IMPROVING TIME-OF-FLIGHT SENSOR FOR SPECULAR SURFACES WITH SHAPE FROM POLARIZATION

Tomonari Yoshida[1,2], *Vladislav Golyanik*[2,3], *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, vladislav.golyanik, oliver.wasenmueller, didier.stricker}@dfki.de

ABSTRACT

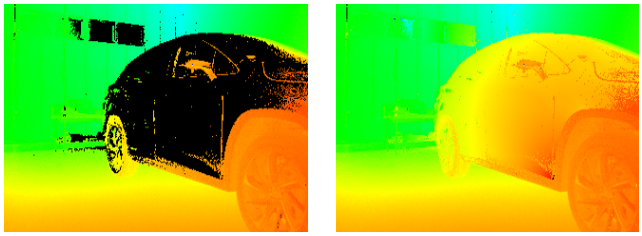
Time-of-Flight (ToF) sensors can obtain depth values for diffuse objects. However, the essential problem is that the sensor can not receive active light from specular surfaces due to specular reflections. In this paper, we propose a new depth reconstruction framework for specular objects that combines ToF cues and Shape from Polarization (SfP). To overcome the ill-posedness of SfP with a single view, we integrate superpixel segmentation with planarity constraints for every superpixel. Experimental results demonstrate the effectiveness of the depth reconstruction algorithm for both controlled environment data and real vehicle data in a parking area.

Index Terms— Time-of-Flight sensor, specular surface, single view reconstruction, Shape from Polarization, superpixel segmentation

1. INTRODUCTION AND RELATED WORK

Time-of-Flight (ToF) sensors have been investigated to adapt automotive environments [1]. The ToF sensor acquires depth by calculating either the time or the phase difference between emitted and received light (direct or indirect ToF). A common problem of these two types is that the sensor can not receive active light from specular surfaces due to specular reflections (Fig. 1(a)). Our goal is to enhance the ToF sensor to estimate depth of specular surfaces.

To achieve the purpose, we utilize a combination of active and passive light, since the sensor is able to receive passive light from specular surfaces, too. Although conventional stereo algorithms [2] reconstruct them by two different views, the stereo camera must have a huge baseline. Multi-view reconstruction like Structure from Motion [3] works well if there are adequate movements. Performing photometric stereo [4] is difficult in automotive scenes due to changing illumination. Shape from Shading [5] reconstructs relative shape from a single image by assuming diffuse reflection. Shape from Polarization (SfP) [6–11] yields relative shape by integrating normals, which are obtained from the Fresnel equations. Inserting micro linear polarizers [12] in front of the photo-receiver chip, it is possible to manufacture a compact ToF sensor which is able to record both ToF depth and polarization cues of a scene.



(a) Raw ToF depth (b) Our enhanced depth

Fig. 1. In this paper, we propose a new algorithm that transforms (a) a raw ToF depth image with huge holes due to specular surfaces to (b) a fully reconstructed depth image utilizing polarizations cues.

In detail, our contributions in this paper are

- to propose a framework reconstructing depth for specular surfaces by combining absolute depth (ToF) and relative shape with polarization cues (SfP),
- to introduce a superpixel segmentation approach to SfP so that normal estimation can be performed by a single view, which handles π-ambiguity correction (Sec. 2.2) and refractive index estimation (Sec. 2.3),
- to evaluate the framework in both controlled environment and real outdoor environment, with a single ToF sensor and a linear polarizer (Sec. 3).

2. PROPOSED METHOD

The main idea of the proposed approach is to propagate few absolute depth values within the specular surface with the support of polarization cues to the whole specular surface (e.g., suppose that depth values in the edge of the door in Fig. 1(a) are propagated to the whole car). An overview of our method is depicted in Fig. 2.

2.1. Ill-posedness of Shape from Polarization

In this section, we briefly explain the principle of SfP. More details can be found in [6, 7]. SfP is divided into two main parts: normal calculation and shape reconstruction by normal integration. The normal is defined by an azimuth angle ψ, which represents an orientation of a plane of incidence, and a zenith angle θ, which represents an angle between the normal and a view axis of a camera. To obtain the normals for each pixel, we change an angle of a linear polarizer mounted in front of the camera and acquire multiple polarization images. A pixel value of the images is defined as

$$I(\Phi) = \frac{I_{max} + I_{min}}{2} + \frac{I_{max} - I_{min}}{2} cos(2\Phi - 2\phi), \quad (1)$$

where Φ is the angle of the linear polarizer, ϕ is a phase angle of polarization, I_{max} and I_{min} express a maximum value and a minimum value of the observation while rotating the linear polarizer respectively. When there are three different polarization images at least, I_{max}, I_{min} and ϕ are calculated by solving a system of linear equations [13]. Here, ψ equals to ϕ or $\phi + \pi$ if the observation is dominated by diffuse reflection. If the observation is dominated by specular reflection, ψ have $\frac{\pi}{2}$-shift from ϕ or $\phi + \pi$. In either case, we can not determine the correct angle ψ (π-flipped or not). This is known as π-ambiguity. To obtain the zenith angle θ, SfP exploits the degree of polarization $\rho = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$. According to Fresnel equations, the relationship between ρ and θ is defined as

$$\rho = \frac{(n - \frac{1}{n})^2 sin^2\theta}{2 + 2n^2 - (n + \frac{1}{n})^2 sin^2\theta + 4cos\theta\sqrt{n^2 - sin^2\theta}}, \quad (2)$$

$$\rho = \frac{2cos\theta sin^2\theta\sqrt{n^2 - sin^2\theta}}{cos^2\theta(n^2 - sin^2\theta) + sin^4\theta}, \quad (3)$$

where n is a refractive index describing the material property, Eqs. (2-3) are diffuse and specular model respectively. In total, three sources of ambiguities can be obsessed in SfP:

- Model selection (diffuse or specular reflection),
- π-ambiguity of the azimuth angle,
- Refractive index is unknown for the zenith angle.

With regard to the model selection, we assume specular reflection because ToF sensor does not receive reflected light for active illumination in target scenes. The π-ambiguity is solved by few ToF depth cues and introducing planarity constraints with the superpixel segmentation approach (Sec. 2.2). Estimation of n is performed under the assumption that one object has the same refractive index for each pixel (Sec. 2.3).

2.2. π-ambiguity correction for azimuth angles

To correct the π-ambiguity, Kadambi *et al.* [8] compared normals from ToF depth values and azimuth angles including the

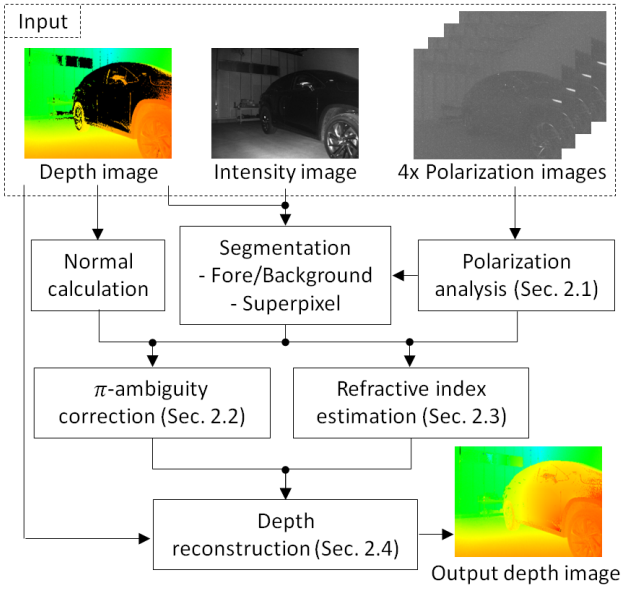


Fig. 2. Overview of our proposed method.

π-ambiguity. This works well for pixels where normals can be derived from ToF depth. However, since only few ToF measurements are available in our case, we can not solve the π-ambiguity without any additional constraints. Considering to propagate corrected azimuth angles over pixels which do not have the ToF normals, there are two main problems: (i) only few seed values are available, (ii) the propagation is easily affected by measurement noise including illumination change and rapid surface change. Therefore, we introduce two corresponding solutions: (i) increase the number of seed values, (ii) assume that the surface can be approximated as a set of planes.

First, we assume that our target object represents a closed surface, which means outward normals of the object contour on a 2D image are sufficient for the π-ambiguity. This is yielded by creating an object mask. To obtain the mask, we perform a segmentation method like GrabCut [14]. In addition to an intensity and a depth image, a degree of polarization and an azimuth angle image are used since the degree of polarization for a specular object takes high values around occluding contour in many cases [15].

Second, we deal with superpixels as a set of small regions like SLIC [16] for a piecewise planar approximation. This constraint leads to a noise-aware result and adaption for different illumination conditions in the outdoor environment.

Thus, we formulate the π-ambiguity correction as a labeling problem

$$\mathbf{E}(\mathbf{l}) = \sum_{u \in \mathbf{\Omega}_{tof}} \mathbf{E}_{data}(l_u) + \lambda \sum_{v \in \mathbf{\Omega}_{con}} \mathbf{E}_{data}(l_v) + \zeta \sum_{p,q \in \mathbf{\Omega}_{nh}} \mathbf{E}_{smooth}(l_p, l_q), \quad (4)$$

$$\mathbf{E}_{data}(l_u) = \begin{cases} \mathrm{diff}(\psi_u^{tof}, \psi_u) & \text{if} \quad l_u = 0 \\ \mathrm{diff}(\psi_u^{tof}, \psi_u + \pi) & \text{if} \quad l_u = 1 \end{cases}, \quad (5)$$

$$\mathbf{E}_{data}(l_v) = \begin{cases} \mathrm{diff}(\psi_v^{con}, \psi_v) & \text{if} \quad l_v = 0 \\ \mathrm{diff}(\psi_v^{con}, \psi_v + \pi) & \text{if} \quad l_v = 1 \end{cases}, \quad (6)$$

$$\mathbf{E}_{smooth}(l_p, l_q) = \begin{cases} \mathrm{diff}(\psi_p, \psi_q) & \text{if} \quad l_p = l_q \\ \mathrm{diff}(\psi_p, \psi_q + \pi) & \text{if} \quad l_p \neq l_q \end{cases}, \quad (7)$$

where $\mathbf{\Omega}_{tof}$ is a set of superpixels for reliable ToF depth, $\mathbf{\Omega}_{con}$ is a set of superpixels for the object contour, $\mathbf{\Omega}_{nh}$ is a set of pairs of neighborhood superpixels for all superpixels, $\mathrm{diff}(\cdot)$ expresses angular distance between two angles. We solve this by a belief propagation algorithm [17]. The result of the π-ambiguity correction is shown in Fig. 3.

2.3. Refractive index estimation with ToF measurements

To obtain the zenith angles, we need a refractive index of the object. Although many works use a fixed value (n=1.3-1.8) as a known refractive index, specular objects in real scenes have a wide range for refractive index, which leads to significantly different zenith angles (Eq. (3)).

Since our assumption is that one object has the same refractive index for the whole object and that depth for parts of the specular object can be obtained, we yield refractive index $\hat{n}$ by solving the non-linear least squares problem

$$\hat{n} = \arg\min_{n} \sum_{u \in \mathbf{\Omega}_{tof}} ||f(n, \theta_u^{tof}) - \rho_u||^2, \quad (8)$$

where $f(\cdot)$ equals to the right-hand side of Eq. (3). Once we earned the refractive index, we recover the zenith angles from degree of polarization by an inverted function $f^{-1}(\hat{n}, \rho)$. Although this inversion has two possible zenith angles, we take a lower value because the range of the higher value is close to 90 degrees – especially for specular surfaces – and this is less probable in practice.

2.4. Absolute depth from SfP normals

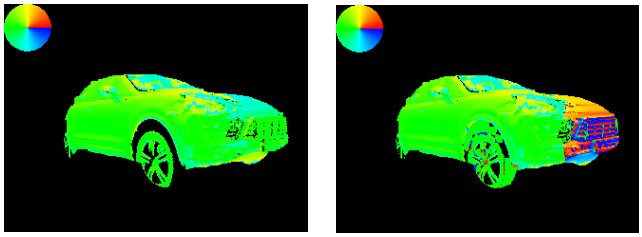
After obtaining the corrected normals, we reconstruct relative shapes by solving the Poisson equation

$$\nabla \cdot \nabla \hat{\mathbf{D}} = \nabla \cdot \mathbf{N}, \quad (9)$$

where $\hat{\mathbf{D}}$ is a set of height values of the reconstructed relative shape and $\mathbf{N}$ is a set of the corrected normals. To solve Eq. (9), many techniques can be found in [18, 19]. We use a method [20] that the reconstructed shape is treated as a set of meshes and is coped with alternating local and global mesh update.

Finally, the relative shapes are scaled to absolute depth values. Again, we take ToF depth into account to define a proper scaling factor s and an offset t

$$\{\hat{s}, \hat{t}\} = \arg\min_{s,t} \sum_{p \in \mathbf{\Pi}} ||\hat{D}_p s + t - D_p||^2, \quad (10)$$



(a) Input ψ (b) Our corrected $\hat{\psi}$

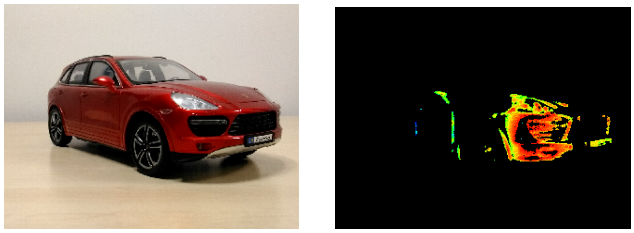
Fig. 3. Exemplary result of our π-ambiguity correction: (a) Input azimuth angles from Sec. 2.1, (b) Output of our algorithm from Sec. 2.2. Focusing on right side of the car, our algorithm effectively performed π-flipping.

where $\mathbf{\Pi}$ is a set of pixels for the specular object, $\hat{D}_p$ is the height of the reconstructed relative shape and D_p is a ToF depth of p. Finally, the absolute depth values are refined with $\hat{\mathbf{D}}$, $\hat{s}$ and $\hat{t}$.

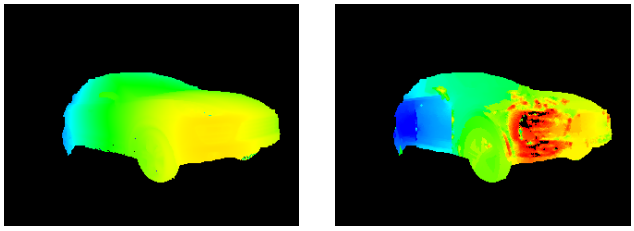
3. EXPERIMENTAL RESULTS

Since we propose to enhance ToF sensor depth, the experiments were performed by the single ToF sensor (TED TB-7Z-TCDK-GC2) [21] with the linear polarizer (Thorlabs LPNIRE100B) [22]. The angles of the linear polarizer were 0°, 45°, 90°, 135°. The datasets are recorded in different environments – a controlled indoor environment and a real outdoor environment. First, we polish a miniature-car to keep specular reflection and record it with unpolarized illumination. Once finished the recording, ground truth is obtained by a precise 3D scanner using structured light [23]. Note that the miniature-car is sprayed for ground truth acquisition to have diffuse reflection since the accuracy of this technique decays for specular surfaces. Second, we present qualitative results for real vehicle data (Fig. 6).

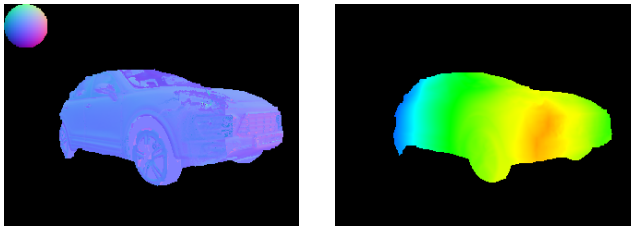
Reconstruction error: We compare our reconstruction against the method of Levin *et al.* [24]. This method uses two input images, a seed image and a guidance image, and propagates few seed values by following the guidance image. In the experiments, depth propagation is performed by an intensity image of the ToF sensor. We use root mean squared error (RMSE) in meter as metric. The error of our proposed method is 0.045, while the error of the method of Levin *et al.* is 0.085. The reconstruction results are shown in Fig. 4. Our proposed method reconstructs more precisely compared to [24]. In regard to the front part of the miniature-car, our result is close to ground truth. Since the proposed reconstruction performs relative shape fitting (Sec. 2.4), an influence of measurement errors due to varying object reflectivity [25] is reduced. In our experiments, we also try to evaluate the neural network based approach of Fan *et al.* [26]. However, it is not able to reconstruct well due to lack of examples in the training set.



(a) Target scene (b) Raw ToF depth



(c) Ground truth (d) Levin *et al.* [24]



(e) Normal map of ours (f) Depth of ours

Fig. 4. Exemplary reconstruction results in controlled indoor environment with ground truth.

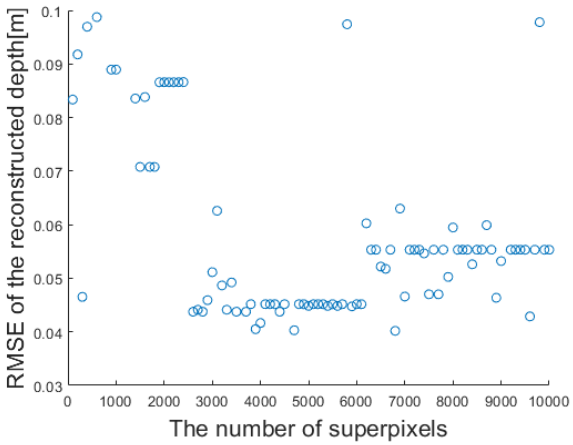


Fig. 5. Influence of the number of superpixels.

Influence of superpixel segmentation approach: To analysis the influence of the segment size, we perform our reconstruction with different segment sizes. The relationship between the segment sizes and reconstruction errors is shown in Fig. 5. When the number of superpixels is small, the large error is caused due to the strong constraint that different azimuth angles have to be taken into account as one single angle. On the other hand, increasing the number of superpixels also leads to higher reconstruction error. In this case, the method is close to the pixel-wise operation, which is affected by noise and rapid surface change.

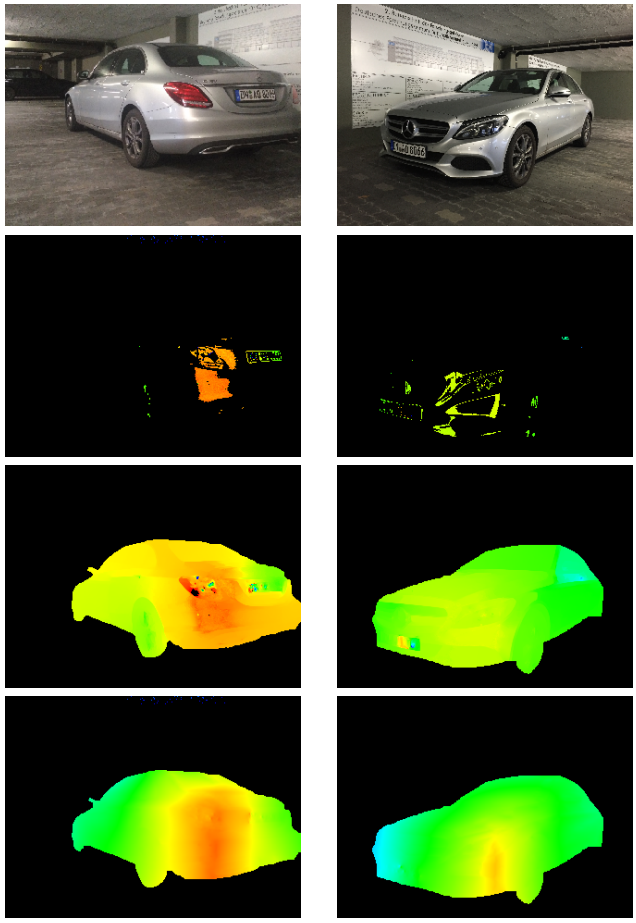


Fig. 6. From top to bottom: target scenes, raw input ToF depth images, outputs of Levin *et al.* [24], results of our proposed method. Although input depth images have few depth data, the proposed method demonstrates adequate reconstruction to a dense depth image.

4. CONCLUSIONS

In this paper, we propose a new depth reconstruction framework for specular objects that combines few absolute depth cues and relative shapes with SfP. Thanks to the planarity constraints implemented with the superpixel segmentation approach, the proposed method can overcome the ill-posedness of SfP with a single view and propagate the depth cues to whole specular objects. Experimental results show that the proposed framework can precisely reconstruct depth values of specular objects in the controlled environment. We also demonstrate the depth reconstruction for the real car in the outdoor environment. Our future work will extend the evaluation dataset to adapt various objects which have different material property as well as moving camera [27].

5. REFERENCES

[1] T. Yoshida, O. Wasenmüller, and D. Stricker, "Time-of-flight sensor depth enhancement for automotive exhaust gas," in *International Conference on Image Processing (ICIP)*, 2017.

[2] D. N. Bhat and S. K. Nayar, "Stereo and specular reflection," *International Journal of Computer Vision (IJCV)*, 1998.

[3] J. J. Koenderink and A. J. v. Doorn, "Affine structure from motion," *Journal of the Optical Society of America A*, 1991.

[4] R. Basri, D. Jacobs, and I. Kemelmacher, "Photometric stereo with general, unknown lighting," *International Journal of Computer Vision (IJCV)*, 2007.

[5] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah, "Shape-from-shading: a survey," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 1999.

[6] D. Miyazaki, M. Kagesawa, and K. Ikeuchi, "Transparent surface modeling from a pair of polarization images," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2004.

[7] G. A. Atkinson and E. R. Hancock, "Multi-view surface reconstruction using polarization," in *International Conference on Computer Vision (ICCV)*, 2005.

[8] A. Kadambi, V. Taamazyan, B. Shi, and R. Raskar, "Polarized 3d: High-quality depth sensing with polarization cues," in *International Conference on Computer Vision (ICCV)*, 2015.

[9] A. H. Mahmoud, M. T. El-Melegy, and A. A. Farag, "Direct method for shape recovery from polarization and shading," in *International Conference on Image Processing (ICIP)*, 2012.

[10] O. Morel, F. Meriaudeau, C. Stolz, and P. Gorria, "Polarization imaging applied to 3d reconstruction of specular metallic surfaces," *Electronic Imaging*, 2005.

[11] Z. Cui, J. Gu, B. Shi, P. Tan, and J. Kautz, "Polarimetric multi-view stereo," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[12] G. P. Nordin, J. T. Meier, P. C. Deguzman, and M. W. Jones, "Micropolarizer array for infrared imaging polarimetry," *Journal of the Optical Society of America A*, 1999.

[13] S. K. Nayar, X.-S. Fang, and T. Boult, "Separation of reflection components using color and polarization," *International Journal of Computer Vision (IJCV)*, 1997.

[14] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM transactions on graphics (TOG)*, 2004.

[15] L. B. Wolff and T. E. Boult, "Constraining object features using a polarization reflectance model," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 1991.

[16] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2012.

[17] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2006.

[18] A. Agrawal, R. Raskar, and R. Chellappa, "What is the range of surface reconstructions from a gradient field?," in *European Conference on Computer Vision (ECCV)*, 2006.

[19] Y. Quéau, J.-D. Durou, and J.-F. Aujol, "Normal integration: a survey," *Journal of Mathematical Imaging and Vision*, 2017.

[20] W. Xie, Y. Zhang, C. C. Wang, and R. C.-K. Chung, "Surface-from-gradients: An approach based on discrete geometry processing," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[21] Tokyo Electron Device, `http://www.teldevice.co.jp/`.

[22] THORLABS, `https://www.thorlabs.com/`.

[23] 3Digify, `http://3digify.com/`.

[24] A. Levin, D. Lischinski, and Y. Weiss, "Colorization using optimization," *ACM Transactions on Graphics (ToG)*, 2004.

[25] M. Lindner, I. Schiller, A. Kolb, and R. Koch, "Time-of-flight sensor calibration for accurate range sensing," *Computer Vision and Image Understanding (CVIU)*, 2010.

[26] H. Fan, H. Su, and L. Guibas, "A point set generation network for 3d object reconstruction from a single image," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[27] O. Wasenmüller, M. D. Ansari, and D. Stricker, "DNA-SLAM: Dense noise aware SLAM for ToF RGB-D cameras," in *Asian Conference on Computer Vision Workshop (ACCV Workshop)*, 2016.