Noise Model for Continuous-Wave Sensors

Tianwen Fu
Carnegie Mellon University
Pittsburgh, PA 15213
tianwenf@andrew.cmu.edu

Abstract

Continuous-wave time-of-flight (CW-ToF) sensors are popular due to their compactness, low cost, and satisfactory reliability. However, the accuracy is limited in low-albedo areas, which is common in tag detection in robotics, as the echo signal does not have sufficient energy to overcome the random noise. In this project, we propose a noise model for the prediction over dark areas and conduct experiments on real devices to verify the model. The results disagree, but we still point out several causes of failure and the next steps to improve our experiments. Solving the discrepancy may enable us to employ HDR techniques in the IR spectrum and fabricate a sensor that performs better over dark areas. Code and related data are available at https://github.com/tianwen-fu/16823-project.

1. Introduction

3D data have a wide range of usages in autonomous driving [2], crack detection [24, 25], novel view synthesis [22], and facial recognition [21] for security purposes such as Face ID. Consequently, a variety of technologies supporting 3D sensors are developed, including LiDARs, continuous wave time-of-flight (CW-ToF) sensors, and stereo cameras. Among the options, a lot of commercial 3D sensors adopt CW-ToF sensors based on their balance of reliability, high resolution, and compactness.

CW-ToF sensors measure depth information by comparing the phase of actively emitted modulated light and its echo from objects. Despite being robust for most convex Lambertian scenes, this system suffers from global illumination and random noise. A multitude of previous work has made great progress in dealing with global illumination [1, 9], but little attention is drawn to methods to cope with the random noise.

The effect of random noise is manifest in regions with low albedo, where an insufficient amount of signal is returned and therefore the depth estimation for points in black regions often deviates from the plane they lie on. However, the robotics community heavily relies on black-and-white tags such as ArUco [7] and AprilTag [17], usually printed on paper, as beacons. Inaccuracies around the black corners may impede the pose estimations of 3D detection systems in these robots.

Nevertheless, in photography, the common way to reduce the noise in low signal regions without saturating the strong signal areas is High Dynamic Range Imaging (HDR), which combines short- and long-exposure images to preserve accurate measurements regardless of scene albedo. However, for robotics uses we cannot afford long exposure that leads to delay and motion blur; instead, with an active camera, we may be able to adjust the signal strength adaptively to the scene measurements.

In this project, we aim to study the noise affecting depth estimation in low albedo regions. We propose both a theoretical result and experiment findings. Although the experiments do not prove the theory, we still hope in the future a more comprehensive theory and more carefully conducted experiments shall fix the discrepancy and enable a realistic simulator, which may further lead to the application of HDR techniques in 3D sensors.

In short, the contributions of the project include:

- A theory that aims to explain the noise;
- Experiment results of a real-world 3D sensor;
- Analysis of the failure and possible future directions.

2. Related Work

ToF Sensors Enhancement. Numerous endeavors have been undertaken to improve the depth estimation performance of Time-of-Flight (ToF) sensors. Godbaz et al. [8] employed blind deconvolution to extend the range of depths and deblur the measurements by estimating focal parameters. Xiao et al. [23] improved by solving a solution to the inverse problem of a blur model to remove the defocus blur and increase the spatial resolution. The proposed method is also applicable to multi-frequency ToF sensors, which have gained widespread use in recent times. To cope with global

illumination, Gupta et al. [9] proposed phasor imaging to model the light transport events by capturing several measurements at high temporal frequency. Reynolds et al. [20] concentrate on a more relevant research area by addressing the issue of flying pixels in materials with low infrared reflectivity. To remove these invalid points, they employ Random Forests to estimate a per-point confidence assignment. However, their work does not involve improving the point estimations. Mask-ToF [5] also aims to mitigate the flying pixels problem, but it models the issue as a consequence of background-foreground interference.

Commercial CW-ToF Sensors. Common commercial sensors include Azure Kinect DK camera [15] and Camboard pico flexx [18], whose depth sensor is Infineon IRS1145C[1] [10]. Both cameras project a modulated infrared (IR) signal and capture the reflected light. Then, the time-of-flight information is obtained by comparing the phase difference. Azure Kinect is equipped with two IR laser diodes, one of which is the near field-of-view mode, with resolution 640×576, measuring $0.5 \sim 3.9$m at 30FPS [14]; the other is the wide field-of-view mode with resolution 512×512, measuring $0.3 \sim 2.9$m at 30FPS. Both modes send 9 IR pulses, each 125us wide. Pico flexx is capable of taking 224×171 depth images at 45 FPS. Along with the depth information, both cameras also have a coupled image sensor: pico flexx captures in grayscale while Azure Kinect can take color images. Although there are no trivial ways to modify the firmware to the best of our knowledge, both brands provide software development kits (SDKs) for processing the depth information. Figure 1 shows a screenshot of Royale Viewer, a built-in application in pico flexx's SDK.

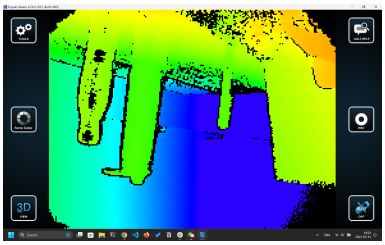


Figure 1. Screenshot of the depth map in a scene with two bottles and a marker on the table.

Multi-exposure Fusion. Generating high dynamic range photographs by fusing images captured with different exposure is a mature technology that can be dated over 2 decades. Debevec and Malik [6] proposed an algorithm for recovering the non-linear correspondence from radiance to pixel values, thus enabling merging multiple captures from an arbitrary camera without knowing the radiance a priori. On the other hand, Mertens et al. [13] avoided the computation of the response curve and computed the tone-mapped images directly by weighting pixels with well-engineered quality measures. Deepfuse [19] replaced the engineered features with deep features that are more robust to extreme exposure pairs.

Low-light Imaging. Besides taking HDR images with multiple exposures, another line of work aims to directly remove the noise from a single shot with a short exposure and high gain. Chen et al. [4] presented a dataset of short-exposure low-light images paired with long-exposure ones as reference images and trained a convolutional neural network to process the noisy images. Monakhova et al. [16] further extend to extremely low light under starlights and train generative adversarial networks as the noise model from high-gain short-exposure images to high-exposure ones. The noise model can create realistic noisy videos to train a deep denoiser. Liba et al. [12] presented a comprehensive system capable of motion-adaptive burst capture, denoising, and white balance adjustments. The system is specialized for hand-held devices equipped with motion sensors.

3. Technical Section

3.1. Scope of the Project

In this project, we define the "flying pixels" as the inaccurate depth measurements over low-albedo regions in the scene where reliable depth information can be obtained for bright pixels under the same condition. Therefore, we limit the scope of our project to only direct diffusion reflections and ignore all the global illumination effects and related post-processing techniques. Furthermore, following the practices in phasor-based works [9], we also assume that we may separate the modulation frequency from the received signal perfectly (by suitable band filters) and thus the measurements are only affected by sinusoidal of the same frequency.

3.2. Noise Model

As shown by Lange and Seitz [11], the time of flight is proportional to the phase difference, which can be recovered by a cross-correlation operator of the emitted and the received signal. Specifically, denote by $s(t) = \cos(2\pi\omega t)$ the modulated signal and $r(t) = L_{DC} + a\cos(2\pi\omega t - \varphi)$, where φ is the quantity we need to compute. The cross-correlation of two signals can be computed by

$$c(\tau) = \int_{-\infty}^{\infty} s(t)r(t+\tau)dt = \frac{a}{2}\cos(2\pi\omega\tau + \varphi).$$

[1]The specific model "IRS1145C" is not in the datasheet cited; however, the datasheet, with chip models IRS1125C, IRS1645C, and IRS1615C, include the pico flexx as a "3D reference camera".

Let $\tau^* \in [-\pi, \pi)$ be the maximizer of $c(\tau)$, the estimated phase shift would be $\hat{\varphi} = -2\pi\omega\tau^*$.

Now, let the noise be $n(t) = b\cos(2\pi\omega t - \varepsilon)$, then the signal we receive would be $\tilde{r}(t) = L_{DC} + a\cos(2\pi\omega t - \varphi) + b\cos(2\pi\omega t - \varepsilon)$ has the Fourier transform

$$\begin{aligned}\tilde{R}(f) =& L_{DC}\delta(f) + \frac{1}{2}(ae^{-j\varphi} + be^{-j\varepsilon})\delta(f-\omega) \\ &+ \frac{1}{2}(ae^{j\varphi} + be^{j\varepsilon})\delta(f+\omega)\end{aligned}$$

The Fourier transform of the cross-correlation function, which is the convolution of the signal and the flipped received signal is

$$\begin{aligned}\tilde{C}(f) = S(f)\tilde{R}(-f) =& \frac{1}{2}(ae^{-j\varphi} + be^{-j\varepsilon})\delta(f+\omega) \\ &+ \frac{1}{2}(ae^{j\varphi} + be^{j\varepsilon})\delta(f-\omega)\end{aligned}$$

and the correlation function

$$\tilde{c}(\tau) = \frac{a}{2}\cos(2\pi\omega\tau + \varphi) + \frac{b}{2}\cos(2\pi\omega\tau + \varepsilon)$$

which gives estimation

$$\hat{\varphi} = -2\pi\omega\tau^* = \tan^{-1}\left(\frac{a\sin\varphi + b\sin\varepsilon}{a\cos\varphi + b\cos\varepsilon}\right). \quad (1)$$

With uncorrelated ambient noise, its phase ε should be distributed evenly across $[0, 2\pi)$. In Figure 2[2], we show the distribution of $\hat{\varphi} - \varphi$ with different ratio $\frac{a}{b}$ for two choices of the ground truth phase φ. Note from the figure that by our noise model, the error should be concentrated around $\pm\frac{b}{a}$ and the ground truth depth only affects the shape but not the mode.

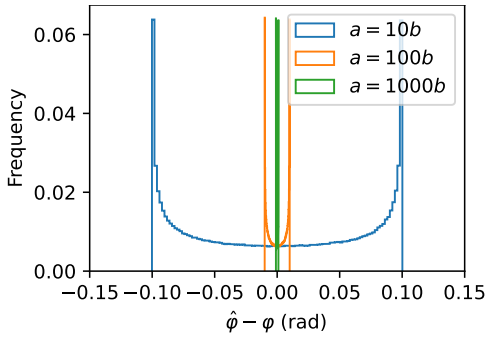
3.3. Validation of Noise Model

Collection of Data. We plan to print a set of checkerboard patterns, where black and white tiles show up alternately, and collect depth measurements by Pico Flexx camera from an orthogonal position. This should minimize the errors from other sources, such as a tilting angle and consequent global illumination effects. With the depth estimation, we can compute the phase from the depth by

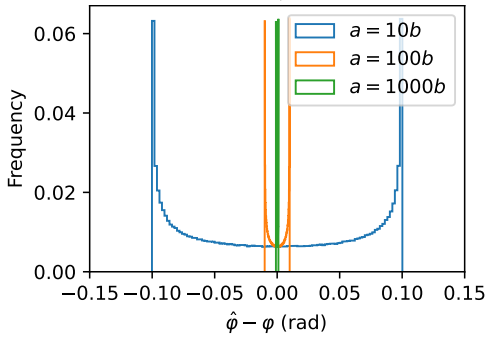
$$\varphi = \frac{4\pi\omega z}{c}$$

where w is the frequency of the IR with wavelength 850nm. In the collected data, we use all the corners to perform Procrustes transformation estimation to get the extrinsic matrix and then transform the entire image to $z = 0$ plane. The depth estimation over black tiles will be converted into data points for $\hat{\varphi}$. To validate our noise model with respect to different relative albedo $\frac{a}{b}$, we also tone map the checkerboard to different grayscales.

[2]I do believe there should be a closed-form PDF for $\hat{\varphi}$ as an expression involving the PDF of ε, but I don't know how to derive that.



(a) $\varphi = \pi/8$



(b) $\varphi = \pi/4$

Figure 2. Distribution of $\hat{\varphi} - \varphi$ according to our noise model with different values of ground truth φ. Each line stands for a histogram of $\hat{\varphi}$'s computed by Equation 1 with 1000000 samples of ε from $Uni[0, 2\pi)$. The y axis is normalized such that the area under each curve is 1.

Verifying the Qualitative Characteristics of the Noise Model. From the simulated densities in Figure 2, there are some key insights into the distribution of our noise that we need to validate:

- The distribution is symmetric and bi-modal, and almost all the densities are on the modes;
- The mode of phase difference $\hat{\varphi} - \varphi$ does not depend on the ground truth depth, but coincides with the ratios $\pm\frac{b}{a}$.

Estimating the Distribution of Noise. From Equation 1, let $r = \frac{b}{a}$, we may derive an alternative form that expresses ε as a function of $\hat{\varphi}$:

$$\sin(\varepsilon - \hat{\varphi}) = \frac{1}{r}\sin(\hat{\varphi} - \varphi) \quad (2)$$

From our data collection, we can acquire φ and $\hat{\varphi}$. Then we also estimate the ratio by

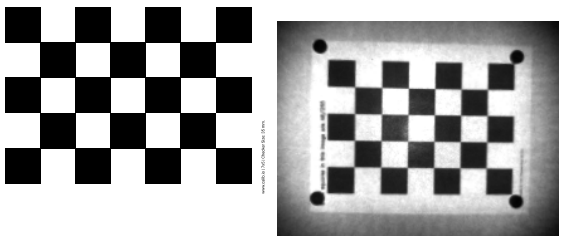
$$\hat{r} = \arg\max_{\Delta\varphi > 0} \tilde{p}(\Delta\varphi)$$

Then, we can obtain an empirical posterior distribution of ϵ given our observation.

4. Empirical Observations

4.1. Implementation Details

We use the bundled SDK of Pico Flexx to capture our data. The use case is set to `MODE_5_45FPS_500` which is suitable for depth measurements from 0.1m to 1m. We print Figure 3a onto a letter-sized paper, put it onto a plane, and capture it from an orthogonal angle like in Figure 3b. Afterward, we detect the checkerboard points by OpenCV [3] and transform the point cloud by orthogonal Procrustes transformation to the $z = 0$ plane. Then, in the transformed world frame, we compute the mean of depth measurements over white blocks z, and report the depth measurements over the black blocks $\hat{z}$ minus z.



(a) The Pattern to Print (b) Grayscale Image Captured

Figure 3. Checkerboard Pattern

4.2. Overview of the Noise Distribution

Figure 4 shows a 3D view of the flying pixel effect in a captured point cloud by our sensor. We also plot the histogram of the distribution of $\hat{z}$ in Figure 5 and per-block histograms in Figure 6; however, it is biased towards the positive direction of $\hat{z}$ and does not match our theoretical results in Figure 2.

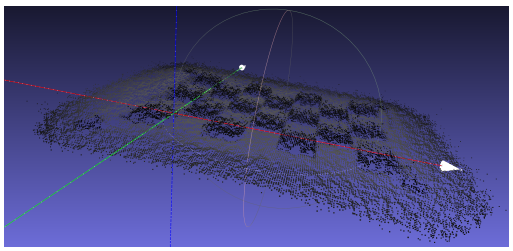


Figure 4. 3D View of the Flying Pixels. The red, green, and blue arrows point to the positive directions of x, y, z-axes respectively. Points corresponding to black points are floating over the plane.

Nevertheless, we still want to explore: *to what extent do the qualitative characteristics hold?* In the next sections, we will report how the distribution of $\hat{z}$ changes with respect to depth and albedo.

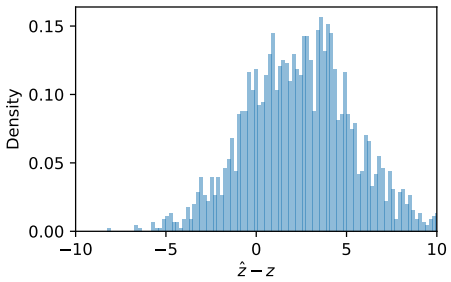


Figure 5. Histogram of the Distribution of $\hat{z} - z$ in the Point Cloud Shown in Figure 4.

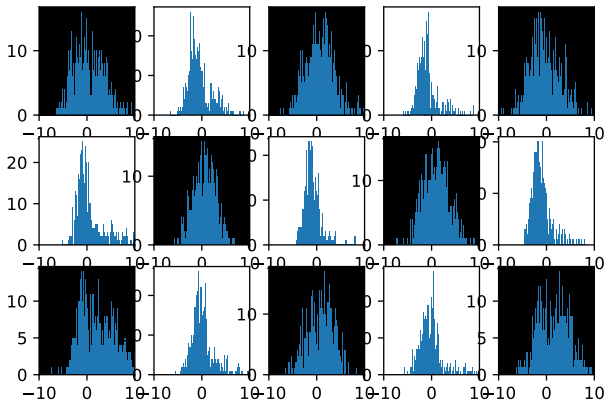


Figure 6. Per-block Distribution of $\hat{z}$. The blocks with black backgrounds correspond to black blocks in the checkerboard pattern.

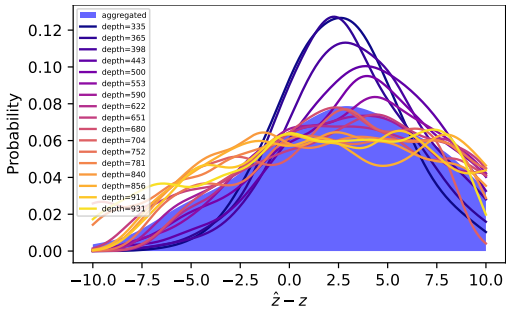
4.3. Variance of Errors at Different Depths

In this section, we take captures while moving the sensor away from the fixed pattern on the wall. For the purpose of visualization, we use a Gaussian kernel with bandwidth 1.0 to estimate a smooth curve of distribution. We report the results at different depths in Figure 7. The curve shifts a little bit to the increasing side of $\hat{z} - z$ at further depths, but is rather stable; this can also be confirmed by the trend of the mode we show in Figure 8.

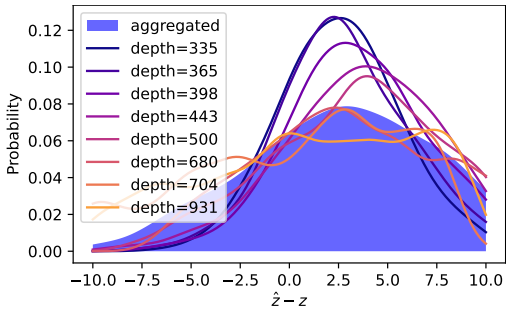
4.4. Variance of Errors with Different Albedo

In this section, we aim to verify whether we can achieve a linear relationship of the mode with respect to the albedo changes. We change the gray value g of black blocks ($g = 0$ for completely black, $g = 255$ for completely white) in the checkerboard patterns from 0 to 255 with step size 16, shown in Figure 9.

Figure 10 and 11 show the error distribution and trend of the mode with linearly varying albedo. We can see an almost linear trend, but the linearity is not strong enough to draw a conclusion. A possible reason may be that the albedo does not linearly vary with the RGB value, which we would discuss in detail in Section 5.2.



(a) Distribution Curves for all the Depths Collected.



(b) Some Selected Curve for Better Visualization.

Figure 7. Change of Error Distribution with Respect to Depth. The blue shaded area represents the estimated distribution by collecting all the errors in different depths.

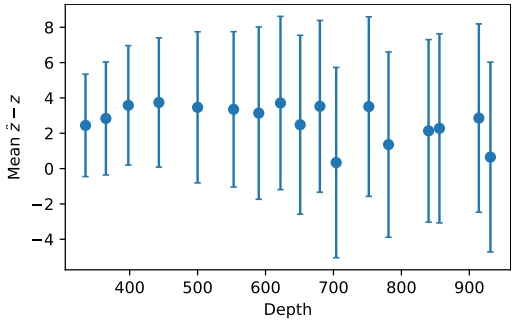
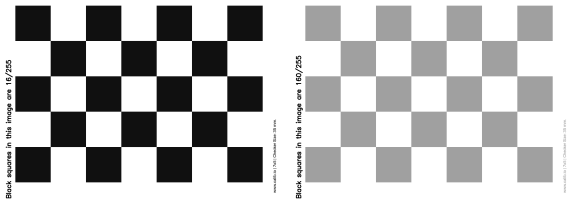


Figure 8. Trend of the $\arg\max \hat{z} - z$. The error bars show one standard deviation of the samples. Small changes with respect to depth can be inferred from the plot.

4.5. Estimated Distribution of ε

A natural question here is, *what if we apply our model and find the distribution of ε anyway?* Figure 12 shows the distribution of $\varepsilon - \varphi$ derived by Equation 2, which is similar across the depths, but completely different from our uncorrelated noise assumption. Moreover, the density of ε is significantly higher around 0, which may suggest that the noise is correlated to the measurement signal.



(a) $g = 16$ (b) $g = 160$

Figure 9. Checkerboard Pattern with Different Gray Value g. We print a series of these modified patterns to simulate linearly varying albedo.

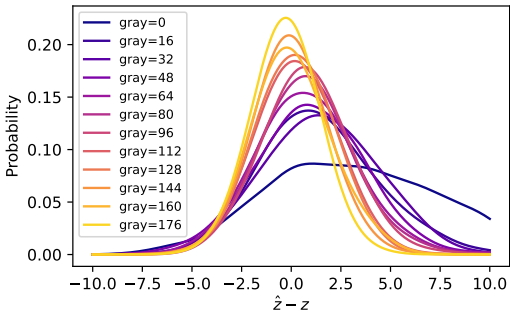


Figure 10. Error Distribution for Patterns with Different Albedo.

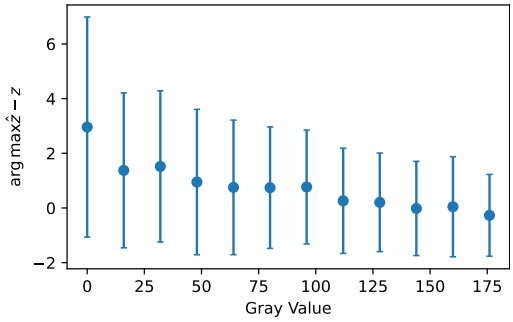


Figure 11. Trend of the Mode with Varying Albedo.

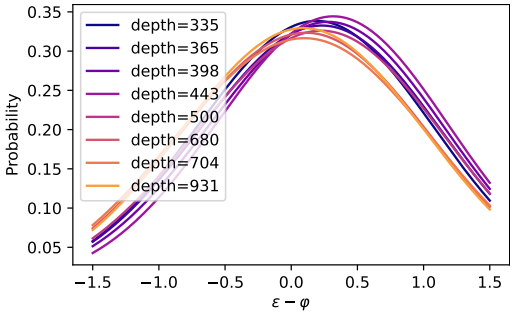


Figure 12. Distribution of $\varepsilon - \varphi$ Derived from Our Observations on Different Depths.

5. Possible Reasons of Failure

5.1. Unideal Camera Setup

For this project, we are hand-holding the camera, and thus cannot get reliable ground truth information. Therefore, we had to resort to checkerboard patterns to estimate camera pose, which is not accurate and may induce unmodelled noise.

Besides, printed patterns unavoidably cause glare, especially for grayscale cameras, and such glare brings about errors in corner identification, shown in Figure 13, and possible noise source in depth estimation. With a fixed camera position and suitable material that reduces glare, we may be able to get more reliable measurement estimation.

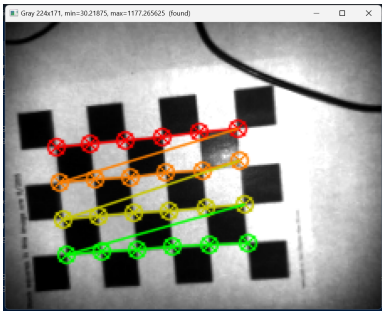


Figure 13. Screenshot of Glare Resulting in Erroneous Calibration. One of the orange corners is off the correct position.

Moreover, we are assuming that the albedo of the black blocks is linear to the gray value. The validity of this assumption depends on the specific printer and ink. We anticipate a more conclusive relationship than Figure 11 when we can have targets with ground-truth albedos.

5.2. Post-processing

Almost all cameras employ some proprietary algorithms to reduce noise and global illumination, including the model we are using. In the internal algorithm of data processing, new noise may be induced and the existing noise may be distorted. We may expect a cleaner observation of the error distribution if we access the raw data captured.

6. Conclusion

This project discusses a theoretical model and some empirical analysis of the noise that occurs in capturing low-albedo regions with continuous-wave time-of-flight sensors. The models don't agree, and we still identify some possible causes of failure on which future work can improve.

Theoretically, uncorrelated noise should lead to a distribution of noisy measurements that is symmetric with two clear spikes, the absolute values of which are the same as the ratio of the amplitude of the signal to the amplitude of the ambient noise. By identifying the spikes, we may estimate the signal strength ratio and then derive the corresponding noises for each observation data point. Then, we can find an empirical distribution of the noise and construct a simulation environment to try out ideas.

However, our empirical results show that the noise distribution is biased towards further depth and unimodal. Moreover, the relationships of the mode in the error to the true depth and albedo are indecisive. If we compute the distribution of the ambient noise phase by our theoretical equations, the results show a strong correlation with the true depth. This may imply noise sources caused by the internal components of the camera, or missing parts in the theory.

Possible reasons causing the discrepancy might be the unideal camera setup, which lacks ground-truth depth and albedo information. Post-processing by the camera manufacturers may also induce or distort noise.

7. Future Steps

In this section, we provide some ideas that we can do immediately, and some ideas that we may not be able to implement.

7.1. More Accurate Camera Setup

Since the dimension of the Pico Flexx is given, we may 3D-print a frame to fix the camera, and thus provide us a ground-truth depth measurement. This further allows us to drop the checkerboard pattern and use a purely black target.

For the board material, as we have already shown that printed paper causes glare, we propose to use alternative materials such as cloth. We hope dyed threads can avoid retroreflections.

7.2. Minimize Post-processing

The work of this project is built upon the Python SDK, which itself consists of wrapped C code. We plan to take a deeper look at all the source codes they provide and see if we can minimize the post-processing steps.

7.3. Ideas not Yet Able to Implement

Varying Albedo. It is non-trivial to measure the albedo of a real-world material or construct those with known albedo.

Controlled or Known Noise. It would be extremely useful if there were some ways to measure the ambient noise or get the ground truth waveform at a particular location.

Single-pixel Continuous Wave Sensor. Figure 12 describing the distribution of ε suggests there may be correlated noise, which might be due to the internal source of cross-pixel inference. It would be most desirable to construct a continuous wave sensor that has only one pixel or sends only one beam of signal at the same time, to minimize the interference.

References

[1] S. Achar, J. R. Bartels, W. L. Whittaker, K. N. Kutulakos, and S. G. Narasimhan. Epipolar time-of-flight imaging. *ACM Transactions on Graphics (ToG)*, 36(4):1–8, 2017. 1

[2] E. Arnold, O. Y. Al-Jarrah, M. Dianati, S. Fallah, D. Oxtoby, and A. Mouzakitis. A survey on 3d object detection methods for autonomous driving applications. *IEEE Transactions on Intelligent Transportation Systems*, 20(10):3782–3795, 2019. 1

[3] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. 4

[4] C. Chen, Q. Chen, J. Xu, and V. Koltun. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3291–3300, 2018. 2

[5] I. Chugunov, S.-H. Baek, Q. Fu, W. Heidrich, and F. Heide. Mask-tof: Learning microlens masks for flying pixel correction in time-of-flight imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9116–9126, 2021. 2

[6] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes*, pages 1–10. 2008. 2

[7] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47(6):2280–2292, 2014. 1

[8] J. P. Godbaz, M. J. Cree, and A. A. Dorrington. Blind deconvolution of depth-of-field limited full-field lidar data by determination of focal parameters. In *Computational Imaging VIII*, volume 7533, pages 92–103. SPIE, 2010. 1

[9] M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin. Phasor imaging: A generalization of correlation-based time-of-flight imaging. *ACM Transactions on Graphics (ToG)*, 34(5): 1–18, 2015. 1, 2

[10] Infineon Technologies. Real3 image sensor family. https://www.infineon.com/dgdl/Infineon-REAL3+Image+Sensor+Family-PB-v01_00-EN.PDF?fileId=5546d462518ffd850151a0afc2302a58. Accessed Mar 30, 2023. 2

[11] R. Lange and P. Seitz. Solid-state time-of-flight range camera. *IEEE Journal of quantum electronics*, 37(3):390–397, 2001. 2

[12] O. Liba, K. Murthy, Y.-T. Tsai, T. Brooks, T. Xue, N. Karnad, Q. He, J. T. Barron, D. Sharlet, R. Geiss, et al. Handheld mobile photography in very low light. *ACM Trans. Graph.*, 38(6):164–1, 2019. 2

[13] T. Mertens, J. Kautz, and F. Van Reeth. Exposure fusion. In *15th Pacific Conference on Computer Graphics and Applications (PG'07)*, pages 382–390. IEEE, 2007. 2

[14] Microsoft Corporation. Azure kinect dk hardware specifications. https://learn.microsoft.com/en-us/azure/kinect-dk/hardware-specification?source=recommendations, . Accessed Mar 30, 2023. 2

[15] Microsoft Corporation. Azure kinect DK depth camera. https://learn.microsoft.com/en-us/azure/kinect-dk/depth-camera, . Accessed Mar 30, 2023. 2

[16] K. Monakhova, S. R. Richter, L. Waller, and V. Koltun. Dancing under the stars: video denoising in starlight. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16241–16251, 2022. 2

[17] E. Olson. Apriltag: A robust and flexible visual fiducial system. In *2011 IEEE international conference on robotics and automation*, pages 3400–3407. IEEE, 2011. 1

[18] PMDTechnologies AG. Camboard pico flexx development kit brief. https://media.automation24.com/datasheet/en/PMD_DevKit_Brief_CB_pico_flexx_CE_V0218.pdf. Accessed Mar 30, 2023. 2

[19] K. Ram Prabhakar, V. Sai Srikar, and R. Venkatesh Babu. Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 2

[20] M. Reynolds, J. Doboš, L. Peel, T. Weyrich, and G. J. Brostow. Capturing time-of-flight data with confidence. In *CVPR 2011*, pages 945–952. IEEE, 2011. 2

[21] A. Scheenstra, A. Ruifrok, and R. C. Veltkamp. A survey of 3d face recognition methods. In *Audio-and Video-Based Biometric Person Authentication: 5th International Conference, AVBPA 2005, Hilton Rye Town, NY, USA, July 20-22, 2005. Proceedings 5*, pages 891–899. Springer, 2005. 1

[22] M. Tancik, V. Casser, X. Yan, S. Pradhan, B. Mildenhall, P. Srinivasan, J. T. Barron, and H. Kretzschmar. Block-NeRF: Scalable large scene neural view synthesis. *arXiv*, 2022. 1

[23] L. Xiao, F. Heide, M. O'Toole, A. Kolb, M. B. Hullin, K. Kutulakos, and W. Heidrich. Defocus deblurring and superresolution for time-of-flight depth cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2376–2384, 2015. 1

[24] S. Xu, R. Wang, and H. Zheng. Road curb extraction from mobile lidar point clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):996–1009, 2016. 1

[25] Y. Zhang, C. Chen, Q. Wu, Q. Lu, S. Zhang, G. Zhang, and Y. Yang. A kinect-based approach for 3d pavement surface reconstruction and cracking recognition. *IEEE Transactions on Intelligent Transportation Systems*, 19(12):3935–3946, 2018. 1