

PCR-Pro: 3D Sparse and Different Scale Point Clouds Registration and Robust Estimation of Information Matrix For Pose Graph SLAM

M. Usman Maqbool Bhutta¹ and Ming Liu²

Abstract—For both indoor and outdoor environments, we propose an efficient and novel method for different scales and sparse 3D point clouds registration that cannot be handled by the current popular ICP approaches. Our algorithm efficiently detects the scale difference between point clouds and uses the keyframes to estimate the relative pose for calculating the scale difference. The algorithm applies a filter and computes the final transformation which coverages to a global minimum. The good estimation of transform and scale helps in the calculation of the covariance matrix using a closed form solution efficiently. This covariance between point clouds helps in the estimation of information matrix for pose-graph SLAM.

I. INTRODUCTION

In the last decade, simultaneous localization and mapping (SLAM) has received a large amount of intention from researchers all over the world due to its vast applications in self-driving cars and indoor/outdoor autonomous robotics employment. In SLAM, the graph-based technique [5] has become a very popular choice due to its robustness and efficiency for large-scale robotics purposes. The main goal of these graph-based approaches is to set up the state variables or parameters that maximally characterizes the effect of Gaussian noise on measurements, where robots poses are used in the optimization and minimization of the error function. In graph-based SLAM, the state variables are the state of the robot and position of the landmarks. These parameters can be estimated with the sensors of the robot. Therefore, the measurement of state variables depends only on the relative poses of the robots. For instance, in visual odometry, measurement depends only on the connected poses.

In graph SLAM, state variables are defined as a node in the graph for optimization, and the pairwise information between two nodes defines the edge. After estimating the edge information, many algorithms have been described in the literature for further solving such a problem. A simple implementation can be done using some standard approaches such as Gauss-Seidel relaxation, Gauss-Newton

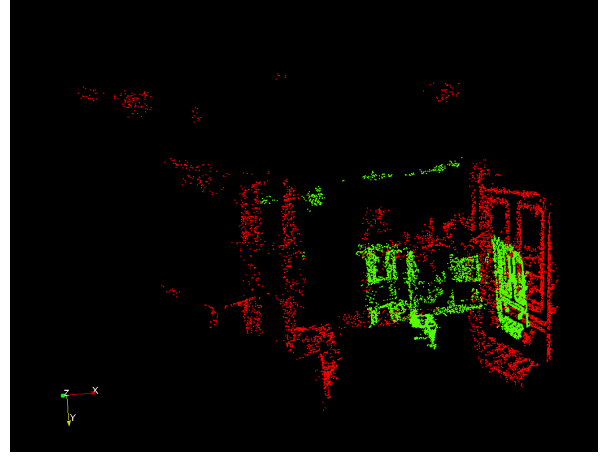


Fig. 1: Scenario of the registration of sparse point clouds with scale difference. The green point cloud is the source point cloud while the target point cloud is shown in red.

and Levenberg-Marquardt (LM), which typically provide meaningful results for many applications. For problems that can be described as a graph, such as nonlinear least square problems, a general framework g^2o [7] is used for optimization. This algorithm produces excellent results because it can exploit the sparsity between adjacent nodes of the graph and uses robust methods to solve the sparse linear system.

Given two point clouds with no scale difference, we can easily apply ICP to calculate the transformation, which basically maps all the points data of the source point cloud to the target point cloud. For aligning point clouds, multiple methods have been proposed [10], [14]–[16]. Once we have a large point cloud map, we can use it in different applications, such as in autonomous navigation [8], [11] and apply processing techniques such as segmentation within a point cloud [9], [12], [20].

To find the edge information between two poses without any external sensors, such as an IMU, or wheels rotatory information is a challenging task in large-scale 3D mapping. Another difficult task is to find out the scale difference when two or more graphs are totally separated and have been created by different cameras and approaches, as shown in Fig. 1. Current ICP techniques are not sufficient to deal with such kinds of target problems. As multi-agent SLAM is the next interest for computer vision researchers, this creates a need for such a framework that actually accepts mapping

*This paper is supported by Shenzhen Science, Technology and Innovation Commission (SZSTI) JCYJ20160428154842603 also supported by the Research Grant Council of Hong Kong SAR Government, China, under Project No. 11210017 and No. 16212815 and No. 21202816, NSFC U1713211 awarded to Prof. Ming Liu

¹BHUTTA, M. Usman Maqbool is a Ph.D. candidate at Robotics and Multi-Perception Lab, Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, HK mumbhutta@ust.hk

²Ming Liu is an Assistant Professor at Department of Electronics and Computer Engineering and also serves as Director of Robotics and Multi-Perception Lab, The Hong Kong University of Science and Technology, HK eelium@ust.hk

information from different approaches and camera systems. For this scenario, point cloud registration of different scales is required. We thus propose a framework which will help deal with such kinds of problems. We strongly believe that it will help the computer vision community in many multi-agent scenarios. Our contribution to this framework is as follows;

- It introduces a way of using direct SLAM approaches for multi-agent SLAM systems where we do not have information of the scale difference.
- It welcomes the features data to the featureless SLAM approaches.
- It presents a robust method to find an accurate transformation between sparse point clouds where the system converges to a global minimum.
- The covariance is very small, which will help in the usage of the information matrix and will solve many existing problems.

II. OVERVIEW OF EXISTING WORK

For large scale 3D reconstruction difficulties, point cloud registration is a very popular choice. ICP is a general and common approach to point cloud registration, which works interactively to find out pairs of closed points in a point clouds scan. The affine transformation (4x4) T is calculated between two point clouds. This iterative algorithm minimizes a loss function $f(p_x, p_y, \omega)$ which is define as,

$$f(p_x, p_y, \omega) = \sum_{i=1}^n \| T(a_i, (p_x, p_y, \omega)) - b_{\theta(i)} \|^2, \quad (1)$$

where T shows the affine transformation from the source scan to the target scan, ω is the rotation value, and the translation in x and y are defined by p_x and p_y , respectively. θ interprets the index value of a point in the target point cloud, which is adjacent to the index i of the source point cloud. In a 2D case, the minimization of the error function is done in the x, y direction, and the rotation angle of the source scan to the target scan is optimized [17].

For SLAM using laser data, [1] proposed an efficient algorithm that firstly pre-processes the point cloud by removing the remote points in the point cloud data. After that, an assignment is done by finding out pairs of the closest points. Following, rejection is done by removing the pairs that have a long distance (by choosing the short-length pairs), and then using equation (1), and the loss function is calculated. After applying the ICP transformation to source point cloud PC_s , the variance (average distance) between the aligned points is then estimated. Sometimes this variance is multiplied by the identity matrix to calculate the information model [3].

A modular ICP chain has been developed by [15]. This approach is known as libpointmatcher, which is open source software for ICP estimation. In simple terms, it calculates the translation and rotation matrix between two

scans by minimizing the error function. The 3D source and target scans are sometimes named reference and reading, respectively, and the main goal is to align the reading with the reference. Libpointmatcher firstly applies different filtration algorithms to both the reference and the reading 3D scans and then starts the iteration by correlating points of the reading and reference point clouds. In each iteration, it finds out the transformation that reduces the alignment error.

The system iterates from the data filters to the error minimizer until the transformation conditions are satisfied or the alignment converges to their global minimum. A transformation checker is used for controlling the number of iterations. For instance, the fixed number of iterations or minimum error function threshold information can be used by the transformation checker. It can be used multiple times like the data filters in a chained form. The key idea for providing an interface for each step is to allow different algorithms at any step for evaluating the impact of the ICP behavior.

Point cloud registration is a point cloud aligning problem that is used to merge different point clouds to get a large scale map if some matches are found. The libpointmatcher and 3D ICP closed form solution [16] performed very well and produced excellent results. But these approaches use dense point clouds as data with the same scale (taken from the depth camera or lidar scan). In the 3D closed form solution, for extraction of uniform key points, SHOT [19] feature descriptors were used, which estimate nearest neighbor matches of the uniform keypoints features. This makes the ICP slightly slow compared to libpointmatcher. While using the [15] libpointmatcher and [16] 3D ICP closed form solution on sparse point clouds for registration using their respective ICP techniques, both the algorithms tried to align both point clouds and were unable to converge to a global minimum.

III. FRAMEWORK OF THE PROPOSED APPROACH

In SLAM using the camera, the point cloud is not as dense as it was when using lidar. Due to much fewer features in sparse point clouds, the chances for global convergence becomes very low using these methods. The scale problem is another factor, but our system (PCR-Pro) helps in the very efficient manipulation of sparse and different scale point clouds fusion. The framework of the PCR-Pro is shown in Fig. 2.

For estimating the robustness of the proposed algorithm, we consider the case of sparse points clouds, where the source 3D point cloud and target 3D point cloud are shown in green and red, respectively. The scale problem between them can be shown very clearly in Fig. 1.

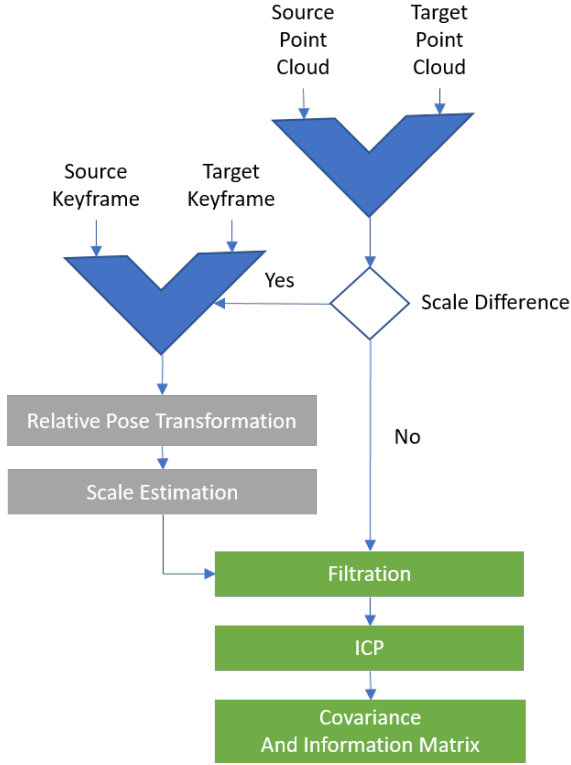


Fig. 2: Overview of the proposed system (PCR-Pro) framework

A. Scale Detection

Our approach is to make the system work for both scenarios, whether they have a scale difference or they are almost equal. The first task is to find the difference using 3D sizes of both point clouds. If there is a difference detected in the volumes, then the system will first calculate the scale factor between them.

B. Scale Estimation

As discussed earlier, the scale plays a very important part in the alignment of point clouds. For the estimation of scale, 3D-point-clouds-related keyframes are used.

1) *Relative Pose estimation*: OpenGV [6] has a generalized description of sensors which makes it more reliable for central and non-central problems of the multi-camera system. It also supports estimation of absolute pose and relative pose for both central and non-central problems. It directly explores 3D information and estimation of the camera projection system using the image information without using the camera intrinsic and extrinsic properties. In relative pose estimation, OpenGV basically first creates the RANSAC object based on keypoint descriptors and their respective correspondences, and by using the threshold mentioned in equation (2) and maximum iteration information, it calculates the relative transformation:

Algorithm 1: Proposed Approach

Input: keyframes, Point cloud of each keyframe

Output: 3D Transformation, Information Matrices initialization;

for each point clouds **do**
 └ detect scale difference

if scale difference **then**

for each keyframe **do**

 compute SIFT features and use FlannBased descriptor to find matches

 calculate matches between keyframes;

for all matches **do**

 filter good matches

if good matches **then**

 by using opengv [6], create RANSAC object;
 set threshold;
 set max iteration;
 compute central relative pose OT between keyframes

for all good matches **do**

 compute scale difference SC using kalman filter
 and relative pose OT

 Transform source point cloud using scale transformation SC ;

 apply filter to crop lower area of source and target point clouds;

 Estimate transformation T of filtered source and target point clouds using ICP [15];

 Transform original source point cloud using rigid transformation RT will map on the target point cloud;
 Now calculate of information matrix;

function 3D ICP COVARIANCE [18] (Source Point Cloud, Target Point Cloud, Final transformation T)

 calculate the covariance using final transformation;

return 6x6 covariance matrix;

 Information Matrix = (covariance matrix) $^{-1}$

$$\epsilon_{threshold} = 1 - \cos\theta_{threshold} = 1 - \cos(\arctan\psi/l), \quad (2)$$

where ψ is the classical reproduction error-threshold which is in pixels. The values are taken from the camera intrinsic parameters.

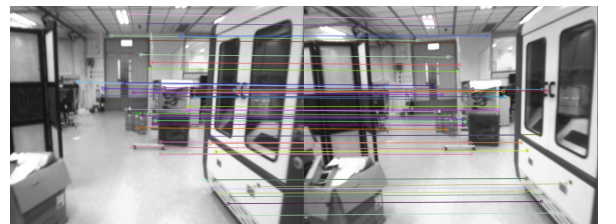


Fig. 3: Feature matching between two keyframes

2) *Scale calculation using Kalman filter*: Camera intrinsic parameters are very important to find out the exact scale difference between two different keyframes of scenarios. These two keyframes can belong to two different cameras, but here we include the results of different sessions using the same camera. Firstly we calculate the feature matches between these keyframes,. We use a Flann-based descriptor [13] for SIFT feature matching, then filter good matches from all the feature matches, as shown in Fig. 3. After filtering the good matches, the main goal is to find the transformation that can align both point clouds so that after applying the ICP, the system will converge to a global minimum, and minimize the covariance as well.

For all matches, we calculate the 2D points on the image, and then we map 2D points to 3D points using equation (3).

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} X * \frac{1}{f_x} + \frac{-c_x}{f_x} \\ Y * \frac{1}{f_y} + \frac{-c_y}{f_y} \\ 1 \end{bmatrix}, \quad (3)$$

where $X, Y \in \mathbb{R}^2$, $x, y, z \in \mathbb{R}^3$ and f_x, f_y, c_x and c_y are taken from the camera intrinsic parameters.

Now that we have 3D points of the source and target keyframes, we apply relative pose transformation to all 3D points of the source keyframes. Using the Kalman filter, we increment the scale until all the transformed 3D points of the source keyframe are almost equal to the 3D points of the target keyframe. When the system converges to a global minimum, it will give us the scale difference, which we will then apply to their related point clouds.

C. Filtration

The source point cloud is transformed by the scale factor to remove the scale issue. Then both the source and target 3D scans are in almost the same scale form ready for the ICP for the indoor environment case. For the outdoor environment, a lot of noise is affected on the upper part of the point clouds where the depth is infinite (towards the sky). Therefore, there is a high error chance while in alignment and then applying the ICP. Our algorithm applies filtration to crop the lower 25% of the point clouds. Most of the features are on the ground side which is sufficient for better alignments of both point clouds.

D. Exact Transformation Computation

Using the filtered source and target point clouds, the libpointmatcher algorithm is applied to obtain the ICP transformation to get the global minimum. The system applies this transformation to the original source point cloud, which basically fuses with the target point clouds and gives excellent transformation, as shown in Fig. 4.

E. Estimation of the information matrix

For pose graph SLAM, optimization is done between two nodes by using the information matrix. The information matrix can be calculated by taking the inverse of the covariance

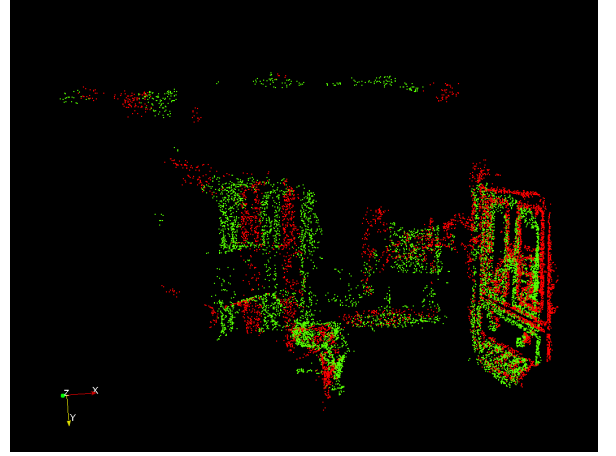


Fig. 4: Source and target point clouds after scaling and registration using the proposed method

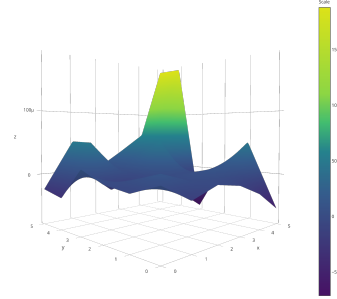


Fig. 5: Covariance calculation

matrix, which is a 6x6 matrix. The covariance matrix tells us about the covariance between two nodes and is usually, it is calculated from the sensors of robots. Recently [16] developed a robust estimate of 3D ICP covariance that is a RANSAC-based approach. This method uses the concept of covariance as constraints that minimize the objective function [2].

Firstly a closed-form expression for the Jacobian-based objective function is calculated, then the algorithm calculates the covariance using equation (4):

$$cov(x) \approx \left(\frac{\partial^2 J}{\partial x^2} \right)^{-1} \left(\frac{\partial^2 J}{\partial z \partial x} \right) cov(z) \left(\frac{\partial^2 J}{\partial z \partial x} \right)^T \left(\frac{\partial^2 J}{\partial x^2} \right)^{-1}, \quad (4)$$

In pose graph SLAM, for each pose, the keyframe transformation can be found out using OpenGV [6] between the two poses of the unknown robots (without knowing the camera matrix). The key idea is to find the covariance matrix by using the above closed-form method on the adjacent keyframes, and information for the pose graph SLAM.

For ICP, the problem can be written as

$$J = \text{minimize} \sum_{i=1}^n \| R P_i + T - Q_i \|^2. \quad (5)$$

Equation (5) represents the objective function of the ICP for registration of point cloud P and Q . P_i and Q_i show the 3D points. Furthermore, R and T are the rotation and

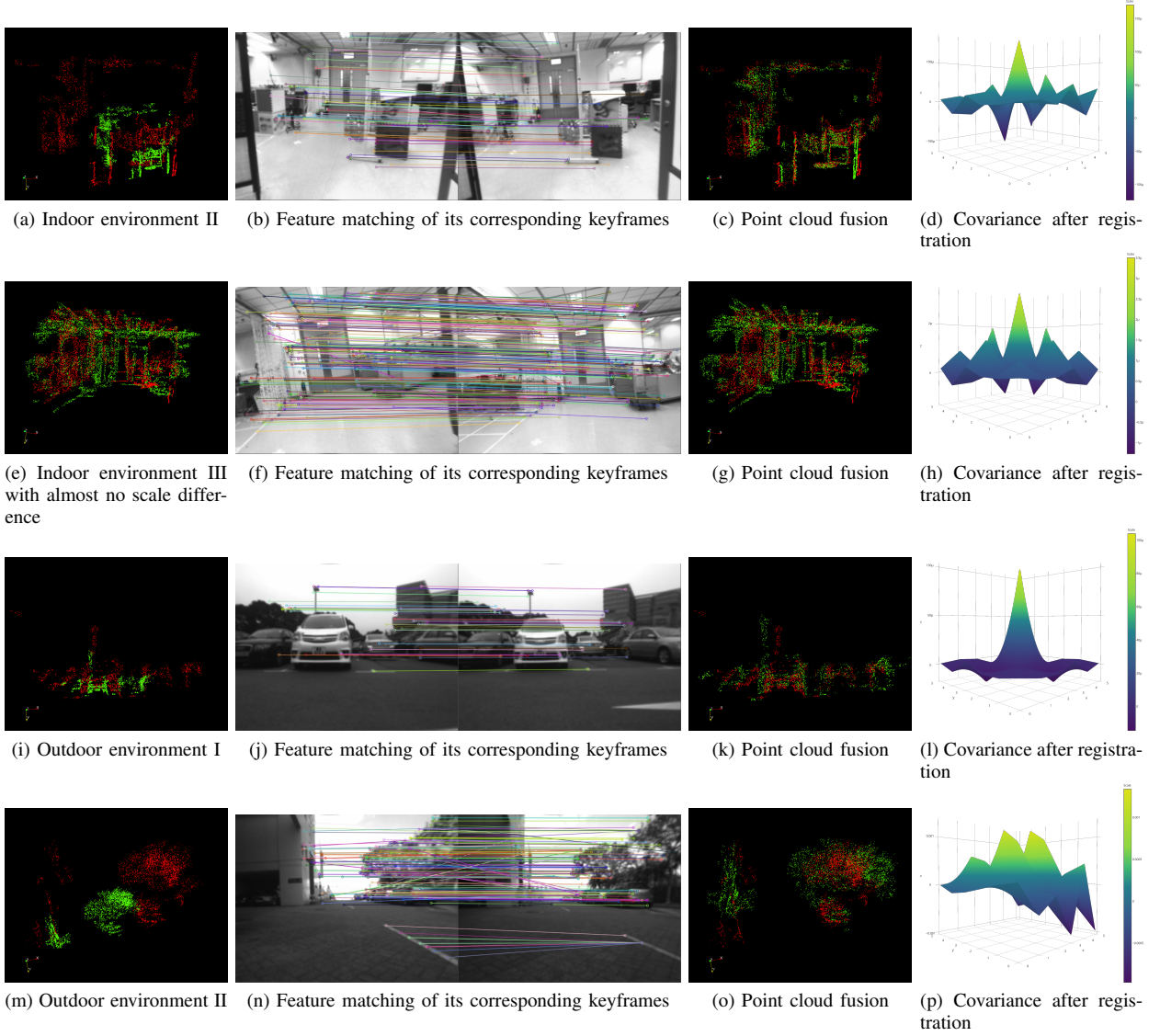


Fig. 6: Indoor and outdoor point cloud fusion and covariance results.

translation, respectively, of homogeneous transformation H . The main goal is to estimate the covariance 6×6 matrix of the ICP transformation. Point to point matrix-based error function will be used, as mentioned in equation (5).

Each set of correspondence $\{P_i, Q_i\}$ calculates $\left(\frac{\partial^2 J}{\partial x^2}\right)$ in which J_i depends on the value of the i^{th} correspondence. After calculating each correspondence, it sums them to form the final estimate of $\left(\frac{\partial^2 J}{\partial x^2}\right)$:

$$\left(\frac{\partial^2 J}{\partial x^2}\right) = \sum_{i=1}^n \left(\frac{\partial^2 J_i}{\partial x^2}\right). \quad (6)$$

The author in [16] proposed a robust technique to calculate the covariance using the point clouds and can be calculated as;

$$J = \text{minimize} \sum_{i=1}^n F^2, \quad (7)$$

where $F = \|G\|$, and $G = RP_i + T - Q_i$.

Our target is to find out the covariance using equation (4), which will use J . We efficiently calculate J using equation (7) and then successfully calculates the covariance matrix of size 6×6 . As the algorithm converges to global minimum, so the resulting covariance is too small. The calculated values are shown in Fig. 5. By taking the inverse transform, the information matrix is calculated.

IV. NUMERICAL RESULTS AND DISCUSSIONS

Our main goal is to finalize a method for sparse point clouds registration to a global minimum. This also lessens the covariance between the correspondence of two 3D scans.

The uncertainty in the transformation from one pose to another can be neglected if the transformation is correct, and an identity matrix can be used in that case while applying the pose graph optimization.

Using the feature matching technique, we aligned the source and target point clouds together. Then the PCR-Pro algorithm estimated the accurate transformation for point cloud registration, which minimized the point to point error between the correspondences (nearest neighbor points). Our algorithm produces excellent results and successfully converges to a global minimum, for both indoor and outdoor environments, and with or without scale difference, as shown in Fig. 6.

As the system converges successfully for different sparse point clouds, as compared to other methods mentioned above, our estimated covariance is minimum. In pose graph SLAM, in this case, we can also take the identity matrix as the information matrix, and multiply it by a large number because the pose transformation is almost correct. Otherwise, we will estimate the covariance using equation (4). We have used LSD-SLAM [4] for the collection of the dataset (keyframes and their respective point clouds) for validation of the algorithm.

After the fusion of point clouds in different indoor and outdoor cases, the covariance estimation results showed the robustness of PCR-Pro in all cases. Due to the best transformation and scaling for all scenarios, the estimated covariance is too small, as shown in Fig. 6. If the measurements from sensors z are uncorrelated with each other, then the covariance matrix will be diagonal. The information matrix is just an inverse of the 6x6 covariance matrix in 6DOF space, where these are the upper triangular values of the 6x6 information matrix, i.e., the inverse of the covariance matrix, representing the uncertainty between two nodes (edges).

With an increase in the number of correspondences, the computation time to estimate the covariance also increases. Applying the certain threshold limit in the number of correspondences minimizes the computation time, which enhances the efficiency of the covariance estimation technique. Our proposed algorithm uses a C++ implementation and the source code can be accessed from <https://sites.google.com/view/pcr-pro>.

V. CONCLUSIONS

We discussed in detail the existing ICP algorithms and covariance estimation between two point clouds for pose graph SLAM. Our algorithm works excellently as compared to other ICP algorithms, regardless of the sparsity of the point cloud. Furthermore, the estimated covariance is also very small as the point cloud registration converges to the global optimal. For pose graph SLAM, estimation of the uncertainty using the covariance, and then calculation of the information matrix is quite difficult for a monocular camera SLAM system. We proposed an efficient algorithm that gives excellent results for transformation estimation, which can

work for both real-time and off-time large-scale mapping applications.

REFERENCES

- [1] Antoni Burguera, Gabriel Oliver, and J.D. Tardos. Robust scan matching localization using ultrasonic range finders. In *2005 IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pages 1367–1372. IEEE, 2005.
- [2] Andrea Censi. An accurate closed-form estimate of ICP's covariance. In *Proc. 2007 IEEE Int. Conf. Robot. Autom.*, pages 3167–3172. IEEE, April 2007.
- [3] Felix Endres, Jurgen Hess, Jurgen Sturm, Daniel Cremers, and Wolfram Burgard. 3-D Mapping With an RGB-D Camera. *IEEE Trans. Robot.*, 30(1):177–187, February 2014.
- [4] Jakob Engel, Thomas Schöps, and Daniel Cremers. *LSD-SLAM: Large-Scale Direct monocular SLAM*, volume 8690 LNCS, pages 834–849. Springer International Publishing, Cham, 2014.
- [5] Giorgio Grisetti, R Kummerle, Cyrill Stachniss, and Wolfram Burgard. A Tutorial on Graph-Based SLAM. *IEEE Intell. Transp. Syst. Mag.*, 2(4):31–43, 2010.
- [6] Laurent Kneip and Paul Furgale. OpenGV: A unified and generalized approach to real-time calibrated geometric vision. In *2014 IEEE Int. Conf. Robot. Autom.*, pages 1–8. IEEE, May 2014.
- [7] Rainer Kummerle, Giorgio Grisetti, Hauke Strasdat, Kurt Konolige, and Wolfram Burgard. G2o: A general framework for graph optimization. In *2011 IEEE Int. Conf. Robot. Autom.*, pages 3607–3613. IEEE, May 2011.
- [8] M Liu. Robotic Online Path Planning on Point Cloud. *IEEE Trans. Cybern.*, 46(5):1217–1228, May 2016.
- [9] Ming Liu. Efficient segmentation and plane modeling of point-cloud for structured environment by normal clustering and tensor voting. In *2014 {IEEE} Int. Conf. Robot. Biomimetics, {ROBIO} 2014, Bali, Indones. December 5-10, 2014*, pages 1805–1810, 2014.
- [10] Ming Liu, Francois Pomerleau, Francis Colas, and Roland Siegwart. Normal Estimation for Pointcloud using GPU based Sparse Tensor Voting. In *IEEE Int. Conf. Robot. Biomimetics*, 2012.
- [11] Ming Liu and R Siegwart. Navigation on point-cloud #x2014; A Riemannian metric approach. In *Robot. Autom. (ICRA), 2014 IEEE Int. Conf.*, pages 4088–4093, May 2014.
- [12] Ming Liu and Roland Siegwart. Information Theory based Validation for Point-cloud Segmentation aided by Tensor Voting. In *Int. Conf. Inf. Autom.*. IEEE, 2013.
- [13] Marius Muja and David G Lowe. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. In *Int. Conf. Comput. Vis. Theory Appl. Viss.*, pages 331–340. INSTICC Press, 2009.
- [14] François Pomerleau, Ming Liu, Francis Colas, and Roland Siegwart. Challenging data sets for point cloud registration algorithms. *Int. J. Rob. Res.*, 31(14):1705–1711, 2012.
- [15] François Pomerleau, Stéphane Magnenat, Francis Colas, Ming Liu, and Roland Siegwart. Tracking a depth camera: Parameter exploration for fast ICP. In *2011 IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pages 3824–3829. IEEE, September 2011.
- [16] Sai Manoj Prakhya, Liu Bingbing, Yan Rui, and Weisi Lin. A closed-form estimate of 3D ICP covariance. In *2015 14th IAPR Int. Conf. Mach. Vis. Appl.*, number 3, pages 526–529. IEEE, May 2015.
- [17] Szymon Rusinkiewicz and Marc Levoy. Efficient variants of the ICP algorithm. In *Proc. Third Int. Conf. 3-D Digit. Imaging Model.*, pages 145–152. IEEE Comput. Soc, 2001.
- [18] Radu Bogdan Rusu and Steve Cousins. 3D is here: Point Cloud Library (PCL). In *2011 IEEE Int. Conf. Robot. Autom.*, pages 1–4. IEEE, May 2011.
- [19] Samuele Salti, Federico Tombari, and Luigi Di Stefano. SHOT: Unique signatures of histograms for surface and texture description. *Comput. Vis. Image Underst.*, 125:251–264, August 2014.
- [20] S Wang, H Huang, and M Liu. Simultaneous clustering classification and tracking on point clouds using Bayesian filter. In *2017 IEEE Int. Conf. Robot. Biomimetics*, pages 2521–2526, December 2017.