Extrinsic Calibration of Lidar and Camera with Polygon

Qinghai Liao[1], Zhenyong Chen, Yang Liu, Zhe Wang and Ming Liu[1]

***Abstract*— Fusion of heterogeneous exteroceptive sensors is the most efficient and effective path to the representation of the environment precisely, as it can compromise various drawbacks of each homogeneous sensor. The rigid transformation (aka. extrinsic parameters) of heterogeneous sensory systems is the prerequisite of fusing the multi-sensor information. Researchers have proposed several approaches to estimate the extrinsic parameters. However, these approaches neither rely on human interventions or specifically designed auxiliary object or do not provide the library which makes it hard to test or benchmark. In this paper, we propose a novel extrinsic calibration approach for the extrinsic calibration of a Lidar (Laser Range Finder) and a camera which only based on a polygon board and we offer the relevant tools.**

In this paper, we firstly track and extract the target polygon from both the image and point-cloud. Then we try to match the polygon between the 2D and 3D feature spaces. With the associated polygon, we are able to get multiple constraints to optimize the extrinsic parameters.

At the end, we validate our approach by four configurations, including the simulation, 16/32-beam Lidar and 100-line MEMS-Lidar. The outcome indicates high-precision extrinsic calibration performance.

I. Introduction

A. Background

In recent years, 3D sensing system has aroused increasing attention due to its widely potential applications, such as autonomous driving and mobile robotics. These tasks have high demands on the environment perception and environment modeling. However, it is either imprecise or requires high computation power to model the environment with homogeneous sensing modalities like cameras. Cameras are so far the most widely used sensors which offer a wealth of information on bearing information and color, but it is heavily affected by the light condition and may not be able to work in many scenarios. People aware that vision sensor's flaw has indeed caused some problems such as autonomous driving accidents [1]. One of the important solution is to combine various sensing modalities to enhance the model. The common practice is to utilize the 3D range finder and camera together to make complementary. To make use of the information derived from heterogeneous modalities, we have to fuse the information and present them in a single reference frame. The goal is to compute the rigid transformation including the relative rotation and translation of different sensor coordinate systems. The 6-DoF rigid transformation is called *extrinsic parameters* and the process to estimate the extrinsic parameters is called *extrinsic calibration*.

[1]Qinghai Liao and Ming Liu are with Department of Electronics and Computer Engineering Engineering, Hong Kong University of Science and Technology.{qhliao,eelium}@ust.hk

In this paper, we propose a generic method for the extrinsic calibration between camera and Lidar, without specific calibration object such as chessboard or prisms.

B. Related works

It is a common agreement that the fundamental theoretical problems have been solved for the extrinsic calibration, along with the development of theories in multi-view geometry [2] and optimization [3]. However, the integration of algorithms and automated calibration in practical cases is still a challenge for us. In these cases, the dependence on facilities or human interventions should be minimized. By this work, we would like to tackle several last-mile problems to realize automatic *feature association* and *robust regression*, leading to high-precision extrinsic calibration over heterogeneous sensors.

Researchers have proposed many methods for extrinsic calibration of 3D Lidar and camera . We classify these methods into three categories.

1) Feature Based: These method first try to find the correspondences from image pixels captured by cameras to the point-cloud captured by Lidar. If we have the correspondences, this problem will be further solved as a PnP problem [4], optimization problem [3], even as an active calibration perception which attempted to find the next-best-view [5], which have been already well-studied. However, images and point-cloud are hard to match due to the inherent representation difference: The images captured from cameras are *dense* representations, for which each pixel has a proper definition. Point-clouds are *sparse* representation. Regardless the density, the space between any two observed points has no definition at all. Thereby, currently there are no such generic feature description across heterogeneous sensors that we can directly use to match the features from images with that from point-clouds.

Therefore, to get the feature correspondence, researchers utilize external artificial calibration objects which could be observed simultaneously from the camera and Lidar. Usually more than one views from different poses are necessary to perform the extrinsic calibration. The most often used calibration object is the chessboard originally for the camera intrinsic calibration [6]. Zhang [7] first put forward the extrinsic calibration approach which used the chessboard to calibrate a 2D Lidar and a camera. Unnikrishnan *et al.* [8] extended Zhang's approach for 3D Lidar with a chessboard, whose approach was the first published method for a 3D Lidar and a camera. Pandey *et al.* [9] extended the above approach for omnidirectional cameras. Rodriguez *et al.* [10] presented an extrinsic and intrinsic calibration approach with

circle based calibration target. Besides, Gong*et al.* [11] used an arbitrary trihedron to assist the calibration and Dhall*et al.* [12] used the aruco QR code and rectangular board to perform calibration. In this paper, we require certain help for feature extraction, but the only calibration object is simply with arbitrary polygonal shape.

Another way to find the feature correspondence is manual selection but provide primary filtering to narrow down the candidates. They do not require artificial calibration target any more but use the features from the natural scene. Scaramuzza *et al.* [13] first tried this method by converting the visually ambiguous 3D range information into a 2D map where natural features of a scene are highlighted. Moghadam *et al.* [14] selected all 3D lines and 2D lines as candidates for human selection which made the algorithm more robust and precise since it greatly reduces the probability of human errors. These methods require massive human attention for feature association.

2) Motion-Based Methods: Motion-based methods take the extrinsic calibration problem as a well-researched hand-eye calibration problem [15], [16] which could be referred as: to recover the rigid-transformation between the two rigidly connected points with series of transformations. Mathematically, to compute an unknown but fixed X from series of $A_iX = XB_i$ where A_i, B_i are the transformation of the points(sensors) and X is the transformation between points. The solution to this problem has been developed for more than 20 years and lots of improved approaches have been proposed. Z.Taylor *et al.* [17], [18] used this method for multimodal sensors extrinsic calibration. We previously utilized this method for extrinsic parameter initialization for automatic Lidar-Camera calibration [19].

Motion-based method does not require any calibration object or rely on any specific sensor measurement infomation which makes it more general than other approaches. But this technique has several disadvantages: 1) the movement of the sensor system is restricted, e.g. the X can not be solved if the system do plane motion. 2) the accuracy is subject to motion estimation of every single sensor 3) higher computational cost and 4) not applicable to static installed sensor system.

3) Mutual Information Based: There is another group of methods, for which the calibration parameters are estimated by maximizing the mutual information obtained between the sensor measurements. Pandey *et al.* [20] and Z.Taylor *et al.* [21] proposed to automatically calibrate the extrinsic parameter using such a method in 2012 separately. Pandey used the measured surface intensity and Taylor used the surface normal. However, the intensity obtained from the Lidar was not very reliable.

Our proposed approach utilizes the polygon board to assist the feature association and optimize the calibration parameter by the feature specific constraints. The most related to our work is Dhall*et al.* [12] who also used a polygon board to partially help associate features. However, their method mainly relied on the QR code and utilize different features.

C. Contributions

In this paper, we propose a novel automated extrinsic calibration algorithm for a 3D Lidar and a camera with arbitrary polygon board. Our proposed algorithm tracks and extracts the target polygon and then optimize the extrinsic parameter. Our approach is also able to perform calibration with single shot. Besides that an open-ware library and GUI tool including code and test scenarios data is to be released along with the publication of the paper which could be directly used to test or benchmark.

D. Organization

The rest of the paper is organized as follows. Section II introduces the methodology of our proposed algorithm. Section III shows the calibration experiments and the simple application. Finally, the conclusion based on our test is discussed in Section V.

II. Methodology

Fig. 1 shows the block diagram of the proposed algorithm which includes sensor data synchronization, polygons extraction and match and optimization.

A. Data Pre-Processing

Sensor data pre-processing includes synchronization and image rectification. The synchronization of point-cloud and image is based on their time-stamp. Out of this concern, we use the Lidar UDP packet time-stamp instead of the point-cloud time-stamp as reference. As for other pre-processes, on one hand, the RS-Lidar we used can operate at maximally 10 frame-per-second (fps) and every frame (point-cloud) is constituted with more than 850 UDP packets which contains two measurements. Thus, the frame rate of Lidar UDP packet is higher than 8500fps which greatly help to reduce the temporal misalignment. On the other hand, the resolution of the camera is 1920x1080 and the field of view (FOV) is 120°, working at 120 fps. We adopted the OpenCV camera model to calibrate the intrinsic parameters and undistort the image. The synchronization was guaranteed via hardware, which we cannot introduce in too much detail considering the scope of the paper and limited space.

B. Polygon Extraction & Match

Our proposed approach adopted an arbitrary polygon object as calibration auxiliary which is very handy to acquire. In our implementation, we use a rectangular board as calibration object which share exactly the same constraints with other polygon and does not affect the generality of our method. The polygon brings three types of constraints:

1) *vertex constraint* is the most intuitive constraint, as the problem to PnP problem which is already well-studied and could be solved. However, vertexes are not always available or could be precisely estimated since the natural characteristic of Lidar that it has different resolution in vertical and circumferential direction. The extreme case is shown in Fig. 2a that all four vertexes

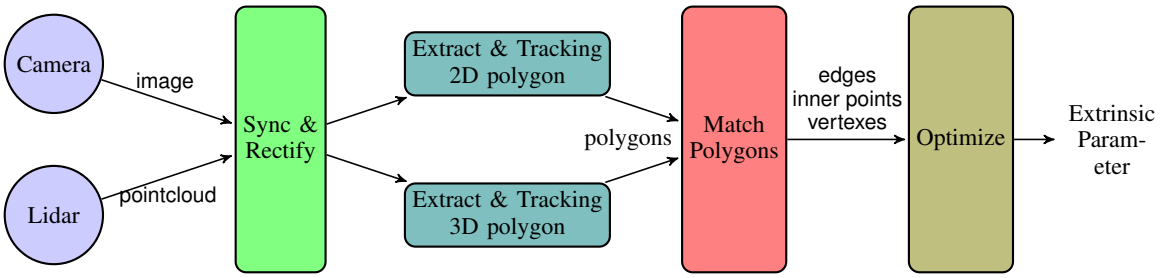


Fig. 1: Block diagram of the proposed extrinsic calibration algorithm

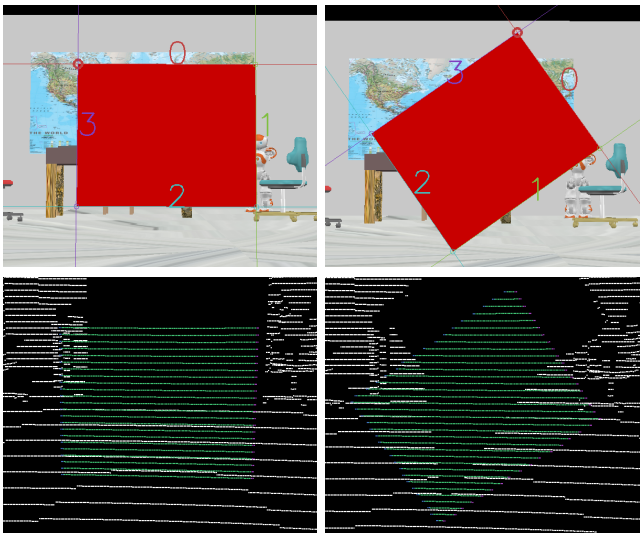
of the polygon in point-cloud could not be precisely located.

2) *edge constraint* is another good and strong constraint. It is also as line features that have been used in [14], [19]. Similarly, part of edges of the polygon may be occluded, invalid or comprise large error in some cases, such as Fig. 2a. Thus, the problem is under constrainted and the 6-DoF extrinsic parameters are unbounded and have infinite solution. Fig. 2b shows the well constrainted case in which all edges and vertexes could be extracted. 6-DoF extrinsic parameters could be precisely recovered from this single frame.
3) *inner-point constraint* is a weak constraint. Calibration cannot be solved by taking only this constraint but mass points narrow down the solution domain and guarantees the reliability of the final solution.

To get reliable and accurate extrinsic parameters, we use both edge and inner-point constraints and to prevent the failure caused by invalid edge constraints in degenerate case our approach utilize a sequence of frames even it will well converge with proper input, such as Fig. 2b.

As shown in Fig. 2, we extract the polygon from the image and point-cloud separately, and then match the edges and vertexes. For the i-th image, we firstly extract lines by EDLines [22] and get the edges E_i^j by filtering the lines with polygon geometry constraints and the vertexes V_i^j, which are the intersections of edges. For the i-th pointcloud, we firstly extract the polygon board by plane model and RANSAC method, and then split the points into several rings (horizontal lines) whose end points must located on the edges. The points of the polygon are classified into two categories:inner points P_i and edge points Q_i^k. Where $j \in \{0,1,2,3\}$ is the index of the edge and vertex of 2D polygon and k is the index of the 3D polygon edge. However, we have to keep in mind that not all 3D edges are always available. To improve the reliability and efficiency of extraction algorithm, we also keep tracking the polygon across the sequence of images and point-clouds.

After extracting the polygon from the image and point-cloud, we match the 2D edges and 3D edges. We simplify this problem to a finding index offset problem by sorting the edges in an clockwise direction. For ith frame, there is an constant *offset* between j and k. We utilize two longest(most points) 3D edges, Q_i^a and Q_i^b, which should always be valid



(a) under-constraint pose (b) well-constraint pose

Fig. 2: Extracted 2D polygon(1st row) and 3D polygon(2nd row) from V-REP simulation data

and find the $offset$ by:

$$offset = \arg\min_n \{f(Q_i^a, E_i^{a+n}) + f(Q_i^b, E_i^{b+n})\} \quad (1)$$

$$f(Q,E) = \sum_{p_i \in Q} g(E, K(R'p_i + t') \quad (2)$$

where $n \in \{0,1,2,3\}$, $f(Q,E)$ is the 3D-2D line error function, $g(\cdot)$ is the 2D point-line distance function, R' and t' are the rotation and translation of the coarse initial extrinsic parameter which could be represented as $T' = \begin{bmatrix} R' & t \\ \mathbf{0} & 1 \end{bmatrix}$.

C. Optimization

To precisely calibrate the extrinsic parameter, We aim to minimize a global penalty function, which is a sum of the re-projection point-to-line error and point-inside-polygon error. As discussed before, for every frame we have the matched edges(Q,E), polygon vertexes(V) and inner 3D points(P). A generic form of the penalty function is:

$$\text{minimize} \quad \sum_i \phi(r_i) \quad (3)$$

$$\text{subject to} \quad r_i = \sum_k \underbrace{f(Q_i^k, E_i^{k+offset_i})}_{edge\ constraint} + \underbrace{w_i h(P_i, V_i)}_{inner\ point\ constraint} \quad (4)$$

where $f(\cdot)$ is the 3D-2D line error function defined at Eq. 2, w_i is scalar weight and $h(\cdot)$ is the points-inside-polygon error function defined by the Algorithm.1. $\phi(\cdot) : R \rightarrow R$ is a convex loss function. Note that $\phi(\cdot)$ is often taking a quadratic form, i.e. $\phi(u) = u^2$, namely Ordinary Linear Square (OLS) as used by most implementations. However, such a quadratic form will introduce high sensitivity to outliers. In this work, we utilize the Soft-L1 loss function, as:

$$\phi(u) = 2(\sqrt{1+u^2} - 1) \tag{5}$$

With such a loss function, the effect of the outliers drops from quadratic to linear form. It leads to enhanced performance in extrinsic parameters calibration which is validated in our experiments.

Algorithm 1 Point-inside-polygon error function $h(\cdot)$

Input: P,V
Output: error
Initialization : error = 0
$V_c \leftarrow$ center of V
for for 3d point p_i in point set P **do**
  $p' \leftarrow K(R'p_i + t')$ {project 3d point to image}
  if (p_i not inside the polygon formed by V) **then**
    $error \mathrel{+}= distance(p_i, V_c)$
  end if
end for
return $error$

III. Experiment

Our proposed algorithm has been validated by both V-REP [23] simulation environment and three type Robosense[1] Lidar-Camera devices. The experiment process and results show that our approach is flexible, accurate and robust.

A. V-REP Simulation

V-REP is an popular robot simulator which is used for fast algorithm development, fast prototyping and verification. It provides various out-of-box sensors and models, and whats more important are the precise ground truth and eliminating temporal synchronization problem. The ground truth of extrinsic nor intrinsic parameter is impossible to measure in real environment and temporal synchronization is hard to solve. V-REP simulation makes us focus on the extrinsic parameter calibration without being affected by the intrinsic parameter and synchronization problem.

The Fig. 3 shows the sensor configuration and experiment scene. We use the 64 lines Lidar which simulate the Velodyne HDL-64E and camera whose resolution is 1280x720 and perspective angle is 45°. The rectangular board for calibration is 1.0x0.8 meter. After calibration, we use the result to project and back-project one frame to indicate the calibration precision which are shown in Fig. 4. Note that part of the wall behind the board has the board color (red) which is not caused by the extrinsic parameter error but the parallax caused by the misalignment of the camera frame and the Lidar frame.

[1] http://robosense.ai/

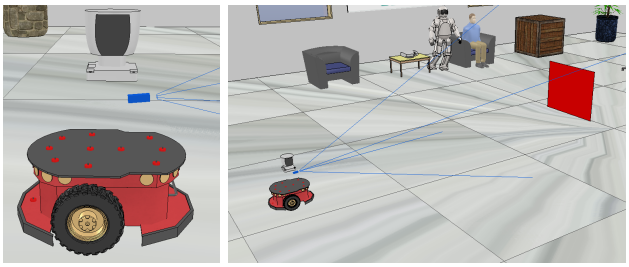


Fig. 3: Simulation with V-REP, left:sensor configuration, right: simulation scene

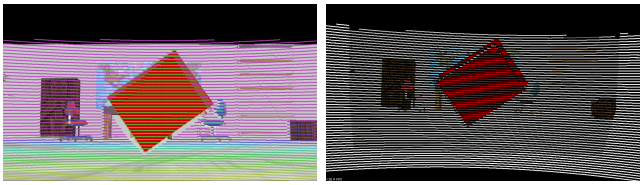


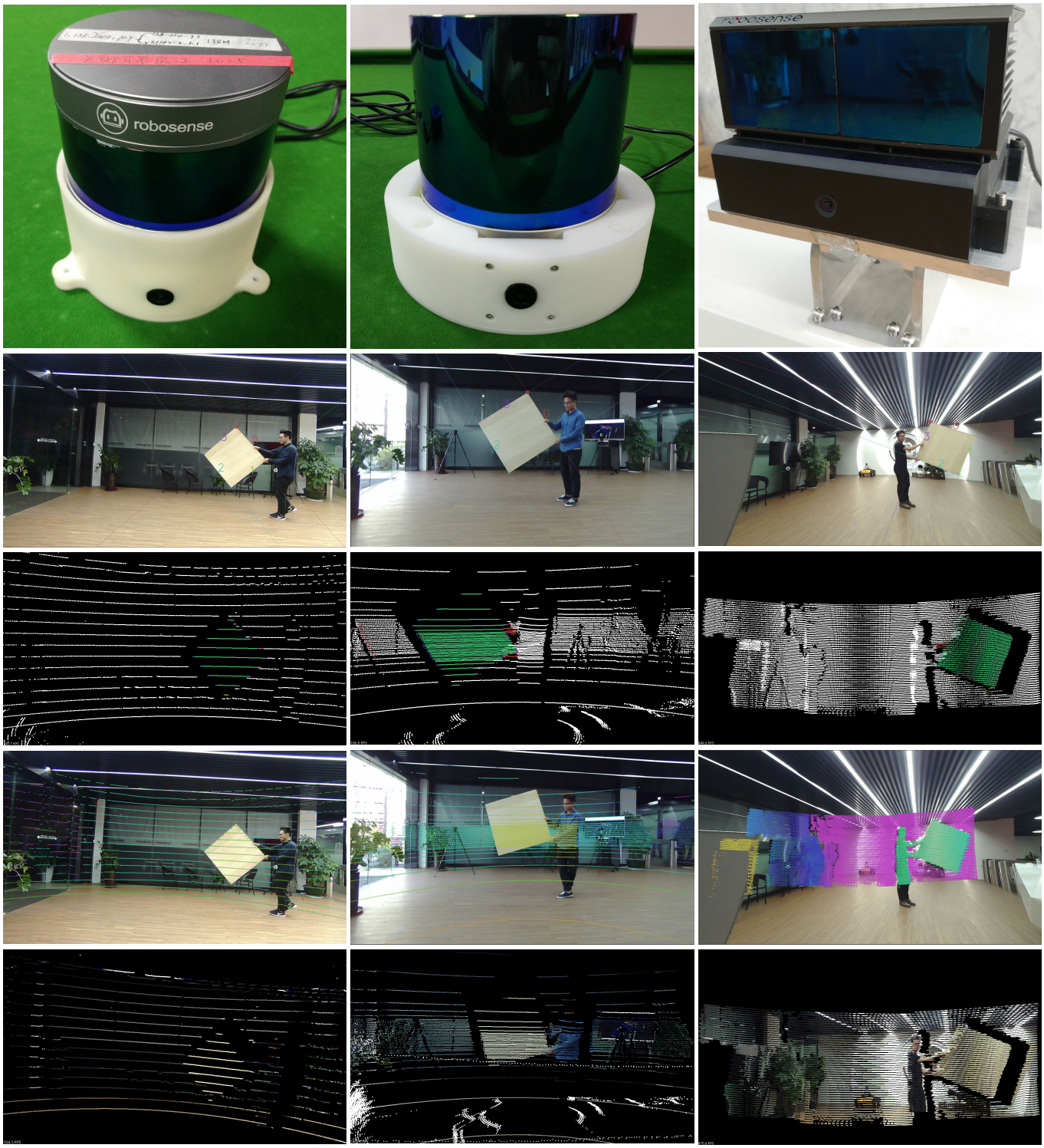
Fig. 4: Calibration result of V-REP simulation. On the left is the depth image, and on the right is the colored pointcloud

B. RS-Lidar

As shown in the 1st row of Fig. 6, we test our approach with three type Lidar in which RS-16 and RS-32 are the normal mechanical scanning Lidar and RS-M1 is newly released 100 line MEMS Lidar. All Lidar-Camera devices are equipped with the same rolling shutter camera which has 1920x1080 resolution, 30 fps and 120° FOV lens. And all camera are calibrated by the ros calibration tool in advance.

The calibration effect is shown in the 4th and 5th row of Fig. 6. Not like the V-REP simulation result, fewer points in 5th row are dyed wrong color since we make the origin of camera frame and Lidar frame close and align their axis by designing the holders. Besides, since the RSLidar-M1 is still in development the calibration effect of RSLidar-M1 and camera is not as good as another two. However, the last figure of Fig. 6 is still impressive.

Fig. 5: Experiment with RSLidar-16



(a) 16 lines Lidar(RSLidar-16) (b) 32 lines Lidar(RSLidar-32) (c) 100 lines Lidar(RSLidar-M1)

Fig. 6: Test devices: three type Lidars with same camera. 1st row shows the test device, 2nd and 3rd row show the polygon extraction, 4th and 5th row show the calibration result, depth image and colored pointcloud

IV. Discussion

Note that the evaluation of the calibration is not feasible to present in quantitative results, as the groundtruth is missing due to the unknown mechanical parameters of the sensors. We instead present the results in a qualitative manner by checking the visualized results along the series of the input sensory data. Further results, library and toolkits are available, in both video and report, at: `www.ram-lab.com/`

research.

V. Conclusion

In this paper, we presented an extrinsic calibration approach for 3D Lidar and camera, which only depends on a polygon board, so that as precision is guaranteed the reliance on the outside is greatly minimized. We also provide a validated calibration library and GUI tool which will be available online. Despite the limitation that the method needs polygon board which is easy to acquire, it still by far a convenient plug-and-play extrinsic calibration system to the community.

For future work, we would improve the 3D polygon extraction algorithm which does not work well for MEMS Lidar currently and more comparisons are to be made. Dataset for bench-marking will also be provided in the near fut re.

References

[1] "Autopilot cited in death of chinese tesla driver," http://www.nytimes.com/2016/09/15/business/fatal-tesla-crash-in-china-involved-autopilot-government-tv-says.html, accessed: 2015-09-14.

[2] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[3] R. Kümmerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g 2 o: A general framework for graph optimization," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 3607–3613.

[4] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o (n) solution to the pnp problem," *International journal of computer vision*, vol. 81, no. 2, pp. 155–166, 2009.

[5] G. Xie, T. Xu, C. Isert, M. Aeberhard, S. Li, and M. Liu, "Online active calibration for a multi-lrf system," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. IEEE, 2015, pp. 806–811.

[6] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, 2000.

[7] Q. Zhang, R. Pless, S. Louis, and M. O. U. States, "Extrinsic Calibration of a Camera and Laser Range Finder (improves camera calibration) ," pp. 2301–2306, 2004.

[8] R. Unnikrishnan and M. Hebert, "Fast Extrinsic Calibration of a Laser Rangefinder to a Camera," *Robotics*, no. July, p. 23, 2005.

[9] G. Pandey, J. McBride, S. Savarese, and R. Eustice, *Extrinsic calibration of a 3D laser scanner and an omnidirectional camera*. IFAC, 2010, vol. 7, no. PART 1. [Online]. Available: http://dx.doi.org/10.3182/20100906-3-IT-2019.00059

[10] S. A. Rodriguez F., V. Frémont, and P. Bonnifait, "Extrinsic calibration between a multi-layer lidar and a camera," *IEEE Int. Conf. Multisens. Fusion Integr. Intell. Syst.*, pp. 214–219, 2008.

[11] X. Gong, Y. Lin, and J. Liu, "3D LIDAR-Camera Extrinsic Calibration Using an Arbitrary Trihedron," *Sensors*, vol. 13, no. 2, p. 1902, 2013. [Online]. Available: http://www.mdpi.com/1424-8220/13/2/1902

[12] A. Dhall, K. Chelani, V. Radhakrishnan, and K. M. Krishna, "LiDAR-Camera Calibration using 3D-3D Point correspondences," *ArXiv e-prints*, May 2017.

[13] D. Scaramuzza, A. Harati, and R. Siegwart, "Extrinsic self calibration of a camera and a 3D laser range finder from natural scenes," *IEEE Int. Conf. Intell. Robot. Syst.*, pp. 4164–4169, 2007.

[14] P. Moghadam, M. Bosse, and R. Zlot, "Line-based extrinsic calibration of range and image sensors," *Proc. - IEEE Int. Conf. Robot. Autom.*, no. c, pp. 3685–3691, 2013.

[15] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 345–358, Jun 1989.

[16] Y. C. Shiu and S. Ahmad, "Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form ax=xb," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 1, pp. 16–29, Feb 1989.

[17] Z. Taylor and J. Nieto, "Motion-based calibration of multimodal sensor extrinsics and timing offset estimation," *IEEE Transactions on Robotics*, vol. 32, no. 5, pp. 1215–1229, Oct 2016.

[18] ——, "Motion-based calibration of multimodal sensor arrays," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 4843–4850.

[19] Q. Liao, M. Liu, L. Tai, and H. Ye, "Extrinsic calibration of 3d range finder and camera without auxiliary object or human intervention," *arXiv preprint arXiv:1703.04391*, 2017.

[20] G. Pandey, J. R. Mcbride, S. Savarese, and R. M. Eustice, "Automatic Targetless Extrinsic Calibration of a 3D Lidar and Camera by Maximizing Mutual Information," *Proc. Twenty-Sixth AAAI Conf. Artif. Intell.*, pp. 2053–2059, 2012. [Online]. Available: http://robots.engin.umich.edu/publications/gpandey-2014b.pdf

[21] Z. Taylor and J. Nieto, "A Mutual Information Approach to Automatic Calibration of Camera and Lidar in Natural Environments," pp. 3–5, 2012.

[22] C. Akinlar and C. Topal, "Edlines: A real-time line segment detector with a false detection control," *Pattern Recogn. Lett.*, vol. 32, no. 13, pp. 1633–1642, Oct. 2011. [Online]. Available: http://dx.doi.org/10.1016/j.patrec.2011.06.001

[23] M. F. E. Rohmer, S. P. N. Singh, "V-rep: a versatile and scalable robot simulation framework," in *Proc. of The International Conference on Intelligent Robots and Systems (IROS)*, 2013.