

Denoising 3D Time-Of-Flight Data

Kapil Gupta

Department of Computational Media
University of California, Santa Cruz
kgupta15@ucsc.edu

Yanwen Xu

Department of Computer Science and Engineering
University of California, Santa Cruz
yxu83@ucsc.edu

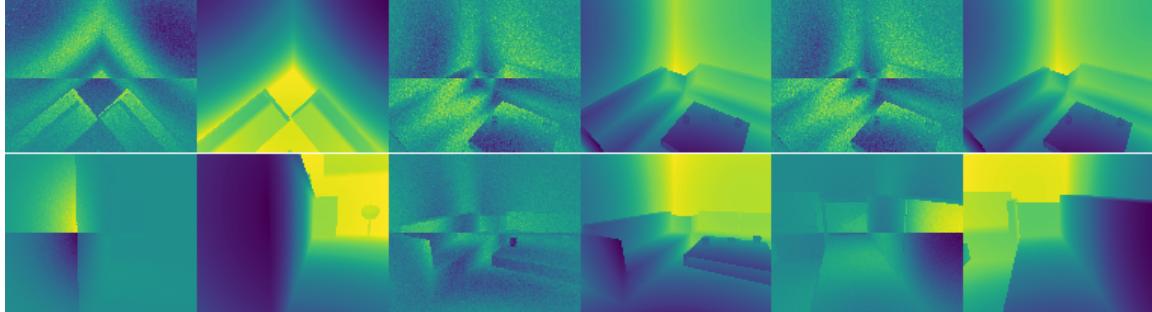


Figure 1: Pairs of False Depth Map with MPI Error and True depth map generated by our model.

ABSTRACT

Depth Perception is widely used in several computer vision tasks. The most common method for depth calculation are range imaging cameras such as Time of Flight (ToF) cameras. However, ToF cameras suffer from non-negligible noise caused by Multi-Path Interference (MPI). While various classical methods exist, machine learning techniques haven't been explored as much due to lack of ground-truth data. In this paper, we propose a novel method for MPI noise removal using a two-part convolutional neural network. The first model learns about the core properties of the reflective objects in the scene, such as the reflectivity of the scene and local ambient light density. The second model learns to map such properties along with a False Depth Map to create a True Depth Map of the scene. We demonstrate and validate our results on synthetic dataset.

CCS CONCEPTS

• Computing methodologies → Computer vision; • 3D imaging;

KEYWORDS

time-of-flight, multi-path interference, deep learning

ACM Reference Format:

Kapil Gupta and Yanwen Xu. 2017. Denoising 3D Time-Of-Flight Data. In *Proceedings of Conference Name*. ACM, New York, NY, USA, 4 pages.
<https://doi.org/10.1145/8888888.7777777>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Conference Name, Conference Date and Year, Conference Location

© 2017 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-1234-5/17/07.

<https://doi.org/10.1145/8888888.7777777>

1 INTRODUCTION

3D imaging systems provide depth information that is crucial in diverse range of computer vision applications. Depth Sensors using Time-of-Flight(ToF) imaging are the most popular amongst such cameras due to cheap cost. Range Imaging and Depth Perception that is available from such cameras can be used for several computer vision tasks such as gesture recognition, image quality improvement, and obstacle avoidance. Nevertheless, ToF cameras suffer from various sources of noise caused by the following: the presence of ambient light, the interference due to different modulating frequencies, the noise caused by multiple reflections, the motion introduced by the subject, and the shot noise by the camera's electronics. In this paper, we focus on the noise caused due to multiple reflections that makes accurate depth calculations difficult.

ToF cameras send a pulsed ray of light and record the time taken for the light ray to reach the sensor. The time taken is used to calculate the distance of an object in the scene from the camera. The camera's depth map acquisition is based upon the assumption that the ray of light is reflected only once in the given scene before reaching the sensor. But objects with different specularity can scatter light differently, causing multiple reflections with different objects present in the scene. The delay caused by multiple reflections causes error in depth calculation. This is an inherently hard problem because the level of MPI error depends on the properties of objects in the scene as well as the ambient lighting of the scene.

Previous solutions [Tomasi and Manduchi 1998; Zhang et al. 2014] to this issue are performing traditional approaches such as Bilateral filtering on the depth image. Bilateral filtering is suitable for Photon Shot Noises because of its nature of taking the average intensity of nearby pixels. However, traditional approaches are not necessarily optimal for MPI noises because of its complexity. Particularly for MPI denoising, attempts have been made, such as modifying the reflection to get away with errors. However, this approach is neither generic nor practical in real-world applications.

Convolutional Neural Networks (CNN) were commonly used in image denoising. Other data-driven approaches [Bolsee and Munteanu 2018; Marco et al. 2017] takes advantage of machine learning, which is to train deep neural networks with synthetic data generated by ToF simulators. While the network can somewhat reduce artifacts, it does not entirely remove them. Moreover, it still reminds uncertain that if the model trained from synthetic data would apply to real-world data.

We propose a novel method for removing noise caused due to Multi-Path Interference. We use a two-part Convolutional Neural Network that is trained on synthetic data. The first CNN is used to learn the essential properties of the scene such as specularity of objects present in the scene, ambient light, and camera position. The second CNN takes the given parameter values from the first model, along with a noisy depth map to generate the true depth map. Our method is based on the observation that MPI noise is directly dependent upon properties of the scene, and hence learning those properties can help determine the depth map with accurate depth information.

We describe our contribution as following:

- A two-stage training strategy, which first extracts important scene parameters from a noisy depth map and then uses those parameters along with a noisy depth map to create a highly-accurate depth map.
- A practical trained network that removes MPI noise from a single ToF image in real-time and outperforms the state-of-art methods.
- A method that is able to process highly specular objects in a scene that may introduce high MPI noise.

2 RELATED WORK

3D cameras send a modulated wave of light that may hit objects in an area and get reflected back with a phase shift in the wave. When the reflected wave of light reaches the camera, the phase shift between the original and reflected wave can be used to get the depth value of the object. When there is a phase shift in the light received by the camera that was not caused by a single reflection with an object, this causes noise/error in the depth calculation, also called MPI noise. This noise can occur due to reflective objects in the scene, that disperses light in a way that multiple ways of light intersect with each other.

2.1 Classical approaches

MPI Error is caused due to multiple photon intersections, leading to an error in measurements. If multiple modulation frequencies of light are used, the time delay operator can be used to process the measurements in the Fourier domain[Ayush Bhandari 2016]. In a controlled setting, $2K+1$ measurements are needed to recover depth information from K interfering rays of light[Ayush Bhandari 2015]. Local and global illumination information has also been used along with the depth information recovered by the ToF camera to overcome MPI based noise[Nikhil Nailk 2015]. Light Transport methods that are based on the idea that direct light follows the epipolar line and indirect light does not have been successfully used to remove indirect light from the scene[Matthew O'Toole

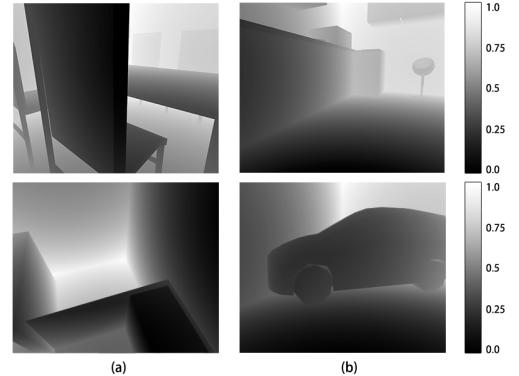


Figure 2: Transient Rendered Scene

2014a]. Further work has been done in separating various transformations indirect light caused by the path of light in the scene in the temporal domain[Matthew O'Toole 2014b] For several other interesting approaches to MPI based noise correction, the reader should refer to the survey paper by Whyte et al. [Refael Whyte 2014].

2.2 Learning based Methods

Due to substantial data requirements for this approach, and lack of available datasets with high-quality ground-truth, Deep Learning approach has not been widely used yet. Qi et al. used Transient Rendering method to generate scenes with varying specular properties and ambient lighting conditions. Along with a simulator for introducing noise into the scenes, Qi et al. created a pipeline method [Guo et al. 2018] to correct various noises simulated by the dataset, including shot noise, multi-path inference noise, noise introduced by dynamic scenes. A two-network adhoc method for depth refinement has been used, where a coarse network first learns the global structure and then another network attempts to remove MPI noise while preserving details. GAN[Agresti et al. 2019] was used to perform domain adaptation from synthetic data to real data after learning to remove MPI noise using a CNN based model. A self supervised method [Vladimiros Sterzentsenko 2019] that uses deep autoencoder along with geometric priors taken from differentiable rendering was also used to remove MPI error. Even though there have been several papers using synthetic data, the performance on the real dataset has not been that great. This is possibly because the statistical patterns of noise seen in real-world data cannot be patterned well by the synthetic dataset.

3 DATASET

We used Nvidia's synthetic data for initial training of our deep learning model. We captured 30 images from different positions of more than 900 different scenes with reflective objects placed in various positions in a room. The scenes were first rendered using transient rendering method[Adrian Jarabo 2014] using same camera FOV as Kinect. The noise was then introduced to these scenes using FLAT's Kinect sensor model. The model is trained on Kinect data and hence understands the statistical properties of noise from the sensor. The model was then retrained on the data

that we collected using a Canesta D350 3D Time-of-Flight camera. We collected 30 images each from 20 different scenes. All scenes had objects with varied specular properties, thus introducing MPI with varying amount. The ambient light was not controlled, providing very close to a real-world conditions.

4 PROPOSED METHOD

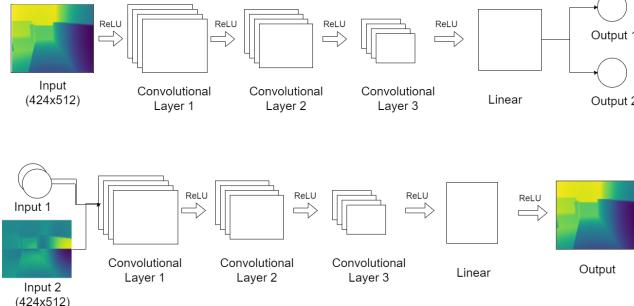


Figure 3: Proposed CNN architecture. The first model takes the false depth map of shape (424x512) as input, and outputs the position of camera and specular value of objects in the scene. The second model takes as input the false depth map as well as the output values of first model and provides the accurate depth map as output.

We propose a two-part convolutional neural network model (Figure 3) to eliminate MPI noise from images captured using a Time-of-flight camera. We demonstrate the use of this model on synthetic data created using the FLAT dataset by Nvidia, which generates Kinect camera-based data.

The first CNN is used to predict scene parameters from a noisy depth map input. The ground-truth parameters are available in the synthetic dataset. The first CNN outputs the correct parameters, which are used as input, along with the noisy depth map, for the second CNN. The input shape of image is (424x512) and the output values are (4x1) vectors. We used three Convolutional layers with kernel size of 5, stride value of 1 and padding value of 2. The output channels were respectively 16, 32 and 64. The convolutional layers were followed by a linear layer that provides vector outputs.

The second CNN is trained to output the True Depth Map. The input to this model are the false depth map image (424x512) and (4x1) parameter vectors. The layers of the model have same hyper-parameters as the first model. Using such a two-part model gives better control over the training, and provides the model with more information to create accurate depth maps.

5 EVALUATION

The first model is able to correctly predict the scene parameters to the accuracy of 66.4%, when the noise (Base Error) present in the data is 0.022050m. The second model is able to predict a True Depth Map using the parameters from the first model as input.

We evaluated our model's output with respect to NVidia's general pipeline for denoising 3D ToF camera data. We also performed comparisons with DeepToF and Phasor. We found that our model is

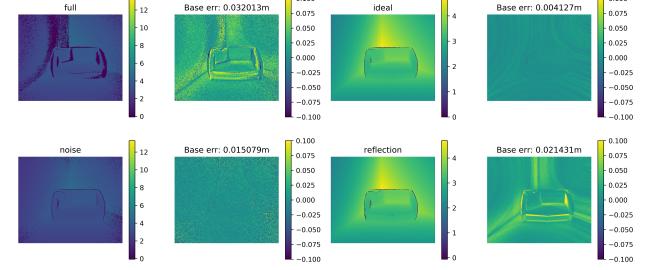


Figure 4: Error introduced in ideal images using Kinect mask. Possible noises can be all possible noises(full), sensor noise(noise) and MPI based noise(reflection).

able to perform better than Nvidia's Static Multi-Reflection Module (Static MRM).

6 LIMITATIONS

The model has been trained on synthetic dataset. The performance on real-dataset is not as good as seen in testing, because of difference in noise statistics between real data and data simulated using the NVidia FLAT simulator. Since model uses noise from kinect camera, results for canesta camera's input depth maps are not of very high accuracy.

We haven't included enough scenes with high specularity, and hence aren't able to handle objects present in a scene with high reflectance such as mirrors or metal surfaces.

7 CONCLUSIONS

Multi-Path Interference based noise is difficult to eliminate because it is difficult to model all possible light paths in a complex scene. We have shown a deep learning technique that can be used to remove MPI based noise using two part deep learning network.

ACKNOWLEDGMENTS

The authors would like to thank

REFERENCES

- Raul Buisan Wojciech Jarosz Diego Gutierrez Adrian Jarabo, Julio Marco1 Adolfo Munoz ~. 2014. A framework for transient rendering. In *SIGGRAPH Asia*.
- Gianluca Agresti, Henrik Schaefer, Piergiorgio Sartor, and Pietro Zanuttigh. 2019. Unsupervised Domain Adaptation for ToF Data Denoising with Adversarial Learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5584–5593.
- Ramesh Raskar Ayush Bhandari. 2016. Signal Processing for Time-of-Flight Imaging Sensors. In *IEEE Signal Processing Magazine*.
- Shahram Izadi Christoph Rhemann Mirko Schmidt Ramesh Raskar Ayush Bhandari, Michal Feigin. 2015. Resolving Multipath Interference in Kinect: An Inverse Problem Approach. In *CVPR*.
- Quentin Bolsee and Adrian Munteanu. 2018. Cnn-based Denoising of Time-Of-Flight Depth Images. In *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE, 510–514.
- Qi Guo, Iuri Frois, Orazio Gallo, Todd Zickler, and Jan Kautz. 2018. Tackling 3D ToF Artifacts Through Learning and the FLAT Dataset. In *The European Conference on Computer Vision (ECCV)*.
- Julio Marco, Quercus Hernandez, Adolfo Munoz, Yue Dong, Adrian Jarabo, Min H Kim, Xin Tong, and Diego Gutierrez. 2017. DeepToF: off-the-shelf real-time correction of multipath interference in time-of-flight imaging. *ACM Transactions on Graphics (ToG)* 36, 6 (2017), 219.
- Kiriakos Kutulakos Matthew O'Toole, John Mather. 2014a. 3D Shape and Indirect Appearance By Structured Light Transport. In *CVPR*.

- Lei Xiao Matthias Hullin Wolfgang Heidrich Kiriakos Kutulakos1 Matthew O'Toole, Felix Heide. 2014b. Temporal Frequency Probing for 5D Transient Analysis of Global Light Transport. In *SIGGRAPH*.
- Christoph Rhemann Shahram Izadi Ramesh Raskar Sing Bing Kang Nikhil Naik, Achuta Kadambi. 2015. A Light Transport Model for Mitigating Multipath Interference in Time-of-flight Sensors. In *CVPR*.
- Michael Cree Adrian Dorington Refael Whyte, Lee Streeter. 2014. Review of Methods for Resolving Multi-path Interference in Time-of-Flight Range Cameras. In *IEEE*.
- Carlo Tomasi and Roberto Manduchi. 1998. Bilateral filtering for gray and color images.. In *Iccv*, Vol. 98. 2.
- Anargyros Chatzitofis Spyridon Thermos Nikolaos Zioulis Alexandros Doumanoglou Dimitrios Zarpalas Petros Daras Vladimiros Sterzentsenko, Leonidas Saroglou. 2019. Self-Supervised Deep Depth Denoising. In *ICCV*.
- Qi Zhang, Xiaoyong Shen, Li Xu, and Jiaya Jia. 2014. Rolling guidance filter. In *European conference on computer vision*. Springer, 815–830.