https://doi.org/10.2352/ISSN.2470-1173.2018.13.IPAS-462

Combining Local and Global Optical Flow for RGB-D Point Cloud Alignment

Sunho Kim, Yo-Sung Ho
Gwangju Institute of Science and Technology (GIST)
123 Cheomdangwagi-ro, Buk-gu, Gwangju, 61005, South Korea
E-mail : {sunhokim, hoyo}@gist.ac.kr

Abstract

3D scene reconstruction using RGB-D camera-based Simultaneous Localization and Mapping (SLAM) is constantly studied today. KinectFusion, GPU-based real-time 3D scene reconstruction framework, is mainly used for many other algorithms of RGB-D SLAM. One of the main limitation of KinectFusion depends only on geometric information in the camera pose estimation process. In this paper, we utilize both geometric and photometric information for point cloud alignment. To extract photometric information in color image, we combine local and global optical flow method, such as Lucas-Kanade and Horn-Schunck, respectively, and make not only dense but also robust to noise flow field. In experimental results, we show that our method can use dense and accurate photometric information.

Keywords: Simultaneous localization and mapping, Iterative closest point, Data association, Optical flow, Lucas-Kanade, Horn-Schunck

1. Introduction

As we can utilize in Augmented Reality, Virtual Reality, and Robotics, 3D object and scene reconstruction techniques are studied continuously. 3D reconstruction is mainly divided two kinds – sparse reconstruction based on feature points, such as Structure from Motion, and dense reconstruction based on depth information, such as Multi-View Stereo. And we can also divide this area in detail based on the objective of reconstruction. Simultaneous Localization and Mapping (SLAM), one of the 3D scene reconstruction technique, can make a 3D scene and localize the position of robot/camera using over hundreds of image frames.

KinectFusion [1] is one of the most commonly used real-time 3D dense reconstruction technique. The main advantage of KinectFusion is that we can utilize depth map captured by Microsoft Kinect, instead of computing specific feature point extraction algorithms in color image. Thus, KinectFusion can make more accurate 3D models than any other SLAM algorithms. In KinectFusion, they use Iterative Closest Point (ICP) method to compute 6 DoF camera pose, find the correspondences between two sequential image frames and remove outliers.

In data association process, however, KinectFusion depends only on geometric information to compute transformation matrix. Recently, many types of research start to utilize not only geometric but also photometric information. Shin and Ho [7] uses the feature of an image based on the SIFT feature extractor to utilize the photometric information. In this method, they gained the weight based on the feature point to get a more accurate registration result, and they show the proposed method decreases the absolute trajectory error and reduces the object drift problem in the reconstructed 3D object model. Like this method, if we make use of the photometric information, we can make more accurate registration result.

Peasley and Birchfield [5] proposed Lucas-Kanade-based data association method. In this method, they made use of the local optical flow method to compute the photometric information. Using the Lucas-Kanade, they compute the local flow field robust to noise and utilize a data association process. Unlike this, Horn-Schunck, global optical flow method, can make denser flow field than the Lucas-Kanade optical flow [2].

In this paper, we propose a novel data association method using the combination of local and global optical flow to improve conventional Lucas-Kanade data association. Our proposed method mainly follows the method proposed by Peasley and Birchfield [5]. In Section 2, we introduce conventional projective data association and Lucas-Kanade data association methods and then propose our improved method in Section 3. Based on our method, we show our experimental results in Section 4.

2. Data association

Basically, ICP algorithm proceeds as the flowchart of Figure. 1. Data association includes point selection, matching, and weighting processes, while alignment includes rejection, error metric, and minimize processes. This Section introduces each process of projective data association and Lucas-Kanade data association, and compare them.

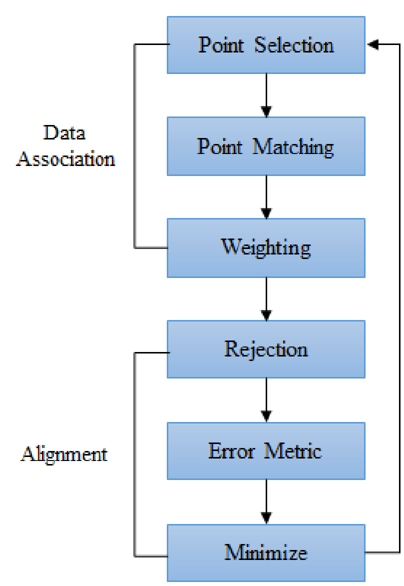

Figure 1. Flowchart of the general iterative closest point algorithm

2.1 Projective data association (PDA)

To perform the data association, we first determine which points from the two clouds to use. In KinectFusion, we use all of the points to perform the dense reconstruction. Two point clouds are used in point matching process to compute the correspondence between the points in the two clouds. First, we project the points in point cloud P onto the camera. Then we can compute projection line and find the closest point in point cloud Q using projection line. The normal of each closest points are used in point-to-plane error metric process. In point matching process, we can get a transformation matrix.

$$T = \begin{bmatrix} R & t \\ \boldsymbol{0}_3^T & \boldsymbol{1} \end{bmatrix} \tag{1}$$

This transformation is 4 x 4 matrix consists of the rotation matrix (R) and the translation vector (t), where 0_3 is 3 x 1 vector of all zeros. After point matching, generally, the weighting process is required. In KinectFusion, however, there is no weighting process for the data association.

The next process removes an outlier. In KinectFusion, we check following two conditions to find the outlier in every corresponding pair,

$$\left\| \hat{p}_i - q_i \right\| > \tau_{distance} \tag{2}$$

$$\frac{\left\| n_i \times \hat{p}_i \right\|}{\left\| \hat{p}_i \right\|} > \tau_{angle} \tag{3}$$

where $\hat{p}_i = [I_3|0_3]T\tilde{p}_i$ is a 3 x 1 vector, I_3 is the 3 x 3 identity matrix, $\tilde{p}_i = [p_i^T \quad 1]^T$ and $\tilde{q}_i = [q_i^T \quad 1]^T$ are the homogeneous coordinates of the points, and $\tau_{distance}$ and τ_{angle} are a predefined distance and angle threshold, respectively.

After the rejection process, we compute error metric using the sum of the Euclidean distance between corresponding points. There are two types of error metric methods: point-to-point and point-to-plane method. We use point-to-plane error metric method in KinectFusion. After that, we perform least square optimization process, then repeat every ICP process iteratively until we align two point clouds.

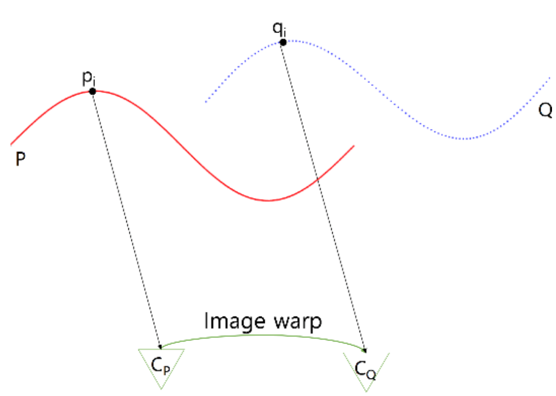

Figure 2. Computing correspondences between two point clouds P and Q.

2.2 Lucas-Kanade data association (LKDA)

Peasley and Birchfield [5] use local optical flow method to use both geometric and photometric information to compute relative transformation. Unlike PDA, which performs entire process iteratively, LKDA method finishes data association process first using Lucas-Kanade optical flow, then perform the alignment process iteratively.

In the point matching, LKDA uses two positions of RGB cameras. Figure. 2 shows the point matching process in LKDA. To match two point clouds P and Q, we first project pi in point cloud P to first camera C_P. Second, projected points are warped to second camera C_Q using specific warping function based on Lucas-Kanade optical flow. Third, we compute projection line from C_Q, then find nearest projection point q_i in point cloud Q. We can represent this process to function below,

$$E_{LK} = \sum_x \sum_y (I_Q(W^{-1}(x, y; \xi)) - I_P(x, y))^2 \tag{4}$$

where $\xi = [r_{xx} \quad r_{xy} \quad r_{yx} \quad r_{yy} \quad a_x \quad a_y]$ is a warping parameter consists of projective warp parameters (r) and affine warp parameters (a). And W is a warping function.

$$W(x, y; \xi) = \begin{bmatrix} x(r_{xx}+1) + yr_{xy} + a_x \\ xr_{yx} + y(r_{xx}+1) + a_y \end{bmatrix} \tag{5}$$

In the rejection process, we have to consider not only distance and angle, but we also consider color similarities between two correspondence points. Color similarity check is performed using the predefined color threshold τ_{color}.

$$\left\| I_Q(W^{-1}(x, y; \xi)) - I_P(x, y) \right\| > \tau_{color} \tag{6}$$

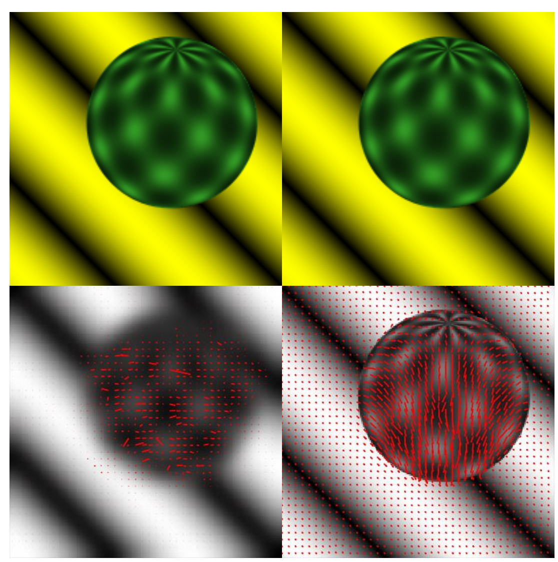

Figure 3. Comparison between Lucas-Kanade (Bottom-left) and Horn-Schunck (Bottom-right) optical flow result using two sequential frames.

3. Proposed method

Bruhn, et. al. [3] proposed an improved method that combines the advantage of local optical flow, robust to noise, and global optical flow, generating dense flow field. In this paper, they combined two optical methods, Lucas-Kanade and Horn-Schunck.

Unlike the Lucas-Kanade, Horn-Schunck method makes an optical flow field throughout the entire pixel of the image. This method can generate denser flow field than Lucas-Kanade method, but it spends more time to compute flow field. Figure. 3 shows the comparison between the results of Lucas-Kanade and Horn-Schunck. The Horn-Schunck method can represent the following function.

$$E_{HS} = \iint ((I_x u + I_y v + I_t)^2 + \alpha |\nabla w|^2) dxdy \quad (7)$$

where $|\nabla w|^2 = |\nabla u|^2 + |\nabla v|^2$. The purpose of Horn-Schunck is to minimize functional, while Lucas-Kanade minimizes quadratic form.

In this paper, we combined Lucas-Kanade and Horn-Schunck to match two point clouds. We modified Horn-Schunck function to replace the first term to Lucas-Kanade function as represented in (4) based on the method proposed by Bruhn, et. al. [3]. Our proposed function as follows.

$$E_{CLG} = \iint ((I_Q(W^{-1}(x,y;\xi)) - I_P(x,y))^2 + \alpha |\nabla w|^2) dxdy \quad (8)$$

Like a Horn-Schunck, the main objective of (8) is to minimize the functional. So minimizing (8) satisfies following two Euler-Lagrange equations.

$$\Delta u = \frac{1}{\alpha}(x(r_{xx}+1) + yr_{xy} + a_x) \quad (9)$$

$$\Delta v = \frac{1}{\alpha}(xr_{yx} + y(r_{yy}+1) + a_y) \quad (10)$$

where $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is the Laplacian operator. (9) and (10) is based on a warping function (5) and warping parameter that used in LKDA. Weighting parameter α is user-defined. Figure. 4 shows a result of 2-dimensional combining local and global optical flow method. In this figure, an input image is degraded by Gaussian noise, as shown in the top-right of Fig. 4.

Figure. 5 shows an entire process of our proposed method. Typical ICP algorithm based on the PDA performs all processes iteratively by selecting and matching point to removing outliers and minimizing the point-to-plane error metrics. But the LKDA-based algorithm finds a rotation and affine warping parameters first based on the optical flow result and the photometric warping function, as shown in Eq. (5), then compute a correspondence map for the point matching process using a warping parameter. Based on a matching result, outlier rejection and error metric minimization processes are performed iteratively.

To perform the proposed method, warping parameter should have computed first. Unlike [5], which computes the warping parameter based on the Lucas-Kanade optical flow only, our method combines the Lucas-Kanade optical flow to the Horn-Schunck, one of the global optical flow method, to compute the warping parameter. Equation (9) and (10) show final equations. That is, we should make the difference between the Laplacian of an optical flow of x-axis and y-axis directions and the warping function approximately 0. So we first calculate the Laplacian of an optical flow, then adjust parameter values to minimize the difference. In this paper, we applied [9] to perform our method.

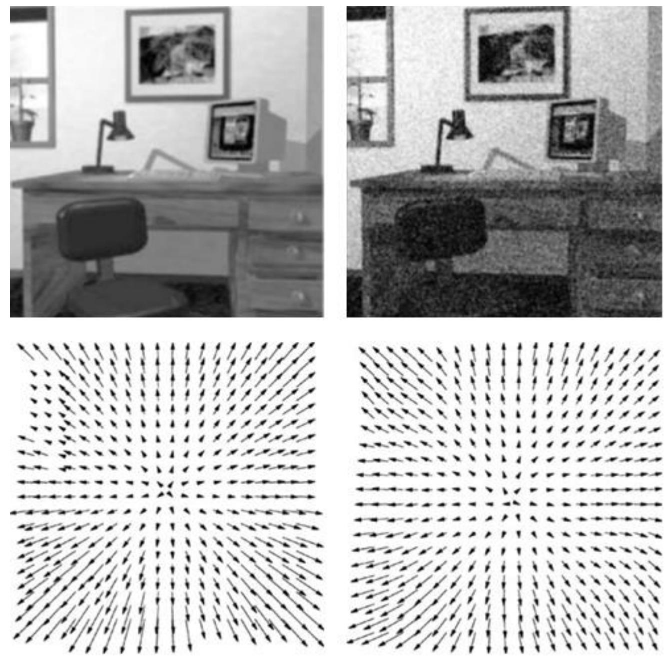

Figure 4. 10 frames of an input image (Top-left), degraded by Gaussian noise with $\sigma_n = 20$ (Top-right), ground truth (Bottom-left), and the result of a combining local and global optical flow method (Bottom-right).

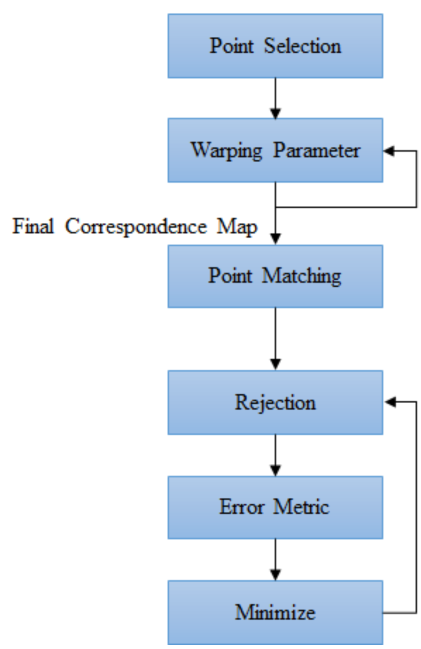

Figure 5. Flowchart of our proposed data association method

The main objective of an optimal warping parameter is to minimize the difference between warped input image based on the warping function and the target image. If not, we compute the variation of the warping parameter Δp using Eq. (11) to update the warping parameter. In Eq. (11), H represents the Hessian matrix, based on the gradient of an input image and the Jacobian of the warping function, as shown in Eq. (12). After calculating the value of Δp, we update the warping parameter, warp the input image again, and compare it to the target image. This process is performed iteratively.

$$\Delta p = \sum_x H^{-1}\left[\nabla I \frac{\partial W}{\partial p}\right]^T [T(x) - I(W(x;p))] \tag{11}$$

$$H = \sum_x \left[\nabla I \frac{\partial W}{\partial p}\right]^T \left[\nabla I \frac{\partial W}{\partial p}\right] \tag{12}$$

$$\mathrm{Proj}^{-1}(x,y,z) = \left[\frac{(x-c_x)}{f_x}z \quad \frac{(y-c_y)}{f_y}z \quad z\right]^T \tag{13}$$

Once the optimal warping parameter was found, we compute the correspondence map based on the warping parameter. In this process, unlike the PDA-based ICP algorithm, we have already found the optimal warping parameter, so we don't need to update the correspondence map iteratively. That is, we can proceed the point matching process only once.

After the data association, an iterative alignment process is required. In this process, similar to previous methods, we remove outliers and minimize the error metric. Previous PDA-based ICP algorithm compares geometric thresholds, such as distance and angle threshold value, to remove outliers. But the proposed method should compare not only geometric thresholds but we also compare the color threshold, because it uses both geometric and photometric information. Therefore, like an LKDA-based method, we should check three inequalities, as shown in Eq. (2), (3), and (6). If one of three inequalities is satisfied, the pair of two points is regarded as an outlier and remove it.

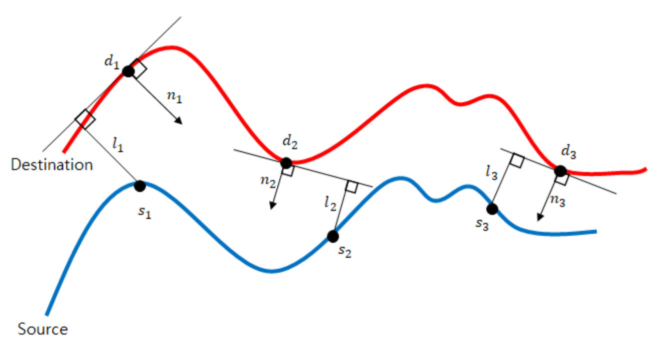

Figure 6. Point-to-plane error between two surfaces.

The remaining process is error metric minimization. Like the KinectFusion, our method applies the linear least-square optimization method to minimize the point-to-plane error metric [4]. Figure 6 briefly shows a point-to-plane error between two surfaces. Our method performs the alignment process iteratively based on the result of a point matching. Algorithm 1 shows the pseudo code of an entire proposed iterative closest point process, where Proj^{-1} is described in Eq. (13).

Algorithm 1. Pseudo code of the proposed method

Select all the point clouds
// Perform the combining local and global optical flow
initialize warping parameter
while iteration number > $\tau_{iteration}$ **do**
  compute *u* and *v*
  compute the Laplacian of *u* and *v*
  $W_x = x(r_{xx}+1)+yr_{xy}+a_x$
  $W_y = xr_{yx}+y(r_{yy}+1)+a_y$
  if $\Delta u = \frac{1}{\alpha}W_x$ and $\Delta v = \frac{1}{\alpha}W_y$ **then**
    break
  else
    calculate Δp
    update warping parameter
  end if
end while
// Make correspondence map
for each pixel *(x,y)* in image domain **do**
  $C_{map}(x,y) = W(x,y;\xi)$
end for
// Compute vertex and normal maps
for each pixel *(x,y)* in the depth map **do**
  $V_P(x,y) = Proj^{-1}(x,y,d_P(x,y))$
  $V_Q(x,y) = Proj^{-1}(x,y,d_Q(x,y))$
  $N(x,y)$ = normal vector of $V_Q(x,y)$
end for
// Compute alignment
while not aligned **do**
  for each pixel *(x,y)* in depth map d_P **do**
    $p = TV_P(x,y)$
    $q = V_Q(C_{map}(x,y))$
    $n = N(C_{map}(x,y))$
    if $\|p-q\| > \tau_{distance}$ **or** $\frac{\|n \times p\|}{\|p\|} > \tau_{angle}$ **or**
    $\|I(x,y) - J(W(x,y;\xi))\| > \tau_{color}$ **then**
      reject the point correspondence
    end if
  end for
  $T^{(k)} \rightarrow$ solve linear system
  Update $T = T^{(k)}\, T^{(k-1)} \ldots T^{(2)}\, T^{(1)}$
end while

4. Experimental results

To implement the proposed method, we utilized an open source implementation version of the KinectFusion, called SLAMBench [6]. SLAMBench is implemented not only the single processor-based C++ language, but it also implemented by the parallel programming library, such as OpenMP, OpenCL, and CUDA. Our method modified a C++-based implementation source code. It is implemented the previous KinectFusion framework, so this code is implemented a projective data association for the iterative closest point. This paper sets the original SLAMBench implementation code as a comparison target, implements a combining local and global optical flow-based data association to the SLAMBench source code, and compares the experimental results. And we also implement a Lucas-Kanade data association to confirm how the proposed method shows similar performance. To evaluate experimental results, we make use of the RGB-D SLAM dataset provided by the Imperial College London [10], as shown in Fig. 7.

Figure 7. ICL-NUIM RGB-D SLAM datasets [10].

Figure 8 shows experimental results of the proposed method. To verify the qualitative evaluation, we use a variety of dataset sequences. In Fig. 8, our method shows a robust result in terms of the color, because our method uses both geometric and photometric information. Based on this, we make a 3D scene reconstruction result, as shown in Fig. 9, and compare the result to the reconstruction result of KinectFusion. As a result, our method makes more robust reconstruction result than the KinectFusion, because our method can remove the more outliers and inaccurate registrations based on the photometric information-based additional constraints.

Table 1 shows a mean and maximum absolute trajectory error (ATE) of the KinectFusion, Lucas-Kanade data association, and proposed method. In this table, we compare quantitative results using four datasets. In KinectFusion, median ATE values ranged from 0.066 to 0.301 and maximum ATE values ranged from 0.236 to 1.038. These error values are relatively higher than the photometric information-based data association method. In the Lucas-Kanade data association, on the other hand, median ATE values ranged from 0.026 to 0.214 and maximum ATE values ranged from 0.045 to 0.567. And a combining local and global optical flow-based data association shows similar quantitative results as the Lucas-Kanade data association. As a result, the proposed method can reduce the ATE compared by the KinectFusion.

5. Conclusion

In this paper, we proposed more efficient iterative closest point algorithm using both geometric and photometric information-based data association method. Previous KinectFusion method makes use of the geometric information only in the projective data association method in point matching process, so it takes a lot of computational costs because this method should perform an entire iterative closest point process in each iteration. To improve this, Peasley and Birchfield [5] makes a correspondence map based on the Lucas-Kanade local optical flow, then just perform the point cloud alignment process iteratively. In this process, warping parameter is required. The warping parameter is updated iteratively by minimizing the difference between the optical flow value and the result of warping function value. In the proposed method, we combine the previously proposed local optical flow equation into a global optical flow equation called Horn-Schunck. In this way, we updated the warping parameters while considering the regularization term.

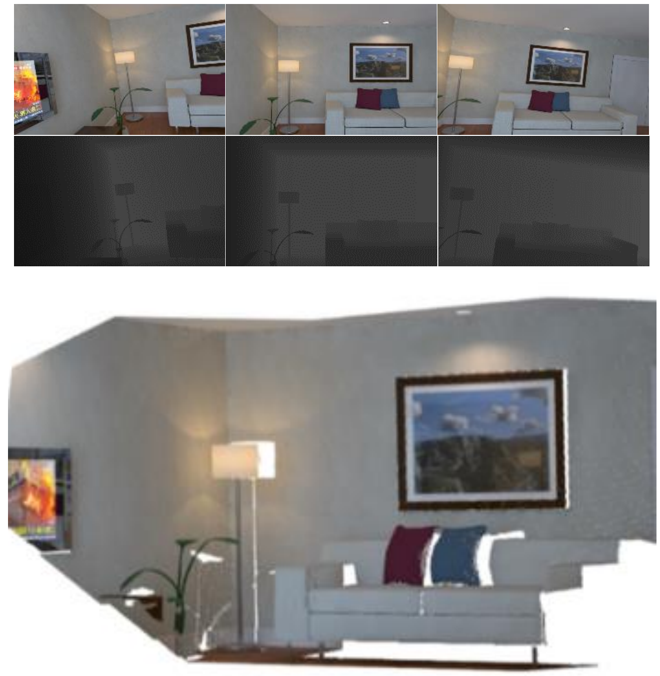

Figure 8. Point cloud registration results using the proposed method and ICL-NUIM datasets.

Table 1. Comparison of the quantitative results

		living room 0	living room 1	living room 2	living room 3
Frames		1510	967	882	1242
Volume size(m^3)		4.9^3	5.2^3	4.8^3	5.0^3
Initial pose		0.3391 0.4931 0.3700	0.5301 0.3439 0.6012	0.3400 0.5000 0.2400	0.2685 0.5012 0.4000
Integration rate		1	1	1	2
PDA [1]	ATE (avg)	0.036	0.008	0.021	0.070
	ATE (max)	0.865	0.015	0.049	0.191
LKDA [5]	ATE (avg)	0.018	0.005	0.014	0.059
	ATE (max)	0.424	0.012	0.030	0.089
Proposed method	ATE (avg)	0.016	0.006	0.013	0.063
	ATE (max)	0.419	0.012	0.027	0.087

Experimental results show that our method makes a similar result as the previous Lucas-Kanade data association method. In table 1, we compare the ATE values. Our method reduces the mean of ATE value approximately 0.02 in average and maximum of ATE value approximately in half respectively compared by the KinectFusion.

Our remaining issues as follow. First, the proposed method makes a slower result than the previous Lucas-Kanade data association method, because our method combines the global optical flow, which is slower than the local method. Therefore, improving the computational speed is one of the remaining issues. Second, our method is highly depended on a local optical flow. To utilize the global optical flow properly, we should adjust the weight of a regularization term. Third, our method just adapted the previous point-to-plane error metric minimization method, but there are several works to improve the previous method. For example, automatic selection of the error metric between the point-to-point and point-to-plane way. So our method will improve the error metric minimization using this kind of method.

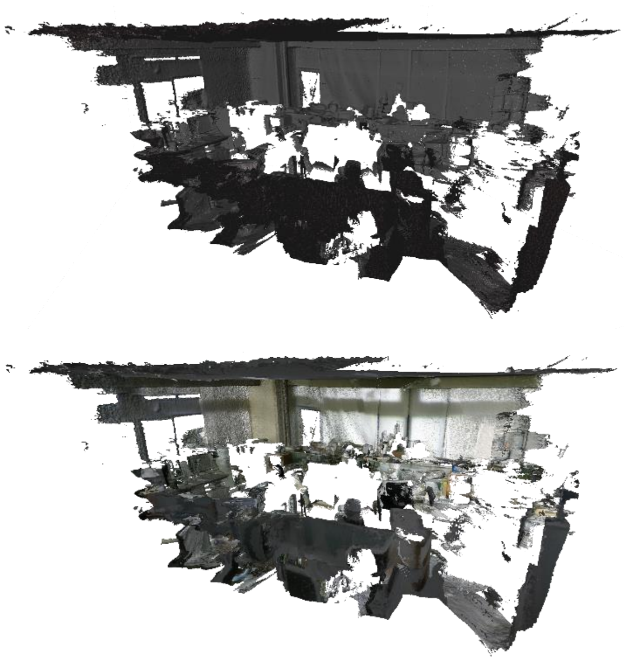

Figure 9. 3D reconstruction result using the proposed method

Acknowledgement

This work was supported by 'The Cross-Ministry Giga KOREA Project' grant funded by the Korea government(MSIT) (GK17C0100, Development of Interactive and Realistic Massive Giga- Content Technology)

References

[1] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon, "KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera," Proceedings of the 24th annual ACM symposium on User interface software and technology. ACM, pp. 559-568, October 2011.

[2] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of Optical Flow Techniques," International Journal of Conputer Vision, Vol. 12, No. 1, pp. 43-77, February 1994.

[3] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade meets Horn/Schunck: Combining Local and Global Optic Flow Methods," International journal of computer vision, Vol. 61, No. 3, pp. 211-231, 2005.

[4] K. L. Low, "Linear Least-Squares Optimization for Point-to-Plane ICP Surface Registration," Chapel Hill, University of North Carolina, pp. 1-3, February 2004.

[5] B. Peasley and S. Birchfield, "Replacing Projective Data Association with Lucas-Kanade for KinectFusion," Robotics and Automation (ICRA), 2013 IEEE International Conference on. IEEE, pp. 638-645, May 2013.

[6] L. Nardi, B. Bodin, M. Z. Zia, J. Mawer, A. Nisbet, P. H. Kelly, A. J. Davidson, M. Lujan, M. F. O'Boyle, G. Riley, N. Topham, and S. Furber, "Introducing SLAMBench, a Performance and Accuracy Benchmarking Methodology for SLAM," Robotics and Automation (ICRA), 2015 IEEE International Conference on. IEEE, pp. 5783-5790, May 2015.

[7] D. W. Shin and Y. S. Ho, "Iterative Closest Points Method based on Photometric Weight for 3D Object Reconstruction," Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2016 Asia-Pacific. IEEE, pp. 1-4, December 2016.

[8] G. Blais and M. D. Levine, "Registering Multiview Range Data to Create 3D Computer Objects," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 8, pp. 820-824, August 1995.

[9] S. Baker and I. Matthews, "Equivalence and Efficiency of Image Alignment Algorithms," Proceedings of the Computer Vision and Pattern Recognition, Vol. 1, pp. 1-8, December 2001.

[10] A. Handa, T. Whelan, J. McDonald, and A. J. Davison, "A Benchmark for RGB-D Visual Odometry, 3D Reconstruction and SLAM," Robotics and Automation (ICRA), 2014 IEEE International Conference on. IEEE, pp. 1524-1531, June 2014.

Author Biography

Sunho Kim received his B.S. degree in computer engineering from Hanbat National University, Korea in 2015. He is currently working towards his M.S. degree in the school of electrical engineering and computer science at Gwangju Institute of Science and Technology (GIST), Korea. His research interests include 3D computer vision, computer graphics, 3D scene reconstruction, simultaneous localization and mapping (SLAM), and augmented reality (AR).

Yo-Sung Ho received his B.S. and M.S in electronic engineering from the Seoul National University, Seoul, Korea (1981) and his Ph.D. in electrical and computer engineering from the University of California, Santa Barbara (1990). He worked in Philips Laboratories from 1990 to 1993. Since 1995, he has been with the Gwangju Institute of Science and Technology, Gwangju, Korea, where he is currently a professor. His research interests include image analysis, 3D television, and digital video broadcasting.