLAM always depend on images of stereo cameras, the method is almost invalid when rover located at poorly textured areas.

RGB-D camera, consisting of a RGB camera and a depth camera, is a new type of sensor that can directly and simultaneously obtain both visual texture information and per-pixel depth information. It allows the capture of depth and color information at high data rates. Therefore, RGB-D camera has a natural advantage for visual-based localization and mapping. Recently, many researches are focused on dense mapping and simultaneous localization and mapping (SLAM) with RGB-D camera. (Dryanovski et al., 2013; Hu et al., 2012; Whelan et al., 2013, 2015). The Kinect-Fusion algorithm of Newcombe et al. (2011) is one of the most notable RGB-D-based 3D reconstruction systems of recent times. Huang et al. (2011) developed a RGB-D SLAM method in which sparse bundle adjustment (SBA) is used for global consistency by minimizing the matching errors of the visual FAST feature correspondences between frames. Henry et al. (2012) introduced a RGB-D

Mapping framework that can generate dense 3D maps of indoor environments despite the limited depth precision and field of view provided by RGB-D cameras, the system utilizes a novel joint optimization algorithm combining visual features and shape-based alignment. Hu et al. (2012) proposed a switching based algorithm to heuristically choose between RGB-BA and RGBD-BA based local maps building, and a low cost and consistent optimization approach is used to join these maps. based on a novel GPU implementation of an existing RGB-D visual odometry algorithm. Dryanovski et al. (2013) presented a real-time visual odometry and mapping system for RGB-D cameras, which used a probabilistic framework to optimize the camera pose estimation. Whelan et al. (2013) presented a system for improved RGB-D camera pose tracking that yields high quality color surface models with low visual artifacts. Further, Whelan et al. (2015) presented a real-time dense SLAM system which makes use of a dense every-frame volumetric fusion frontend for camera pose estimation and surface reconstruction in combination with a non-rigid map deformation backend to correct the mapped dense surface upon loop closure. Moreover, a new method of RGB-D camera SLAM is proposed based on extended bundle adjustment with integrated 2D and 3D information on the basis of a new projection model (Di et al., 2016).

The remarkable advantages of these systems lie in the high mobility and low cost. However, RGB-D sensors have some significant drawbacks with respect to dense 3D mapping. These sensors only allow measurement ranges of a limited distance and a limited field of view, making the point cloud noisy and loss of details. This may cause details of far range loss due to lack of the depth information needed to constrain ICP (iterative closest point) alignments. To determine how the overlap of

* Corresponding author.

several infrared beams affects the tracked position of the user, Oliver et al. (2016) studied the precision of collected data using multiple RGBD sensors. Tang et al. (2016) presented a novel approach to integrate the depth scene and RGB scene geometrically to enlarge the measurement distance of RGB-D sensors and enrich the details of model generated from depth images. To enhance the quality of depth efficiently, Lee et al. (2017) proposed a method which consists of an image segmentation algorithm to extract object regions and a weighted linear combination of spatial filtering algorithms.

In planetary exploration, the complex conditions of illumination and surface texture may make the traditional methods which only depend on visual feature tracking infeasible. In this paper, we proposed a method based on RGB-D camera by integrating depth and image sequences for planetary rover mapping. Results of field experiment is given to verify the accuracy and effectiveness of this new method.

The rest of this paper is structured as follows: Section 2 presents and specifies the proposed method; Experimental result is presented in Section 3. Finally, conclusions and suggestions for future work are given in Section 4.

2. METHOD

The flowchart of our method is shown in Figure 1. First, the strict projection relationship among 3D space, depth data and RGB data is established based on the imaging principle of RGB-D camera. As there exist holes in depth images, depth images are enhanced. And based on a precise calibration for both of IR and RGB cameras, the relationship between the depth and RGB camera are obtained. Then, an extended bundle adjustment (BA) based SLAM method with integrated 2D and 3D measurements is applied to the image network for high-precision pose estimation, the method takes depth measurements as independent observations and integrates them with image measurements through the projection model and error model. Next, as the interior and exterior elements of RGB images sequence are available, dense matching is completed with CPMVS tool, and the dense model can be generated. Finally, according to the registration parameters after ICP, the 3D scene from RGB images can be registered to the 3D scene from depth images well, and the fused point cloud can be obtained.

2.1 Camera calibration

In order to obtain mapping results accurately, the RGB image and the simultaneously acquired depth image should be registered. As there exist errors of several pixels of the register function given in the Microsoft Kinect SDK, precise camera calibration is performed by the camera calibration model of Di et al.(2016), the lens distortion model can be represented by Equation(1) based on the pinhole camera model

$$\begin{cases} x^d = x + \delta_x, y^d = y + \delta_y, r^2 = x^2 + y^2 \\ \delta_x = x(k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 xy + p_2(r^2 + 2x^2) \\ \delta_y = y(k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1(r^2 + 2y^2) + 2p_1 xy \end{cases} \quad (1)$$

where $(k_1, k_2, k_3, p_1, p_2)$ = lens distortion coefficients
 (δ_x, δ_y) =the camera distortion along x direction and y direction,

(x^d, y^d) = the original image coordinates of an image point
 (x, y) = the image coordinates after distortion correction.

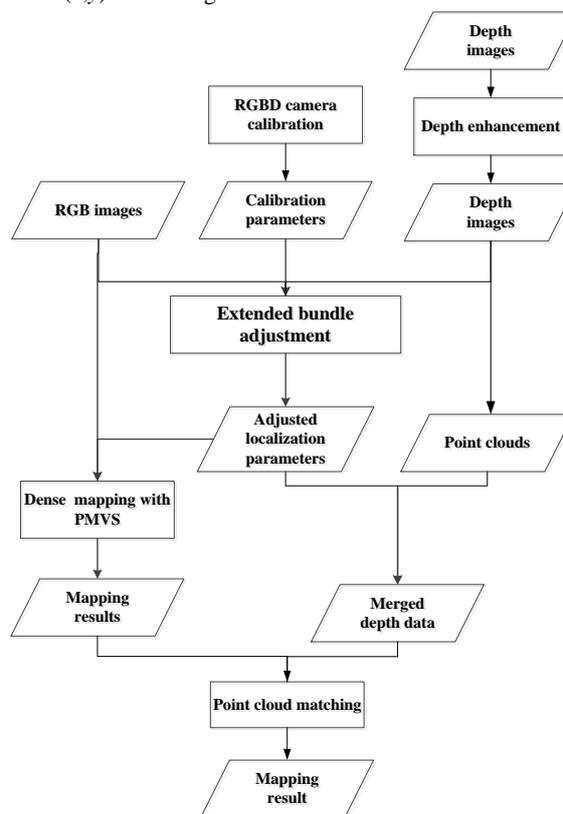


Figure 1. Flowchart of the integrating mapping method

Coordinate systems $o-xyz$ and $o_r-x_r y_r z_r$ of the depth camera and the RGB camera are defined such that the $z(z_r)$ axis points along the optical axis, the $x(x_r)$ axis is horizontal to the right (perpendicular to the plane $z = 0$), and the $y(y_r)$ axis forms the right-hand coordinate system, a 3×3 rotation matrix R and a translation vector $T(X_S, Y_S, Z_S)$ are used to represent the relative relationship between the RGB camera and the depth camera.

Supposing that there is an object point (X, Y, Z) , the projected point on the depth image is (x_D, y_D) the depth value obtained by depth camera is d

$$X = \frac{(x_D - x_{0D}) \cdot d}{f_{Dx}}; Y = \frac{(y_D - y_{0D}) \cdot d}{f_{Dy}}; Z = -d \quad (2)$$

Taking the depth camera coordinate system as the reference coordinate system, the projected points on the RGB image is (x_R, y_R) , so that the imaging geometric models of RGB camera is:

$$\begin{cases} x_R - x_{0R} = -f_{Rx} \frac{R_{11}(X - X_S) + R_{12}(Y - Y_S) + R_{13}(Z - Z_S)}{R_{31}(X - X_S) + R_{32}(Y - Y_S) + R_{33}(Z - Z_S)} \\ y_R - y_{0R} = -f_{Ry} \frac{R_{21}(X - X_S) + R_{22}(Y - Y_S) + R_{23}(Z - Z_S)}{R_{31}(X - X_S) + R_{32}(Y - Y_S) + R_{33}(Z - Z_S)} \end{cases} \quad (3)$$

where (f_{Dx}, f_{Dy}) = the focal lengths of the depth camera
 (f_{Rx}, f_{Ry}) = the focal lengths of the RGB camera
 (x_{0D}, y_{0D}) = principal points of the depth camera
 (x_{0R}, y_{0R}) = principal points of the RGB camera
 $R_{ij} (i, j, = 1, 2, 3)$ = the elements of the rotation matrix R of the RGB camera with respect to the depth camera

Based on equations (1), (2) and (3), the rotation matrix R and the translation vector T can be determined through camera calibration.

The Kinect cameras are calibrated together using a planer checkerboard. Based on images taken from different distances and orientations, Camera Calibration Toolbox for Matlab(http://www.vision.caltech.edu/bouguet_j/calib_doc/) is used to complete the calibration with the images. The calibration results include external and internal parameters, lens distortion coefficients as well as the relative pose between RGB camera and IR camera.

2.2 Depth data enhancement

The depth values returned by the Kinect suffer from noise, and sometimes are absent for some points depending on the reflectance properties of the scene. Therefore the depth map will be enhanced.

Assuming that at regions around the depth holes, there are valid depth values to fill the holes. A depth value histogram of neighboring pixels of depth holes is computed and the dominant peaks are detected in the histogram. The proposed method assumes that a valid depth value to fill a depth hole corresponds to a dominant peak that is the most distant depth value among multiple peaks. The detail of depth data enhancement are shown as follows:

Step 1: Depth holes are detected to make a binary map, B that indicates pixels with unknown depth information as 0 (black).

Step 2: 8-connectivity is used to label depth holes. To decide a neighboring region of depth holes, validation region V is generated of region label map using morphological operation.

Step 3 : The histogram of valid regions of each labeled region is computed, dominant peaks are experimentally chosen as peaks that have pixels more than 10% of the total number of pixels in the region.

Step4 : An average value of dominant depth values is used as threshold value, a median of depth values which are greater than the threshold is used to fill depth holes.

The depth hole filling is a block-wise process. A uniform depth value on invalid region may cause natural artifact. Therefore, we filter the initial depth map by a depth color cross-bilateral filter to remove the artifact and refine the initial depth map. The final depth map is filtered as

$$\hat{D}(p) = \frac{1}{k(p)} \sum_{q \in \Omega_p} f(p, q) g(\|\hat{D}_m(p) - \hat{D}_m(q)\|) h(\|I(p) - I(q)\|) \quad (4)$$

where $f(p, q)$ = domain term that measures the closeness of the pixels, p and q

$g(\cdot)$ = an depth range term that measures the pixel similarity of the modeled depth map

$h(\cdot)$ = an intensity term that measures the intensity similarity.

Ω_p =denotes the spatial neighborhood of position p .

2.3 Extended bundle adjustment

Bundle adjustment of the image network, which is the technique of refining a visual reconstruction to produce jointly optimal 3D

structure and orientation parameters estimated by using accurate projection model, statistical error models and well-developed quality control methodology [35, 36], is used to optimize the initial exterior orientation result. Constructing the projection model and error model of RGB-D camera is the key to achieve optimal estimation.

In order to calculate the initial exterior orientation calculation, firstly 2D and 3D feature points are automatically extracted and matched between consecutive frames to build a continuous image network. 2D features are detected in the registered images and 3D features are re-projected to the registered images, so that all the 2D features and 3D features have the image plane coordinates values and the depth measurement values, and matching errors are detected using RANSAC. These matched feature points are used as tie points to link consecutive frames to build a continuous image network. Then based on the 2D and 3D features, the rotation matrix R and translation vector T between the first frame and next frame are recovered through by SVD.

The projection model of a RGB-D camera represents the relationship of an object point in the real world and its measurements in the RGB-D images. There are two types of measurements: image coordinates from the RGB-image and depth values from the depth image. The camera pose is R and $T=[X_s, Y_s, Z_s]$ which express the relationship between the world coordinate system and the local camera coordinate system For an object point $p=[X, Y, Z]$, its image plane coordinates is (x, y) and the depth value is d .

According to the collinearity equation model, the following equation is obtained

$$\begin{cases} x - x_0 = -f_x \frac{a_1(X - X_s) + b_1(Y - Y_s) + c_1(Z - Z_s)}{a_3(X - X_s) + b_3(Y - Y_s) + c_3(Z - Z_s)} \\ y - y_0 = -f_y \frac{a_2(X - X_s) + b_2(Y - Y_s) + c_2(Z - Z_s)}{a_3(X - X_s) + b_3(Y - Y_s) + c_3(Z - Z_s)} \\ d = -[a_3(X - X_s) + b_3(Y - Y_s) + c_3(Z - Z_s)] \end{cases} \quad (5)$$

where (x_0, y_0) = the coordinate of principal point
 (f_x, f_y) = the focal length of the depth camera.

In the projection model, the relationship between the measurement data and the unknowns is nonlinear. The linearized equation can be represented as:

$$\begin{cases} v_x = a_{11}\Delta X_s + a_{12}\Delta Y_s + a_{13}\Delta Z_s + a_{14}\Delta\omega + a_{15}\Delta\phi \\ + a_{16}\Delta\kappa + a_{17}\Delta X + a_{18}\Delta Y + a_{19}\Delta Z - l_x \\ v_y = a_{21}\Delta X_s + a_{22}\Delta Y_s + a_{23}\Delta Z_s + a_{24}\Delta\omega + a_{25}\Delta\phi \\ + a_{26}\Delta\kappa + a_{27}\Delta X + a_{28}\Delta Y + a_{29}\Delta Z - l_y \\ v_d = a_{31}\Delta X_s + a_{32}\Delta Y_s + a_{33}\Delta Z_s + a_{34}\Delta\omega + a_{35}\Delta\phi \\ + a_{36}\Delta\kappa + a_{37}\Delta X + a_{38}\Delta Y + a_{39}\Delta Z - l_d \end{cases} \quad (6)$$

Rewrite equation (6) into a matrix form:

$$V = AX - L, P \quad (7)$$

where P = the weight matrix of the image plane coordinates and the depth measurement values.

The weight of the measurement data is inversely proportional to the variance of its measurement accuracy. The measurement accuracy of the image plane coordinates depends on the matching accuracy of the SIFT keypoints which is up to sub-pixel level (0.3 pixel is taken in this research). The depth measurement accuracy can be computed by the equation given by Di et al. (2016).

Through the above steps, the error model of RGB-D camera is built and the camera pose of each frame can be refined by least squares solution of Equation (7).

2.4 Dense mapping with sequence images

After bundle adjustment, the point cloud from depth data can be merged. However, the depth sensors capture depth information based on the concept of structured light pattern and time-of-flight, and the measurement is highly related to the material and structure of objects. There would be some holes when modeling objects with smooth surfaces or low reflection certain materials or scene structures which do not reflect infra-red (IR) light. Moreover, the point cloud at far range are of low accuracy.

The sequence images can be used to for dense matching with CMPMVS tool. Camera positions derived from Bundler are used as input, CMVS then decomposes overlapping input images into subsets or clusters of manageable size, while PMVS is used to reconstruct 3-D data from these individual clusters (Furukawa and Ponce, 2007). The core idea of PMVS can be understood as the following steps: Firstly, the feature points in each image are extracted using Harris + DoG operator, and the feature point pairs are matched. Under the condition that the epipolar geometry constraint is satisfied, the triangulation method is adopted. A rectangular patch centered on a three-dimensional space point is generated. The patch is called a seed patch with a normal vector as a specific direction; then two similar characteristics of the adjacent point cloud patch are used as conditions (normal vector and space point location), step by step to diffuse and reconstruct the similarity point clouds of the seed point cloud piece; finally, after the diffusion is completed, a filtering operation is required to remove the error patch that does not meet the consistency. The result of this processing is a significant increase in point density;

2.5 Point cloud matching

Two points pertaining to a point cloud taken by Kinect can have a minimal distance of about a few millimeters. It is acceptable to lose some precision in order to reduce all point clouds' size, reducing considerably both computational time and memory usage. Firstly, the number of points is down sampled by using Voxel Grid. It divides a point cloud in boxes ("voxels") with the desired width. Then, all points within a box are reduced to an unique point corresponding to their centroid.

Point clouds taken from Kinect can have measurement errors that create sparse outliers. Such points can lead to errors while estimating local point clouds features such as surface normals. The outliers are removed as follows: first, mean distance from each point to all its neighbors is calculated. Next, assuming a Gaussian distribution, with a mean and a standard deviation, all points having their mean distances outside a certain interval are considered as outliers and are removed from the dataset. In our work, 30 neighbours are used for each point to analyse its status, removing all points having a distance larger than 1 standard deviation of the mean distance to the analysed point.

We use the iterative closest point (ICP) algorithm (Besl and McKay, 1992) for registration of point cloud by PMVS to depth point cloud. As represented in Equation (8), the goal of the ICP algorithm is to minimize the sum of square errors with respect to the PMVS points and the corresponding closest depth points. In each iteration step, the algorithm selects the closest points as the correspondences and re-calculates the rotation and translation parameters (R , t) of the rigid transformation to minimize the equation

$$F = \min_{R,t} \sum_{i=1}^{N_m} \sum_{j=1}^{N_d} \|d_i - (Rm_j + t)\|^2 \quad (8)$$

Where m_i = coordinates of the RGB dense points

d_i = coordinates of the depth points

t = a 3*1 vector that describes the translations between the two datasets,

R = a 3*3 matrix that describes the rotations between the two datasets.

3. EXPERIMENT

To verify the actual performance of the proposed method, a field experiment has been performed. Figure 2 shows the moving platform (model rover) used in these experiments. A Microsoft Kinect V2 camera which has a RGB image resolution of 1920*1080 pixels and a horizontal field of view of 70 degrees was attached on the top of the camera mast. The camera is about 100cm above the ground.



Figure 2. The RGB-D camera mounted on the moving platform

Experiment was performed in an outdoor field to simulate the lunar surface. The rover travelled along a loop path with the origin set at [0,0]. The same image was used for the first and last positions to ensure that the true last camera pose was exactly the same as where the first image was recorded. Given that the loop is closed, we can use it to evaluate the accuracy. Figure 3 shows some typical images acquired by the RGB camera and the depth camera.



Figure 3. Typical RGB images

Figure 4(b) shows depth images after enhancement, compared to the original depth images in Figure 4(a), the proposed method uses depth distributions of neighboring regions of depth holes. Due to some invalid areas, there still exist holes in enhanced images.

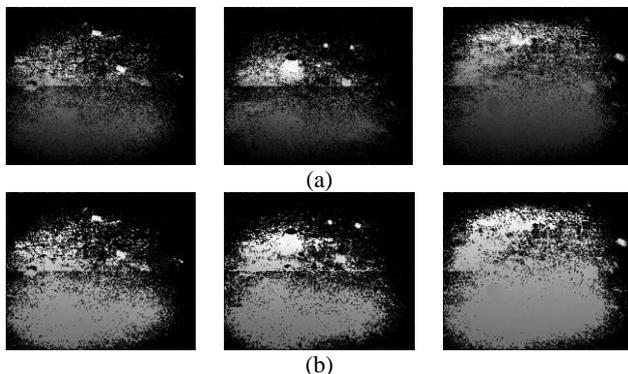
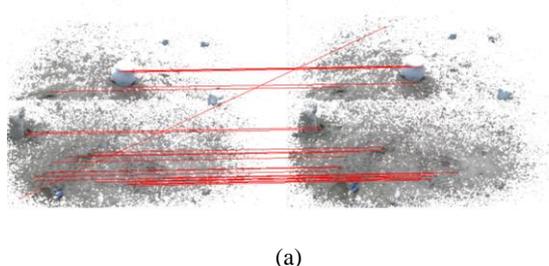


Figure 4. (a) Initial depth images (b) depth images after enhancement

Based on the calibrated parameters, we register the depth image and RGB image. The result of features detection and matching between registered images is shown in Figure 5.

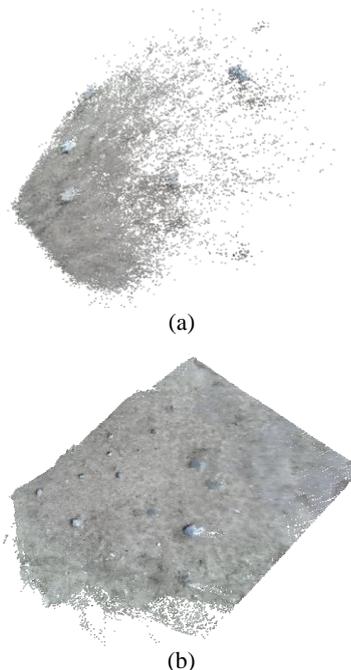


(a)



(b)

Figure 5. Features detection and matching results. (a) initial matching results (b) matched points after RANSAC



(a)

(b)

Figure 6. Point cloud results. (a) initial matching results (b) matched points after RANSAC

Figure 6(a) and (b) show the original 3D scene generated from depth image and the enhanced 3D scene fusing 3D information from RGB images and from depth images. Only a close-range scene with about 3.6 m maximum length can be obtained from the depth images. After point cloud fusing, the vertices number have a significant increase from 62,979 to 147,964 and the measurement distance can be extended to about 8 m. Thus the information from the RGB image sequences both enriched the details for the close-range model from the depth images and greatly broadened the mapping range of the RGB-D camera.

4. CONCLUSION

In this paper, we proposed a method based on RGB-D camera by integrating depth and image sequences for planetary rover mapping. We focused on verifying mapping ability of RGB-D camera onboard a rover that could be used in a GPS-denied environment such as lunar and Martian surface. In outdoor environment with sunlight, the depth sensor may have weak performances, hole filling strategy and cross-bilateral filter were used for depth enhancement. Based on the extended bundle adjustment accurate pose estimation results are refined. The 3D scene from RGB images with PMVS tools were registered to the 3D scene from depth images well after ICP. Our method

provides an effective way to enlarge mapping distance of a RGB-D camera. Further research will be performed to improve the computational efficiency of the method. And more data will be used for validation of the new method.

ACKNOWLEDGEMENTS

This study was supported in part by National Natural Science Foundation of China under Grants 41671458 and 41771488.

REFERENCES

Besl, P., N. McKay, 1992. A Method for Registration of 3-d Shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), pp. 239-256.

Dryanovski, I., Valenti, R.G., and Xiao, J. 2013. Fast visual odometry and mapping from RGB-D data. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation (ICRA), Karlsruhe, Germany, 6–10 May 2013, pp. 2305–2310.

Furukawa Y, Ponce J. Accurate, dense, and robust multi-view stereopsis, 2007: IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA, USA: IEEE Computer Society, 1-8.

Henry, P.; Krainin, M.; Herbst, E.; Ren, X.; Fox, D., 2012. RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments. *Int. J. Robot. Res.*, 31, 647–663.

Huang, A.S.; Bachrach, A.; Henry, P.; Krainin, M.; Maturana, D.; Fox, D.; Roy, N. Visual odometry and mapping for autonomous flight using an RGB-D camera. In Proceedings of the 15th International Symposium on Robotics Research (ISRR), Flagstaff, AZ, USA, 28 August–1 September 2011.

Hu, G.; Huang, S.; Liang, Z.; Alempijevic, A.; Dissanayake, G. A robust RGB-D slam algorithm. In Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura, Algarve, Portugal, 7–12 October 2012; pp. 1714–1719

Lee, K., Ban, Y., & Lee, S. (2017). Efficient Depth Enhancement Using a Combination of Color and Depth information. *Sensors*, 17(7), 1544.

Newcombe, R.A.; Izadi, S.; Hilliges, O.; Molyneaux, D.; Kim, D.; Davison, A.J.; Kohi, P.; Shotton, J.; Hodges, S.; Fitzgibbon, A. Kinectfusion: Real-time dense surface mapping and tracking. In Proceedings of the 2011 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR), Basel, Switzerland, 26–29 October 2011; pp. 127–136.

Oliver M, Montero F, Molina J P, et al. Multi-camera systems for rehabilitation therapies: a study of the precision of Microsoft Kinect sensors. *Frontiers of Information Technology & Electronic Engineering*, 2016, 17(4): 348-364.

Rusu R B, Blodow N, Beetz M. Fast point feature histogram for 3D registration. *Robotics and Automation, IEEE International Conference*, 2009: 3212-3217

Tang, S., Zhu, Q., Chen, W., Darwish, W., Wu, B., Hu, H., & Chen, M. (2016). Enhanced RGB-D mapping method for detailed 3D indoor and outdoor modeling. *Sensors*, 16(10), 1589.

Whelan, T.; Johannsson, H.; Kaess, M.; Leonard, J.J.; McDonald, J. Robust real-time visual odometry for dense RGB-D mapping. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation (ICRA), Karlsruhe, Germany, 6–10 May 2013; pp. 5724–5731.

Whelan, T.; Kaess, M.; Johannsson, H.; Fallon, M.; Leonard, J.J.; McDonald, J. Real-time large-scale dense RGB-D slam with volumetric fusion. *Int. J. Robot. Res.* 2015, 34, 598–626.

Yang, M., Cui, J., Jeong, E. S., & Cho, S. I. (2017). Development of High-resolution 3D Phenotyping System Using Depth Camera and RGB Camera. In 2017 ASABE Annual International Meeting. American Society of Agricultural and Biological Engineers.