HiToF: A ToF Camera System for Capturing High-Resolution Textures

Zhiyuan Xie[1], Xiaomin Ouyang[1], Li Pan[1], Wenrui Lu[2], Xiaoming Liu[3], and Guoliang Xing[1,*]
[1]The Chinese University of Hong Kong, [2]University of Michigan, Ann Arbor, [3]Michigan State University
xavier_ie@link.cuhk.edu.hk,xmouyang@link.cuhk.edu.hk,lipan@cuhk.edu.hk
wenruilu@umich.edu,liuxm@cse.msu.edu,glxing@cuhk.edu.hk

ABSTRACT

We present a demonstration of an enhanced Time-of-Flight (ToF) depth system named HiToF, which can expose high-resolution textures from captured depth maps. By design, a ToF camera can easily capture the depth maps of a scene while largely omitting the corresponding texture information, which is often critical for the performance of many depth applications. HiToF is developed to address this issue by generating enhanced depth maps with high-resolution textures. The key idea is to manipulate the phase components used in the measurement of time-of-flight for the received IR light. In this demo, we showcase our implementation using off-the-shelf ToF cameras and engage audience with an interactive experience in various scenarios, which illustrates the system's effectiveness in improving the performance of ToF cameras in depth applications.

CCS CONCEPTS

• **Computer systems organization → Sensors and actuators**; • **Computing methodologies → Computer vision**.

KEYWORDS

ToF depth camera, Texture exposure, Depth data processing, Image enhancement

ACM Reference Format:
Zhiyuan Xie[1], Xiaomin Ouyang[1], Li Pan[1], Wenrui Lu[2], Xiaoming Liu[3], and Guoliang Xing[1,*]. 2022. HiToF: A ToF Camera System for Capturing High-Resolution Textures. In *The 28th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '22), October 17–21, 2022, Sydney, NSW, Australia.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3495243.3558747

1 INTRODUCTION

Time-of-Flight (ToF) depth cameras capture the depth information of the scene by measuring the time-of-flight of IR light. It has attracted significant attention in various real-world applications. For example, ToF camera is widely used together with RGB camera in human detection and tracking [4], daily activity recognition [8, 9], pose estimation [3, 7], etc., where the distance information obtained from ToF cameras is fused with detailed textures from RGB cameras. However, due to the increasing privacy concerns, RGB cameras may not be an option in many privacy-sensitive scenarios. Although depth-only systems are gaining popularity [5, 10], their sensing performance is often unsatisfactory in real-world complex environments such as those with extensive occlusion [11], due to the lack of detailed textures.

To address these issues, we propose HiToF, a pure depth system that utilizes the phase components from ToF measurements to generate high-quality and tunable maps for enhancing the performance of ToF cameras in various depth applications. Specifically, HiToF can provide detailed texture information of the scene without using RGB cameras, which can largely alleviate privacy concerns. Moreover, HiToF can precisely and dynamically adjust the distance range of interest where detailed textures need to be enhanced. This ability can not only improve the performance of various depth applications but also protect sensitive information outside the enhanced area. Moreover, HiToF can expose textures in very challenging settings, such as long distance, dynamic range, and overexposure.

The design of HiToF is based on the key idea that the phase components of ToF measurements can be manipulated (which we refer to as "phase manipulation") to generate high-quality maps. The distance measurement of most ToF cameras relies on two critical phase values of the camera's receiver circuit, i.e., the In-phase (I) and Quadrature-phase (Q) measurements. Our key observation is that, sophisticated manipulation on the I/Q phase maps besides calculating the time-of-flight is promising to generate more texture information of the scene.

2 SYSTEM DESIGN

ToF cameras obtain the depth measurement of an entire scene by transmitting an IR light and measuring the time needed for the IR light to reach targets in the scene and return. Specifically, the time-of-flight t can be calculated using the in-phase (I) and (Q) components [6] of ToF measurements for most ToF cameras in the market:

$$t = T \cdot \tan^{-1}(\frac{Q}{I})/(2\pi),$$

where T is the period of the modulated IR.

The manipulation on I/Q components will result in value changes in the calculated depth maps. Here we apply an affine transformation to do the manipulation because affine transformation can help expose textures of a certain range of the scene in the calculated depth map. The transformed I/Q components can be written as:

$$\begin{bmatrix} Q' \\ I' \end{bmatrix} = \mathbf{A}\left(\begin{bmatrix} Q \\ I \end{bmatrix} + \begin{bmatrix} Q_s \\ I_s \end{bmatrix} \right),$$

where A is a 2×2 matrix, (Q_s, I_s) is the vector representing the translation transform, and (Q', I') is the I/Q components after

---

*Corresponding author.

ACM MobiCom '22, October 17–21, 2022, Sydney, NSW, Australia

ACM ISBN 978-1-4503-9181-8/22/10.
https://doi.org/10.1145/3495243.3558747

transformation. Such an affine transformation is applied to each pixel of the depth map and changes the depth measurement of each pixel. Although the affine transformation is identical for all pixels, the depth change is not uniform across the entire image, because it not only depends on the reflectivity of objects (which leads to texture exposure of objects), but also is determined by the actual distance of objects (which can be controlled to expose objects within a certain distance range). Therefore, our proposed system features a novel approach called *phase manipulation*, which exploits an effective affine transformation on the in-phase and quadrature (I/Q) components of ToF measurements to enhance the detailed texture within a certain distance range of interest in depth maps generated by ToF cameras.

For those ToF cameras from which we cannot obtain I/Q components directly from APIs, we can extract I/Q data from the Depth and IR maps using the following formulas:

$$\mathbf{I} = \mathbf{R}\cos\left(4\pi f \mathbf{D}/c\right), \quad \mathbf{Q} = \mathbf{R}\sin\left(4\pi f \mathbf{D}/c\right),$$

where f is the modulation frequency of the ToF camera, c denotes the speed of light, $\mathbf{R}$ and $\mathbf{D}$ are the IR and depth maps, respectively. Preliminary experiments indicate that both I/Q acquisition approaches have similar performance in enhancing ToF cameras.

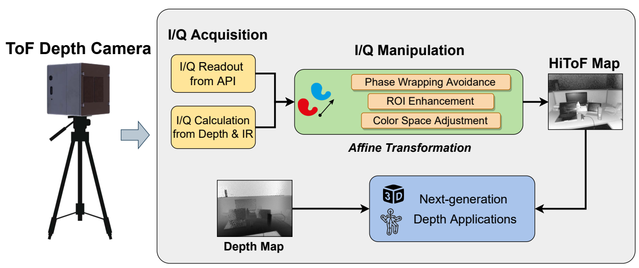

Figure 1: System Overview of HiToF.

Figure 1 depicts the system architecture of HiToF. Unlike other depth camera systems that read out depth maps, HiToF reads out the I/Q raw data directly or calculates I/Q from the corresponding depth map and IR map, which can be applied to almost all the ToF cameras on the market. After the acquisition of I/Q data, the I/Q manipulation approach is applied to enhance the detailed texture within the specified distance range of interest for different depth application scenarios. Moreover, this approach is extremely lightweight and thereby can be conducted in real-time by low-power embedded platforms.

3 DEMONSTRATION

We implement the proposed system with a DepthEye ToF depth camera [1] and an edge device Nvidia Jetson Xavier [2]. The DepthEye camera is developed based on IMX556PLR CMOS from Sony, from which we can easily obtain the in-phase (I) and quadrature (Q) components via APIs. The Nvidia Jetson Xavier processes the obtained I/Q components to generate enhanced depth maps. We implement HiToF using Python in Ubuntu 18.04. The basic hardware setup of the system is shown in Figure 2.

The demonstration consists of a ToF camera, an edge device, and a monitor. During the demonstration, the ToF camera will capture the scene and people in front of the ToF depth camera. The edge device will process the captured data, and the monitor will

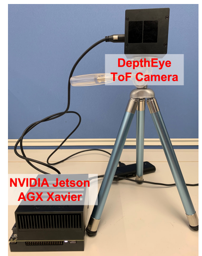

Figure 2: Setup.

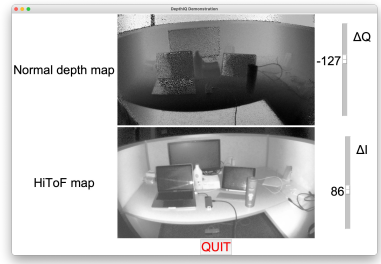

Figure 3: HiToF UI

display the enhanced depth map in real-time. Specifically, a normal depth stream and a HiToF stream will be shown simultaneously in the HiToF UI (Figure 3). Moreover, the parameters Q_s and I_s for generating different enhanced depth maps can be adjusted. This indicates that the enhanced depth maps are tunable and are fit for a wide range of application scenarios. The audience can interact with this demo by setting different manipulation parameters and observing the resulted outputs displayed in the UI. Furthermore, the real-time demonstration shows that our system does not incur additional overhead compared with off-the-shelf ToF cameras.

ACKNOWLEDGEMENT

This work is funded in part by the Research Grants Council of Hong Kong under General Research Fund grants 14209619 and 14203420

REFERENCES

[1] 2022. DepthEye Wide. https://www.seeedstudio.com/DepthEye-Wide-ToF-Camera-with-Sony-IMX556PLR-DepthSense-p-4809.html.
[2] 2022. Nvidia Jetson Xavier. https://developer.nvidia.com/embedded/jetson-agx-xavier-developer-kit.
[3] Pratik Chattopadhyay, Aditi Roy, Shamik Sural, and Jayanta Mukhopadhyay. 2014. Pose Depth Volume extraction from RGB-D streams for frontal gait recognition. *Journal of Visual Communication and Image Representation* 25, 1 (2014), 53–63.
[4] Jungong Han, Eric J Pauwels, Paul M de Zeeuw, and Peter HN de With. 2012. Employing a RGB-D sensor for real-time tracking of humans across multiple re-entries in a smart environment. *IEEE Transactions on Consumer Electronics* 58, 2 (2012), 255–263.
[5] Carlos A Luna, Cristina Losada-Gutierrez, David Fuentes-Jimenez, Alvaro Fernandez-Rincon, Manuel Mazo, and Javier Macias-Guarasa. 2017. Robust people detection using depth information from an overhead time-of-flight camera. *Expert Systems with Applications* 71 (2017), 240–256.
[6] Ahmed Makki, Abubakr Siddig, Mohamed Saad, and Chris Bleakley. 2015. Survey of WiFi positioning using time-based techniques. *Computer Networks* 88 (2015), 218–233.
[7] Xiaomin Ouyang, Zhiyuan Xie, Jiayu Zhou, Jianwei Huang, and Guoliang Xing. 2021. Clusterfl: a similarity-aware federated learning system for human activity recognition. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*. 54–66.
[8] Xiaomin Ouyang, Zhiyuan Xie, Jiayu Zhou, Guoliang Xing, and Jianwei Huang. 2022. ClusterFL: A Clustering-based Federated Learning System for Human Activity Recognition. *ACM Transactions on Sensor Networks (TOSN)* (2022).
[9] Linlin Tu, Xiaomin Ouyang, Jiayu Zhou, Yuze He, and Guoliang Xing. 2021. Feddl: Federated learning via dynamic layer sharing for human activity recognition. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 15–28.
[10] Zhiyuan Xie, Xiaomin Ouyang, Xiaoming Liu, and Guoliang Xing. 2021. UltraDepth: Exposing High-Resolution Texture from Depth Cameras. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 302–315.
[11] Mao Ye, Qing Zhang, Liang Wang, Jiejie Zhu, Ruigang Yang, and Juergen Gall. 2013. A survey on human motion analysis from depth data. In *Time-of-flight and depth imaging. sensors, algorithms, and applications*. Springer, 149–187.