International Journal of Sensors and Sensor Networks
2022; 10(1): 1-6
http://www.sciencepublishinggroup.com/j/ijssn
doi: 10.11648/j.ijssn.20221001.11
ISSN: 2329-1796 (Print); ISSN: 2329-1788 (Online)

SciencePG
Science Publishing Group

Multi-frame Point Cloud Fusion Method Based on Depth Camera Sensors

Yang Zhongfan[1, 2], **Wang Xiaogang**[1, 2, *], **Hou Jing**[1, 2]

[1]School of Automation & Information Engineering, Sichuan University of Science & Engineering, Yibin, China
[2]Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science & Engineering, Yibin, China

Email address:

yang56535@163.com (Yang Zhongfan), wxg_zf@163.com (Wang Xiaogang)
[*]Corresponding author

To cite this article:

Yang Zhongfan, Wang Xiaogang, Hou Jing. Multi-frame Point Cloud Fusion Method Based on Depth Camera Sensors. *International Journal of Sensors and Sensor Networks*. Vol. 10, No. 1, 2022, pp. 1-6. doi: 10.11648/j.ijssn.20221001.11

Received: December 22, 2021; **Accepted**: January 7, 2022; **Published**: January 15, 2022

Abstract: As a consumer-grade portable depth image data acquisition device, the depth camera is widely used in the field of computer vision, such as slam, autonomous driving, environment perception, etc. However, due to the limitation of the device angle, the complete 3D point cloud of the target cannot be obtained at one time. Point cloud registration can complete the overlap of two frames of point clouds. Therefore, a multi-frame point cloud fusion method based on key points and registration is proposed. First, the point cloud is calculated on the depth map obtained by the depth camera, and then an improved point cloud filtering algorithm based on the normal vector inner cumulus is used to remove the background and noise points. Secondly, four key point detection algorithms and three registration algorithms with different principles are applied to the point cloud data obtained by the depth camera, and the applicable scenarios and limitations of each algorithm are analyzed. Finally, a multi-frame point cloud fusion algorithm is used to splice the point clouds, and the redundant points after splicing are filtered out to obtain a complete point cloud of the object. The experimental verification of the target object using the depth camera shows that the proposed method can obtain the complete point cloud data of the target object robustly.

Keywords: Point Cloud, Kinect, Filter, Registration

1. Introduction

One of the mainstream applications of 3D vision is 3D reconstruction. A variety of 3D reconstruction technologies have been developed based on different principles, such as binocular stereo vision [1], structured light [2, 3], Lidar [4, 5], TOF depth camera [6], etc. The reconstruction technology used is different according to the difference of reconstruction scenes, but the common purpose is to obtain depth information. In the absence of marker positioning, it is difficult to solve the external parameter matrix between the camera coordinate system and the world coordinate system. There are direct solutions, such as three-dimensional coordinate measuring machines, three-dimensional target direct calibration methods, and two-dimensional plane target calibration methods, etc. The coordinate measuring machine is expensive and complicated to operate. The three-dimensional target direct method and the two-dimensional plane target calibration method need to solve overdetermined equations, iterative optimization, and a large amount of calculation. There are also restrictions on a certain dimensional coordinate, which can be solved by restriction conditions, but it needs to have restrictions on the light plane or the position of the camera, and it is easy to be disturbed.

At present, the mature portable 3D vision devices on the market mainly include 3D laser scanners, depth cameras, and lidars, each of which has its advantages and disadvantages. The 3D laser scanner has the highest accuracy, but because it is based on optical principles, the design of hardware chips and algorithms is more complicated, and the cost of a single unit is too expensive for ordinary users to bear. The depth camera was originally developed by Microsoft. Its products are Kinect v1 and kinect v2, which are mainly used in 3D somatosensory games, and their accuracy is lower than that of laser scanners. Lidar is mainly used in the field of autonomous driving, and its imaging range can reach tens of meters, so the application scene is outdoors and the accuracy is lower.

With the rapid development of sensors in recent years, many researchers have been able to solve problems with the help of three-dimensional vision. Ji Baibing [7] designed a monocular three-dimensional reconstruction system that can provide pose information for robot grasp tasks. Sun Shuo [8] designed a three-dimensional facial reconstruction system for virtual facial plastic surgery to provide data for medical plastic surgery. Huang Zhiming [9] uses a depth camera and an ultrasonic sensor to detect transparent obstacles and provides visual assistance to the visually impaired. Dai Wen [10] used a depth camera to perform three-dimensional reconstruction of complex workpieces and applied the three-dimensional reconstruction technology to the field of industrial spraying.

It can be seen that the 3D reconstruction method of the depth camera has gradually been applied in various industries and has a large application scenario. In this paper, the KinectV2 depth camera method is used to study the point cloud algorithm, and the reconstruction is aimed at a single target. The reconstructed object is a workpiece or other object with a radius of half a meter. The purpose is to reconstruct the complete point cloud of the target object to verify the improved filtering algorithm, the comparison of multiple registration algorithms, and the multi-frame fusion registration method.

2. Methods and Results

2.1. Point Cloud Data Acquisition

When using Kinect to acquire depth images, due to the limitations of the device itself, only one depth image of the object to be measured can be captured at a time. To get the complete information of the object, we need to shoot from multiple angles. Generally, we can shoot by keeping the object still and moving the camera or keeping the camera still and rotating the object for shooting.

The SDK officially provided by Microsoft cannot directly output raw infrared data. Instead, after the infrared camera obtains the data, it will perform calculation processing in its built-in chip to obtain the depth data of the object.

The specific steps to obtain the depth image are:

(1) Check whether the Kinect equipment is operating normally;

(2) Initialize the device and set the depth data stream of the device to a usable state;

(3) Call the internal acquisition function, extract the depth data and store it;

(4) Release the data frame and Kinect device.

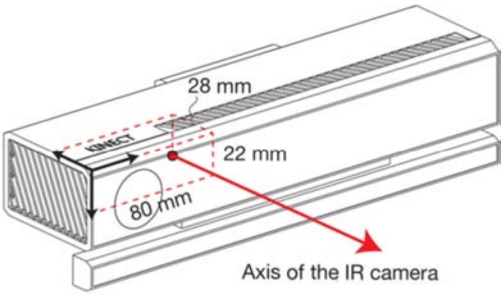


Figure 1. *Kinect v2 camera.*

The world coordinate system in Figure 1 takes the KinectV2 device as the origin of the coordinates, while in the coordinate system of the depth image, the origin of the depth image is used as the origin of the coordinates. Therefore, it is necessary to calculate the X and Y coordinates in the real-world coordinate system based on the measured depth z coordinate information.

The camera model can be regarded as an ideal pinhole imaging model. The depth image resolution of KinectV2 is 512x424. Based on the triangle similarity principle, the XY coordinates can be calculated as:

$$X = (u - 256) \times Z \times \frac{1}{f} \tag{1}$$

$$Y = (v - 212) \times Z \times \frac{1}{f} \tag{2}$$

In the formula, u, v, Z, and f are known.

2.2. Point Cloud Filtering

After the depth image in the previous section is converted into a point cloud, due to factors such as target material, ambient light, and calculation errors, a lot of noise will be generated in addition to the background, as shown in Figure 2.

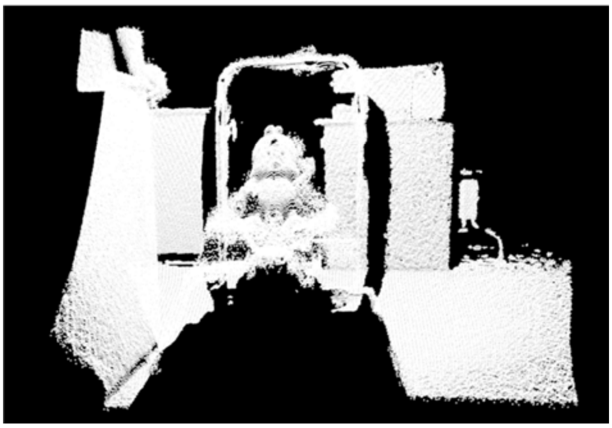


Figure 2. *Initial data with background noise.*

Generally, the point cloud data obtained by the point cloud acquisition device is very large, which not only contains the environmental information around the target but also has a large number of outliers. If only the target is required to be processed, it is necessary to filter the original point cloud to remove redundant point clouds. This paper uses conditional filtering and an improved radius removal algorithm to remove environmental and noisy point clouds in turn.

The improved filtering algorithm process is:

1. Determine the lower limit of the Z-axis of conditional filtering according to the distance between the origin of the camera coordinate system of the adjusted depth camera and the target object;
2. Determine the upper limit of the Z-axis, X-axis interval, and Y-axis interval of the filter based on the size of the target object itself;
3. Calculate the normal vector for each point of the point cloud initially processed in steps 1-2;

4. Calculate the inner product α of the normal vector of each point and the normal vector of the adjacent point. If α>0 and α<ε (ε>0), then the point is reserved; if α<0 or α>ε (ε>0), then Discard this point.

For the background environment, conditional filtering can be used to remove. The idea of conditional filtering is relatively simple. According to the approximate position of the target in the point cloud, other points can be removed by the threshold limit. The processed result is shown in Figure 3:

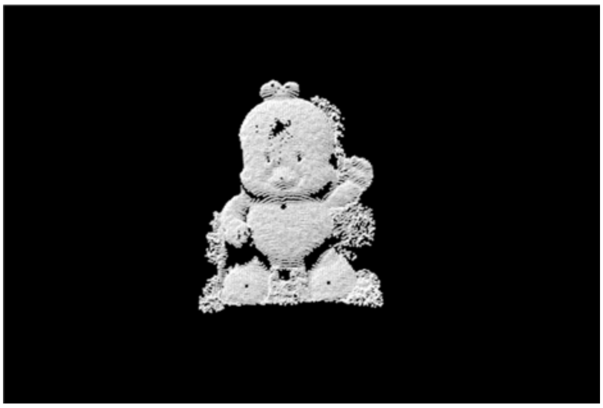


Figure 3. *Remove background noise.*

As can be seen from the above figure, there are still irregular and disorderly patchy point clouds after preliminary filtering. The difference between this noise and an object is that the object's point cloud is a relatively smooth surface, and the noise is an irregular noise. This noise can be removed by calculating the plane normal vector of all points in a certain radius area, and then setting the threshold of the normal vector inner product. The rule is: if the normal vector inner product between adjacent points is positive and less than a certain threshold, the point is considered to be the target object (inner point); if the normal vector inner product between adjacent points is negative or greater than a certain threshold Threshold, the point is considered a noise point (outside point). Applying the filter designed by this rule, the final target point cloud is shown in Figure 4.

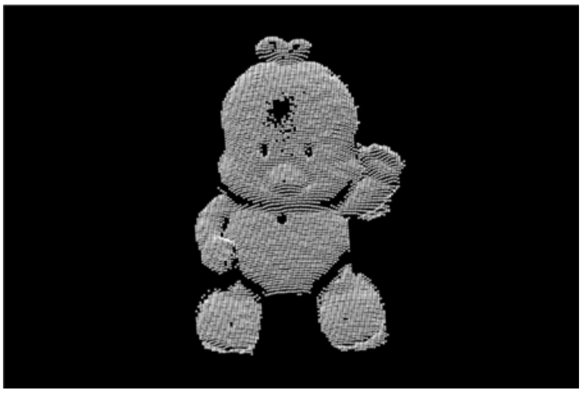


Figure 4. *Further denoising.*

Applying the above filtering algorithm to the point cloud with 12 frame intervals of 30°, the obtained point cloud is shown in Figure 5.

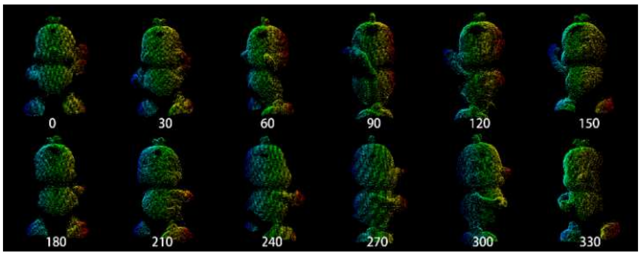
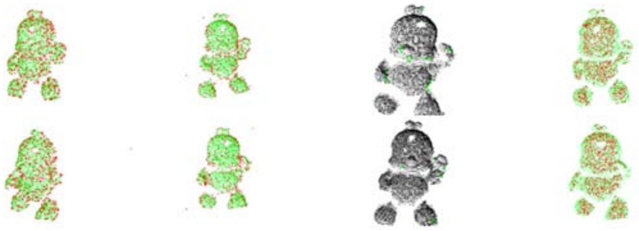


Figure 5. *Frame point cloud from various angles.*

2.3. Keypoint Extraction

Key points are also called feature points. The points with special information calculated by specific rules are further calculated with the surroundings and called descriptors. The registration algorithm based on feature points performs registration on these matched descriptors. This article has studied SIFT3D [11], Harris3D [12], NARF [13], ISS3D [14], the main key point extraction algorithms. I will not repeat the principle here but focus on the analysis of what they extract. The characteristics and application scenarios of the key points.

Among several key points, SIFT3D is similar to SIFT2D, and the extracted key point characteristics are scale-rotation-invariant, but they have more dimensions than two-dimensional images and are computationally expensive. Harris3D is a corner detection algorithm, which is also developed from the Harris2D algorithm, and mainly extracts corners. Although NARF is also a key point detection algorithm, it mainly detects special points on the edge, that is, key points are mainly distributed on the edge, which is suitable for scenes that identify objects from the background. ISS3D establishes a local coordinate system and performs feature value analysis, and the extracted key points have local features.



（From left to right, SIFT3D, Harris3D, NARF, ISS3D）

Figure 6. *Four key points extraction and comparison.*

Figures 2-6 shows the results of four key point algorithms detecting point clouds in two adjacent frames. According to the analysis of literature [15], the key points obtained by the ISS algorithm can represent local information, the number of matching points is the largest, and it is not interfered by noise. Therefore, this paper chooses the ISS algorithm to extract the key points.

2.4. Point Cloud Registration

To stitch the point clouds between the various perspectives together, it is necessary to perform pairwise registration between the point clouds of two adjacent frames. In this paper, the registration algorithm based on three different principles of feature matching, random search, and statistical model is tested on actual data, and the corresponding representative registration algorithms are selected respectively: SAC-IA [16], ISS-4PCS, and NDT [17] are compared.

The algorithm steps of SAC-IA are as follows:

1. Calculate the FPFH feature descriptors of the source point cloud and target point cloud separately;
2. Match the points in the two-point clouds based on the FPFH feature descriptor;
3. Randomly select n (n≥3) pairs of matching points;
4. Solve the rotation vector and translation vector in this matching case by SVD;
5. Calculate the root mean square error;
6. Repeat steps 3-5 until the conditions are met, and take the rotation vector and translation vector corresponding to the minimum root mean square error as the final result.

The algorithm steps of ISS-4PCS are as follows:

1. Extract the ISS key points of the source point cloud P and the target point cloud Q respectively, and express them as P',Q';
2. Randomly select 3 points from P' and Q' respectively;
3. Then according to the source point cloud P and the overlap rate f of the target point cloud Q, select the fourth coplanar point far enough from the other 3 points (also selected from P', Q') to form a coplanar four-point base B;
4. Then according to the affine invariant ratio, extract all the 4-point sets $\mathrm{U} = U_1, U_2, U_{3,\ldots}$ that may be consistent with B within a certain distance δ from the point set Q. For anyU_i, calculated by the relationship between B and U_i rigid transformation T;
5. Test the different bases of the L group according to the overlap ratio. When a constant number of random sampling points in P have enough corresponding points in Q, the best rigid transformation matrix T_{best} for rough registration is obtained.

The NDT algorithm steps are as follows:

1. Divide the space occupied by the source point cloud into a specified size grid or voxel (cell);
2. Calculate the multi-dimensional normal distribution parameters of each grid cell:

Calculate the center of the included points in the grid (the average of each axis) and the covariance matrix (similar to the ISS keypoints):

$$q = \frac{1}{n}\sum_i x_i \tag{3}$$

$$\Sigma = \frac{1}{n}\sum_i (x_i - q)(x_i - q)^T \tag{4}$$

3. Transform the target point cloud to the source point cloud coordinate system:

$$\mathrm{T}: \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R\begin{pmatrix} x \\ y \\ z \end{pmatrix} + t \tag{5}$$

4. Calculate the probability of each conversion point falling in the corresponding grid according to the normal distribution parameters:

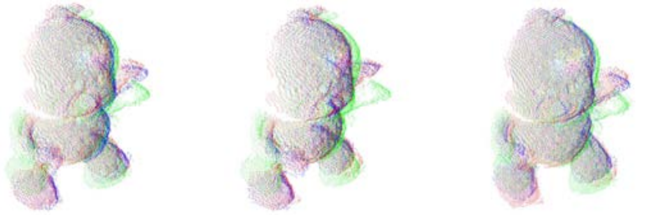
$$\mathrm{p(x)}{\sim}\mathrm{exp}(-\frac{(x-q)^T\Sigma^{-1}(x-q)}{2}) \tag{6}$$

5. NDT registration score: Calculate the sum of the probability that the corresponding point falls in the corresponding grid:

$$\mathrm{score(p)} = \sum_i \mathrm{exp}(-\frac{(x_i'-q_i)^T\Sigma_i^{-1}(x_i'-q_i)}{2}) \tag{7}$$

6. Optimize the objective function according to the Newton optimization algorithm, that is, find the transformation parameters to maximize the value of score;
7. Jump to step 3 and continue execution until the convergence condition is reached.

Figure 7 shows the comparison of the registration results of the three algorithms for 0-30° point cloud. Green is the source point cloud, red is the target point cloud, and blue is the transformed point cloud. The higher the degree of coincidence between the blue point cloud and the red point cloud, the better the effect of the registration algorithm. From the point of view of the details of the point cloud, the ISS-4PCS has the best registration result. Table 1 shows the transformation matrix and the configuration calculated by the three algorithms. The root means square error of the corresponding point after accurate, the registration time. From Table 1 again, the root means square error of ISS-4PCS is the smallest and the registration time is the shortest.



（From left to right are SAC-IA, ISS-4PCS, NDT）

***Figure** 7. Comparison of the effects of three-point cloud registration algorithms.*

***Table 1.** Root mean square error and registration time.*

	RMSE (m)	Registration time (s)
SAC-IA	0.00421947	0.905539
ISS-4PCS	0.00387012	0.60386
NDT	0.00679893	14.8562

2.5. Point Cloud Fusion

In the experiment in the previous section, we selected the ISS-4PCS algorithm with the best registration effect for pairwise registration and obtained the transformation matrix of the point cloud of two adjacent frames. Then the ICP algorithm is used for fine registration. The following figure shows the result of the target point cloud registration. Then the point clouds of 11 frames of other perspectives are transformed to the same coordinate system through the transformation matrix. The result is shown in Figure 8.

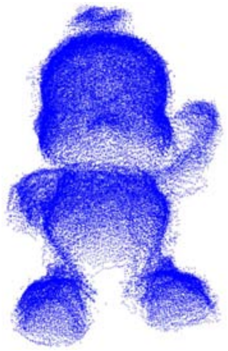


Figure 8. *Target point cloud fusion result.*

It can be seen that after multi-frame point cloud fusion, there are still ghosting phenomena and extra points on the edge. It can be further optimized to remove redundant points. First, the nearest point approximate center voxel filtering algorithm is used to remove ghosts, and then the radius filtering algorithm is used to remove isolated points. The result after processing is shown in Figure 9.

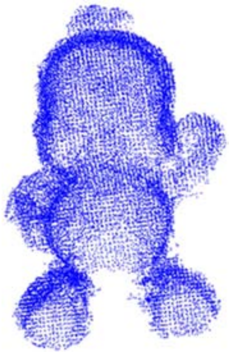


Figure 9. *Final result.*

3. Conclusion

The 3D point cloud reconstruction of the target object is one of the key technologies of 3D vision. In this paper, a depth camera is used to complete the 3D point cloud reconstruction of the target object. Focusing on the analysis of common algorithms for point cloud registration, the ISS-4PCS algorithm with the best comprehensive performance is selected, and various filtering algorithms are used to complete the fusion of multi-frame point clouds and achieve better point cloud results. The reconstruction error is at the millimeter level. Although the accuracy is lower than that of a 3D scanner, the reconstruction efficiency is high, suitable for rapid reconstruction, and low cost. It can provide data support for related point cloud algorithm research.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China (grant no. 61902268), Sichuan Science and Technology Program (grant no. 2018JY0197, 19ZDZX0037, 2019YFSY0045, 20ZDYF0919, 21ZDYF4052, 2020YFH0124, 2021YFSY0060).

References

[1] Liu Lexuan. Lane Offset Survey for One-Lane Horizontal Curvatures Using Binocular Stereo Vision Measurement System [J]. Journal of Surveying Engineering, 2021, 147 (4).

[2] Zhao Jianping, Feng Chang, Cai Gen, Zhang Run, Chen Zhibo, Cheng Yong, Xu Bing. Three-dimensional reconstruction and measurement of fuel assemblies for sodium-cooled fast reactor using linear structured light [J]. Annals of Nuclear Energy, 2021, 160.

[3] Zhenzhou Wang, Qi Zhou, YongCan Shuang. Three-dimensional reconstruction with single-shot structured light dot pattern and analytic solutions [J]. Measurement, 2020, 151 (C).

[4] Li Xingdong, Gao Zhiming, Chen Xiandong, Sun Shufa, Liu Jiuqing. Research on Estimation Method of Geometric Features of Structured Negative Obstacle Based on Single-Frame 3D Laser Point Cloud [J]. Information, 2021, 12 (6).

[5] Yevgeny Milanov, Vladimir Badenko, Vladimir Yadykin, Leonid Perlovsky. Method for clustering and identification of objects in laser scanning point clouds using dynamic logic [J]. The International Journal of Advanced Manufacturing Technology, 2021, 117 (7-8).

[6] M. Ruiz-Rodriguez, V. I. Kober, V. N. Karnaukhov, M. G. Mozerov. Algorithm for Three-Dimensional Reconstruction of Nonrigid Objects Using a Depth Camera [J]. Journal of Communications Technology and Electronics, 2020, 65 (6).

[7] Baibing Ji, Qixin Cao. A monocular real-time 3D reconstruction system for robot grasping [J]. Machinery Design and Manufacturing, 2021 (09): 287-290. DOI: 10.19356/j.cnki.1001-3997.2021.09.064.

[8] Shuo Sun, Xiaoqiang Ji, Dan Liu. Design of three-dimensional face reconstruction system for facial virtual plastic surgery [J]. Science Technology and Engineering, 2021, 21 (25): 10806-10813.

[9] Zhiming Huang. Research on Transparent Obstacle Detection Technology Used for Visual Aid for the Blind [D]. Zhejiang University, 2020. DOI: 10.27461/d.cnki.gzjdx.2020.003506.

[10] Wen Dai. 3D measurement of complex workpiece based on depth camera [D]. Hunan University, 2019. DOI: 10.27135/d.cnki.ghudu.2019.003470.

[11] B. Rister, M. A. Horowitz and D. L. Rubin, "Volumetric Image Registration from Invariant Keypoints," in IEEE Transactions on Image Processing, vol. 26, no. 10, pp. 4900-4910, Oct. 2017. DOI: 10.1109/TIP.2017.2722689.

[12] Ivan Sipiran, Benjamin Bustos. Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes [J]. The Visual Computer, 2011, 27 (11).

[13] B. Steder, R. B. Rusu, K. Konolige and W. Burgard, "Point feature extraction on 3D range scans taking into account object boundaries," 2011 IEEE International Conference on Robotics and Automation, 2011, pp. 2601-2608, DOI: 10.1109/ICRA.2011.5980187.

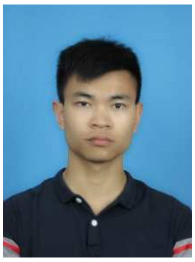
[14] Zhong Y. Intrinsic shape signatures: a shape descriptor for 3D object recognition [C]. IEEE International Conference on Computer Vision Workshops, 2009: 689-696.

[15] Z. Yang, X. Wang and J. Hou, "A 4PCS Coarse Registration Algorithm Based on ISS Feature Points," 2021 40th Chinese Control Conference (CCC), 2021, pp. 7371-7375, DOI: 10.23919/CCC52363.2021.9549486.

[16] RUSU R B, BLODOW N, Fast point feature histograms (FPFH) for 3D registration [C] // IEEE International Conference on Robotics and Automation. Kobe: IEEE, 2009.

[17] P. Biber and W. Strasser, "The normal distributions transform: a new approach to laser scan matching," Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No. 03CH37453), 2003, pp. 2743-2748 vol. 3, DOI: 10.1109/IROS.2003.1249285.

Biography



Yang Zhongfan was born in Dazhou City, Sichuan Province, China in 1997. He received a bachelor's degree from Sichuan University of Science and Engineering, and now he is a graduate student in the school of Automation and Information Engineering of Sichuan University of Science and Engineering. His main research direction is 3D computer vision and point cloud processing.

Wang Xiaogang was born in Baoji City, Shanxi Province, China in 1984. He received his Ph. D degree from the Chongqing University, Chongqing, China. He is currently an associate professor, with Artificial Intelligence Key Laboratory of Sichuan Province, School of Automation and Information Engineering, Sichuan University of Science and Engineering. His current interests are in the area of wireless sensor network and security, IoT, artificial intelligence.

Gan Shuchuan was born in 1967. He received his Master's degree from Sichuan University in 2003. He is currently a professor school of Automation and Information Engineering of Sichuan University of Science and Engineering. He is currently mainly engaged in Integration of management and control, intelligent measurement and control technology, intelligent building technology, chemical instrument corrosion prediction technology and other aspects of scientific research.