TIME-OF-FLIGHT SENSOR DEPTH ENHANCEMENT FOR AUTOMOTIVE EXHAUST GAS

Tomonari Yoshida[1,2], *Oliver Wasenmüller*[2], *Didier Stricker*[2,3]

[1] DENSO CORPORATION, Kariya, Japan
[2] DFKI - German Research Center for Artificial Intelligence, Kaiserslautern, Germany
[3] University of Kaiserslautern, Kaiserslautern, Germany
{tomonari.yoshida, oliver.wasenmueller, didier.stricker}@dfki.de

ABSTRACT

The Time-of-Flight (ToF) sensor has been envisioned as a candidate of next generation sensors for intelligent vehicles. One of the problems in automotive environment is that the sensor outputs wrong values if exhaust gas exists in the scene. In this paper, we provide two new contributions to the signal processing aspects of the ToF sensor for automotive use. First, we present the sensor characteristics and models for exhaust gas to cope with them. Second, we develop a depth enhancement algorithm to reject the influence of exhaust gas from multiple images. Experimental results demonstrate the effectiveness of the depth enhancement algorithm for both static data (including ground truth) and on-vehicle data acquired by the sensor mounted on a car.

Index Terms— Time-of-Flight sensor, exhaust gas, depth enhancement, Gaussian Mixture Models, image processing

1. INTRODUCTION

In recent years, the Time-of-Flight (ToF) sensor has been receiving greater attention as being one of prospective sensors for future intelligent vehicles. In order to mount the sensor on a vehicle, it must be small due to limited available space. However, there is generally a trade-off between depth accuracy of the sensor and its size. High accuracy is yielded by high power illumination that the sensor needs huge cooling devices to emit [1–3].

One solution for this problem is to enhance the ratio by using sequential depth images from the point of view of image-processing. In this field, the various existing methods are widely known as the solutions of the depth enhancement problem [4–10]. The solutions can be subdivided into two categories: reconstruction with voxel space like KinectFusion [4] and accumulation in image space [5–7, 10]. Wasenmüller et al. [6] propose an image fusion algorithm using per-pixel median to reduce the random noise. Cui et al.'s approach [5] finds optimal surfaces by a combination of ToF super-resolution approach [7] with a ToF-specific probabilistic method. From several experimental results, the effectiveness for indoor scenes has been reported [11, 12].
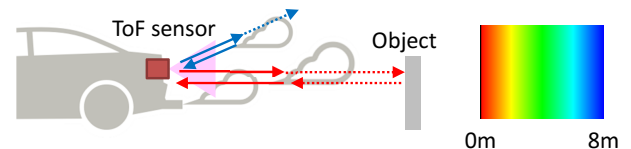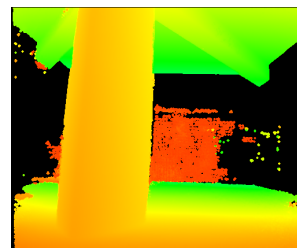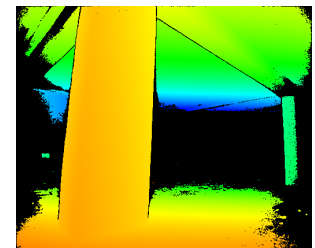

(a) Problem
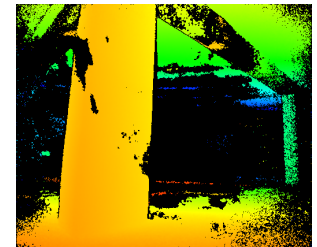
(b) Target scene (c) Input

(d) KinectFusion [4] (e) Proposed

Fig. 1. Automotive scene including exhaust gas: (a) ToF sensor problem for gas (b) a reference image from a RGB camera; (c) an example of input depth images from the Kinect v2 sensor; (d,e) are created by each method for 100 depth images.

However, while these algorithms consider artifacts in indoor scenes [5, 13], they do not consider influences in automotive environments like rain, snow and exhaust gas. In this paper, we focus on exhaust gas. Figure 1(c) and supplemental video represent exhaust gas influences. Figure 1(d) shows a result of KinectFusion [4] as one of the state-of-the-art algorithms. The algorithm outputs some wrong depth values, which orange pixels of Fig. 1(d) represent 0.5-1m from the sensor, despite the fact that there are no objects. This error is caused by accumulating incorrect depth values

due to the measurement error that exhaust gas is recognized as objects by reflecting the irradiation light from the sensor as shown in Fig. 1(a). In addition, the voxel-based method suffers from the limitation of the voxel volume as shown in Fig. 1(d). Therefore, we follow image-based methods. Although several state-of-the-art algorithms use joint bilateral filter [14, 15], they are not suited to remove exhaust gas.

In this paper, we propose a novel depth image processing algorithm to reject the influence of exhaust gas by using characteristics of ToF sensor outputs. On the basis of the analysis for exhaust gas and other rigid objects, we present a model based on Gaussian Mixture Models (GMM) [16] to cope with the sensor characteristics and the solution based on Expectation Maximization (EM) algorithm [17].

2. PROPOSED METHOD

2.1. Basic idea based on analysis of sensor output

According to the principle of the sensor, it outputs two images at the same time without any optical axis alignment, an infrared intensity image and a depth image [1, 2]. We present their relationship and distributions to integrate the characteristics into our reconstruction algorithm. Figure 2 shows two experimental results to clarify the behavior of the ToF sensor. In these experiments, we use the Kinect v2 sensor [18].

The first experiment data consists of depth and infrared images with the same conditions while changing only the distance between the sensor and a target from 0.5m to 6.5m in 0.5m steps. The data has 500 images for each distance. As illustrated in Fig. 2(a), the depth accuracy is inversely proportional to the intensity and has a slight offset. Therefore, we extend an equation of the depth deviation in [1] to

$$\Delta L(u) = \alpha \cdot \frac{\sqrt{B(u) + I(u)}}{I(u)} + \beta, \qquad (1)$$

where α is a constant value including sensor modulation frequencies, $B(u)$ is a term for background light and other noises, $I(u)$ is an intensity at a pixel $u = (x, y)$, β is an offset value ($\alpha = 11$, $B(u) = 30000$, $\beta = 1$ in all experiments).

The second experiment data consists of 30 datasets that each dataset has 100 depth and infrared images to clarify the depth distribution in the scene including exhaust gas. For each dataset, we calculate the histogram of time-direction at the same points of a target object. As illustrated in Fig. 2(c,d), since the distribution without exhaust gas follows a normal distribution, we can enhance the depth accuracy by calculating a mean value for multiple images. On the other hand, the distribution with exhaust gas does not follow a normal distribution because of the fluctuation of the gas. Therefore, we can reject influences of exhaust gas by focusing attention on the distribution and detecting exhaust gas. In order to cope with these distributions simultaneously by one model, we utilize Gaussian Mixture Models:
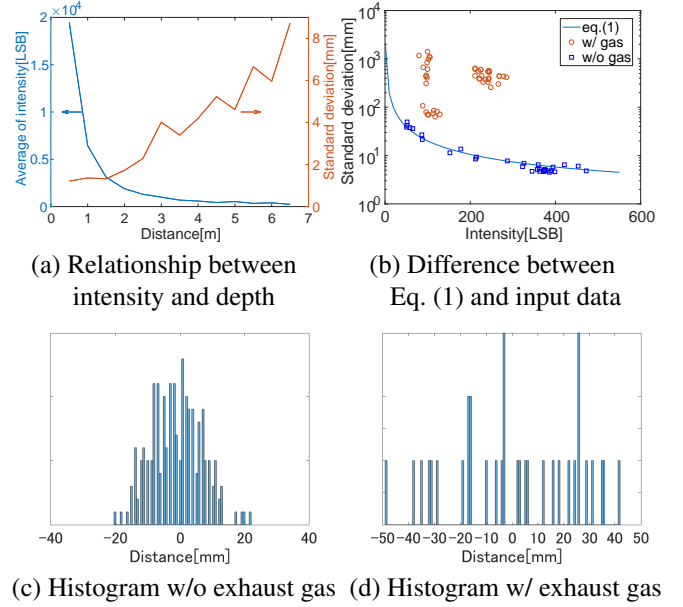

(a) Relationship between intensity and depth

(b) Difference between Eq. (1) and input data

(c) Histogram w/o exhaust gas (d) Histogram w/ exhaust gas

Fig. 2. Analysis of the sensor output

$$GMM = \sum_{k=1}^{K} \pi_k \mathcal{N}(x \backslash \mu_k, \Sigma_k), \qquad (2)$$

where $\mathcal{N}(x \backslash \mu_k, \Sigma_k)$ is a normal distribution with a mean μ_k, a variance Σ_k. π_k is a mixture weight at a k-th distribution. Since GMM is the sum of normal distributions, we can distinguish exhaust gas by considering the adaptability to the model and can accumulate depth information simultaneously.

Our goal is to estimate accurate depth information from multiple depth images acquired with automotive movement. Therefore, the proposed method is composed of two processes; the alignment process and the accumulation process, like the state-of-the-art algorithms [5, 6]. In the alignment process, we use the relationship between the intensity and depth accuracy to perform precise registration. In the accumulation process, exhaust gas is rejected by considering the relationship and the depth distribution. Algorithm 1 summarizes the proposed method.

2.2. Alignment considering infrared intensity

In order to align between input depth images to compensate position aberrations by car movement, we use Iterative Closest Point (ICP) [19]. For the accurate alignment, we introduce a simple and effective pre-filter for input sources. Since the relationship between the infrared intensity and the depth accuracy is presented by Eq. (1), we define the pre-filter extended voxel grid as used in [8] :

$$V_s(q) = \begin{cases} 0 & if \ \min(I_q(v)) < \delta \\ centroid & else \end{cases}, \qquad (3)$$

where $V_s(q)$ is a voxel value of q-th, s is scale factor of voxel, $I_q(v)$ is values of $I(v)$ and v represents 2D points corresponding to 3D points within a voxel q. This pre-filter leads to high accurate alignment by ignoring points with large dispersion of depth values due to high-dynamic ranges of depth accuracy caused by materials composed of various reflectivity and different ambient light in outdoor scene [20, 21]. Although this is a simple algorithm, it achieves comparable results to an conventional algorithm. Moreover, it can be applied to any state-of-the-art registration algorithms without any changes because of the pretreatment for input sources.

2.3. Accumulation by Gaussian Mixture Models

We can classify sensor outputs for time-direction at a pixel into three cases,
  (i) reflection from a solid object
  (ii) reflection from a solid object and gas
  (iii) reflection from gas.
In the case (i), we can calculate the mean value for each pixel in order to enhance the depth accuracy because of the characteristic that the outputs follow a normal distribution. In the case (ii), the proposed method performs to distinguish the solid object and gas simultaneously, and to estimate the mean value only for the solid object. In the case (iii), we can reject the pixel by performing distribution analysis.

The proposed method is composed of two steps. First, Equation (1) is used to distinguish the cases (i) and (ii)(iii) since a distribution included gas does not follow the equation as shown in Fig. 2(b). Second, Equation (2) is integrated into the process to cope with the cases (ii) and (iii). Therefore, the final depth map $R(\cdot)$ is

$$R(u) = \begin{cases} \bar{\mu} & if \ |\bar{\sigma} - \Delta L(u)| < \epsilon \\ G(u) & else \end{cases}, \quad (4)$$

where $\bar{\mu}, \bar{\sigma}$ is the mean and the standard deviation for input depth values, ϵ is a scholar value ($\epsilon = 10$), $G(u)$ presents the case (ii)(iii) and is calculated on the basis of the estimation of GMM. The estimation is performed by calculating following equations alternatively,

E step:

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(x_n \backslash \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n \backslash \mu_j, \Sigma_j)}$$

$$\ln p(X \backslash \pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n \backslash \mu_k, \Sigma_k) \quad (5)$$

M step:

$$\mu_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) x_n$$

$$\Sigma_k^{new} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k^{new})(x_n - \mu_k^{new})^T$$

$$\pi_k^{new} = \frac{N_k}{N}, \quad (6)$$

Algorithm 1 Depth processing for an automotive scene

for each infrared image I_i and depth image D_i, $i \in N$ **do**
    $V_{si} \Leftarrow I_i, D_i$
    $I_{j \leftarrow i}, D_{j \leftarrow i} \Leftarrow ICP(V_{si}, V_{sj})$
end for
For pre-processing, we apply median filter [22]
for each pixel set $u_p, p \in$ the number of pixels **do**
    if $|\bar{\sigma} - \Delta L(u_p)| < \epsilon$ **then**
        $R(u) \Leftarrow \bar{\mu}$ ················· (case (i))
    else if $\ln p(X \backslash \pi, \mu, \Sigma) < \zeta$ **then**
        $R(u) \Leftarrow 0$ ················· (case (iii))
    else if $\arg \max_{k \in K} \mu_k \cap F_k < \epsilon$ **then**
        $R(u) \Leftarrow \mu_k$ ················· (case (ii))
    end if
end for

---

where $\gamma(z_{nk})$ is the membership weight of depth value x_n in cluster k, $N_k = \sum_{n=1}^{N} \gamma(z_{nk})$, N is the number of input depth values, K is the number of normal distributions. Since our goal of this estimation is to separate the cases (ii)(iii) and to estimate the correct value in the case (ii), $K = 2$ is used. The log-likelihood function of Eq. (5) depends on whether the estimated model fits to the data. In this case, the value of the function generated from sources in the case (ii) is relatively higher than the case (iii), since a distribution of depth values for a solid object follows a normal distribution while a distribution for gas does not follow a normal distribution. In order to estimate the correct value in the case (ii), we calculate differences between the estimated standard deviation and Eq. (1). By using the membership weight $\gamma(z_{nk}) > 0.8$, depth values can be allocated to either class k. Physically, when the sensor records both gas and a solid object, gas is close to the sensor against a solid object. Therefore, $G(\cdot)$ is

$$G(u) = \begin{cases} 0 & if \ \ln p(X \backslash \pi, \mu, \Sigma) < \zeta \\ \mu_k & else \ \arg \max_{k \in K} \mu_k \cap F_k < \epsilon \\ & F_k = \left| \sqrt{\Sigma_k} - \Delta L_k(u) \right| \end{cases}, \quad (7)$$

where ζ is a scholar value ($\zeta = 10$), $\Delta L_k(u)$ represents $\Delta L(u)$ corresponding to k-th class.
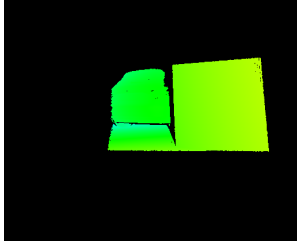
3. EXPERIMENTAL RESULTS

We evaluate performances of the alignment process and the accumulation process individually in order to obtain meaningful results. Moreover, we present the supplemental video for real on-vehicle data.
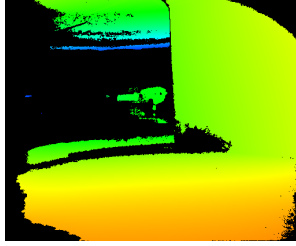
Alignment evaluation: We make use of a public dataset, CoRBS [23], which provides the combination of real depth and infrared intensity data from Kinect v2 sensor [18] with a ground truth trajectory of the camera and a ground truth 3D model of the scene. We compare our proposed method against voxel grid as used in [8]. ICP [19] is used as the align-
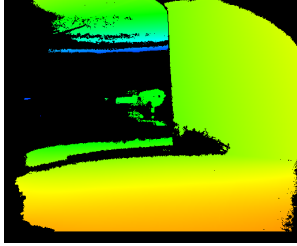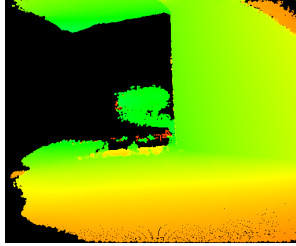
(a) Target scene

	ICP w/ Voxel Grid	ICP w/ Ours ($\delta = \acute{\delta} \times 10^4$)		
		$\acute{\delta} = 1$	$\acute{\delta} = 1.5$	$\acute{\delta} = 2$
E1(s=0.5)	0.185	0.168	0.149	**0.125**
E1(s=0.1)	0.173	0.172	0.153	**0.131**
D1(s=0.5)	0.136	0.104	0.076	**0.074**
D1(s=0.1)	0.127	0.097	0.077	**0.073**
H1(s=0.5)	0.177	0.123	0.104	**0.070**
H1(s=0.1)	0.132	0.098	0.096	**0.087**

Table 1. RMSE for CoRBS dataset [23]. For each object, we randomly select 10 dataset that consists of sequential 50 images. The center image is selected as the target.

Fig. 4. RMSE for the number of frames.
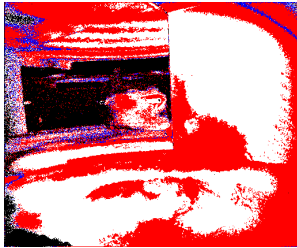

(b) Ground truth

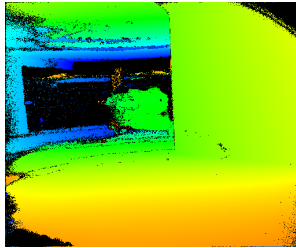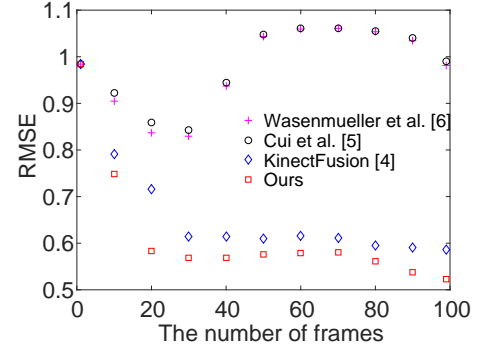(c) Wasenmüller et al. [6]

(d) Cui et al. [5]

(e) KinectFusion [4]

(f) Case map

(g) Ours

Fig. 3. Accumulation results for 100 images included exhaust gas. Although the region of (b) is limited due to the limitation of the measuring environment, the region includes influences of exhaust gas. (f) is the case map of our method, where white pixels are the case (i), red pixels are the case (ii), blue pixels represent the case (iii) or points that the sensor does not provide enough information to estimate the distribution.

ment method. Table 1 shows the results, where the evaluation metric is the root mean squared error (RMSE) of translational components. The proposed filter demonstrates good performance compared to voxel grid, where especially the RMSE is reduced with the increase of δ.

Accumulation evaluation: Our dataset is composed of Kinect v2 sensor depth and infrared images and ground

truth by a precise 3D scanner using structured light [24]. The scene is a parking area with assuming automotive parking scenario. The sensor is fixed at the same pose to isolate this accumulation evaluation from the alignment effect. We compare our proposed accumulation against state-of-the-art algorithms [4–6]. As shown in Fig. 4, our algorithm essentially reduces RMSE compared to other methods. Figure 3 visualizes that the comparative methods is not able to accumulate the several parts correctly. Focusing on the center of the images, Fig. 3(c) and 3(d) shows the black pixels that the value is zero. In Fig. 3(e), some pixels of the region output wrong depth values which are closer than the actual depth. These errors are caused by the high dynamic range of the depth values due to exhaust gas. On the other hand, our method is able to accumulate those parts correctly.

4. CONCLUSIONS

In this paper, we consider the depth enhancement problem in automotive scene with exhaust gas. We present a novel approach integrated with the characteristics of the ToF sensor, focusing on the relationship between infrared images and depth images and on the depth distribution. Experimental results show that the proposed method provides precise depth information for automotive scenes compared to state-of-the-art algorithms. Our future work will extend this framework for other automotive problems such as rain, snow and more complex environment including multiple moving objects.

5. REFERENCES

[1] R. Lange and P. Seitz, "Solid-state time-of-flight range camera," *IEEE Journal of Quantum Electronics*, vol. 37, no. 3, pp. 390–397, 2001.

[2] A. Kolb, E. Barth, R. Koch, and R. Larsen, "Time-of-flight cameras in computer graphics," *Computer Graphics Forum*, vol. 29, no. 1, pp. 141–159, 2010.

[3] P. Fürsattel, S. Placht, M. Balda, C. Schaller, H. Hofmann, A. Maier, and C. Riess, "A comparative error analysis of current time-of-flight sensors," *IEEE Transactions on Computational Imaging*, vol. 2, no. 1, pp. 27–41, 2016.

[4] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2011, pp. 127–136.

[5] Y. Cui, S. Schuon, D. Chan, S. Thrun, and C. Theobalt, "3D shape scanning with a time-of-flight camera," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 1173–1180.

[6] O. Wasenmüller, M. Meyer, and D. Stricker, "Augmented reality 3D discrepancy check in industrial applications," in *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2016, pp. 125–134.

[7] S. Schuon, C. Theobalt, J. Davis, and S. Thrun, "Lidar-Boost: Depth superresolution for ToF 3D shape scanning," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 343–350.

[8] T. Whelan, M. Kaess, M.F. Fallon, H. Johannsson, J.J. Leonard, and J. McDonald, "Kintinuous: Spatially Extended KinectFusion," in *RSS Workshop on RGB-D: Advanced Reasoning with Depth Cameras*, 2012.

[9] L. Sheng, K. N. Ngan, and S. Li, "Temporal depth video enhancement based on intrinsic static structure," in *IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 2893–2897.

[10] R. Maier, J. Stückler, and D. Cremers, "Super-resolution Keyframe Fusion for 3D Modeling with High-Quality Textures," in *International Conference on 3D Vision (3DV)*, 2015, pp. 536–544.

[11] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, pp. 1437–1454, 2012.

[12] D. Anderson, H. Herman, and A. Kelly, "Experimental characterization of commercial flash ladar devices," in *International Conference of Sensing and Technology*, 2005, pp. 17–23.

[13] O. Wasenmüller, M. D. Ansari, and D. Stricker, "Dnaslam: Dense noise aware slam for tof rgb-d cameras," in *Asian Conference on Computer Vision Workshop (ACCV workshop)*, 2016.

[14] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint bilateral upsampling," *ACM Transactions on Graphics (TOG)*, vol. 26, no. 3, 2007.

[15] A. Vianello, F. Michielin, G. Calvagno, P. Sartor, and O. Erdler, "Depth images super-resolution: An iterative approach," in *IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 3778–3782.

[16] G.J. McLachlan and K.E. Basford, *Mixture Models: Inference and Applications to Clustering*, Marcel Dekker, Inc., 1988.

[17] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., 2006.

[18] Microsoft, "Kinect v2," http://www.microsoft.com/en-us/kinectforwindows/.

[19] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *Third International Conference on 3D Digital Imaging and Modeling (3DIM)*, 2001, pp. 145–152.

[20] D. Falie and V. Buzuloiu, "Wide range time of flight camera for outdoor surveillance," in *Microwaves, Radar and Remote Sensing Symposium (MRRS)*, 2008, pp. 79–82.

[21] W. Kazmi, S. Foix, G. Alenyá, and H. J. Andersen, "Indoor and outdoor depth imaging of leaves with time-of-flight and stereo vision sensors: Analysis and comparison," *Journal of Photogrammetry and Remote Sensing*, vol. 88, pp. 128 – 146, 2014.

[22] I. Pitas and A. N. Venetsanopoulos, "Order statistics in digital image processing," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1893–1921, 1992.

[23] O. Wasenmüller, M. Meyer, and D. Stricker, "CoRBS: Comprehensive rgb-d benchmark for slam using kinect v2," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016.

[24] 3Digify, http://3digify.com/.