

A WEIGHTED CLOSED-FORM SOLUTION FOR RGB-D DATA REGISTRATION

K. M. Vestena^a, D. R. Dos Santos^a, E. M. Oilveira. Jr, N. L. Pavan, K. Khoshelham^b

^aDept. of Geomatic, University Federal of Parana, Curitiba, Parana, Brazil – kauemv2@gmail.com, danielsantos@ufpr.br,
elizeuoliveirajr@gmail.com
nadissonluisp@gmail.com

^bDept of Infrastructure Engineering, University of Melbourne, 3010 VIC, Parkville, Australia - k.khoshelham@unimelb.edu.au

Commission III, WG III/2

KEY WORDS: RGB-D data, dual-number quaternions, registration, loop closing, globally consistent maps

ABSTRACT:

Existing 3D indoor mapping of RGB-D data are prominently point-based and feature-based methods. In most cases iterative closest point (ICP) and its variants are generally used for pairwise registration process. Considering that the ICP algorithm requires an relatively accurate initial transformation and high overlap a weighted closed-form solution for RGB-D data registration is proposed. In this solution, we weighted and normalized the 3D points based on the theoretical random errors and the dual-number quaternions are used to represent the 3D rigid body motion. Basically, dual-number quaternions provide a closed-form solution by minimizing a cost function. The most important advantage of the closed-form solution is that it provides the optimal transformation in one-step, it does not need to calculate good initial estimates and expressively decreases the demand for computer resources in contrast to the iterative method. Basically, first our method exploits RGB information. We employed a scale invariant feature transformation (SIFT) for extracting, detecting, and matching features. It is able to detect and describe local features that are invariant to scaling and rotation. To detect and filter outliers, we used random sample consensus (RANSAC) algorithm, jointly with an statistical dispersion called interquartile range (IQR). After, a new RGB-D loop-closure solution is implemented based on the volumetric information between pair of point clouds and the dispersion of the random errors. The loop-closure consists to recognize when the sensor revisits some region. Finally, a globally consistent map is created to minimize the registration errors via a graph-based optimization. The effectiveness of the proposed method is demonstrated with a Kinect dataset. The experimental results show that the proposed method can properly map the indoor environment with an absolute accuracy around 1.5% of the travel of a trajectory.

1. INTRODUCTION

RGB-D cameras have received considerable attention of robotic, vision and photogrammetric researches and new challenges for 3D mapping of indoor environment have been developed. Such sensors can provide a colored 3D point cloud, quite useful for on line maps and their advantages compared with laser scanning and ToF cameras are the lightweight, low cost, faster and highly flexible and also it does not need expertize human interaction. In the actual applications, we need to track features in texture-less regions, or with limited texture, without the need for additional sensors and it requires robust registration of long sequences of color/depth images. Systems using RGB-D cameras also are quite useful because can exploit both visual and depth information to handle the problem.

The most important existing 3D mapping methods using RGB-D cameras firstly perform a feature-based tracking technique for extract and matches 2D feature in the RGB images. Algorithm developed by Lowe (2004) known as scale-invariant feature transform (SIFT), is often used. Then, their associated depth values are used for a pairwise registration. The registration goal is to compute the rotation and translation parameters of the separately acquired 3D point clouds in a global coordinate framework. The most prominent registration method is known as iterative closest point (ICP) algorithm and it was originally developed by Besl and McKay (1992). Finally, a global consistency of the complete data sequence is performed to minimize the registration errors, which can be solve using methods based on extended kalman filtering (EKF), maximum

likelihood estimation, expectation maximization (Thrun, 2003) and pose graphs (Grisetti et al., 2010).

In literature most of the proposed works, has despite of the limited depth precision provided by the RGB-D cameras taking as an implicit assumption that errors in XYZ coordinates are independent and equally weighted. According to Khoshelham and Elberink (2012); and Park et al. (2012) the quality of the point cloud from the Kinect devices is influenced by: the low resolution of the depth measurements; by the lighting condition; properties of object surfaces in the disparity data and also due to sensor noise, occlusions and changes in the viewpoint direction achieving low accuracy for 3D mapping (Henry et al., 2012). In order to utilize RGB-D data, the information regarding uncertainty of the sensor is essential for a reliable 3D registration.

As described, in most cases, ICP and its variants are generally used for pairwise registration process. Considering that the ICP algorithm requires an relatively accurate initial transformation and high overlap a weighted closed-form solution for RGB-D data registration is proposed. In our solution, we have weighted and normalized the 3D points based on the theoretical random xyz errors and the dual-number quaternions are used to represent the 3D rotation. Basically, dual-number quaternions provide a closed-form solution by minimizing a cost function. The most important advantage of the closed-form solution is that it provides an optimal transformation in one-step and it does not need to calculate a good initial estimate and expressively decreases the demand for computer resources, in contrast to the iterative method. We also describe a new RGB-D

loop-closure approach, based on the uncertainty of the sensor and also on the volumetric information between overlap areas of the pair of point clouds.

The paper proceeds with related work in the Section 2. In section 3, the methods for weighting of closed-form solution and loop-closure detection as presented. Section 4, describes the experiments and its results using the proposed method. The paper concludes with final remarks, in Section 5.

2. RELATED WORK

Since 90s several approaches have been proposed to register 3D point clouds for 3D mapping applications. Although the ICP algorithm is quite useful for point cloud registration, its performance on a dense point cloud is a time-consuming procedure and also requires a relatively accurate initial transformation. Chen and Medioni (1992), Masuda and Yokoya (1994), Eggert et al. (1998), Okatani and Deguchi (2002), Park and Murali (2003), Segal et al. (2009) and others have proposed some variations and improvements for the ICP algorithm, using point-to-plane and plane-plane approaches. Other researchers addressed methods that integrate both depth and intensity information for the 3D registration (Godin et al., 1994; Weik, 1997; Stamos and Leordeanu, 2003; Barnea and Filin, 2008; Swadzba et al., 2007; May et al., 2009; Ye and Bruch, 2010; Khoshelham, 2010). The aforementioned approaches discussed there, make use of the depth image information supplied by a laser scanning device or range cameras, to transfer the selected 2D features to 3D points. They also report some aspects of their matching algorithm design and registration method proposed. It should be important to note that, the majority of the approaches consist of feature detection and matching on arbitrary images.

However, due to the low resolution of intensity images delivered by both laser scanning and range devices, providing only a few features for certain scene and a low accuracy and precision in the depth value at the feature location, the automatic detection and matching algorithms, such as SIFT and SURF, cannot correctly localize the key points on the actual corner location. Driven by this issue, Al-Manasir and Fraser (2006) suggested the use of the calculated transformation parameters via a bundle adjustment of images taken by a digital camera to register the laser scanning data. In Ellekilde et al. (2007) a data association, performed by a combination of SIFT 2D matching plus a 3D outlier removal and a least square 3D point set fitting was proposed for successfully provide the means to accurately establish an initial local registration. The distinctive features are detected in the images taken by a conventional camera mounted and they discuss the use of initial local registration as input for globally estimation process. Prusak et al. (2008), suggested a joint approach for robot navigation employing ToF camera combined with a fish eye camera. They also report the use of a Trimmed ICP algorithm proposed by Chetverikov et al. (2002) for improve the registration, affected by many errors. The authors generate a 3D-depth-panorama from many ToF images to figure out the very few 2D-3D correspondences caused due to the small field-of-view (FoV) covered by ToF camera, for the registration problem. Nonetheless, the registration using the 3D-panorama lacks absolute accuracy, due to the incompleteness of the panorama at occlusions and unmapped objects. In Huhle et al. (2008) the local registration is performed based on the normal distribution transform (NDT) and the SIFT 2D feature detector is applied to the RGB images. They combine the globally feature-based approach and the local fitting technique and

report some results qualitative results of scene reconstruction. Image-based registration group (IBR) methods for 3D point cloud also were proposed by Bendels et al. (2004), Dold and Brenner (2006), Barnea and Filin (2007), Ellekilde et al. (2007), Prusak et al. (2008), Huhle et al. (2008) and Kang et al. (2009).

In the ideal case, one would like to have added images that could have improved the registrations performed previously, to the meta-view. The RGB-D cameras potentially can also be used for that purpose.

In Henry et al. (2010), Du et al. (2011), Engelhard et al. (2011), Steinbrucker et al. (2011), Dryanovski et al. (2012), Endres et al. (2012) they all have discussed the use of both rich visual features and shape matching for pairwise registration and globally consistent alignment into ICP optimization. In Henry et al. (2012) a new ICP variant (RGB-D ICP) with re-projection RANSAC algorithm is proposed to exploit the advantage of both color and depth information contained in RGB-D, to categorize the sparse point feature in each frame. Recently, RGB-D cameras have been widely utilized in simultaneous localization and mapping (SLAM) problems, in which the globally consistent alignment is one of the most important methods for 3D registration. The SLAM problem became ubiquitous. Accordingly to Grisetti et al (2010), "*Lu and Milios (1997) were the first to refine a map by globally optimizing the system of equations to reduce the error introduced by constraints*".

As described before, in indoor environments, is very common the occurrence of texture-less or repetitive pattern. Taguchi et al. (2013) was the first to propose plane-based SLAM for RGB-D sensors using both points and planes as primitives for registration of 3D data. Raposo et al. (2013) proposed a new visual odometry method which uses both depth and color information for relative pose estimation between consecutive frames. In the absence of the minimum number of required planes, 2D point correspondences are extracted for finding the remaining degrees of freedom. Ataer-Cansizoglu et al. (2013) presented a SLAM system that exploits planes in conjunction with points, as primitives in order to minimize failure cases, due to texture less regions. In Khoshelham et al. (2013) a weighting scheme to adjust the contribution of the 3D point correspondences for estimation of the transformation parameters is proposed. The obtained results demonstrated that weighting the 3D points improves the accuracy of sensor pose estimation along the trajectory. However, the before mentioned and proposed method, are not implemented to be robust into texture less regions. However, such approach fail if there are no a minimum of 3 non-parallel planes or too few point correspondences to estimate the parameters.

In Chow et al. (2014) a new point-to-plane ICP, that minimizes the reprojection error of the infrared camera and projector pair in an implicit iterative extended Kalman filter (IEKF) to account for the texture-less regions is presented. They also integrated observations from an IMU to provide changes in rotation and translation for initializing 2D/3D matching. According the authors, the point-to-plane ICP only minimize the 1D orthogonal distance. For a side-looking Kinect it not prevents it from sliding along the wall as desired. An inertial navigation system is used to help push the solution forward in the along-track direction while the ICP corrects the across-track position and the two orientations that are not parallel to the normal of that wall. They also included weighting scheme

describing the uncertainties in the point cloud, due to the baseline-to-range ratio and parallax angle over the Kinect's full frame. Nevertheless, the developed approach depends on additional sensors (i.e, IMU) to provide a good initial alignment for the ICP.

Dos Santos et al. (2016) presented an adaptive coarse-to-fine registration method for 3D indoor mapping using RGB-D data. They have weighted the 3D points based on the theoretical random error of depth measurements, such as in Khoshelham et al. (2013) They also have introduced a novel disparity-based model for an accurate and robust coarse-to-fine registration.

3. WEIGTHED CLOSED-FORM SOLUTION

The generic structure of our proposed method framework, shown in Fig. 1, is as follows. Given a pair of RGB frames, first, we extract and match sparse visual features in successive RGB frames. After the outliers are detected and removed using the RANSAC algorithm and a statistical dispersion called interquartile range (IQR), then we create the successive 3D point clouds by projecting the depth map into 3D coordinates and associate the inlier points with their corresponding depth values. After that, we employ a registration approach to find the transformation parameters. The loop-closure detection is applied based on the volumetric information between pair of point clouds. Finally, the global adjustment must be executed.

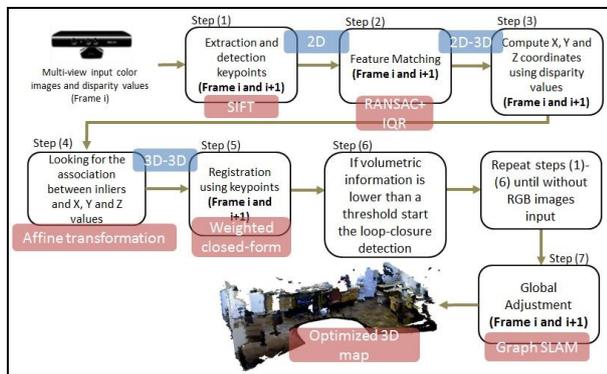


Figure 1. Generic structure of the proposed method.

Note that we have previously calibrated the Kinect device using the MATLAB Camera Calibration Toolbox (Bouquet, 2004). The Bouquet method improved the technique originally proposed by Zhang (2000), and also estimated the relative translations and rotations between the RGB and IR sensors. The following sections present the adopted strategy for 3D indoor mapping in this paper.

3.1 3D point matching procedure

Since we have RGB images, we can exploit the advantages of the color information and extract sparse visual features from the two consecutive frames and after associate them with their corresponding depth values. In this paper we used the well-known SIFT algorithm to produce a feature descriptor that allows quick and highly discriminatory assessments with other features. SIFT keypoints are detected along with their descriptors, the process was done using the OpenCV library.

Once that exact corresponding points cannot be assumed in the dense point clouds captured by the Kinect, the cost function is highly sensitive to outliers in the approximated correspondences. In order to minimize the possibility of outliers

we used the RANSAC algorithm (Fischler and Bolles, 1981) with a 2D-affine transformation as underlying mathematical model, instead of the usual model, the fundamental matrix estimation, that becomes unreliable because of the very small baselines caused by the data high acquisition rate. The RANSAC-step is followed by a statistical dispersion technique that is called interquartile range (IQR). Given a data set, the IQR goal is to find the difference between the upper and low quartiles (Graham and Ian, 1996). Basically, IQR divide the data in four equal parts, being Q1 and Q3 the median values of low (25%) and upper (75%) quartiles, while Q2 represents the median calculated value (50%), as follows:

$$IQR = Q3 - Q1 \quad (1)$$

where n = number of coordinates

$$Q1 = \frac{1}{4}(n + 1)$$

$$Q3 = \frac{3}{4}(n + 1)$$

The maximum (*OutlierMax*) and minimum (*OutlierMin*) values can be obtained as follows:

$$Outlier_{Max} = Q3 + Coef_{IQR} * IQR \quad (2)$$

$$Outlier_{Min} = Q1 - Coef_{IQR} * IQR \quad (3)$$

where $Coef_{IQR}$ = the half upper value.

The points between *OutlierMax* and *OutlierMin* are inliers, otherwise should be eliminated. These issues are of extreme importance with respect to the quality of the registration process.

For the estimation of initial transformation parameters, the 2D inliers should be transformed to 3D space by using the depth data. The 3D points (x^k, y^k, z^k) are computed as in Khoshelham and Elberink (2012):

$$x_i^k = \frac{z_i^k}{f} x_i^{kv} \quad (4)$$

$$y_i^k = \frac{z_i^k}{f} y_i^{kv} \quad (5)$$

$$z_i^k = \frac{1}{c_0 + c_1 d_i^{kv}} \quad (6)$$

where x_i^{kv}, y_i^{kv} = 2D coordinates for each SIFT matched current images

f = calibrated focal distance

c_0, c_1 = slope and the intercept of the line

d_i^{kv} = denormalized disparity value.

As the shift between the coordinates of the conjugate points in the RGB frame and depth image has a large variance, even when the image coordinates are corrected for the lens distortions, we use the method proposed by Khoshelham et al. (2013) to generate 3D correspondences from the 2D inliers.

The only required inputs are the interior and relative orientation parameters between the RGB and IR cameras, which are previously obtained in a calibration procedure. Then, we realize an image normalization procedure in the RGB frame, and the epipolar geometry is used to transform the visual features to the 3D space (Dos Santos et al., 2016).

3.2 Pairwise registration

Once the corresponding 3D associated points are obtained, the point clouds of two or more successive frames can be registered. In order to speed up the registration step, we use

dual number quaternions including a normalized weighted scheme for the closed-form solution. Walker and Shao (1991) presented a single-cost function to represent rotation and translation, which enables the simultaneous determination of transformation parameters without initial values. Then, first we calculate the quaternion \hat{q} using the equation as:

$$\hat{q} = -\frac{1}{\sum_{i=1}^n \beta_i} \left(\sum_{i=1}^n (W(p'_i) - Q(p_i)) \right) q \quad (7)$$

where, \hat{q} and q are the parts with dual and real coefficients of a dual quaternion, respectively

- n = number of points used for each pair of images
- β_i = normalized weight associated to the pair point factor
- p'_i, p_i = set of corresponding points
- W, Q = product of a right and left quaternion transformation.

According Khoshelham et al. (2013) since the Kinect depth images are captured typically at a frame rate of 20 to 30 fps, we can approximate our observation equations with $v_i = x_{ij-1} - x_{ij}$, for which the weight can be defined inversely proportional to the variance of the observation, as follows:

$$\beta_i^{x^k y^k z^k} = \frac{k}{\sigma_{v_i}^2} = \frac{k}{\sigma_{x_{ij-1}}^2 + \sigma_{x_{ij}}^2} \quad (8)$$

where σ_x^2 = variance of point x and k is an arbitrary constant.

As the theoretical precision of point coordinates x^k, y^k, z^k is different, we have defined the weights of each coordinate separately. Using the equations 4-6 the variance of image measurements can be propagated to the 3D coordinates in object space. For the depth coordinate we have:

$$\sigma_{z^k}^2 = c_1^2 \sigma_d^2 z^{k4} \quad (9)$$

where σ_d = precision of disparity measurements. Doing the same for x^k yields:

$$\sigma_{x^k}^2 = \frac{z^{k2}}{f^2} \sigma_{z^k}^2 + \frac{z^{k2}}{f^2} \sigma_x^2 \quad (10)$$

Substituting equations (5) and (9) in (10) gives us the following equation:

$$\sigma_{x^k}^2 = c_1^2 \sigma_d^2 x^{k2} z^{k2} + \frac{\sigma_x^2}{f^2} z^{k2} \quad (11)$$

And similarly for y^k we can obtain:

$$\sigma_{y^k}^2 = c_1^2 \sigma_d^2 y^{k2} z^{k2} + \frac{\sigma_y^2}{f^2} z^{k2} \quad (12)$$

where σ_x, σ_y = precision of measuring pixel coordinates in x and y respectively.

Using equations (9), (11) and (12) in (8) we can now define the weights for x^k, y^k, z^k coordinates as follows:

$$\beta_i^{z^k} = \frac{p}{z^{k4}_{ij-1} + z^{k4}_{ij}} \quad (13)$$

$$\beta_i^{x^k} = \frac{p}{z^{k2}_{ij-1}(x^{k2}_{ij-1} + q) + z^{k2}_{ij}(x^{k2}_{ij} + q)} \quad (14)$$

$$\beta_i^{y^k} = \frac{p}{z^{k2}_{ij-1}(y^{k2}_{ij-1} + q) + z^{k2}_{ij}(y^{k2}_{ij} + q)} \quad (15)$$

where $p = kc_1^{-2} \sigma_d^{-2}$

c_1 = depth calibration parameter as described in Khoshelham and Elberink (2012)

k = an arbitrary constant

σ_d^2 = variance of the disparity

z = depth information

Note that c_1 is obtained by a simple least-squares linear regression of the disparity-depth relation, and to avoid round-off errors, it is useful to scale the weight values, by tuning the parameter r , and keeping it constant in all pairwise registrations.

Normalizing the weight associated to the pair point yields:

$$\beta_i = \frac{\beta_i^{x^k} + \beta_i^{y^k} + \beta_i^{z^k}}{\|\beta_i^{x^k} + \beta_i^{y^k} + \beta_i^{z^k}\|} \quad (15)$$

Then, we can calculate the translation vector t from the 3D weighting based on the theoretical random error of xyz measurements as

$$t = W(q)^T \hat{q} \quad (16)$$

where $q = [q_0 \ q_1 \ q_2 \ q_3]^T$

Note that, the 3D rotation matrix (R) can be obtained using the q elements.

3.3 RGB-D loop-closure solution and global optimization

The global optimization is achieved using the obtained transformation parameters as the initial alignment along the trajectory. One camera pose is represented as one node in the graph pose, and the edge linking two nodes represents the transformation parameters from one pose to another. For each pairwise registration, the edges are obtained using the loop-closure detection solution. The loop-closure detection that relies on the idea that one interesting event occurs when the sensor has returned to a past position.

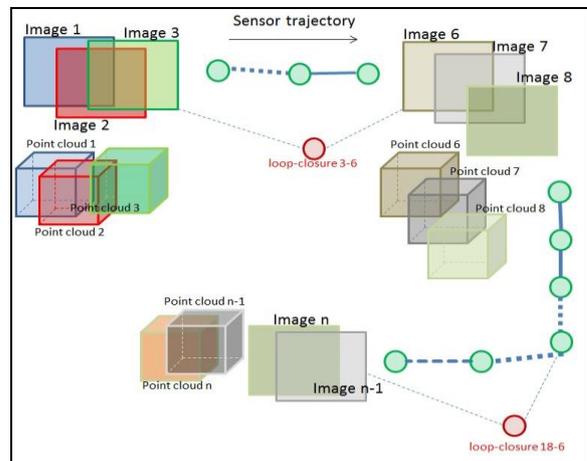


Figure 2. A loop-closure for a sequence of RGB-D data.

Our method addresses the loop-closure problem as an image and depth (RGB-D loop-closure) retrieval task: the algorithm seeking for the past image and point cloud, which looks similar enough to the current one.

In order to select the past frames, that are more likely to match with the current frame, we use the method proposed by Engelhard et al. (2011). In Fig. 2, as the sensor moves through its work space and it creates a sequence of RGB-D data $\vartheta = [D_1, D_2, \dots]$. Then, the loop-closure problem should find two subsequence of ϑ , $U = [u_1, u_2, \dots]$ and $V = [v_1, v_2, \dots]$

being u_i and v_i the index variables (the weighted 3D inliers and the volume of overlap area of a pair of point clouds).

We use a modified form of the proposed method by Ho and Newman (2007). First, the most recent RGB-D data (sorted after a pre-defined number of RGB-D data collected by the sensor) is compared with the previous one. The number of matched inliers, a histogram that represents the cumulative weighted inliers and the volume for the pair of point clouds are all computed. Then, the algorithm proceeds by generating a symmetric matrix D . Each element D_{ij} is the maximal cumulative u_i and v_i score of a pairing of RGB-D data ending with pairing D_i and D_j .

In Fig. 2 the green nodes denote the transformation parameters between the pair of the RGB-D data, the red nodes represent a detected loop-closure, blue dashed edges denote a link with high probability of the current sensor is revisiting a past position based on before mentioned variables. While the fill edges represent a link with high volume value, however with the low number of inliers (probably it found a low textured region, then the ICP algorithm can be applied to solve the registration problem). Moving the sensor from position 0 to n each D_{ij} is described by a probability mass function giving priority to the before mentioned variables as follows. Let H_l be the random variable denoting loop-closure hypotheses at location l : $H_l = n$ is the event that current sensor $RGBD_n$ close the loop with past sensor $RGBD_{n-1}$.

$$\rho(h_n) = h_n/n \quad (17)$$

where n the number of variables, $h_1 = \begin{cases} 1 & \text{if } \sum \beta_i^{x^k} + \beta_i^{y^k} > \text{threshold}_\beta \\ 0 & \text{otherwise} \end{cases}$ denotes the distribution of weighted inliers and

$$h_2 = \begin{cases} 1 & \text{if } \text{height} * \text{width} * \text{length} > \text{threshold}_{vol} \\ 0 & \text{otherwise} \end{cases}$$

represents the volume calculated into overlap area of the point cloud.

In other words, we look for pair of the RGB-D data whose sum of the probabilities is above the threshold. If the motion transformation satisfies the specific threshold constraints, it is thought that there is a valid transformation parameter between the above two observations. Hereafter, the whole pose graph can be optimized using a tree-based network optimizer (Grisseti et al., 2007).

4. EXPERIMENTS

As a proof-of-concept of our proposed method, we achieved a relative accuracy assessment of the obtained results through the root mean square error (rmse) of the mean discrepancies for each pairwise registration into indoor environment and an absolute accuracy with some visual landmarks. The sensor travelled a distance of around 14 m before returning back to a previously visited location. The algorithm is implemented using C++ on Linux. Table 1 summarizes the results obtained using the proposed method.

Experiment	Accuracies (cm)	
	Relative	Absolute
A	14.5	21.1

Table 1. Relative and absolute accuracy obtained with proposed method

The residual distributions were centered approximately at zero, implying that the weighted closed-form solution is appropriate and that the overall adjustment estimation is probably reliable. The assignment of unequal uncertainties to points is essential for registering datasets acquired with RGB-D sensors as errors increase significantly with increasing range. The absolute accuracy is around 1.5% with respect to the expected errors. Figure 3 shows the trajectory of sensor poses maintained before and after loop-closure. Note that the estimated position of the sensor is more than 2 meters off its actual position when it has completed a loop.

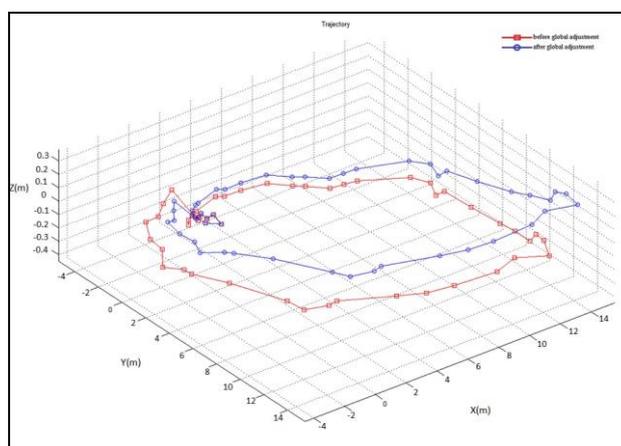


Figure 3. The trajectory of sensor poses maintained before and after loop-closure

In order to evaluate the benefits of RGB-D loop-closure proposed solution, we collected a challenging dataset with less textured surfaces. This is an advantage of our solution - when there is enough overlap between the current and past RGB-D data the volume computed is high, however, keypoints cannot be detected. In this case, the volume value scores significantly enough to trigger up a loop-closure. Then, we used ICP algorithm to compute the transformation parameters. This is a disadvantage of our approach – a certain amount of overlap between the point clouds must exist for less textured regions and we have to use an iterative solution to compute the parameter transformation instead a closed-form solution as proposed in this paper.

5. CONCLUSIONS

When iterative methods are used for pairwise registration, it requires a relatively accurate initial transformation and a high overlap between consecutive point clouds. We investigate how to weight and normalize the 3D points based on the theoretical random xyz errors into a closed-form solution for pairwise registration. The key insights of this investigation are, first, the closed-form solution provides an optimal transformation in one-step. It does not need to calculate good initial estimates and expressively decreases the demand for computer resources in contrast to the iterative methods; second, the xy errors also can

be used for a more efficient loop-closure solution; third, the computed volume value between the current and past RGB-D data also must become more efficient with the the loop-closure detection, mainly into less texturized regions.

The most important characteristic of the RGB-D loop-closure proposed solution is that, because $\beta_i^{x^*}$ and $\beta_i^{y^*}$ are highly influenced by x^* and y^* image coordinates, because of the lens distortions, the dominant weight in the histogram suggests if the sensor is revisiting a past position or not, jointly with both before mentioned information.

In order to avoid incorporate all the information from each RGB-D data into a concise 3D map representation, our algorithm only use low overlap areas based on volume computed.

A drawback of registration by using RGB-D sensors is the influence of low resolution of the depth measurements, lighting condition, properties of object surfaces in the disparity data, sensor noise, occlusions and changes in viewpoint direction. All this factors contribute to achieve low accuracy for 3D mapping (Henry et al., 2012). This emphasizes the importance of more sensors capturing RGB-D data along the trajectory of the sensor.

ACKNOWLEDGEMENTS

This work was supported in part by the CNPq (*Conselho Nacional de Pesquisa e desenvolvimento*) under Grant no. 302644/2013-0”.

REFERENCES

Al-Manasir, K., and Fraser, C.S., 2006, Registration of terrestrial laser scanner data using imagery: The Photogrammetric Record, v. 21, p. 255-268.

Ataer-Cansizoglu, E.; Taguchi, Y.; Ramalingam, S.; Garaas, T., "Tracking an RGB-D Camera Using Points and Planes," Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on , vol., no., pp.51,58, 2-8 Dec. 2013.

Barnea, S., and Filin, S., 2007, Registration of terrestrial laser scans via image based features, ISPRS Workshop on Laser Scanning 2007 and SilviLaser 2007, Volume XXXVI: Espoo, Finland, p. 32-37.

Barnea, S., and Filin, S., 2008, Keypoint based autonomous registration of terrestrial laser point-clouds: ISPRS Journal of Photogrammetry and Remote Sensing, v. 63, p. 19-35.

Bendels, G. H., Degener, P., Wahl, R., Körtgen, M., & Klein, R., 2004. Image-Based Registration of 3D-Range Data Using Feature Surface Elements. In VAST (pp. 115-124).

Besl, P.J., and McKay, H.D., 1992, A method for registration of 3-D shapes: Pattern Analysis and Machine Intelligence, IEEE Transactions on, v. 14, p. 239-256.

Bouguet, J. Y., 2004. Camera calibration toolbox for matlab.

Chen, Y., and Medioni, G., 1992, Object modelling by registration of multiple range images: Image Vision Comput., v. 10, p. 145-155.

Chetverikov, D., Svirko, D., Stepanov, D., & Krsek, P., 2002. The trimmed iterative closest point algorithm. In Pattern Recognition, 2002. Proceedings. 16th International Conference on (Vol. 3, pp. 545-548). IEEE. Chow J., Lichti D., Hol J., Bellusci G., Luinge H., 2014. "IMU and multiple RGB-D camera fusion for assisting indoor stop-and-go 3D terrestrial laser scanning", Robotics: Robot Vision, vol. 3, no. 3, pp.247-280, Jul. 2014.

Dold, C., and Brenner, C., 2006, registration of terrestrial laser scanner data using planar patches and image data, in Maas, H.-G., and Schneider, D., eds., ISPRS Comm. V Symposium "Image Engineering and Vision Metrology", Volume XXXVI Part. 5, 25-27: Dresden, p. 78-83.

dos Santos, D.R., Basso, M. A., Khoshelham, K., de Oliveira, E., Pavan, N. L., Vosselman, G. Mapping Indoor Spaces by Adaptive Coarse-to-Fine Registration of RGB-D Data, in *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 2, pp. 262-266, Feb. 2016. doi: 10.1109/LGRS.2015.2508880

Dryanovski, I., Jaramillo, C., and Jizhong, X., 2012, Incremental registration of RGB-D images, Robotics and Automation (ICRA), 2012 IEEE International Conference on, p. 1685-1690.

Du, H., Henry, P., Ren, X., Cheng, M., Goldman, D. B., Seitz, S. M., & Fox, D., 2011. Interactive 3D modeling of indoor environments with a consumer depth camera. In Proceedings of the 13th international conference on Ubiquitous computing (pp. 75-84). ACM.

Eggert, D.W., Fitzgibbon, A.W., and Fisher, R.B., 1998, Simultaneous registration of multiple range views for use in reverse engineering of CAD models: Comput. Vis. Image Underst., v. 69, p. 253-272.

Ellekilde, L.-P., Huang, S., Valls Miro, J., and Dissanayake, G., 2007, Dense 3D Map Construction for Indoor Search and Rescue: Journal of Field robotics, v. 24, p. 71-89.

Endres, F., Hess, J., Engelhard, N., Sturm, J., Cremers, D., and Burgard, W., 2012, An evaluation of the RGB-D SLAM system, Robotics and Automation (ICRA), 2012 IEEE International Conference on, p. 1691-1696.

Engelhard N., Endres F., Hess J., Sturm J., Burgard W. 2011, "Realtime 3D visual SLAM with a hand-held RGB-D camera", RGB-D Workshop on 3D Perception in Robotics at the European Robotics Forum, Vasteras, Sweden.

Fischler, M. A., & Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), 381-395.

Godin, G., Rioux, M., and Baribeau, R., 1994, Three-dimensional registration using range and intensity information, p. 279-290.

- Grisetti, G., Grzonka S., Stachniss C., Pfaff P., Burgard W., 2007, "Estimation of accurate maximum likelihood maps in 3D", *In Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 65-72.
- G. Grisetti, Kummerle R., Stachniss C.; Burgard W., 2010. "A tutorial on graph-based SLAM", *IEEE Intell. Transp. Syst. Mag.*, vol. 2, no. 4, pp. 31-34, 2010
- Henry, P., Krainin, M., Herbst, E., Ren, X., and Fox, D., 2010, RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments, *Symposium on Experimental Robotics (ISER)*.
- Henry, P., Krainin, M., Herbst, E., Ren, X., and Fox, D., 2012, RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments: *The International Journal of Robotics Research*, v. 31, p. 647-663.
- Huhle, B., Jenke, P., and Strasser, W., 2008, On the fly scene acquisition with a handy multi sensor system: *Int. J. Intell. Syst. Technol. Appl.*, v. 5, p. 255-263.
- Kang, Z., Li, J., Zhang, L., Zhao, Q., and Zlatanova, S., 2009, Automatic Registration of Terrestrial Laser Scanning Point Clouds using Panoramic Reflectance Images: *Sensors*, v. 9, p. 2621-2646.
- Khoshelham, K., 2010, Automated localization of a laser scanner in indoor environments using planar objects, *Indoor Positioning and Indoor Navigation (IPIN)*, 2010 International Conference on, p. 1-7.
- Khoshelham, K., and Elberink, S.O., 2012, Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications: *Sensors*, v. 12, p. 1437-1454.
- Khoshelham, K., dos Santos, D.R. and Vosselman, G., 2013, Generation and weighting of 3D point correspondences for improved registration of RGB-D data. In: *ISPRS Annals Volume II-5/W2 : ISPRS Workshop laser scanning*, 11-13 November 2013, Antalya, Turkey / ed by M. Scaioni et al. Antalya: ISPRS, 2013. ISSN: 2194-9050. pp. 127-132.
- Ho, K. L., & Newman, P., 2007. Detecting loop closure with scene sequences. *International Journal of Computer Vision*, 74(3), 261-286.
- Lowe, D., 2004, Distinctive Image Features from Scale-Invariant Keypoints: *International Journal of Computer Vision*, v. 60, p. 91-110.
- Lu, F., & Milios, E., 1997. Globally consistent range scan alignment for environment mapping. *Autonomous robots*, 4(4), 333-349.
- Masuda, T., and Yokoya, N., 1994, A robust method for registration and segmentation of multiple range images, *CAD-Based Vision Workshop, 1994.*, *Proceedings of the 1994 Second*, p. 106-113.
- May, S., Droschel, D., Holz, D., Fuchs, S., Malis, E., Nüchter, A., & Hertzberg, J., 2009. Three-dimensional mapping with time-of-flight cameras. *Journal of Field Robotics*, 26(11-12), 934-965.
- Okatani, I.S., and Deguchi, K., 2002, A Method for Fine Registration of Multiple View Range Images Considering the Measurement Error Properties: *Computer Vision and Image Understanding*, v. 87, p. 66-77.
- Park, J.-H., Shin, Y.-D., Bae, J.-H., and Baeg, M.-H., 2012, Spatial Uncertainty Model for Visual Features Using a Kinect™ Sensor: *Sensors*, v. 12, p. 8640-8662.
- Park, S.Y., and Murali, S., 2003, A fast point-to-tangent plane technique for multi-view registration, *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*, p. 276-283.
- Prusak, A., Melnychuk, O., Roth, H., Schiller, I., and Koch, R., 2008. Pose estimation and map building with a Time of flight camera for robot navigation: *Int. J. Intell. Syst. Technol. Appl.*, v. 5, p. 355-364.
- Raposo C., Lourenco M., Antunes M., Barreto J. P., 2013, Plane-based Odometry using an RGB-D Camera. In *BMVC*.
- Segal, A., Haehnel, D., and Thrun, S., 2009, Generalized-icp, In *Proc. of Robotics: Science and Systems (RSS)*, Volume 25, 26, 27.
- Stamos, I., and Leordeanu, M., 2003, Automated feature-based range registration of urban scenes of large scale, *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, Volume 2, p. II-555-II-561 vol.2.
- Steinbrucker, F., Sturm, J., and Cremers, D., 2011, Real-time visual odometry from dense RGB-D images, *Computer Vision Workshops (ICCV Workshops)*, 2011 IEEE International Conference on, p. 719-722.
- Swadzba, A., Liu, B., Penne, J., Jesorsky, O., and Kompe, R., 2007, A Comprehensive System for 3D Modeling from Range Images Acquired from a 3D ToF Sensor *The 5th International Conference on Computer Vision Systems: Bielefeld*.
- Taguchi, Y., Jian, Y. D., Ramalingam, D., Feng, C., 2013, Point-Plane SLAM for Hand-Held 3D Sensors. *IEEE International Conference on Robotics and Automation (ICRA)*, p.5182-5189.
- Thrun S., Martin C., Liu Y., Hähnel D., Emery-Montemerlo R., Chakrabarti D., Burgard W., 2003. A real-time expectation maximization algorithm for acquiring multi-planar maps of indoor environments with mobile robots. *IEEE Transactions on Robotics*, 20(3):433-443, 2004.
- Weik, S., 1997, Registration of 3-D partial surface models using luminance and depth information, *3-D Digital Imaging and Modeling. Proceedings.*, *International Conference on Recent Advances in*, p. 93-100.
- Walker M.W., Shao L., Volz R.A., 1991. Estimating 3-D location parameters using dual number quaternions, *CVGIP: Image Understanding*, vol. 54, no. 3, pp. 358-367, 1991.
- Ye, C., and Bruch, M., 2010, A visual odometry method based on the SwissRanger SR4000, p. 769211-769211.
- Zhang, Z., 2000. A flexible new technique for camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(11), 1330-1334.