Fusion Algorithm of Laser-Point Clouds and Optical Images

Yuehai Wang, Si Li, and Na Xing
North China University of Technology, Beijing, China
Email: wangyuehai@ncut.edu.cn, 1294899761@qq.com, xingna@ncut.edu.cn

***Abstract*—The fusion of laser-point clouds and optical images has been increasingly used in the field of environmental sensing and intelligent devices. It has boosted the development of areas such as entertainment, online house viewing, virtual reality and driverless cars. Traditional fusion algorithms of laser-point clouds and images require a lot of interactive operations such as artificial calibration of sensors and setting of control points. To solve the above problems, this paper proposed an automatic fusion algorithm of laser-point clouds and optical images based on depth images. In this algorithm, the laser-point clouds are converted into depth images, and the mapping relationship between depth images and optical images will be solved. Then the laser-point clouds and the optical images can be fused. The experimental results show that this method can not only ensure the accuracy, but also realize the automation.**

***Index Terms*—LIDAR (Light Detection and Ranging), point clouds, depth images, optical images**

I. Introduction

LIDAR can range objects with high accuracy and obtain spatial geometric information of objects quickly and accurately. However, the point clouds model cannot express the real color information of the object, resulting in the false color display of the point clouds model. It seriously affects the realistic representation of 3D model. Color cameras can obtain details and color information of objects or geographical scenes, but they cannot be directly used to obtain 3D information of target objects. By fusion, the advantages of laser-point cloud and optical images can be combined. Then the point clouds data has the real color information in the objective world.

In recent years, scholars' research directions on the fusion of point clouds and images can be roughly divided into two categories: One is to carry out feature point matching between point clouds data and color images in different dimensions so as to give color to the point clouds data. The representative one is the point clouds and images fusion method based on 3D-2D feature matching proposed by Wang X. [1] *et al*. However, this method relies on manual visual method to select point clouds and color images feature points, and the accuracy is not ideal. The other is to find the relationship between the coordinate system of depth camera and color camera, so as to realize the fusion of point clouds and images. A more representative one is the point clouds and images fusion method based on camera position relationship proposed by Zhang X. [2]. This method still requires manual setting of control points to determine the position relationship between the depth camera and the RGB camera. In addition, the relation matrix is ideal, and this method needs post-processing to obtain the high precision fusion effect.

To solve the above problems, this paper proposes a laser- point clouds and optical images fusion algorithm based on depth images. The method maps the laser-point clouds into a two-dimensional space and generates a depth image, and it uses the depth images and optical images to match the feature points, and obtains the mapping relationship between the depth images and optical images, so as to realize the fusion of laser- point clouds and optical images.

II. Algorithm Implementation

In this paper, the fusion of laser-point clouds and optical images algorithm is used to obtain the rotation and translation matrix, and then the color information of the corresponding pixel of the optical images is given to the laser-point clouds to generate the color point clouds. Depth images is generated by perspective projection of laser-point clouds, and the depth information of point clouds data becomes the gray value of pixel points corresponding to the depth images. The gray value of the depth images is the value mapped by the depth information of the corresponding point clouds in the interval [0, 255]. Then the depth images and optical images are matched to find out the feature points. And the mapping matrix of point clouds and images can be solved, so as to realize the fusion of laser- point clouds and optical images.

A. Conversion of Laser-Point Clouds to Depth Images

The point clouds data (X, Y, Z) are transformed into depth images (x, y, Z') according to the projection model. Z' is the gray value of the depth images.

Projection model is a kind of mapping method. Projection model can be roughly divided into three categories: panoramic images projection, perspective projection and center projection. Panoramic images projection will cause distortion in large angle area [3]. Perspective projection will lead to pixel void and spatial

Manuscript received January 15, 2022; revised June 25, 2022.

doi: 10.18178/joig.10.3.127-131

occlusion [4]. But central projection will not cause the above problems. Therefore, central projection is adopted as the projection model of depth images generation in this paper.

The central projection model is shown in Fig. 1.

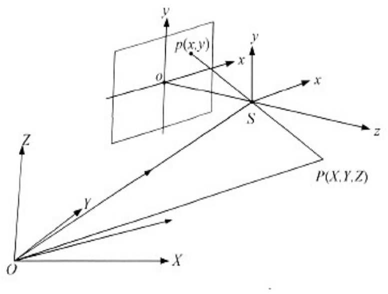


Figure 1. Central projection model.

where *XYZ-O* is the laser-point clouds coordinate system, and $\mathrm{P}(X,Y,Z)$ is a point in the coordinate system. *xyz-s* is the coordinate system of the objective world; *xy-o* is the projected plane coordinate system.

The coordinates projected from the point clouds to the two-dimensional plane can be expressed by (1).

$$\begin{cases} x = -\frac{DX}{AX+BY+CZ} \\ y = -\frac{DY}{AX+BY+CZ} \\ z = -\frac{DZ}{AX+BY+CZ} \end{cases} \tag{1}$$

In general, choosing the projection plane perpendicular to the Z axis can avoid mistake. The normal vector can be expressed as $n = (A,B,C) = (0,0,-1)$, and then the coordinates after projection can be summarized into (2):

$$\begin{cases} x = \frac{DX}{Z} \\ y = \frac{DY}{Z} \\ z = D \end{cases} \tag{2}$$

The step for generating a depth image is shown in Fig. 2.

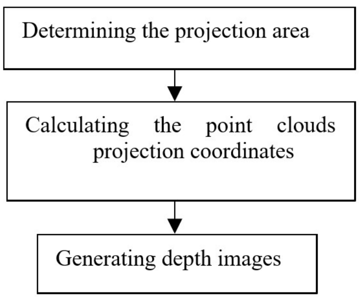


Figure 2. Depth images generation step diagram.

The specific steps are as follows:

- The determination of projection plane: we can Select the farthest point from the scanner as the point on the projection plane and let the normal vector of the projection plane be $n(a,b,c)$. Then the plane equation is:

$$aX+bY+cZ+d=0 \tag{3}$$

where a,b,c and d are the parameters of the plane equation and a,b and c cannot be 0 at the same time.

- Calculating the coordinates of point cloud after projection: We can project the point cloud data along the z-axis onto the plane in step (1) and record the projected point $(X,Y,Z)_i$ as $(x,y)_i$.
- Generating depth images: Because the number of laser points mapped to the image may be relatively small, the effective pixels of the processed depth image will be less. In order to solve this problem, this paper deals with the pixels. we can divide the point could into m×n small rectangular grids according to the scanning resolution of the point cloud. The average depth of the point cloud in the grid is taken as the gray value of the grid, which is finally normalized to [0~255]. The diagrammatic sketch is shown in Fig. 3.

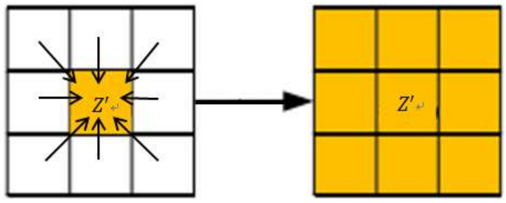


Figure 3. Schematic diagram of taking the average depth of point cloud in the grid.

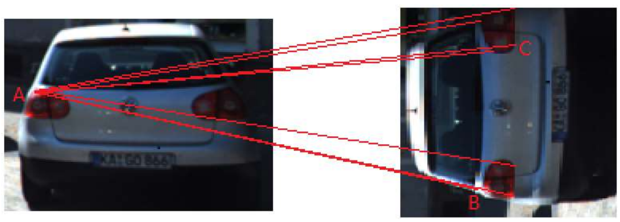
B. *Feature Points Matching between Depth Images and Optical Images*

The main work of this section is to obtain the mapping matrix of depth images and optical images. The mapping matrix can be calculated by matching feature points. And the feature points can be found accurately by Scale-Invariant Feature Transform (SIFT) [5] algorithm.

SIFT algorithm uses Euclidean distance of feature points to determine the matching degree. The shorter the distance, the higher the matching degree.

$$d(A,B) = d(B,A) = \sqrt{(\mathrm{u_1}^A - \mathrm{u_1}^B)^2 + (\mathrm{v_1}^A - \mathrm{v_1}^B)^2}$$

where d is the Euclidean distance between A and B. And $(\mathrm{u_1}^i, \mathrm{v_1}^i)$ is the 2-D coordinates of point i.



(1) (2)

Figure 4. Schematic diagram of feature point matching criteria.

As shown in the Fig. 4, AB and AC respectively represent the minimum and second smallest distances from point a on image (1) to all feature points on image (2). The feature point matching criterion is as follows:

$$\frac{d(A,B)}{d(A,C)} < t \tag{4}$$

That is, the ratio of the minimum distance to the second minimum distance is less than a certain threshold $t(t = 0.8)$. This method can find feature points efficiently and quickly.

After determining the feature points, we can calculate the mapping matrix. If $\mathrm{n}(\mathrm{n} \geq 8)$ and feature points $\mathrm{A}_i(\mathrm{x}, \mathrm{y}, 1)$ and $\mathrm{A}_i'(\mathrm{x}', \mathrm{y}', 1)$ $i = (1,2, \ldots \mathrm{n})$ are known, when n=8, the linear 8-point algorithm [6] can be used to solve the mapping relationship between images. When n>8, the least square method can be used to solve the mapping matrix.

Let F be the mapping matrix between 2-dimensional images, and the vectorization [7] of the coordinate of matching points and the fundamental matrix can be expressed as (5).

$$\begin{cases} U = [\mathrm{x}_i\mathrm{x}_i' \ \ \mathrm{y}_i\mathrm{y}_i' \ \ \mathrm{x}_i' \ \ \mathrm{x}_i'\mathrm{y}_i \ \ \mathrm{x}_i\mathrm{y}_i' \ \ \mathrm{y}_i' \ \ \mathrm{x}_i \ \ \mathrm{y}_i \ \ 1] \\ F = [F(1,1) \ \ F(1,2) \ \ F(1,3) \ \ F(2,1) \ \ F(2,2) \ \ F(2,3) \ \ F(3,1) \ \ F(3,2) \ \ F(3,3)] \end{cases} \tag{5}$$

It can be sorted into:

$$\sum_{i=1}^{n}(\mathrm{A}_i F \mathrm{A}_i'^T)^2 \ FU^T U \mathrm{F}^T \tag{6}$$

According to the polar geometric constraint theory [8], the F that minimizes the cost function C is the mapping matrix between 2-dimensional images. Usually, the cost function C [9] is:

$$C = \frac{F^T U U^T F}{\|F\|^2} \tag{7}$$

The cost function C has the minimum value when $\frac{\partial C}{\partial F} = 0$, that is:

$$\frac{\partial C}{\partial F} = 2\frac{UU^T - \mathrm{C}I_{\mathrm{m}\times\mathrm{n}}}{\|F\|^2} F{=}0 \tag{8}$$

According to the above formula, the singular vector corresponding to the minimum characteristic value of the UU^T is the mapping matrix F. Decomposition UU^T by SVD [10]. We can gain that:

$$[x, d, y^T] = SVD(UU^T) \tag{9}$$

In this way, the requirement to minimize the cost function C under the premise of $\|F\| = 1$ can be satisfied.

Since the rank of the basic matrix F is 2, and its degree of freedom is 7, and the res-ult of linear 8-point algorithm is 8 parameters, which will make some parameters not independent, and also don't satisfy the polar geo-metric constraint theory. Therefore, additional coercion is required.

Let the singular value of the fundamental matrix F be decomposed into $x_F d_F {y_F}^T$, where $d = diag(r, s, t), r \geq s \geq t$ is a diagonal matrix, and $r(d)$=2 regard as a mandatory constraint condition [11]:

$$\mathrm{F}' = \mathrm{x} \ \mathrm{diag}(r, s, 0) \ \mathrm{y}^T \tag{10}$$

F' is the approximate matrix of the fund-mental matrix F which makes $\|F - F'\|$ the smallest.

Although SIFT algorithm has good effect on feature description and matching in theory, SIFT does not retain object boundary information. The details and noises are smooth to the same degree at all scales, which affects the accuracy and uniqueness of localization.

In the case of the existence of false matching, we can use Random Sample Consensus (RANSAC) [12] to improve the accuracy of matching.

C. Point Clouds and Images Fusion

The core of laser-point clouds and optical images fusion is to accurately calculate the mapping matrix between the two corresponding pixel coordinates. The direct linear trans-formation method can take the coordinates of several known pixels and their corresponding 3-dimensional space coordinates as reference points [13]. In other words, using the feature points of depth images and optical images, the parameters of the collinear equation bet-ween the images plane coordinate of the optical images and the object space coordinate were solved by DLT [14]. And the mapping matrix between point clouds and images is determined by least square method.

The pixel coordinate system and the images coordinate system are not the same coordinate system. And the coordinate origin of the images coordinate system is the center of the images plane. The relationship between the images coordinate system and the pixel coordinate system is shown in Fig. 5.

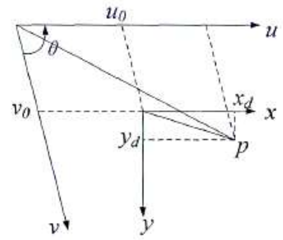


Figure 5. Relationship between images coordinate system and pixel coordinate system.

where *uv-o* is the images coordinate system and *xy-o* is the pixel coordinate system.

The relationship between pixel coordinates and images coordinates is:

$$\begin{cases} u = u_0 + \frac{x_d}{dx} \\ v = v_0 + \frac{y_d}{dy} \end{cases} \tag{11}$$

where *dx*, *dy* are the physical dimensions of pixels on the light screen; (x_d, y_d) is the position of point P in the pixel coordinate system; (u_0, v_0) is the position of the center point of pixel coordinates in the images coordinate system.

The mapping relationship between the images coordinate system and the object coordinate system can be expressed as:

$$\begin{bmatrix} u \\ v \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{12}$$

where *u*, *v* are the coordinates in the images coordinate system, R represents the 3×3 rotation matrix, and T is the

3×1 translation matrix. Equation (13) can be summarized as:

$$\begin{bmatrix} x \\ y \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{dx} & 0 & 0 & x_0 \\ 0 & \frac{1}{dy} & 0 & y_0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & & & \\ & f & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (13)$$

where x, y are the coordinates in the pixel coordinate system, and f is the focal length of the camera.

The pixel coordinate after the point clouds are projected to the images plane is not necessarily an integer, and the nearest pixel is taken as the projection images point coordinate. The RGB value of the corresponding optical images pixel after mapping is assigned to the point clouds. The laser- point clouds can obtain the real color information of the experimental target surface, and the fusion of laser- point clouds and optical images is completed.

III. Experimental Results and Analysis

A. *Evaluation*

In this paper, images from different perspectives in the same scene in the KITTI dataset are used as a group of test images. As shown in Fig. 6.

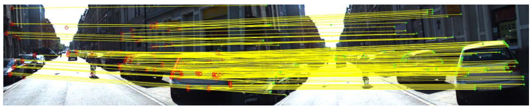


Figure 6. Inspection images.

The same process is used to obtain the fusion result of point clouds and images. The Euclidean distance is obtained for the 3-dimensional coordinates of the feature points in two optical images. The results are shown in Table I. The average error of the $\bar{\delta}$ as indicators of fusion.

$$\bar{\delta} = \frac{\sqrt{\Delta X^2 + \Delta Y^2 + \Delta Z^2}}{n} \quad (14)$$

TABLE I. Fusion Error

The percentage of points with error less than 0.2m	The percentage of points with error greater than 0.2m
95.6%	4.4%

The average error is $\bar{\delta}$=0.00953m. The accuracy is in centimeter level, and the fusion accuracy is high.

B. *Fusion Results*

Some point clouds coordinates and corresponding depth images coordinates and optical images coordinates are shown in Table II.

TABLE II. Some Point Clouds Coordinates, Corresponding Depth Images Coordinates and Optical Images Coordinates

number	X	Y	Z	u_1	v_1	u_2	v_2
1	-10.234	-3.525	0.227	498.41	200.69	475.02	244.05
2	-10.225	-3.392	0.02	468.63	200.72	475.87	246.32
3	-10.22	-3.175	0.815	483.58	200.8	476.33	245.64
4	-10.217	-3.116	2.366	395.35	256.23	379.69	291.87
5	-10.213	-3.033	2.012	206.29	295.45	194.27	319.21
6	-10.21	-3.007	1.007	379.67	296.23	357.02	325.76
7	-10.208	-3.012	0.29	470.37	302.41	443.06	333.33
8	-11.334	-3.212	2.309	211.62	305.01	199.24	327.77
9	-11.567	-3.377	0.25	475.93	310.93	448.09	340.18
10	-11.036	-3.772	2.378	206.75	315.97	193.01	339.05
11	-10.395	-3.774	2.343	207.3	335.13	194.56	355.16
12	-10.131	-3.263	2.248	193.58	335.23	181.64	355.92
13	-10.056	-3.322	2.376	204.95	335.61	192.22	359.64

In the table, X, Y, Z are laser-point clouds coordinates, u_1, v_1 are pixel coordinates of the depth images, u_2, v_2 are feature points of the depth images (u_1, v_1) in the opticalimage.

Compared with the point clouds data without color information, it can be seen that thecolor point clouds with RGB informationafter fusion is more consistent with humansensory experience. The results are shown in Fig. 7-Fig. 10.

Figure 7. Optical image.

Figure 8. Point clouds data.

Figure 9. Fusion results without RANSAC processing.

Figure 10. Fusion results of point clouds and optical images processed by RANSAC.

Since the contrast of the car on the right side of the optical images in the KITTI data is low, this paper uses color processing to increase the contrast of the car on the left side, so as to make the fusion effect more intuitive.

In the fusion results without RANSAC processing, the color of point clouds data that does not correspond to the color of the optical images can be found very obviously.

In the fusion results processed by RANSAC, the details such as vehicle contour, lights and Windows can be identified from the point clouds, and the detail information in the optical images can be accurately mapped to the point clouds data.

IV. Conclusion

After verification by KITTI data, it can be seen that the laser-point clouds and optical images fusion algorithm proposed in this paper can better realize the fusion of point clouds and images. And it can realize the automation of point clouds fusion and enhance the visibility of point clouds without manually setting control points.

However, this fusion method is lacking in the fusion effect of scenes with high reflective rate, and further research is needed to solve this problem in the future work.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Yuehai Wang and Si Li conducted the research; Si Li wrote the paper; Yuehai Wang and Na Xing revised the paper. all authors had approved the final version.

Acknowledgment

The authors wish to thank Zhipeng Zhuang and Xiaoyang Yu. This work was supported in part by a grant from Image and Perception Lab.

References

[1] S. Wang "Three-Dimensional reconstruction of antiques based on LIDAR and high-resolution images," *Engineering of Surveying and Mapping*, vol. 18, no. 6, pp. 53-55, 60, 2009.

[2] X. Zhang, "Post-Calibration fusion and obstacle ranging based on laser point clouds and images," Master's thesis, Dalian University of Technology, 2018.

[3] V. Girbés and L. Armesto, "An active safety system for low-speed bus braking assistance," *IEEE Transactions on Intelligent Transportation Systems*, no. 99, pp. 1-11, 2017.

[4] A. Geiger and P. Lenz, "Vision meets robotics: The KITTI dataset," *International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231-1237, 2013.

[5] F. Vasconcelos, "A minimal solution for the extrinsic calibration of a camera and a laser-rangefinder," *IEEE Transaction on Pattern Analysis & Machine Intelligence*, vol. 34, no. 11, pp. 2097-2107, 2012.

[6] A. Barla, F. Odone, and A. Verri, "Histogram intersection kernel for image classification," in *Proc. International Conference on Image Processing*, 2003.

[7] O. Cordon, S. Damas, and J. Santamar, "A fast and accurate approach for 3D image registration using the scatter search evolutionary algorithm," *Pattern Recogn. Lett.*, vol. 27, no. 11, 2006.

[8] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *Proc. Third International Conference on 3D Digital Imaging and Modeling*, 2001, pp. 145-152.

[9] N. Gelfand, L. Ikemoto, S. Rusinkiewicz, and M. Levoy, "Geometrically stable sampling for the ICP algorithm," in *Proc. Fourth International Conference on 3D Digital Imaging and Modeling*, Oct. 2003.

[10] R. B. Rusu, Z. C. Marton, N. Blodow, and M. Beetz, "Learning informative point classes for the acquisition of object model maps," in *Proc. the 10th International Conference on Control, Automation, Robotics and Vision*, Hanoi, Vietnam, December 17-20, 2008.

[11] L. Silva, O. R. P. Bellon, and K. L. Boyer, "Precision range image registration using a robust surface interpenetration measure and enhanced genetic algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 762-776, 2005.

[12] G. Sharp, S. Lee, and D. Wehe, "ICP registration using invariant features," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 90-102, 2002.

[13] A. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3D scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, May 1999.

[14] A. W. Fitzgibbon, "Robust registration of 2D and 3D point sets," in *Proc. the British Machine Vision Conference*, 2001.

Yuehai Wang was born in Jinan, Shandong Province in March 1975. In 1996 and 1999, he obtained bachelor's degree and master's degree respectively in the computer application department of Harbin University of technology. In November 2003, he obtained doctor's degree in engineering in the computer science and Technology Department of Harbin University of technology.

In December 2003, he taught in North China University of technology, and he is currently a professor and graduate tutor of School of information, North University of technology. His current research interests include fault diagnosis for PCB, pattern recognition, computing intelligence and robot soccer.

Till now, Dr. Wang has published about 40 papers including SCI and EI.