Proceedings of 2016 IEEE
International Conference on Mechatronics and Automation
August 7 - 10, Harbin, China

A Novel Segmentation Correction Method for Fusion of Vision and Laser

Yawei Yuan
Department of Automation
University of Science and Technology of China
Hefei, Anhui Province,230027,China
ywyuan@mail.ustc.edu.cn

Jianghai Zhao*, Shihui Fang and Qunshan Xu
Institution of Advanced Manufacturing Technology
Hefei Institution of Physical Science, CAS
Changzhou, Jiangsu Province 213164, China,
jhzhao@iamt.ac.cn

***Abstract* – In this paper, a method is proposed to solve the displacements and distortions, which are caused by inaccurate calibration in the low-level fusion. Compared with existing methods, the proposed method does not rely on any specified environmental feature and can be applied to a variety of scenarios. To implement it, twice clustering processes are applied to segment the input point cloud, and an iterative closest point (ICP) algorithm is used to iterate and correct the corresponding partitions. Furthermore, we also quantify an index to evaluate the result of correction and provide some simplified constraints to improve the measurement accuracy. Finally, the effectiveness of the proposed methods is verified by the result of 3D reconstruction.**

Index Terms – 2D laser scanner, stereo vision, low-level sensor fusion, clustering algorithm, distortion correction

I. INTRODUCTION

Excellent real-time performance, rich information, and reasonable price have helped cameras are widely used in environment perception. There are a great deal of algorithms based on color image can provide varies of descriptors to detect obstacles and recognize targets. But it is not a good idea to acquire the perception of depth information by using cameras. Take stereo vision as an example, it is strongly depend on the textures of surroundings and the accuracy also can't meet the requirements in many tasks. In contrast, as a high-precision measuring instrument, laser scanner can be used to compensate the weakness of camera. Consequently, fusing the data derived from both sensors can dramatically improve their performance.

Currently, the technology of vision and laser fusion have been hotly discussed in a series of research areas, such as intelligent transportation system, autonomous robotics, multiple objects detecting and tracking, and 3D environment reconstruction [1], [2], [3], [4]. Referred to the three levels of the multiple sensors data fusion [5], [6], the low-level fusion developed slowly. Currently, almost all of the implementations of synchronization between the vision and laser is by computing the extrinsic calibration. However, transforming the point cloud measured by laser to the cameras point of view may lead to a series of problems. For example, Ref [7] introduced the inaccurate transformation lead to the displacements and distortions. Ref [8] analyzed the occlusion phenomenon caused by different viewpoints. Previous literature deals mostly with the calibration between vision and laser. Both optimizing the selected features [9], [10] and modifying the existing calibration models [11] are all relied on the current environment. There is always a methodical error when we transplanted it to another environment. The phenomenon can be observed more easily under wide range scenarios because of the insufficient accuracy. This is why we have to correct the correlation during the fusion. Although, [7] developed a genetic Iterative Closest Point (ICP) algorithm to avoid the process of calibration, the filtering of misplaced points relies on the threshold of the clustering size. Additionally, the ignored noise also can lead to an uncertainty for the fusion result.

One contribution in this paper is that a robust method is proposed to correct the mismatching between stereo point cloud and the laser point cloud. Considering the different degrees of mismatching in different regions, we are not simply correct the entire point cloud. There is a clustering algorithm is applied to a few key data to motivate the next correction.

Another contribution is the handling of the occlusion problem. Unlike [8], it ignored the spot-size and reordered the 2D convex hulls and consecutively checked each point whether can pass the specific constraint. We proposed a simplified constraint based on the redundant depth information to filter out error measurement points. The constraint can avoid the feedback correction which may cost great computation.

Generally speaking, the target of the low-level fusion of vision and laser is to obtain a dense point cloud, which consists of accurate depth information and correct color information [4]. In this paper, we also present the whole fusion framework and complete it by adding some constraints and correction. It is worth to mentioned, our fusion is not simply merge two point clouds to the same space. There is an average processing is used to improve the final ranging accuracy.

The paper is organized as follows: section 2 introduces the raw sensor information acquisition and the whole fusion framework. There is some visualization constraints are proposed during the preprocessing. In section 3, the problems yielded during the fusion will be analyzed and handled. Finally, section 4 describes data fusion and the related experimental results. And the conclusions are given in the section 5.

II. DATA ACQUISITION AND FUSION FRAMEWORK

A. Data Acquirement

The experiment platform we used is Multisense-SL produced by Carnegie Robotics LLC Company. It consists of a 2D laser scanner and a stereo vision system. To make use of cheap 2D laser acquires 3D environment perception information, the laser is mounted on a speed adjusted motor. More details about their relative position and orientation are shown in Fig. 1.

In order to obtain unified fusion represent, we transform laser range data into the left camera optical frame (the system frame) by the extrinsic calibration matrix between vision and laser. Due to the rotation of laser scanner, the extrinsic calibration matrix changes with time. Here, all the calibration matrices are known. Compared with the high frame rate of camera, the laser takes a long time to collect a full revolution and acquired the expected large scene. In this paper, we just discuss the static scenes, and the data examples provided by both sensors are shown in Fig.2.

B. Fusion Framework

Refer to three levels fusion framework [5], [6] and considered the obvious gap between the two FOV shown in Fig.2, we add a visualization constraint into preprocess to reduce the amount of calculation in the next processes. A whole process framework is shown in Fig.3.

The visualization constraint is composed of two parts. The one is based on the physical location where the cameras can only catch the scenes before the system. The other is based on the resolution of camera image, because the correlation just appears in the points that can be remapped into the camera image. Let

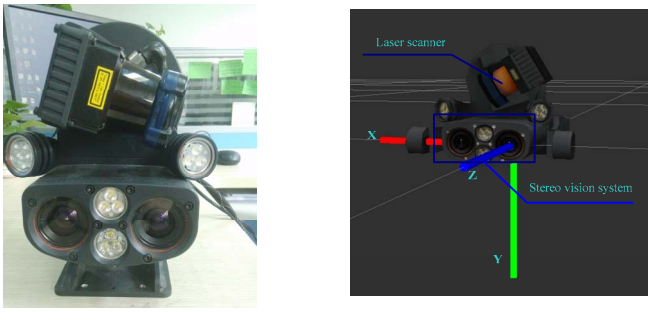
$$P = \{p_1, p_2, \ldots, p_n\} \tag{1}$$

be the set of all 3D points (have been transformed from laser frame to camera frame) generated by the laser scanner through the full revolution.

Then, the two constraints can be described as follows:

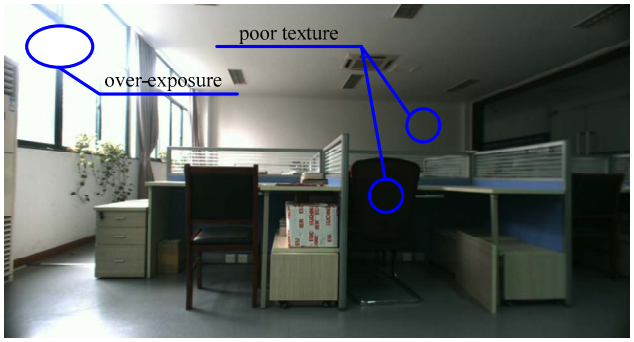
Constraint 1: $$P_{\geq 0} = \{p_i \in P \mid p_{i.x} \geq 0\} \tag{2}$$

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = P \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \tag{3}$$

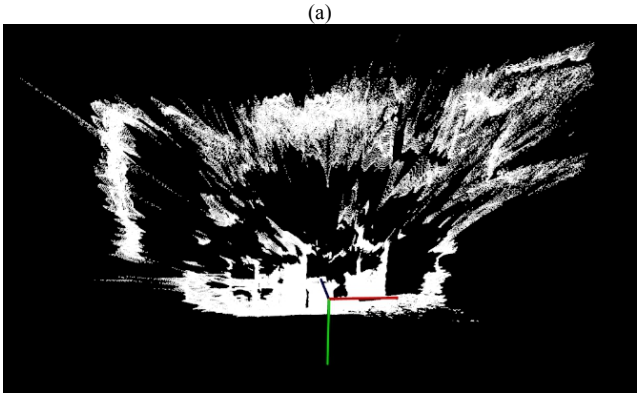


(a) (b)

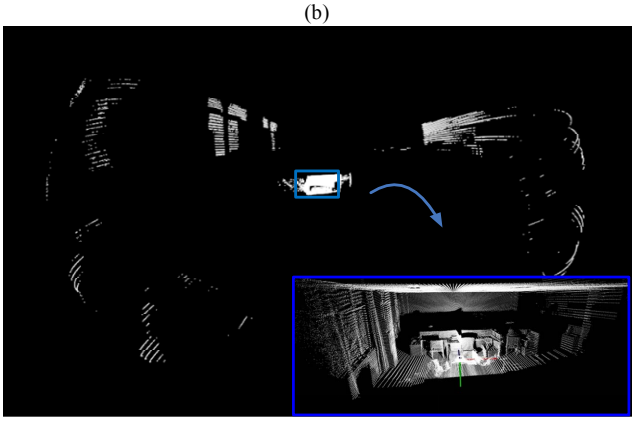
Fig. 1 (a) is the Multisense-SL and (b) is a model shown in Rviz based on Robot Operating System (ROS).



(a)



(b)



(c)

Fig. 2 Original data provided by camera and laser scanner under a scene of office. (a) shows the left camera color image at the resolution of 544×1024, and the corresponding stereo point cloud is (b). Laser point cloud acquired by rotation shows in (c), where the FOV is large to reach outside.

$$u = (f_x x + c_x z)/z \tag{4}$$

$$v = (f_y y + c_y z)/z \tag{5}$$

$$\text{Constraint 2: } \begin{cases} 0 \leq u < width\_of\_image \\ 0 \leq v < height\_of\_image \end{cases} \tag{6}$$

We iterate over the 3D points in $P_{\geq 0}$ and use the typical camera projection matrix (P) to project each point into the corresponding 2D image pixel coordinate (3)(4)(5). Simply remove the points out of the pixel range is reasonable (6). Where (u, v) is the coordinates in the image pixel coordinate system, (X_c, Y_c, Z_c) is the coordinates in the camera coordinate system.

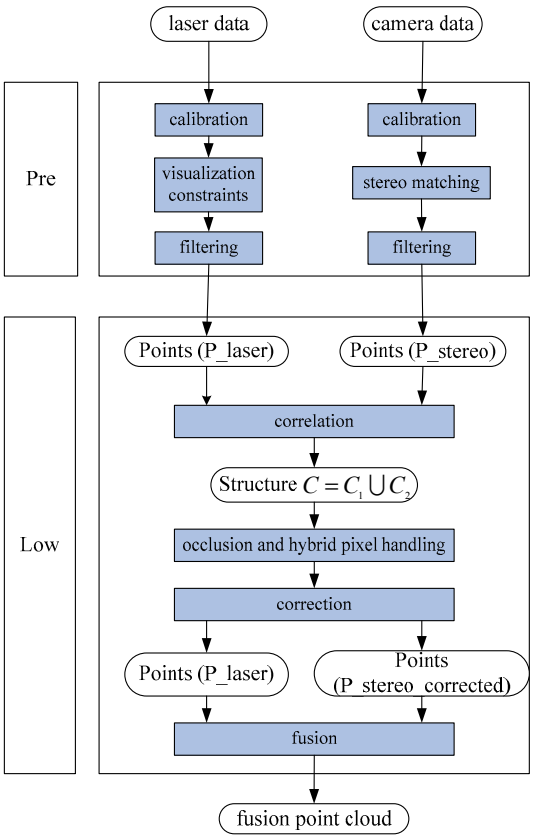


Fig. 3 The low-level fusion framework: more details about correction will be presented in the next section.

III. Proposed Solutions

A. Occlusion and hybrid pixel Handling

The ranging measurement principle of laser scanner is based on the time of flight and the energy of spot. Since the laser beam emitted by the laser scanner diverges in the shape of a cone, the spot produced when it cast onto the surface of object will become larger with increasing range. Once a spot cast onto the edge of object, the spot will be split into two parts. The one on the object, the other arrives to the rear of the object. This is the hybrid pixel phenomenon, and it can cause measurement errors around object borders [12].

As [8] analyzed, the occlusion phenomenon are also occurred on the borders of the object, though the principle is different. In our sensor system, the rotation of laser scanner and the difference of mounted locations between two sensors both can cause the different FOV, and then can cause the occlusion problem. Fig.4 describes the two situations, where the solid blue lines represent the laser beams and the dotted red lines represent the trace ray projected to camera image. Fig. 4(a) demonstrates the first situation, the 3D point①,②,③ are all projected to the same image pixel, but the corresponding pixel only represent the point ①. Fig. 4(b) shows the second situation, it is similar to the first situation that the point ② can't be seen in the camera image.

The result of correlation is stored in a customized image-like structure C, whose size is the same as the camera image. Because of different resolutions, not every cell has only a depth value if we remap the 3D measurement points

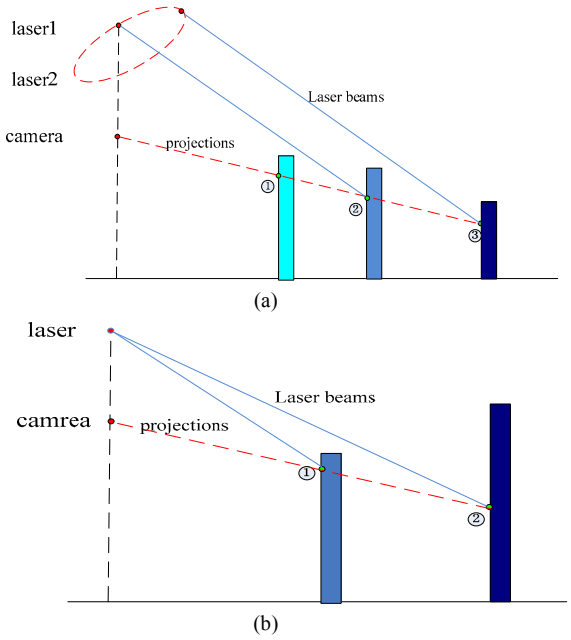


Fig. 4 The occlusion phenomenon: (a) is caused by the changing viewpoint of laser scanner. (b) is caused by the different mounted positions.

into 2D structure by camera projection matrix. According to the source of depth information of each cell, all the cells that have depth information in the structure (collected into set C_1) can be divided into three categories:

$$C_1 = C_{category\,1} \cup C_{category\,2} \cup C_{category\,3} \tag{7}$$

Where:

$C_{category\,1}$: depth values only come from the laser scanner.

$C_{category\,2}$: depth values only come from vision system.

$C_{category\,3}$: depth values are contributed by two sensors.

Moreover, since the hybrid pixel phenomenon and the occlusion problem both can remap the error measurement points and correct measurement points into the same cell. We can simply use a constraint based on depth information to remove error measurement points. The method is independent on any particular feature, and suitable for the $C_{category\,1}$ and $C_{category\,3}$. Let

$$C_{category\,1} \cup C_{category\,3} = \{c_1, c_2, \cdots, c_n\} \tag{8}$$

If the number of depth values come from laser scanner for cell c_i $(i = 1,2,\cdots,n)$ is k, and recorded those 3D depth values as $x_1, x_2, \cdots, x_k$. Providing a threshold value (ε), and we can define:

$$u = x_{k_1} - x_{k_2} \,(1 \le k_1 \ne k_2 < k) \tag{9}$$

$$sign(u) = \begin{cases} 0, |u_d| < \varepsilon \,(d = 1,2,3) \\ 1, else \end{cases} \tag{10}$$

$$f(c_i) = \sum_{1 \le k_1 \ne k_2 < k} sign(x_{k_1} - x_{k_2}) \begin{cases} = 0, pass\ the\ constriant \\ > 0, error\ measurement \end{cases}$$

$$(i = 1,2,\cdots,n) \tag{11}$$

In other words, as for a cell, if all the depth values come from laser scanner can be enveloped by a sphere with a diameter of ε, we say the cell is pass the constraint. Otherwise, we assert the depth value of the cell is vain.

B. Distortion Correction

For stereo vision, the depth information are computed by the disparities, and we can use (12) analyze the resolution. The z is the measured distance, f is the focal length, d is the disparity, Δd is the resolution of disparity, and Δz is the resolution of measurement.

$$\Delta z = (z^2 \Delta d) / (fT) \tag{12}$$

It is obvious that the resolution decreases significantly with increasing range when other parameters are fixed. Take the hardware we used as an example, if set the resolution to 3cm, the laser scanner (Hokuyo UTM-30LX-EW) can reach to 30m, while stereo vision can only reach to 4.5m (here $f = 580, T = 0.07, \Delta d = 0.0625$). Additionally, insufficient texture and poor illumination are (have marked example regions in Fig2. (a)) also result in poor matching between two point clouds. Fig.5 describes the mismatching. The gap between two point clouds is obvious, and the noise around the windows is terrible.

Our correction is based on the segmentation for input point cloud and then iterate over each of the corresponding clustering to individually adjust by ICP algorithm. Here, a novel segmentation method is proposed. Its implementation is through twice clustering processes based on calibration matching degree (η_i) and the neighbor relationship. Let

$$C_{category\,3} = \{c_1, c_2, \cdots, c_n\} \tag{13}$$

for cell c_i $(i = 1,2,\cdots n)$, k_1 depth values come from stereo vision are S_{ir} $(r = 1,2,\cdots,k_1)$, and k_2 depth values come from laser scanner are L_{ir} $(r = 1,2,\cdots k_2)$. We can define:

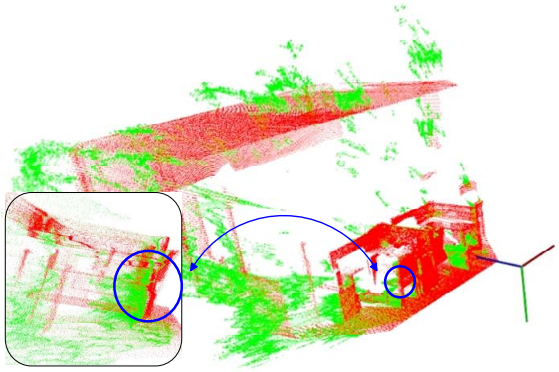


Fig.5 The terrible matching between two sensors: where red points come from the laser scanner and green points come from the stereo vision. More details are shown in blue circular window.

$$S_{avg}(i) = \frac{1}{k_1} \sum_{r=1}^{k1} S_{ir} \tag{14}$$

$$L_{avg}(i) = \frac{1}{k_2} \sum_{r=1}^{k_2} L_{ir} \tag{15}$$

$$\eta_i = S_{avg}(i) - L_{avg}(i) \tag{16}$$

$$\eta = \frac{1}{n} \sum_{i=1}^{n} |\eta_i| \tag{17}$$

Named η_i the calibration matching degree of cell c_i, η the average calibration matching degree. Here, all above varies has three dimensions: x, y, z, and the unit is meter.

As it shown in Fig. 6, we make use of a few key cells $(c_1, c_2, \cdots c_n)$ to acquire a set of η_i as the input of a standard Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm. The result can helped us to determine the final number of clusters and original cluster center elements for Re-clustering. Similar to K-mean algorithm, the customized Re-clustering algorithm search the nearest center element for each non-center element and add it into the cluster that the center element belongs to. Then the non-center element becomes a new center element of chosen cluster. The difference is that determined center elements do not need to optimize clustering by large number of iterations and compute cluster centers.

We offer a small neighborhood threshold (K) for search at the beginning of the Re-clustering and iterate non-center elements to expand the center elements until the algorithm stop convergence. Considering the laser point cloud is sparse, another variable (σ) is determined as the ratio of the number of suspended non-center elements to the total number of input elements. Once stopping convergence, the algorithm will automatically increase K until σ less than the threshold we provided. Finally, all remaining suspended non-center elements are marked as noise.

Additionally, before the correction is executed, an extra calibration matching degree constraint is added to handle the stereo noise caused by illumination. In other words, we mark those points whose calibration matching degree above the set threshold as noise directly.

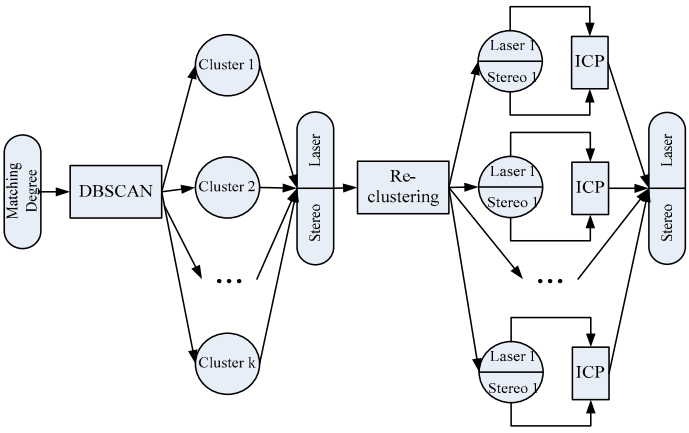


Fig.6 A model of the novel clustering method

IV. Fusion and Experiment

Completed the correction, we conduct the fusion process. Merging the camera color image, the stereo point cloud and the laser point cloud to unified colored point cloud. The result will have the color of the corresponding pixel and the average depth information stored in the corresponding cell.

A. Subjective Evaluation

Taking the office scene as an example, original input data have been shown in Fig. 2. If we fuse the data without constraints and correction, a color point cloud we acquired in Fig.7. There are some obvious noises near the walls and windows, while most of the points have correct color and seemingly correct positions. But an enlarged graph shows the obvious gap between two point clouds (shown in Fig.5).

The result of DBSCAN algorithm is shown in Fig.8, and the result of Re-clustering is shown in Fig.9, where the noise cluster has been removed in the stereo point cloud. To improve the efficiency, we conduct a down-sampling before the first clustering, and conduct an up-sampling for laser point cloud to obtain dense data from sparse point cloud before the final fusion. Fig. 10 shows the resulting fusion point cloud, and Fig. 11 demonstrates the gap between two point clouds after correction.it is almost invisible.

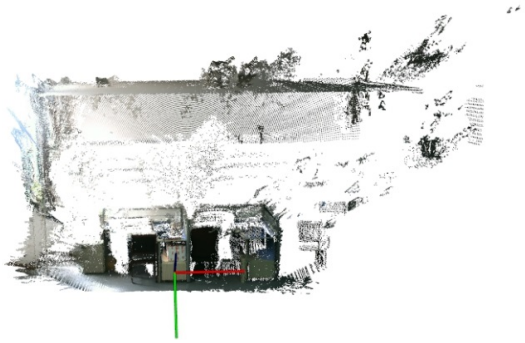


Fig. 7 Fusion point cloud without constraints and correction

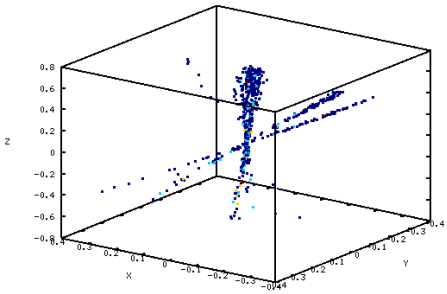


Fig. 8 The result of DBSCAN algorithm, where the different color represent different clusters. Related parameters as follows: the minimum number of points (minPts) is 10, and the neighborhood radius (esp) is 0.01m.

(a) (b)

Fig. 9 The result of Re-clustering algorithm: (a) comes from stereo and (b) comes from laser, where different color represent different clusters, and the corresponding clusters in both figures have the same color.

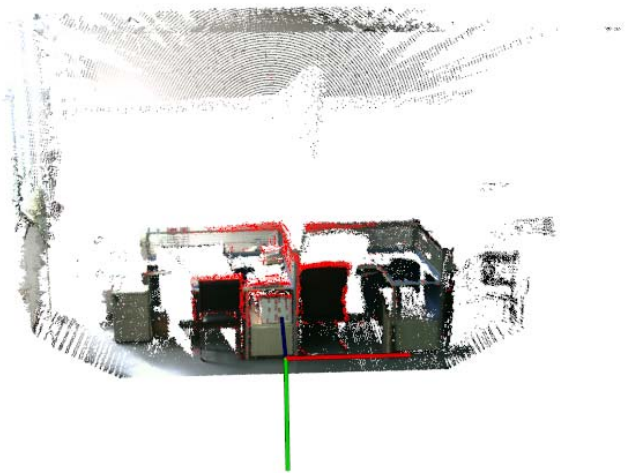


Fig. 10 The Resulting fusion point cloud after constraints and correction, where red points are error measurement judged by occlusion and hybrid pixel handling. To be more clearly, red points are magnified.

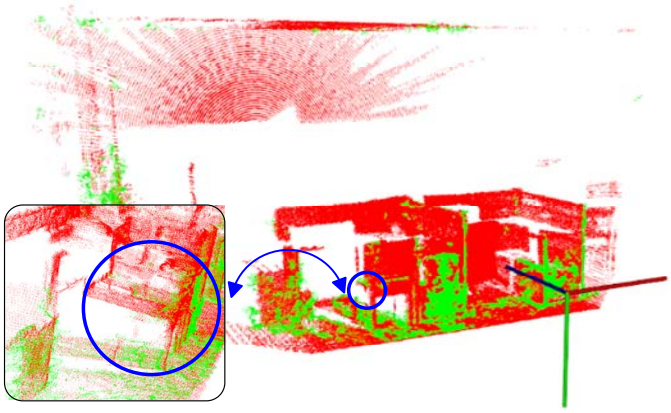


Fig.11 The matching between two sensors: where red points come from the laser scanner and green points come from the stereo vision. More details are shown in blue circular window.

B. Objective Evaluation

Related statistics are listed as follows. TABLE I shows the average calibration matching degree (η), where relative constraints including occlusion handling, hybrid pixel handling, and calibration matching degree constraint. The smaller the value, the better the correlation. Obviously, constraints and correction both can improve the matching performance, but the correction can reduce a magnitude.

TABLE II shows the ratio of the number of positive values to the total number for all η_i . It can imply the distribution of mismatching. An obvious deviation (0.855071) in the z-axis direction can be concerned. We discovered the phenomenon is present in large number of experiments under different real scenes. It is definitely not accidental error. While the constraints can reduce the matching degree but can't avoid the problem (0.871132), the correction performs well in both aspects. More clearly contract comparison is shown in Fig. 12, where we can see the significant improvements in the calibration matching degree. After the correction, almost all of the values are limited to the vibration near 0.

TABLE I
COMPARISON OF AVERAGE CALIBRATION MATCHING DEGREE

	η		
	x (m)	y (m)	z (m)
Traditional method	0.178172	0.0873616	0.622987
Relative constraints	0.0580887	0.0368648	0.160188
Correction process	0.0119064	0.00941743	0.0385223

TABLE II
COMPARISON OF POSITIVE RATIO

	positive ratio		
	x	y	z
Traditional method	0.419087	0.642899	0.855071
Relative constraints	0.407737	0.756871	0.871132
Correction process	0.481455	0.489673	0.490344

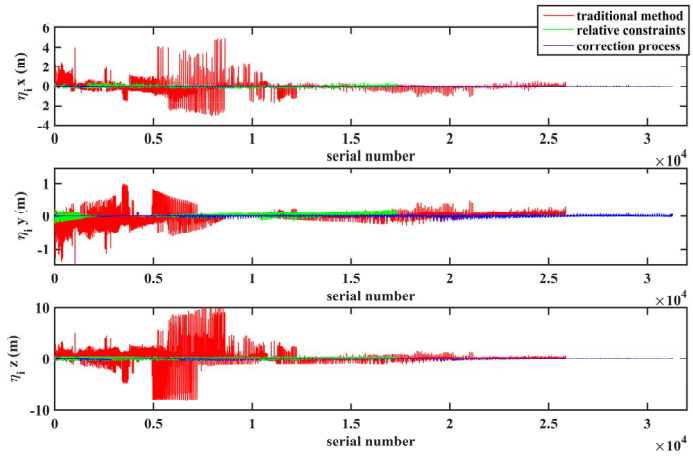


Fig. 12 The comparison of calibration matching degree: from top to down respectively is x, y, z dimension.

C. *Universality*

An extra experiment added to prove that the correction does not affect the well-matched point clouds fusion. Fig.13 shows the resulting fusion point cloud.

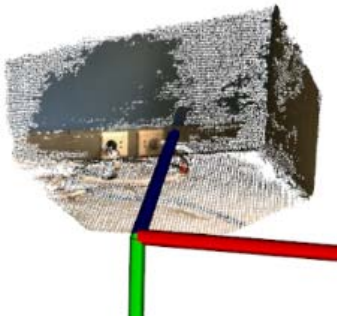


Fig.13 the final fusion result under a good experiment environment

V. CONCLUSION

In this paper, we have shown a complete low-level sensor fusion process, and provided a novel solution to correct the distortion. A novel segmentation method for fusion point cloud based on calibration matching degree is proposed and it does not rely on any specific environment feature. Meanwhile, some constraints for dealing with occlusion problem, hybrid pixel phenomenon and stereo noise have also been presented in the paper. We got a nice fusion results when applying these methods to fuse laser data and stereo data under different environments.

Considering the time of rotating laser scanner to obtain data, we are only able to test the correction method under a static environment. But, in theory, it also can be applied to dynamic scenes if we can optimize the overall runtime and data acquisition process. However, it is still a challenging work.

Our work presented in the paper is just a start for many applications. We can develop it to realize object classification and recognition, or higher levels sensor fusion to perform complex tasks.

ACKNOWLEDGMENT

This research has been supported by National Science and Technology Support Program (No. 2015BAK06B02) and the project of Science and Technology Support Plan of Jiangsu province (No. BE2013003).

REFERENCES

[1] B. Leibe, K. Schindler, N. Cornelis, L. G. Gool, "Coupled object detection and tracking from static cameras and moving vehicles," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 10, pp. 1683-1698, October 2008.

[2] J. A. Castellanos, J. Neira, J. D. Tardós, "Multisensor fusion for simultaneous localization and map building," IEEE Transactions on Robotics and Automation, vol. 17, no. 6, pp. 908-914, December 2001.

[3] X. Song, J. Cui, H. Zhao, H. Zha, "Bayesian fusion of laser and vision for multiple people detection and tracking," SICE Annual Conference, The University Electro-Communications, Japan, pp. 3014-3019, August 20-22, 2008.

[4] J. H. Joung, K. H. An, J. W. Kang, M. J. Chung, W. Yu, "3D environment reconstruction using modified color ICP algorithm by fusion of a camera and a 3D laser range finder," IEEE/RSJ International Conference on Intelligent Robots and Systems, Louis, USA, pp. 3082-3088, October 11-15, 2009.

[5] D. L. Hall, J. Llinas, "An introduction to multisensor data fusion," Proceedings of the IEEE, vol. 85, no. 1, pp. 2220-2225, January 1997.

[6] M. Haberjahn, K. Kozempel, "Multi level fusion of competitive sensors for automotive environment perception," 16th International Conference on Information Fusion, Istanbul, Turkey, pp. 397-403, July 9-12, 2013.

[7] Q. Muhlbauer, K. Kuhnlenz, M. Buss, "Fusing laser and vision data with a genetic ICP algorithm," 10th International Conference on Control, Automation, Robotics and Vision, Hanoi, Vietnam, pp. 2008. 1844-1849, December 17–20, 2008.

[8] S. Schneider, M. Himmelsbach, T. Luettel, H. J. Wuensche, "Fusing vision and lidar-synchronization, correction and occlusion reasoning," 2010 IEEE Intelligent Vehicles Symposium, University of California, San Diego, CA, USA, pp. 388-393, June 21-24, 2010.

[9] C. Zhao, "Laser and Vision Data Fusion for 3D Scene Reconstruction and Surveillance," Dalian University of Technology, 2011.(in Chinese)

[10]G. Li, Y, Liu, L. Dong, X. Cai, D. Dong, "An algorithm for extrinsic parameters calibration of a camera and a laser range finder using line features," Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA, USA, Oct 29 - Nov 2, 2007.

[11]C. Mei, P. Rives, "Calibration between a central catadioptric camera and a laser range finder for robotic applications," Proceedings 2006 IEEE International Conference on Robotics and Automation, Orlando, Florida, pp. 532-537, May, 2006.

[12]F. Zheng, "A Research on Vision based Laser Measurement System Calibration Method," Graduate School of National University of Defense Technology, 2005. (in Chinese)