https://doi.org/10.2352/EI.2024.36.18.3DIA-102

Extending Lidar Depth Range using Stereo Depth Estimation on Intensity Data

Filip Taneski, Institute for Integrated Micro and Nano Systems, University of Edinburgh, Edinburgh, United Kingdom
Tarek Al Abbas, Ouster Automotive, Ouster, Inc., Edinburgh, United Kingdom
Robert Henderson, Institute for Integrated Micro and Nano Systems, University of Edinburgh, Edinburgh, United Kingdom

Abstract

Solid-state lidar cameras produce 3D images, useful in applications such as robotics and self-driving vehicles. However, range is limited by the lidar laser power and features such as perpendicular surfaces and dark objects pose difficulties. We propose the use of intensity images, inherent in lidar camera data from the total laser and ambient light collected in each pixel, to extract additional depth information and boost ranging performance. Using a pair of off-the-shelf lidar cameras and a conventional stereo depth algorithm to process the intensity images, we demonstrate increase of the native lidar maximum depth range by 2× in an indoor environment and almost 10× outdoors. Depth information is also extracted from features in the environment such as dark objects, floors and ceiling which are otherwise not detected by the lidar sensor. While the specific technique presented is useful in applications involving multiple lidar cameras, the principle of extracting depth data from lidar camera intensity images could also be extended to standalone lidar cameras using monocular depth techniques.

Introduction

The evolution of 3D lidar from traditional mechanical scanning lidar to state-of-the-art solid-state "flash" lidar now allows for affordable 3D imaging in a compact form factor. This makes high-performance 3D imaging commercially viable for a whole host of applications including robotics and self-driving vehicles, accelerating the lidar market from $1.8B in 2020 to a predicted $5.7B in 2026 [1]. Solid-state flash lidar cameras gain many advantages over the traditional mechanical scanning approach. These include higher achievable frame rate, low-cost manufacture, improved mechanical robustness and reduced artefacts from point occlusions. However, since the laser power of a flash lidar needs to illuminate and spread across the full field-of-view in contrast to a scanning approach, the maximum sensing range is limited.

An often neglected, yet useful image also produced by lidar cameras is an in-tensity image, given by the total amount of light captured at each point in the scene. Example intensity images are shown below each corresponding depth im-age in Figure 1, uncovering details not captured by the depth images alone. Our work presents the first application of this inherent lidar intensity data to extract additional depth information. This enables an increase in the maximum achievable lidar depth range with no additional power consumption or hardware modification. Additional depth data of features which pose difficulties to lidar ranging such as perpendicular or dark surfaces are also extracted. The method implemented in this work uses a pair of short-range lidar cameras to compute stereo depth from combined intensity images, as illustrated in Figure 1.

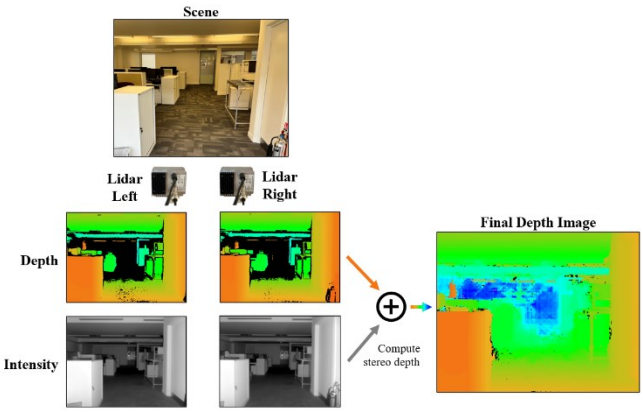

Figure 1. *Illustration of extending lidar depth range by processing intensity images*

Related Work

Extracting additional depth information from lidar intensity images is an approach that is unique to this work. However, a handful of prior studies have explored other means of improving ranging performance from multiple lidars. The refinement of lidar images using additional vision cameras is a more commonly explored topic of research. Both approaches are explored here.

Lidar-Assisted Lidar

The first work to combine multiple lidar cameras to improve depth performance was published by Castañeda et al. [2]. Depth images from two lidar cameras were used to define a cost function at each pixel and optimize the depth value. While this approach improves the accuracy of the acquired depth data, the maximum sensing range remains the same.

In [3], a secondary lidar camera was used to help resolve the number of phase wrappings seen at each point in the primary lidar camera, thereby unlocking additional depth range. Phase wrapping is unique to indirect time-of-flight (iToF) techniques and is inherently overcome in many modern lidar cameras which employ direct time-of-flight (DToF) approaches. Nevertheless, the increased range enabled by this technique is still limited by the maximum laser power.

Vision-Assisted Lidar

While this work is unique in extracting stereo depth estimation from the lidar's own intensity image, the use of an additional pair of vision cameras to enhance the lidar depth image is an alternative and commonly explored approach. These methods can be summarized into two categories: depth completion and error refinement. In the first category, vision camera images are used to upscale the

resolution of typically low resolution lidar cameras. The first published example was proposed in [4] using a Markov Random Fields (MRF) method by matching discontinuities in lidar depth to those in camera brightness. A summary of similar methods extended from this is given in [5]. The MRF method is further improved upon in [6] to work better under dynamic scenes.

In the second category - error refinement - additional data from vision sensors are used to refine the accuracy of generated lidar data. This was performed most notably in [7] where a reliability fusion algorithm is developed using global regularization to improve lidar accuracy. This improves depth accuracy in regions of varying texture which pose challenges to lidar.

It is noted that the opposite approach: lidar-assisted stereo vision e.g. to assist solving disparity estimation [8, 9], is a frequently explored topic beyond the scope of this paper.

Preliminaries

Lidar

Time-of-flight (ToF) lidar sensors measure depth by timing the roundtrip of an emitted laser pulse. This can be determined directly (dToF) or indirectly (iToF) through measuring the different in phase of the returning pulse, typically by measuring four integration phases as illustrated in Figure 2.

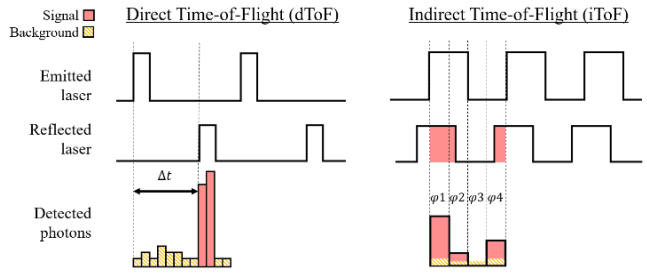

Figure 2. *Illustration of indirect vs direct time-of-flight (lidar) methods.*

In dToF, the target distance d is simply given as a function of the laser return time t and speed of light c by (1):

$$\boldsymbol{d = \frac{\Delta t \times c}{2}} \tag{1}$$

On the other hand, the total integrated signal in each phase φ using a conventional four-phase iToF scheme at a pulse period T can be determined using (2):

$$\boldsymbol{d = \frac{T \times c}{4\pi} \times tan^{-1}\left(\frac{\varphi_3 - \varphi_4}{\varphi_1 - \varphi_2}\right)} \tag{2}$$

In both methods, the total light gathered is used to determine the laser signal arrival time. As a result, an intensity image created by integrating the total light detected at each pixel is readily available.

In both approaches, ambient background (non-signal) photons are also collected by the lidar sensor, impacting the accuracy of the resulting depth measurement. To mitigate this, the presence of narrow bandwidth light filter, centered around the laser wavelength is integrated into the optics. Nevertheless, some ambient background photons are still collected in this wavelength range. The resulting intensity image is therefore often a combination of both the reflected laser (solid red in Figure 2) and ambient background light (dashed orange in Figure 2), a feature which will be exploited in this work.

Stereo Depth

By using a pair of vision cameras and identifying the pixel displacement Δp of common points between both camera images, the distance to each point can be determined. For a camera baseline separation B and focal length f, the distance is given by (3).

$$\boldsymbol{d = \frac{f \times B}{\Delta p}} \tag{3}$$

Method

Apparatus

This work was conducted using a pair of commercially available Basler Blaze-101 iToF lidar cameras mounted on a rigid dovetail rail as shown in Figure 3. The lidar cameras output processed depth and intensity data at a resolution of 640×480 pixels and maximum range of 10 m. They operate at a wavelength of 940 nm with an optical filter allowing only light between 920–970 nm to be detected. A Bosch GLM 250 VF rangefinder (not pictured) is used to assess the ground truth accuracy of processed intensity images.

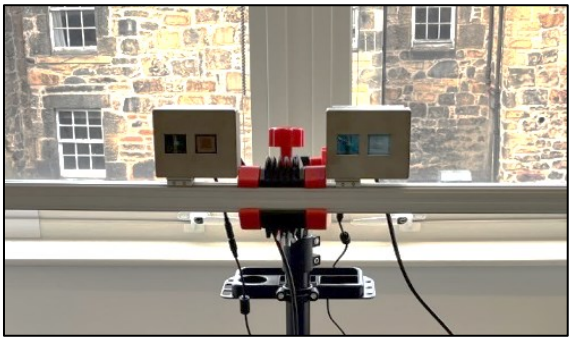

Figure 3. *Camera setup.*

Processing

The processing of lidar intensity images to provide additional depth information is split into two parts, setup and runtime, as illustrated by the flow diagram in Figure 4.

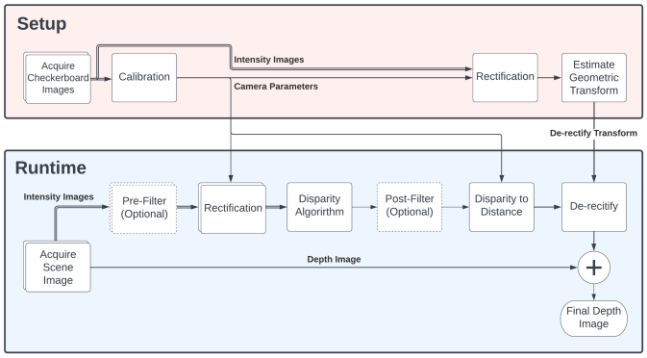

Figure 4. *Processing flow diagram for enhancing depth image from processed intensity images.*

Setup

The first phase of setup consists of camera calibration using a checkerboard. Intensity images of a checkerboard target are taken using both lidar cameras processed using the MATLAB Stereo Camera Calibrator [10]. A unique challenge in calibrating lidar cameras using intensity images compared to conventional imaging cameras is the presence of glare introduced by the lidar laser, as shown in Figure 5(a). This can be overcome by acquiring numerous checkerboard images at a variety of positions and angles (Figure 5 (b)) until a suitable number of glare-free images have been acquired. A calibration of less than 0.1 pixels of reprojection error was found to give sufficiently accurate camera parameters for stereo matching.

In order to assist the disparity algorithms in matching corresponding points in images between the left and right camera, the images are first rectified. This applies a transform to both images such that all matching points line up along the same horizontal plane, thus reducing the search space for disparity algorithms. The drawback of rectification is that the transformed image no longer has a one-to-one pixel correspondence with the original intensity image. Since image rectification is a non-linear transformation, the resulting processed image cannot be directly mapped to and merged with the original lidar depth image. To solve this issue, image registration is used [11] to find a transformation (translation, rotation, scaling, and/or shearing) which aligns a sample rectified image with its corresponding pre-rectified image. Once estimated, the same transformation can be reapplied to any rectified image provided the intrinsic and extrinsic properties of the camera remain unchanged. Since the transform estimation only needs to take place during setup, the additional processing overhead during runtime is reduced to just a single inexpensive image transformation.

***Figure 5**. The issue of flash glare (a) from lidar intensity images which can result in degraded calibration. Use of various calibration target angles to capture images without glare (b).*

Runtime

After acquiring depth and intensity images simultaneously from both lidar cameras, the intensity images are processed to extract further depth information. The narrow bandwidth light filter used in lidar sensors increases the shot noise of the intensity image compared to a conventional camera. This makes it more difficult for stereo disparity algorithms to match points between images. Using an optional pre-filter (median) can help to reduce this effect. After undergoing rectification, the pixel disparity between left and right images can then be resolved. In this work, we use the semi-global matching (SGM) algorithm [12] as a well-established method of solving stereo matching. A second optional post-filter (median) can be used, this time to replace small patches of missing pixel disparities. Pixel disparities are then converted to depth using the camera extrinsic parameters established during setup, along with (**3**). Finally, the stereo depth image is de-rectified using the transform estimated during "setup" and combined with the original lidar depth image. When combining the original lidar depth image with the stereo depth estimates, any points from the original lidar depth image are considered accurate and are therefore not overwritten. The depth estimated from the processed intensity images fills in any missing data points from the original lidar image. A breakdown of runtime processing time is provided in Table 1.

Table 1. Breakdown of processing time during main runtime.

Process	Time (ms)	Notes
Acquire images	33	Limited by lidar
Pre-filter	10	Optional
Rectification	60	-
Disparity algorithm	60	SGM
Post-filter	5	Optional
Disparity to distance	2	-
De-rectify	2	-
Merge lidar depth	1	-
Total	**174**	**Equivalent to 5+ fps**

Results

Indoor

The first set of results is comprised of images taken in an indoor environment. Under these conditions most of the light captured by the sensor is the reflected laser light. To assess the performance of the processed intensity depth within a real-world environment, images of a real human figure have been captured at a varying distance away from the lidar cameras. The captured and processed images are shown in Figure 6.

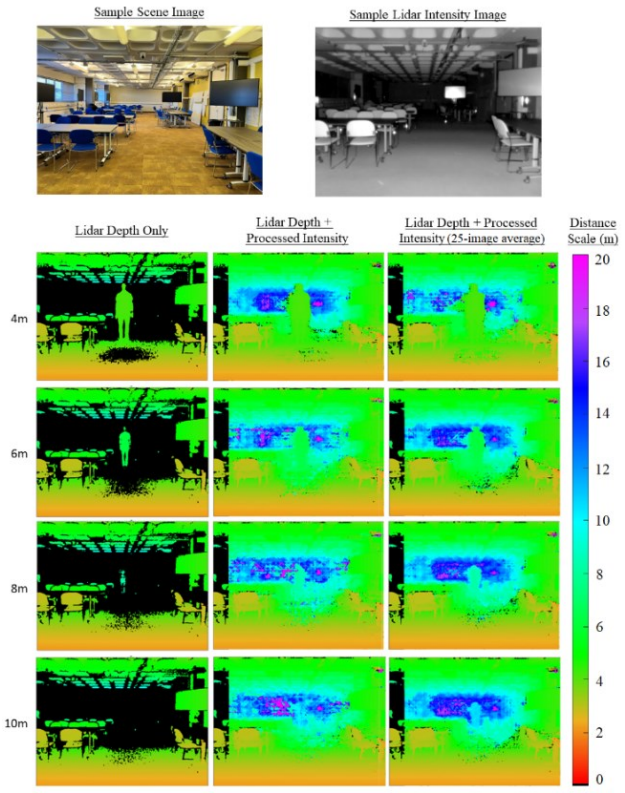

***Figure 6.** Captured lidar and processed intensity images in an indoor environment.*

At the shorter distances, the benefit of the additional processed intensity depth data in situations that are challenging for lidar sensors is apparent. These include surfaces that are near-perpendicular to the sensors such (e.g. floor, ceiling, tables) and narrow objects such as chair legs. As the target human figure approaches 8 m and beyond, the returning laser signal is no longer sufficient for lidar measurements. However, the processed intensity images continue to resolve the distance to the human target.

By trading off frame rate, multiple consecutive intensity images can be taken and averaged to extract further detail. The resulting processed depth from averaging 25 intensity frames is shown in the final column of images in Figure 6. Using this technique, the human figure is able to be resolved to a distance as far as 18 m, increasing the native lidar depth range of the human figure by over 2×. The processed intensity depth accuracy of the human figure at each distance step sampled over a 5×5 window of pixels is summarized in Figure 7. A degradation in accuracy with distance is observed with a root-mean squared error of 1 m at the maximum distance.

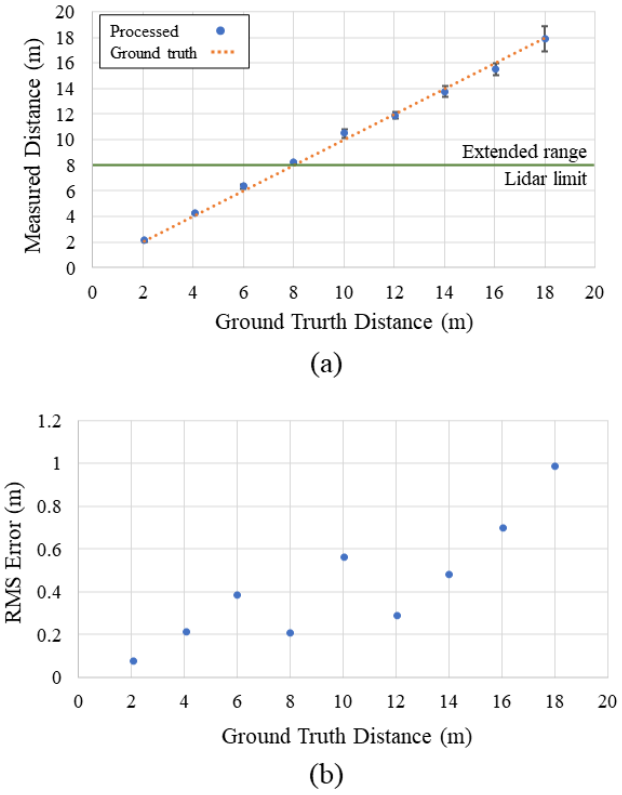

Figure 7. *Processed intensity depth accuracy ranging a human figure indoors using a 25-image average.*

Outdoor

The second set of results is comprised of images taken in an outdoor environment. Under these conditions most of the light captured by the lidar is ambient background light. The resulting processed images from this scene are shown in Figure 8.

The additional challenge presented to the lidar sensor under these conditions compared to the indoor environment is evident by the lack of native lidar depth points beyond even 5 m. The additional challenge presented to the lidar by dark surfaces is also evident in the lack of depth data points present on the foreground plant pot. In this scenario, the additional range data extracted from the intensity images not only fills in the missing plant pot data, but also reveals the presence of a human figure at 11.5 m which was otherwise undetected. Furthermore, buildings as distant as 40 meters from the target are detected, extending the maximum range in this scenario by nearly 10×.

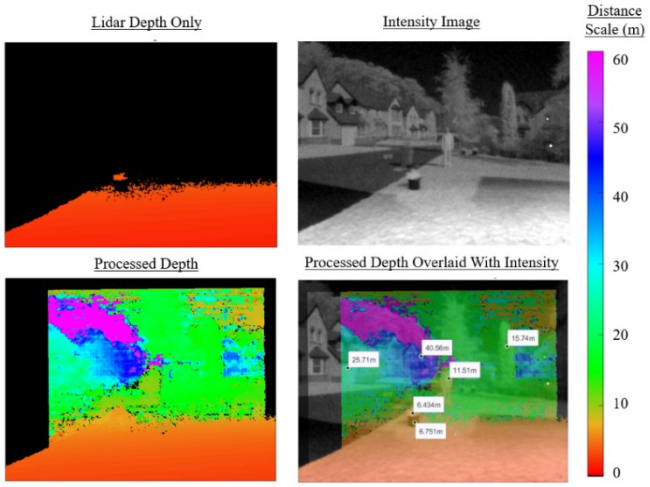

Figure 8. *Captured lidar and processed intensity images in an outdoor environment.*

Discussion

To provide a basis for further discussion, a large-scale sample image from the indoor dataset is provided in Figure 9 with accompanied zoomed in window portions.

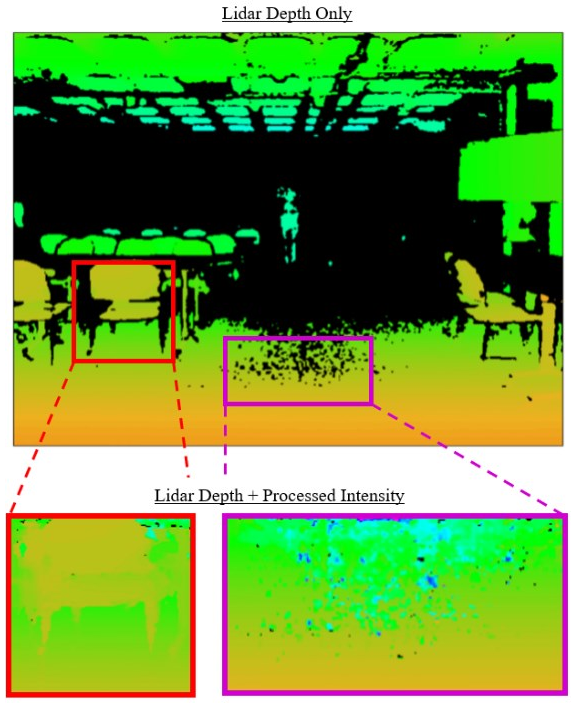

Figure 9. Zoomed in samples windows of the captured indoor dataset

The figure exemplifies the ability of the processed intensity images to fill depth information of narrow objects, in this case the legs of a chair. The limited laser returns from a narrow object, coupled with scattering issues make such objects challenging for lidar but relatively unchallenging for stereo depth estimation. While the missing depth data on the floor is also filled in by the processed intensity images, the contrast between the original depth data and the filled in estimated depth suggests some inaccuracy in the estimated depth from intensity images.

The main limitation of the presented technique in its current form is the requirement to average images over long acquisition times to achieve intensity image quality high enough for stereo depth estimation. In addition, while the intensity images captured indoors are comprised of mainly the reflected laser light, the outdoor scene relies on ambient illumination. As a consequence, this technique is limited to bright ambient conditions when conducted outdoors. However, this is complimentary to the performance of the native lidar depth performance, which improves under low ambient conditions as seen in Figure 6. (indoor) vs. Figure 8 (outdoor). Under these conditions the need for additional depth data from processed intensity images is less critical.

Conclusion

We demonstrate that the native depth range of lidar cameras can be increased through processing the inherent intensity images. This technique could provide additional useful data in applications such as self-driving vehicles which integrate multiple lidar sensors. This technique has been demonstrated using a pair of off-the-shelf solid-state lidar cameras and a basic stereo depth algorithm. Missing depth information has been successfully extracted from features in the environment such as floors, ceiling and dark objects. By averaging multiple frames of intensity images, the native lidar maximum depth range increased by 2× in an indoor environment and almost 10× outdoors.

To develop this work further, future work should focus on reducing the required exposure time/averaging of intensity images. This can be achieved by (i) exploring alternative stereo depth estimation algorithms and (ii) the increased light sensitivity of state-of-the-art single photon avalanche diodes (SPADs) based lidar sensors [13, 14] which will soon improve the quality of the inherent intensity image. Finally, exploring new learning-based monocular depth estimation techniques [15, 16] would enable range extension using only a single lidar camera.

References

[1] Y. Développement, "LiDAR for Automotive and Industrial Applications," 2021. [Online]. Available: https://web.archive.org/web/20221229085924/https://s3.i-micronews.com/uploads/2021/09/YINTR21174-LiDAR-for-Automotive-and-Industrial-Applications-2021-Sample.pdf

[2] V. Castañeda, D. Mateus, and N. Navab, *Stereo Time-of-Flight*. 2011, pp. 1684-1691.

[3] O. Choi and S. Lee, "Wide range stereo time-of-flight camera," in *2012 19th IEEE International Conference on Image Processing*, 30 Sept.-3 Oct. 2012 2012, pp. 557-560, doi: 10.1109/ICIP.2012.6466920.

[4] J. Diebel and S. Thrun, *An Application of Markov Random Fields to Range Sensing*. 2005.

[5] H. Andreasson, R. Triebel, and A. Lilienthal, "Non-Iterative Vision-Based Interpolation of 3D Laser Scans," in *Autonomous Robots and Agents*, S. C. Mukhopadhyay and G. S. Gupta Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 83-90.

[6] J. Dolson, J. Baek, C. Plagemann, and S. Thrun, "Upsampling range data in dynamic environments," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 13-18 June 2010 2010, pp. 1141-1148, doi: 10.1109/CVPR.2010.5540086.

[7] J. Zhu, L. Wang, R. Yang, J. E. Davis, and Z. pan, "Reliability Fusion of Time-of-Flight Depth and Stereo Geometry for High Quality Depth Maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 33, no. 7, pp. 1400-1414, 2011, doi: 10.1109/TPAMI.2010.172.

[8] C. Dal Mutto, P. Zanuttigh, and G. M. Cortelazzo, "A probabilistic approach to tof and stereo data fusion," presented at the 3DPVT, Paris, France, 2010.

[9] V. Gandhi, J. Čech, and R. Horaud, "High-resolution depth maps based on TOF-stereo fusion," in *2012 IEEE International Conference on Robotics and Automation*, 14-18 May 2012 2012, pp. 4742-4749, doi: 10.1109/ICRA.2012.6224771.

[10] MathWorks, "MATLAB Computer Vision Toolbox Reference," *vol. R2022b*, *Stereo Camera Calibrator*, 2022, pp. 390-391.

[11] MathWorks, "MATLAB Image Processing Toolbox Reference," *vol. R2022b*, *imregtform*, 2022, pp. 1900-1907.

[12] H. Hirschmuller, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 30, no. 2, pp. 328-341, 2008, doi: 10.1109/TPAMI.2007.1166.

[13] O. Kumagai *et al.*, "7.3 A 189×600 Back-Illuminated Stacked SPAD Direct Time-of-Flight Depth Sensor for Automotive LiDAR Systems," in *2021 IEEE International Solid- State Circuits Conference (ISSCC)*, 13-22 Feb. 2021 2021, vol. 64, pp. 110-112, doi: 10.1109/ISSCC42613.2021.9365961.

[14] K. Morimoto *et al.*, "3.2 Megapixel 3D-Stacked Charge Focusing SPAD for Low-Light Imaging and Depth Sensing," in *2021 IEEE International Electron Devices Meeting (IEDM)*, 11-16 Dec. 2021 2021, pp. 20.2.1-20.2.4, doi: 10.1109/IEDM19574.2021.9720605.

[15] A. Saxena, S. Chung, and A. Ng, "Learning Depth from Single Monocular Images," *vol. 18*, *Advances in Neural Information Processing Systems 18 (NIPS 2005)*, Y. W. a. B. S. a. J. Platt, Ed., 2005.

[16] D. Eigen, C. Puhrsch, and R. Fergus, "Depth Map Prediction from a Single Image using a Multi-Scale Deep Network," p. arXiv:1406.2283. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2014arXiv1406.2283E

Author Biography

Filip Taneski received an MEng degree in Electrical and Electronics Engineering from the University of Edinburgh (2013). Following this he joined Broadcom as a mixed-signal integrated circuit designer. In 2020, he began pursuing a PhD in Engineering at the University of Edinburgh to develop solid-state lidar technology for the application of self-driving vehicles.