



(51) International Patent Classification:
G06T 15/04 (2011.01)

(21) International Application Number:
PCT/CN2023/113154

(22) International Filing Date:
15 August 2023 (15.08.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/508,741 16 June 2023 (16.06.2023) US

(71) Applicant: **THE CHINESE UNIVERSITY OF HONG KONG** [CN/CN]; Shatin, New Territories, Hong Kong 999077 (CN).

(72) Inventors: **XING, Guoliang**; c/o The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong (CN). **XIE, Zhiyuan**; c/o The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong (CN). **OUYANG, Xi-aomin**; c/o The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong (CN).

(74) Agent: **BEYOND TALENT PATENT AGENT FIRM**; 1005A, 9th Floor, Building 1, No. A-3, Yongandongli, Chaoyang District, Beijing 100022 (CN).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA,

(54) Title: TIME-OF-FLIGHT CAMERA SYSTEM AND METHODS OF PHASE COMPONENTS MANIPULATION FOR TEXTURE EXPOSURE IN TIME-OF-FLIGHT SYSTEM

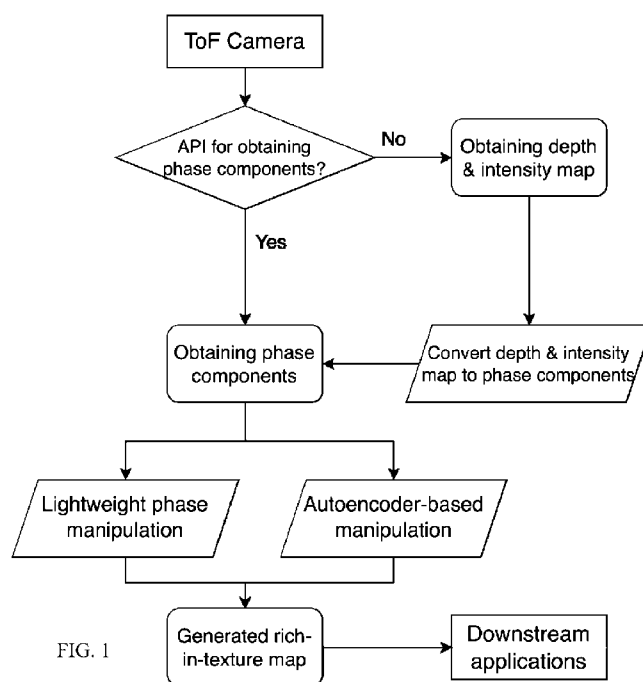


FIG. 1

(57) Abstract: A time-of-flight (ToF) method and systems are provided. The system includes an off-the-shelf indirect time-of-flight (iToF) depth camera configured to provide ToF phase components; a computing platform; a data connector connecting the iToF depth camera to the computing platform. The computing platform includes a driver in communication with the iToF depth camera to control the iToF depth camera for data collection; and one or more processors and co-processors having stored methods and instructions configured for phase components pre-processing, transformation, and manipulation to effectively expose high-resolution textures of scenes under different light conditions, including scenes having details captured in dark scenarios; and a storage media for storing results of phase manipulation; and/or a monitor for demonstrating the results of phase manipulation in real time; and/or a downstream application

NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO,
RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH,
TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS,
ZA, ZM, ZW.

- (84) Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

running on the computing platform that has a manipulated texture map as an input.

TIME-OF-FLIGHT CAMERA SYSTEM AND METHODS OF PHASE COMPONENTS MANIPULATION FOR TEXTURE EXPOSURE IN TIME-OF-FLIGHT SYSTEM

5

CROSS-REFERENCE TO RELATED APPLICATION

This application claims the benefit of U.S. Provisional Patent Application Serial No. 63/508,741, filed June 16, 2023, which is hereby incorporated by reference in its entirety including any tables, figures, or drawings.

10

BACKGROUND OF THE INVENTION

1. Introduction

Sensing in low-light and dark environments has a wide range of applications, such as smart building, smart health, and robot navigation. For example, a highly desirable feature of smart door locks is automatic unlock via face recognition or secret hand gestures in dark environments [31, 43]. Moreover, many health monitoring systems and human activity recognition systems [52, 54, 62] require 7/24 sensing capabilities, for example, detecting sudden infant death syndrome (SIDS) during sleep using a smart baby monitor [63].

As summarized in Table 1, although there already exist many sensing technologies that function to some extent in dark conditions, they cannot meet the requirements of high-resolution sensing applications. RF-based systems such as mmWave radar and Wi-Fi are not interfered with by visible light. Unfortunately, their sensing data are highly sparse [46, 58, 72], making them poorly suited for applications that require high-resolution results such as human faces and hand gestures. Thermal, IR, and depth cameras can work in the dark. However, thermal cameras have limited resolution [28]. Although IR cameras can provide more detailed information, they rely on strong IR emissions, which incur high power consumption ranging from 5 to 20W [15, 61]. Moreover, off-the-shelf commercial IR cameras suffer from the over-exposure effect when objects are too close and can only capture image details within a short distance, for example, up to 5m [71], due to the fast decay of light intensity [65]. Time-of-flight (ToF) depth cameras have a more extended range and lower power consumption and are increasingly embedded in smartphones or used as standalone sensors for 3D applications. However, by design, ToF depth cameras cannot capture most texture information of the scene [66].

Sensing Technologies	Modality	Work in the Dark?	Cost	Power	Noise	Typical Range (with texture details)	Texture Resolution
RGB Camera	Visible Light	No	Low	Low	High	5m	Low
mmWave Radar	Radio Waves	Yes	Medium	low	Very High	0m	Very Low
Thermal Camera	Longwave Infrared	Yes	High	High	Medium	5m	Low
IR Camera	Near Infrared	Yes	Medium	High	Medium	5m	Medium
ToF Camera	Near Infrared	Yes	Medium	Low	High	5m	Medium
<i>Mozart (Ours)</i>	Near Infrared	Yes	Medium	Low	Low	10m	High

Table 1: Comparison of various technologies for sensing in the dark.

2. Previous Related Work

2.1 Sensing Technologies for Applications in the Dark

Sensing in the dark has a wide range of applications such as robot navigation, face authentication, gesture recognition, and surveillance [43, 46, 47, 68]. Most of the current approaches in this area are based on vision or RF sensors. RGB camera is a ubiquitous vision system that cannot work in dark conditions [59]. Other vision sensors such as thermal, IR, and depth cameras have shortcomings such as low resolution [28], high power consumption [15, 61], and limited texture exposure [66]. In particular, the intensity maps [39, 42] collected by ToF cameras are essentially the IR maps collected by IR cameras. RF-based sensing technologies such as mmWave radar and Wi-Fi are not interfered with by visible light. Unfortunately, their sensing results are highly sparse [46, 58, 72].

2.2 Image Enhancement in Low-light Conditions

There have been extensive efforts to improve the RGB image quality in low-light conditions. A multi-task learning framework is proposed in [22] to explore the intrinsic pattern behind illumination translation for object detection in poor light environments. Moreover, in [35], thermal images are synthesized from RGB images with a Generative Adversarial Network to enable monitoring in low-light conditions. Recently, several studies have formulated the light enhancement of RGB images as a deep curve estimation problem, which requires the reference images as the input [29]. These approaches cannot work in fully dark conditions without any illumination. Moreover, designed for RGB cameras, they cannot be directly applied to ToF cameras due to the fundamental differences between the two sensor modalities.

2.3 ToF Augmentation

A family of techniques has been proposed for depth image enhancement, most focused on improving the noise models of ToF cameras for accurate distance measurement [40, 55, 70]. An energy-efficient epipolar imaging approach is proposed in [14] to improve

the robustness of depth measurement in extreme scenarios, and the centimeter-wave and interferometric imaging are utilized in [16] to enhance the precision of iToF cameras. A recent work [66] illustrates the feasibility of extracting rich textures from depth maps. However, it requires additional hardware, such as an external IR emitter and distorts depth measurements during texture exposure, making it incompatible with current depth-based applications.

BRIEF SUMMARY OF THE INVENTION

Embodiments of the subject invention pertain to a method and systems for manipulating phase components of time-of-flight (ToF) measurements to expose textures in the ToF system.

According to an embodiment of the subject invention, a time-of-flight (ToF) system comprises an off-the-shelf indirect time-of-flight (iToF) depth camera configured to provide ToF phase components; a computing platform; a data connector connecting the iToF depth camera to the computing platform; and wherein the computing platform comprises a driver in communication with the iToF depth camera to control the iToF depth camera for data collection; and one or more processors and co-processors having stored methods and instructions configured for phase components pre-processing, transformation, and manipulation to effectively expose high-resolution textures of scenes under different light conditions, including scenes having details captured in dark scenarios; and a storage media for storing results of phase manipulation; and/or a monitor for demonstrating the results of phase manipulation in real time; and/or a downstream application running on the computing platform that has a manipulated texture map as an input. The off-the-shelf iToF depth camera comprises a modulator to modulate emitted infrared light, an emitter to emit IR light, a receiver to collect reflected IR light by objects in the scenes, and a demodulator to determine phase shift between the received light and the emitted light. The phase shift is calculated by the amount of light received in two successive time windows, being defined as the phase components, or the equivalent phase components are calculated to denote the phase shift. Moreover, the iToF depth camera is powered by either the data connector connected to the computing platform or by a corresponding power adapter. The computing platform is powered by a power adapter. The driver is configured to control the iToF camera to measure distances of the scenes and output raw phase components to the processor.

In another embodiment of the subject invention, a time-of-flight (ToF) method comprises a data transformation step for obtaining input data; a step for performing phase

manipulation to generate rich-in-texture maps of scenes from phase components; an autoencoding step with a learning objective for automatically learning an efficient representation of the rich-in-texture information in the scenes from the phase components. The method may further comprise a step of converting a depth map and an intensity map to equivalent phase components. The method may further comprise a step with a judgement criterion for checking whether operations on the phase components generate the rich-in-texture maps. Moreover, the method may further comprise an illumination compensation step to compensate for uneven brightness distributions of entire map. The method may further comprise an outlier redistribution step for eliminating influence of abnormal pixels with total reflection. The method may further comprise an end-to-end autoencoder-based step to automatically learn the texture representation, comprising: an autoencoding step having phase components as the input, and the generated rich-in-texture map as an output; a step having three sub-objectives configured to guide the autoencoding step to effectively learn texture representations; and a step having three tunable coefficients to balance influence weight of different optimization objectives on overall texture exposure.

In another embodiment of the subject invention, a computer program product is provided, comprising a non-transitory computer-executable storage device having computer readable program instructions embodied thereon that when executed by a computer cause the computer to perform time-of-flight (ToF) method for texture exposure in TOF systems, the computer-executable program instruction comprising: a data transformation step for obtaining input data; a step for performing phase manipulation to generate rich-in-texture maps of scenes from phase components; and an autoencoding step with a learning objective for automatically learning an efficient representation of the rich-in-texture information in the scenes from the phase components. The computer program product may further comprise a step of converting a depth map and an intensity map to equivalent phase components. The computer program product may further comprise a step with a judgement criterion for checking whether operations on the phase components generate the rich-in-texture maps. The computer program product may further comprise an illumination compensation step to compensate for uneven brightness distributions of entire map. The computer program product may further comprise an outlier redistribution step for eliminating influence of abnormal pixels with total reflection. The computer program product may further comprise an end-to-end autoencoder-based step to automatically learn the texture representation, comprising: an autoencoding step having phase components as the input, and the generated rich-in-texture map as an output; a step having three sub-objectives configured to guide the autoencoding

step to effectively learn texture representations; and a step having three tunable coefficients to balance influence weight of different optimization objectives on overall texture exposure. The autoencoding step is configured to generate high-resolution texture maps.

5

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is the workflow of the Mozart ToF methods and system, according to an embodiment of the subject invention.

Figure 2 shows schematic representations of applications of the Mozart ToF methods and system in the dark sensing scenarios, wherein in addition to the depth maps from ToF cameras, the Mozart ToF methods and system also generates high-resolution and rich-in-texture maps, which can significantly enhance the performance of sensing tasks in the dark, according to an embodiment of the subject invention.

Figures 3A-3B are schematic representations of ToF depth cameras obtain depth maps by emitting and receiving the IR light to calculate the time of flight, which is unaffected by the ambient light, wherein Figure 3A shows capturing depth maps of the scene with a ToF camera; and wherein Figure 3B shows principles of iToF depth cameras, according to an embodiment of the subject invention.

Figures 4A-4B show that compared to the depth map, the values of phase components have larger fluctuations across the face and adding a shift to $N1$ when calculating depth based on Equation (1) leads to an image with fine-grained textures, wherein Figure 4A shows the original depth map and values of phase components; and wherein Figure 4B shows the new depth map obtained by shifting $N1$, according to an embodiment of the subject invention.

Figures 5A-5B show the IR maps collected by IR cameras and ToF cameras both suffer over-exposure for near objects and under-exposure for distant objects, wherein Figure 5A shows IR maps collected by IR cameras; and wherein Figure 5B shows IR maps collected by ToF cameras, according to an embodiment of the subject invention.

Figure 6 is a schematic representation of the Mozart ToF methods and system designed based on the physics models for exposing and enhancing textures through phase manipulation, wherein the high-resolution textures can be generated by both highly compute-efficient phase manipulation functions and an autoencoder-based approach, according to an embodiment of the subject invention.

Figure 7A shows Lambertian reflection model illustrating that the reflected IR intensity is determined by both reflectivity and incidence angle of light and these two factors

together form the textures of objects, which is referred to as albedo β herein, and **Figure 7B** shows the mapping functions that are albedo-monotonic can expose textures of the scene and vice versa, according to an embodiment of the subject invention.

Figure 8A shows that in a typical texture map, the near region is brighter while the far is dark and **Figure 8B** shows that the averaged intensity of received light decreases drastically with the distance, according to an embodiment of the subject invention.

Figure 9 shows redistribution of outliers reduces the influence of total reflection and reveals more textures, according to an embodiment of the subject invention.

Figures 10A-10C show that the functions with large polynomial degrees turn normal values into outliers, resulting in over-exposure, according to an embodiment of the subject invention.

Figure 11 is a schematic representation of the autoencoder that learns efficient representations from $N1$, $N2$ maps to generate robust Mozart ToF maps for different applications, wherein the loss functions are designed based on the physics models in Section 6.1, according to an embodiment of the subject invention.

Figures 12A-12B show implementation of the Mozart ToF methods and system with smartphones and standalone ToF cameras, wherein Figure 12A shows real-time Android App; and wherein Figure 12B shows three ToF modules, according to an embodiment of the subject invention.

Figures 13A-13B show system overhead on smartphone and edge platforms, wherein Figure 13A shows generating depth, Mozart ToF maps, and inference with Mozart ToF maps on smartphones; and wherein Figure 13B shows generating maps on Jetson Xavier, according to an embodiment of the subject invention.

Figures 14A-14C show experiment settings of different datasets, wherein the photos are taken using iPhone XR camera with default mode, wherein Figure 14A shows human tracking; wherein Figure 14B shows face recognition; and wherein Figure 14C shows gesture recognition, according to an embodiment of the subject invention.

Figures 15A-15C show overall accuracy when the Mozart ToF methods and system, the Mozart-manual methods and system, and baseline methods and system are performed on different datasets, wherein both the Mozart ToF methods and system and Mozart-manual consistently outperform the baselines, wherein Figure 15A shows human tracking, wherein Figure 15B shows face recognition, and wherein Figure 15C shows gesture recognition, according to an embodiment of the subject invention.

Figure 16 shows performance of the Ablation Study, according to an embodiment of the subject invention.

Figure 17 shows performance comparison in different environments, light conditions and distances, respectively, according to an embodiment of the subject invention.

5

DETAILED DISCLOSURE OF THE INVENTION

Embodiments of the subject invention are directed to a time-of-flight (ToF) method and systems.

The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention. As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items. As used herein, the singular forms “a,” “an,” and “the” are intended to include the plural forms as well as the singular forms, unless the context clearly indicates otherwise. It will be further understood that the terms “comprises” and/or “comprising,” when used in this specification, specify the presence of stated features, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, steps, operations, elements, components, and/or groups thereof.

Unless otherwise defined, all terms (including technical and scientific terms) used herein have the same meaning as commonly understood by one having ordinary skill in the art to which this invention pertains. It will be further understood that terms, such as those defined in commonly used dictionaries, should be interpreted as having a meaning that is consistent with their meaning in the context of the relevant art and the present disclosure and will not be interpreted in an idealized or overly formal sense unless expressly so defined herein.

When the term “about” is used herein, in conjunction with a numerical value, it is understood that the value can be in a range of 90% of the value to 110% of the value, i.e. the value can be +/- 10% of the stated value. For example, “about 1 kg” means from 0.90 kg to 1.1 kg.

The Mozart ToF methods and system of the subject invention generate high-resolution texture maps entirely based on device sensing data processing. As a result, they not only can be implemented on mainstream off-the-shelf ToF cameras and ToF-enabled smartphones, but also can obtain high-resolution texture maps and depth maps simultaneously.

Further, the Mozart ToF methods and system of the subject invention are designed to generate rich-in-texture maps with details using iToF depth cameras. The exposed textures from iToF cameras can be used to enhance the performance of sensing tasks in the dark scenarios and enable more sensing applications of iToF cameras. The methods can precisely control the degree of texture exposure for various application scenarios.

The time-of-flight camera methods and system adopt methods of phase components manipulation for texture exposure in the time-of-flight system.

The system comprises an iToF depth camera, a computing platform with the processors, co-processors, memory and other computing resources, and a data connector such as a data cable connecting the ToF camera and the computing platform for data transmissions. A driver for the ToF camera and a series of processing methods are designed to run on the computing platform. The invention can be formed by all systems that contain these modules, including standalone ToF modules and mobile devices with a ToF module, such as a mobile phone or VR headset.

A series of physical models and design principles are provided to connect the software and hardware layers, including a texture exposure model, illumination compensation for texture enhancement, and redistribution on total reflection outliers for texture enhancement. Through analysis of the Lambertian reflection model, a texture exposure principle is provided such that the phase manipulation result must be monotonous relative to the albedo. Illumination compensation for texture enhancement balances the problem of uneven brightness displayed by objects at different distances by eliminating the distance parameter in the phase component after manipulation. Redistribution of total reflection for texture enhancement depresses the value of outliers through a redistribution function, such that most textures are enhanced and clearer.

The method includes several modules, including a data pre-processing module, a lightweight phase manipulation module, and an autoencoder-based phase manipulation module.

Specifically, the data pre-processing module converts data from the ToF camera into unified phase components for use in the next phase of manipulation. The lightweight phase manipulation module pre-designs a series of lightweight phase manipulation functions based on the physical model and texture exposure and enhancement principles to generate detailed texture maps. The autoencoder-based phase manipulation module adopts a self-supervised learning method to automatically learn efficient texture representation from phase components, enabling the generation of high-quality texture maps in various scenarios.

As shown in Figure 2, Mozart ToF maps can significantly enhance the performance of various sensing tasks in the dark. The design of the Mozart ToF methods and system is based on the key observation that the phase components of ToF measurements can be carefully controlled (which is referred to as “phase manipulation”) to generate high-resolution maps with detailed textures [67]. To design the Mozart ToF methods and system, an in-depth analysis of the physical reflection model is provided for exposing texture information through phase manipulation. The key finding is that the textures can be exposed and enhanced using highly compute-efficient phase manipulation functions. Accordingly, an end-to-end autoencoder-based unsupervised learning approach is employed to automatically learn efficient representations from the phase component maps to generate Mozart ToF maps. To train the deep autoencoder, three novel loss functions are designed by exploiting the physics models developed, including the albedo similarity loss, the illumination attenuation loss, and the uniform distribution loss.

In the following section, the basic principles of ToF sensing and study of the impact of phase component manipulation are provided and the manipulated maps are compared with the IR maps.

3.

3.1 Principles of ToF Depth Sensing

3.1.1 Depth Measurement from Time-of-flight

As shown in Figure 3A, a ToF depth camera emits IR light, illuminates the scene to be captured, and receives the IR light reflected by the objects in the scene. The distance is measured based on the fact that the round-trip time-of-flight (t) of the IR signal between the scene and the camera is strictly proportional to the distance. Specifically, $t = 2d/c$, where d is the distance of the scene and c is the speed of light. Off-the-shelf ToF cameras fall into two categories based on how the time-of-flight is measured: direct Time-of-Flight (dToF) and indirect Time-of-Flight (iToF). Compared with dToF, iToF is more suitable for 3D imaging applications due to its low cost and high-resolution [66]. Most of the ToF modules on mobile devices, especially Android smartphones, in the current market adopt the iToF technology [21, 23]. Moreover, the iToF camera is expected to account for the major share of the global ToF market in the next decade [11]. The Mozart ToF methods and system are designed to work with iToF cameras, and all ToF cameras in this paper refer to iToF cameras unless otherwise indicated.

3.1.2 Measuring ToF Based on Received Signal Phase

The iToF camera has two successive windows (in-phase and quadrature) to receive the reflected light and uses the phase shift of the returned light to calculate the time of flight.

Figure 3B illustrates the general principle of calculating the phase shift in a ToF camera. The time of flight, t , can be calculated by Equation (1):

$$t = \frac{N_2}{N_1 + N_2} \cdot T_p, \quad (1)$$

where T_p is the width of the pulse, N_1 , N_2 are the amount of received light in successive in-phase and quadrature windows, which are referred to as phase components.

Besides the basic designs, mainstream off-the-shelf iToF cameras also adopt several advanced techniques to mitigate the influence of ambient light on distance measurement. For example, the continuous-wave iToF cameras take multiple samples per measurement, that is, using more than two windows to calculate the phase shift, which can reduce the energy offset caused by ambient light during the process of each distance measurement [27]. For different types of iToF cameras, the equivalent N_1 , N_2 can always be obtained from their raw phase components.

3.2 Motivation Study

3.2.1 Impact of Phase Components

As shown in Equation (1), the depth measurements are calculated from the phase components (N_1 , N_2). Moreover, calculating the depth of a point in the scene is equivalent to the dimension reduction from 2-D (N_1 , N_2) to 1-D distance d , which inevitably loses other information such as textures. In other words, the phase components contain more information about the captured scene than the depth measurement. This key observation provides opportunities for exposing detailed texture in the calculated map.

To validate this observation, the depth measurements and phase components are compared for the points across a vertical line of a human face in Figure 4A, where the blue curve denotes the normalized depth values, and the orange and green curve denote N_1 and N_2 components, respectively. It is observed that the depth values are less volatile, while the phase components (N_1 , N_2) fluctuate drastically. Moreover, for two points A and B with the same distance from the ToF camera, they have totally different phase components (N_1 , N_2).

Therefore, by exploiting such information encoded in the phase components, it is possible to show detailed texture information of different points in the scene.

Next, a simple manipulation operation can be conducted on the phase components to show the feasibility of exposing more texture information about the scene. Specifically, a slight shift is added to $N1$ when calculating depth using Equation (1). Figure 4B shows the resulting depth map, which exhibits significantly finer-grained textures because shifting $N1$ is equivalent to physically adding a well-modulated interfering signal. Next, a commonly used object detection model [13] is applied to the original depth maps and the new maps with simple phase component manipulation. The results show that the human detection rate increases from less than 20% to more than 90%. The result clearly shows the great potential of exposing high-resolution textures from ToF cameras using phase manipulation. Moreover, the generated maps can significantly improve the performance of perception tasks, especially in the dark.

3.2.2 Limitations of the Existing IR-based Methods

IR-based techniques are mainstream solutions for providing detailed textures and sensing in the dark. Most iToF cameras provide intensity maps [39, 42], which are essentially the IR maps collected by IR cameras. However, IR/intensity maps represent the amplitude of received IR signals and cannot fully expose texture information because other factors, such as distance and scene structure, also affect the received signals. IR maps collected by IR and ToF cameras have the same key drawbacks. When an object is close to/far from the IR/ToF camera, its texture details will be overwhelmed by saturation/lost due to the extremely weak signal strength. In particular, adding IR power to sense distant objects is infeasible on battery-sensitive mobile platforms, such as smartphones.

The face detection rates on the IR images collected by ToF cameras are then examined and compared with the maps generated by phase component manipulation. It turns out the detection rate of IR images is merely 2% while the rate of manipulated maps is more than 80%, which indicates that the performance of IR images is significantly limited by distance. In contrast, the manipulated maps suffer less from distance. Moreover, the IR maps collected by the IR and ToF cameras show the same properties. Therefore, unless otherwise indicated, IR maps collected by IR camera and ToF camera will not differentiate.

The key observations on phase component manipulation during ToF measurement are summarized. First, the original depth maps are calculated using the phase components of the received IR signal. However, the transformation from phase components to depth suffers

dimension reduction. In other words, the phase components contain more information than the depth map. Second, phase component manipulation can exploit such information and therefore expose more textures, which can be used to enhance the performance of various depth applications in the dark. Moreover, phase component manipulation can overcome the key shortcomings of traditional IR images, including short sensing range due to rapid signal decay, significant noises, and the over-exposure effect.

4. Application Scenarios

The Mozart ToF methods and system of the subject invention exploit the phase components of infrared light during ToF measurements to expose high-resolution textures of the scenes as illustrated by the workflow in Figure 1. As current ToF modules adopt various measures to eliminate the interference from ambient light, The Mozart ToF methods and system can work in all light conditions. Nevertheless, the focus is placed on sensing in the low-light and dark environments since there currently does not exist a ubiquitous high-resolution vision technology in these challenging conditions, that is, the counterpart of RGB cameras in good lighting environments. The robust high-resolution sensing in the dark has many applications, for example, longitudinal assessment of physical and mental health of elders or babies. In general, the Mozart ToF methods and system can enable applications mainly in the following two manners.

4.1 Mozart-only Methods and System

Thanks to its high-quality textures, the Mozart ToF map alone can enable or augment various applications in the dark. For example, ToF-based face recognition would typically fail when the user's face is away from the ToF camera more than 0.8 m due to the excessive noise of depth measurement. In such cases, the Mozart ToF maps can be applied to augment face recognition, which is an essential function for smartphones, smart door locks, and smart surveillance systems. In addition, the Mozart ToF methods and system on mobile phones can enable accurate facial expression recognition under all lighting conditions to monitor the user's emotional state, which enables a more natural user interface adaptive to the user's emotions. Other representative applications in the dark include complex gesture recognition, security surveillance, robot navigation. For instance, in a smart building embedded with depth ToF cameras on the wall, users can use gestures to control lights and other appliances, even in low-light and dark conditions.

4.2 Integration with Depth Map

The Mozart ToF methods and system not only provide high-quality input for downstream applications but also enables the integration of depth maps and rich-in-texture Mozart ToF maps for new 3D applications. First, the Mozart ToF maps can provide a new mechanism for training machine learning models for depth data in ToF-only systems. Specifically, the high-quality texture details can generate accurate labels by directly leveraging CV algorithms, which can be used for quick model training without manual labeling. Second, the accuracy of perception tasks can be improved by fusing the features of the Mozart ToF maps and depth maps. Finally, better 3D structures of objects can be captured by combining detailed textures of the Mozart ToF maps and corresponding depth maps for ToF-only modules.

5. System Architecture

The Mozart ToF methods and system utilize phase component manipulation, which exploits effective mapping of the phase components during ToF measurements (that is, N_1 , N_2 in Equation (1)) to generate the high-resolution texture of the scene.

Referring to Figures 5A-5B, the IR maps collected by IR cameras and ToF cameras both suffer over-exposure for near objects and under-exposure for distant objects are shown. In particular, Figure 5A shows IR maps collected by IR cameras and Figure 5B shows IR maps collected by ToF cameras.

Figure 6 shows the system architecture of the Mozart ToF methods and system. Unlike other depth camera systems that obtain depth maps directly, the Mozart ToF methods and system takes advantage of the phase components (N_1 , N_2) during ToF measurements. As the theoretical foundation of the Mozart ToF design, an in-depth analysis of the relationship between phase components and exposed textures of the scene based on the physical reflection model for the received IR light is provided. Then two techniques are provided for enhancing the exposed texture map, including redistribution of total reflection outliers and compensation for illumination attenuation. Based on these analyses, it is found that the textures can be exposed and enhanced using highly compute-efficient phase manipulation functions. Lastly, an end-to-end unsupervised learning approach is adopted, which employs an autoencoder to automatically learn efficient representations from the phase component maps to generate the Mozart ToF maps. Specifically, the autoencoder neural network first converts the phase components into deep latent space and then reconstructs high-dimension Mozart ToF maps

from the deep embeddings. To train the deep autoencoder, three novel loss functions are designed to exploit the physics models for texture exposure and enhancement, including the albedo similarity loss, the uniform distribution loss, and the illumination attenuation loss. Combining these learning objectives, the autoencoder-based Mozart ToF methods and system can effectively generate high-resolution texture maps for various scenes and applications. The approach has several key advantages. First, the autoencoder is trained in an unsupervised manner, which does not require any manual labeling or reference images. Second, the autoencoder is scalable in generating various high-resolution Mozart ToF maps, as it can be directly applied to different applications without manual system tuning.

6. Methodology

This section illustrates how to expose and enhance detailed textures from ToF phase components. To this end, a first-principle physics model is provided in Section 8.3, which lays the theoretical foundation for the texture exposure approaches in the ToF system. Then two specific techniques are utilized to further enhance the exposed texture information, that is, compensation for illumination attenuation and redistribution of total reflection outliers. With the help of the physics texture model, highly compute-efficient phase manipulation functions for exposing and enhancing textures are provided in Section 6.2. Examples of lightweight manipulation functions and guidance in selecting effective functions for different applications are then provided. Lastly, an end-to-end autoencoder-based texture generation implementation by designing highly effective learning objectives according to the physics models is provided in Section 6.3, which automatically learns efficient representations from N_1 , N_2 maps. Even though both implementations are based on the physics model, the lightweight approach is the better choice when computing resources are limited, and the autoencoder-based method performs best in dynamic and complex scenes.

6.1 Physics Model for Exposing Textures

In this section, a physics model is provided to facilitate exposing and enhancing detailed textures from ToF phase components.

6.1.1 Modeling Textures Using Phase Components.

As introduced in Section 3.1, the phase components of ToF measurements may vary for the points at the same distance due to objects' texture, thereby essentially encoding detailed texture information besides the distance. Therefore, a physics model is needed to

guide the manipulation of phase components of ToF measurements for revealing texture information and augmenting ToF sensing in the dark. It is known that the IR light emitted by the ToF camera will be diffusely reflected by the surface of objects in most cases. Therefore, the Lambertian reflection model as shown by Figure 7A is employed to model the process of reflection, in which the amount of received IR light reflected by the object at a distance d can be calculated by Equation (2):

$$E_d = \frac{E_0 \alpha \cos \theta}{8d^2} = \frac{E_0 \beta}{8d^2}, \quad (2)$$

where E_0 is a constant determined by the emission power of the ToF camera, α is the reflectivity of the object and θ is the angle of incidence. It can be seen that the intensity of received light is determined by objects' reflectivity α , the incidence angle θ , and the distance d . The former two factors together form the textures of objects. A new variable, "albedo" $\beta = \alpha \cos \theta$, is defined to quantify the two factors on the object side that have an impact on the intensity of the received light.

Based on Equation (2) and the physical meaning of N_1 , N_2 (see Section 3.1), the relationship between the phase components and the texture information β is established:

$$N_1 = \frac{E_0}{8D} \cdot \frac{\beta(D-d)}{d^2}, \quad N_2 = \frac{E_0}{8D} \cdot \frac{\beta}{d}, \quad (3)$$

where D is the ToF camera's range of measurement. Note that E_0 , D are all constants, thus N_1 , N_2 are functions of albedo β and distance d , namely $N_1 = n_1(\beta, d)$, $N_2 = n_2(\beta, d)$.

Therefore, the goal of phase (that is, N_1 , N_2) manipulation in the Mozart ToF methods and system is to find an optimal way of combining albedo β and distance d to augment texture information as much as possible for applications in the dark.

6.1.2 Exposing Textures: Albedo-monotonic

Given the above physics model, manipulation of the phase components N_1 , N_2 to expose textures effectively is shown. the phase component manipulation problem can be formulated as a mapping from a 2-D vector to a scalar as:

$$f(N_1^i, N_2^i) \rightarrow S_i, \quad (4)$$

wherein i is the index of a pixel in the map, and Si is the corresponding scalar in the resulting map. The function $f(\cdot)$ is applied to every pixel in the whole map.

The original depth maps do not contain detailed texture information because points with the same distance d but different albedos β cannot be differentiated. Therefore, to expose detailed texture information, the phase components mapping $f(\cdot)$ should keep the same order as albedo β , which means that the pixel with larger β will always have a larger value after the mapping. In this way, for any two points A and B with the same distance d in

the scene, if $\beta_A < \beta_B$, the mapping should have $f(N_1^A, N_2^A) < f(N_1^B, N_2^B)$. Therefore, following formula is obtained:

$$\frac{\partial f(N_1, N_2)}{\partial \beta} > 0$$

Combining with Equation (3), the following constraints of the phase manipulation mapping are obtained:

$$\frac{\partial f(N_1, N_2)}{\partial N_1} (D - d) + \frac{\partial f(N_1, N_2)}{\partial N_2} d > 0 \quad (5)$$

Equation (5) ensures that the transformed result $f(N_1, N_2)$ is monotonically increasing in terms of albedo β at a given distance d , which is referred to as albedo-monotonic. The monotonicity keeps the same structure in the transformed map as the albedo map, exposing detailed textures without introducing any artifacts as shown in Figure 7B. It is worth noting that if the inequalities in Equation (5) are completely opposite to the current ones, the monotonicity will still hold. However, the texture structure in the transformed map will be reversed to albedo, resulting in “negative images”. For example, negative images are useful for enhancing white or grey detail embedded in dark regions of an image. Moreover, Equation (5) is a sufficient but not necessary condition, which means all transformations that meet this constraint can effectively expose textures.

6.1.3 Enhancing Textures: Illumination Compensation

According to Equation (3), the phase components $N_1 = n_1(\beta, d)$ and $N_2 = n_2(\beta, d)$ decrease with the distance. Therefore, as shown in Figure 8A, in a typical map generated by

phase manipulation, the near objects are much brighter than distant objects, reducing the utility of distant points of the image. Moreover, the average intensity of received light decreases drastically with the distance as shown in Figure 8B. Therefore, in this section, the illumination attenuation introduced by longer distances is compensated for to further enhance the exposed textures during the phase manipulation. To achieve this objective, d can be removed from Equation (3) to obtain the following function:

$$g(N_1, N_2) = \frac{N_2^2}{N_1 + N_2} = \frac{E_0}{8D^2} \cdot \beta, \quad (6)$$

where E_0 , D are all constants, which means that the transformed result by Equation (5) is not be affected by the distance while only related to the texture information β . Moreover, (N_1, N_2) strictly satisfies the constraints of exposing textures in Equation (5). Therefore, the function (N_1, N_2) can correct illumination attenuation introduced by the distance of objects d while enhancing the detailed textures of the scene.

6.1.4 Enhancing Textures: Redistribution of Outliers

Based on the physics model of Section 6.1.2, texture information can be exposed through ToF phase manipulation. In practice, there will always exist total reflection on the surface of certain objects, such as metals and glass. Therefore, the points with total reflection will not satisfy the Lambertian reflection model in Equation (2), and the corresponding N_1 , N_2 received by the ToF camera will far exceed the normal values. For example, a typical value of N_2 is smaller than 100, while the N_2 value of total reflection can be greater than 1,000. Then the mapping results of these outliers through the non-decreasing functions defined in Equation (5) will also exceed the normal range. As shown in Figure 9, if the texture map is normalized to grayscale, most of the pixels will be squeezed into a small range near 0, while the points with total reflection exhibit isolated bright spots. Therefore, in this section, the method to enhance the exposed textures during phase manipulation by redistributing the total reflection outliers is provided.

To alleviate the impact of outliers introduced by total reflections, the dense values around 0 are expanded and the sparse outliers are compressed with large values. To achieve this, a new mapping (S) outside the function $f(\cdot)$ defined in Equation (5) is added, where S denotes the transformed result of $f(N_1, N_2)$. Here the continuous function (S) must be monotonically increasing, and the rate of increase gets smaller with the pixel value S . Figure

9 shows an example of the redistribution function $r(S)$, where the dense distribution at 0 can be expanded, and the sparse distribution at larger values can be squeezed. Therefore, $r(S)$ can limit the bound of higher outlier values. The lower part of Figure 9 shows the texture map before and after the redistribution of phase manipulation, where the manipulated map after redistribution has a substantially higher contrast and more uniform distribution of pixel values.

6.2 Light-weight Phase Manipulation

Given the physics model and principles of exposing and enhancing textures proposed in Section 6.1, the method to design efficient functions $f(N_1, N_2)$ in practice to expose textures is provided. Selecting functions for texture exposure. The first stage of de-signing efficient phase manipulation functions is to check whether the candidate functions are albedo-monotonic. In Figures 10A-10C, the generated maps are obtained by applying the following functions, respectively:

$$f_0(N_1, N_2) = \frac{N_2}{N_1}, f_1(N_1, N_2) = \frac{N_1 \cdot N_2}{(N_1 + N_2)}, f_3(N_1, N_2) = N_1 N_2^2,$$

where f_1 and f_3 satisfy Equation (5) while f_0 does not. It is observed that f_0 cannot expose textures of the scene while both f_1 and f_3 can, which is consistent with the principle provided in Section 6.1. However, the number of functions that are albedo-monotonic is enormous. To efficiently select proper functions, checking the polynomial degrees of candidate functions is required. It can be easily seen that the polynomial degrees of f_0, f_1, f_3 with respect to N_1, N_2 are 0, 1, 3, respectively. Moreover, the generated map f_3 exhibits an over-exposure effect, reducing the quality of exposed textures, which indicates that the functions with larger polynomial degrees amplify many normal values to larger values. Therefore, besides satisfying Equation (5), an effective function $f(\cdot)$ to expose textures should not have a large polynomial degree with respect to N_1 and N_2 . Selecting functions for illumination compensation. To compensate for the illumination attenuation of the texture maps, the functions $f(N_1, N_2)$ could be the variant of $(N_1, N_2) = \frac{1}{N_1^2} (N_1 - N_2)$ or itself.

Moreover, based on the observations in Section 6.2, the degree of polynomials for the manipulation function should not be too large. Therefore, to compensate for the illumination attenuation during phase manipulation, $f(N_1, N_2) = (N_1, N_2)$ can be chosen to directly or a linearly transforming $g(N_1, N_2)$ in most common scenarios. Note that the Equation (6) is derived under the inverse-square law assumption, which may be distorted in real-world

systems and complex environments. Therefore, (N_1, N_2) is not necessarily the optimal function for illumination compensation. Selecting Redistribution Functions. Now how different redistribution functions $r(S)$ affect the generated maps is shown. Here three examples of redistribution functions that satisfy the requirements defined in Section 6.1.4 are given, including $r_1(S) = \sqrt[4]{S}$, $r_2(S) = \ln(S)$ and $r_3(S) = \arctan(S)$. For the same outlier $S_0 = 3,000$, $r_1(S_0) = 54.7$, $r_2(S_0) = 8.00$, $r_3(S_0) = 1.57$. It can be seen that different functions $r(S)$ have different redistribution performances for larger pixel values introduced by total reflection outliers. For those scenarios where the total reflection is strong, functions such as $r_3(S)$ can be selected to better redistribute the total reflection outliers. On the contrary, a milder function such as $r_1(S)$ can be selected such that the distribution does not affect the pixels with typical values.

6.3 Autoencoder-based Phase Manipulation

The function-based texture generation in Section 6.2 requires careful design and manual tuning for different applications, which is labor-intensive and requires substantial domain expertise. Therefore, an end-to-end autoencoder-based texture generation approach is employed, which automatically learns efficient representations from N_1, N_2 maps. The key concept is to utilize the physics models for texture exposure and enhancement provided in Section 6.1 to design highly effective learning objectives for the autoencoder.

The approach has several key advantages. First, the autoencoder neural network is trained in an unsupervised manner, which does not require manual labeling or reference images, in contrast to previous supervised image enhancement solutions. Second, by exploiting the physics models for texture exposure and enhancement proposed in Section 6.1, several effective loss functions are designed to train the deep autoencoder neural network. As a result, the autoencoder can output high-resolution texture maps without introducing artifacts or large noises. Third, due to the convolutional layers, autoencoder-based methods can capture local spatial information within the receptive field of convolutional kernels. Finally, compared with the manually crafted manipulation functions, the autoencoder network is more scalable in generating high-resolution Mozart ToF maps for different applications. For example, for the two applications (for example, human tracking and gesture recognition) with different outlier distributions or IR illumination attenuation effects, the autoencoder-based texture generation can be directly applied without any modification or manual system tuning.

6.3.1 Autoencoder Design

Autoencoder is a widely used unsupervised learning approach in computer vision tasks that can learn efficient features from unlabeled data. A deep autoencoder neural network is utilized as a generative model to adaptively generate high-resolution Mozart ToF maps from N_1 , N_2 maps. As shown in Figure 11, the autoencoder neural network has two main components: the encoder and the decoder network. During the training of the deep autoencoder, the phase component N_1 , N_2 maps are input to the neural network. Then the encoder network maps the input N_1 , N_2 maps into deep latent space, and the decoder network reconstructs high-dimension Mozart ToF maps from the deep embeddings. Accordingly, the autoencoder neural network can learn invariant features underlying the N_1 , N_2 maps collected from different scenarios. A 3D-CNN is adopted for the encoder and decoder to explore the inter-channel relationships between N_1 , N_2 maps. Finally, the output maps are used to calculate the unsupervised training loss, where the loss functions are designed based on the physics models for texture exposure and enhancement in Section 6.1.

The goal of training the autoencoder neural network is to learn efficient representations automatically from N_1 , N_2 maps to generate high-resolution texture maps. Therefore, similar to the lightweight mapping function in Section 6.2, the autoencoder is trained to exploit efficient manipulations to the phase components for exposing texture information.

6.3.2 Design of Loss Functions

To train an autoencoder neural network that can generate high-resolution texture maps, three loss functions are designed according to the physics models of Section 6.1, including the albedo similarity loss, the illumination attenuation loss, and the uniform distribution loss.

Suppose P denotes the Mozart output of the autoencoder neural network.

Albedo Similarity Loss

Unlike the lightweight phase manipulation functions that can maintain the pixel topology of N_1 , N_2 maps, the neural network-based method is more like a black box, which may generate artificial textures that do not exist in the actual scene. As shown in Section 6.1, the albedo map calculated by the manipulation function $f(\cdot)$ contains textures of the scene. Therefore, the structural similarity between the Mozart output and the corresponding albedo map is employed to guide the training of the Mozart ToF methods and system Autoencoder

model. Suppose B denotes the albedo map, then the albedo similarity loss is calculated as follows:

$$\mathcal{L}_S = 1 - S(P, B), \quad (7)$$

where $S(X, Y) \in [0, 1]$ denote the structural similarity index of two images X and Y . A larger $S(X, Y)$ means more similarity between the map X and Y .

Illumination compensation loss. As shown in Section 6.1.3, the phase components $N1$ and $N2$ decrease with the distance, making the near objects much brighter than distant objects. Therefore, an illumination compensation loss is provided to penalize the high-intensity pixels near the ToF camera.

Phone Model	ToF Camera(s)	ToF API	Resolution
Huawei P30 Pro	rear	AREngine	240×180
Huawei Mate30 Pro	rear & front	AREngine	240×180
Samsung S20 Ultra	rear	ARCore	640×480

Table 2: Summary of the Mobile Phones with the Mozart ToF Methods and System Implementation.

The resolution refers to the typical resolution obtained through the corresponding API (instead of the physical resolution of the ToF sensor on the mobile phone). All resolutions in the table are sufficient for typical applications such as faceID at reasonable distances.

By design, the depth maps have larger distance values for distant objects and smaller distance values for close objects. Therefore, the depth maps (or maps positively correlated to distance) can serve as a reference kernel to correct the uneven light field of Mozart output. Moreover, as the depth maps usually have lots of noises that can affect the quality of generated Mozart ToF maps, the denoised depth maps D after median filter is utilized to calculate the light compensation loss:

$$\mathcal{L}_L = \|P \circ D\|_1, \quad (8)$$

where $\circ$ denotes the Hadamard product.

Uniform Distribution Loss

As shown in Section 6.1, the outlier values introduced by total reflection distort the distribution uniformity. However, when the histogram of a map is uniform across the entire value range, the map has a higher contrast. This feature is also friendly to many existing object detection and tracking algorithms [41, 45]. Therefore, consistent with the redistribution function in Section 6.1.4, a uniform distribution loss is designed by minimizing the negative histogram entropy for the output map:

$$\mathcal{L}_U = \sum_{c=0}^{255} p_c(\mathbf{P}) \log p_c(\mathbf{P}). \quad (9)$$

where the generated map is changed to grayscale with the value range [0, 255] and $p_c(\cdot)$ is the normalized histogram counts of value c for a map.

Overall Training Loss Function

Putting the above three loss functions together, the overall loss function for training the autoencoder neural network is:

$$\mathcal{L} = \lambda_s \mathcal{L}_S + \lambda_l \mathcal{L}_L + \lambda_u \mathcal{L}_U \quad (10)$$

Here, λ_s , λ_l , and λ_u are the coefficients that weigh the contribution of each loss function and can be adjusted easily in different applications. For example, when the depth maps are very noisy, a smaller λ_l can be set to reduce the impact of depth on the light compensation, leading to reduced noise in the generated Mozart ToF maps.

7 System Implementation and Overhead

7.1 Implementation on Various Platforms

Smartphones Platforms

A number of smartphones (for example, Huawei P/Mate series, or Samsung S/Note series) are equipped with ToF cameras for various applications such as FaceID, In-Air Gesturing, and AR/VR. The Mozart ToF methods and system are first implemented on three off-the-shelf smart phones with embedded ToF cameras, whose specifications are shown in Table 2. To demonstrate the real-time performance of Mozart ToF methods and system, an Android App as shown in Figure 12A that can help users identify objects under dark

environments is built. The demo video (https://youtu.be/qBEffXVft_8) shows that, compared with depth maps, the Mozart App can provide substantially more texture details in the dark environment in real time.

The raw phase components from ToF cameras are not available on Android smartphones. To address this challenge, the phase components are calculated indirectly from depth and intensity maps (that is, the confidence maps in Android documentation), which can be obtained from all Android smartphones using Camera2 API. Moreover, smartphones from a few manufacturers provide dedicated ToF APIs to access the ToF data. For example, AREngine available on Huawei devices can provide 3-bit confidence maps and 13-bit depth maps. ARCore available on Google-certified models such as Samsung S20 Ultra can provide 8-bit confidence maps and 16-bit depth maps. After obtaining the phase components, the Mozart ToF maps are calculated using manually designed functions on the phones for downstream applications. The Mozart ToF methods and system on smartphones are implemented using Java on Android Studio. It is noted that the latency of generating a single Mozart ToF map on all three smartphones is merely around 15 ms as shown in Figure 13A, enabling real-time applications. Edge Platforms. The Mozart ToF methods and system can also be implemented with three mainstream standalone ToF depth cameras as shown in Figure 12B, including DMOM2508, Vzense DCAM710, and DepthEye Wide ToF, on Nvidia Jetson Xavier. The DepthEye ToF camera adopts an IMX556PLR CMOS from Sony, from which the phase components of ToF measurements can be easily obtained directly via APIs. The phase components of the other two ToF cameras need to be calculated indirectly from depth maps and intensity maps. The Mozart ToF methods and system operating on Nvidia Xavier (running Ubuntu 18.04) are implemented using C++ and Python.

On edge platforms, two different texture generation approaches are implemented, including Mozart-manual, where the best functions are manually selected for different applications according to the principles provided in Section 6.2, and the Mozart ToF methods and system, where the autoencoder neural network is trained using the loss functions defined in Section 6.3. It is noted that Mozart-manual requires substantial efforts and domain expertise to choose a proper function in a trial-and-error manner. Moreover, such a labor-intensive process must be repeated for different applications. On the contrary, the autoencoder can automatically learn efficient representations from phase components to generate texture maps while achieving similar performance with Mozart-manual.

7.2 System Overhead

In this section, the system overhead of the Mozart ToF methods and system is evaluated on three smartphones and Nvidia Jetson Xavier with three mainstream standalone ToF modules. First, a smartphone App is developed to classify facial expressions from a typical expression set such as expressions of angry, disgust, fear, happy, sad, surprise, and neutral under different illumination conditions². Results show that the Mozart ToF methods and system on smartphones outperforms depth maps by 20% and IR maps by 40% in mean recognition accuracy. An object detection task is further implemented on Nvidia Xavier with the three ToF modules. The results indicate that the performance improvement of the Mozart ToF maps is significant for all ToF cameras. Specifically, on the Vsenze ToF camera, the Mozart ToF methods and system outperform depth and IR maps by 63.76% and 45.28%, respectively.

System Overhead on Smartphones

The system overhead of the native ToF system, Mozart-manual, and inference with the Mozart ToF maps are compared in terms of latency, power consumption, and memory usage. PerfDog is used to measure the overall power consumption and memory usage of each task on three smartphones. The results are shown in Figure 13A. First, the latency of calculating a single Mozart ToF map on mobile phones is smaller than 14.5 ms, which can easily support real-time applications with a frame rate of 30 fps. Moreover, calculating the Mozart ToF maps does not significantly increase any resource consumption, including power consumption and memory usage. It is worth noting that the average latency for each major step (that is, sensor sampling, phase components calculation, and the Mozart ToF map generation) is 0.06 ms, 10.92 ms, and 7.13 ms, respectively. The fact implies that the latency can be significantly improved if the Mozart ToF methods and system can access native phase components through APIs since there is no phase components calculation. Moreover, the image quality of the Mozart ToF methods and system can also be improved thanks to more accurate native phase components.

System Overhead on Edge Devices

The Mozart ToF methods and system are compared against Mozart-manual and the native depth system that generate depth maps from ToF phase components. As discussed earlier, Mozart-manual is designed by an extensive search of compute-efficient texture generation functions, which requires substantial efforts and domain expertise. As a result, the computing overhead of Mozart-manual at runtime is expected to be smaller than the

autoencoder-based Mozart ToF methods and system. During the end-to-end experiments, the averaged computation time and power consumption are measured for obtaining one map frame. The power consumption is obtained using tegrastats provided by Nvidia.

The results are shown in Figure 13B. First, the Mozart-manual out-performs the native depth system in computation time and power consumption. This shows that the phase manipulation functions designed for the Mozart-manual are more efficient than the built-in transformation of phase components to depth maps. Specifically, the Mozart-manual produces each frame in merely 43.5ms, that is, at a rate of about 23 fps, which allows it to be executed in real time on embedded platforms. The autoencoder-based Mozart ToF methods and system take more time and power consumption, while it can still achieve about 12 fps, which is acceptable in most depth applications. It is noted that the system overhead can be further reduced by various existing techniques [49], including optimizing the computation pipeline in a hardware-software co-optimization and adopting a sparse autoencoder, which is left for future work.

8 Comparison with Other Sensor Modalities

In this section, the Mozart ToF methods and system of the subject invention are compared with the baseline systems that employ the following sensor modalities: RGB, RGB-enhanced, Depth, IR and mmWave Radar, using three new real-world datasets collected in dark environments. The RGB-enhanced maps are generated by Zero-DCE, a state-of-the-art image enhancement approach from computer vision literature.

8.1 Data Collection in the Dark

The reasons for collecting the new datasets are as follows. First, existing depth camera-based datasets do not contain corresponding samples of other sensor modalities. Besides, there are few multimodal datasets (including RGB/Depth/IR/Radar) collected under dark environments, which is the main application scenario of the Mozart ToF methods and system. The three datasets include over 1,000,000 frames with a total of 33 subjects.

For a fair comparison, the data of all these sensors and the ToF phase components are simultaneously collected at 10 Hz. The RGB images are collected using the Vzense camera, and the IR/depth images are collected using the DepthEye ToF camera. The radar point clouds are collected using a 60-64GHz mmWave radar TI IWR6843. The IR, RGB, RGB-enhanced, Depth, and the Mozart ToF maps are converted to the same dimension (640, 480) and the same models are applied to them for each task. Each frame of radar points is

converted into voxels of a fixed dimension (2, 16, 32, 16) and existing detection or classification models are applied in each task.

Human Tracking Dataset

5 Continuous human tracking in the dark is a typical task in applications like security surveillance, which need to capture high-resolution textures of the scene. As shown in Figure 14A, a real-world human tracking dataset is collected under low-light conditions in three different environments (that is, square, room, and corridor). The volunteers are asked to walk freely within 10 m away from the sensors. In each environment, data are collected under
10 single-person and multi-person settings (where 2, 3, and 4 persons appear simultaneously). In total, over 8,000 frames are collected from nine subjects for each modality.

Face Recognition Dataset

Face recognition in the dark is important for user authentication, for example, for
15 smartphones or smart doors. However, existing technologies such as Apple's FaceID can only work in close range (for example, 0.8m). In this case, the Mozart ToF methods and system can boost the performance at a significantly longer distance. As shown in Figure 14B, a face recognition dataset is collected in a dark room. 12 volunteers are recruited and are asked to sit in front of the sensors at a distance of 1 m (the near setting) and 2 m (the far
20 setting), respectively. Data are also collected under four different illumination conditions by adjusting the lights in the room. over 15,000 data frames of data are totally collected for each modality.

Hand Gesture Recognition Dataset

25 Hand gesture recognition is important for human-computer interaction applications such as controlling smart home appliances. However, due to the small dimensions of hand gestures, it is extremely difficult to achieve robust performance in the dark. As shown in Figure 14C, a hand gesture dataset is collected in a dark room. 12 volunteers are recruited and asked to perform 20 different gestures, including calling, dislike, like, victory, fist, ok, one,
30 three, four, palm, rock, stop, mute, crossed fingers, no, pause, grabbing, gun, pointing, and holding up. The gestures are collected at a distance of 1.5 m from all sensors.

8.2 Accuracy in Different Applications

The Mozart ToF methods and system are first compared against different sensing approaches for the three datasets in the dark. For the human tracking task, a widely used object detector YOLOv5 is configured to detect the persons in each frame, which will output the predicted objects and the prediction confidence for each object. However, this evaluation is not applicable to radar point clouds as they do not support semantic-based tasks due to the data sparsity. Therefore, the radar baseline in this experiment only tracks moving objects in the scene without detecting the type of objects (that is, the person). For the face recognition task, the faces in the image maps are first detected using a pre-trained RetinaFace detector. Then a pre-trained ArcFace model transforms the detected face areas into 512-dimension feature vectors. For a fair comparison, a neural network is directly trained on the detected face areas in depth maps instead of applying ArcFace. Moreover, the accuracy of depth maps is calculated only based on the maps with face areas detected. For voxels of radar data, a 3D-CNN model is trained to classify the faces of 12 different people. To recognize the hand gestures, the hands are first detected and localized by applying the MediaPipe to the RGB, depth, IR, and the Mozart ToF maps. Then the cropped hands area is input into a lightweight 2D-CNN model to classify the gestures. The radar voxels are trained using a 3D-CNN model directly to classify 20 different gestures.

The results of different datasets are shown in Figures 14A-14C. First, the baselines perform very poorly for applications in the dark. For example, the RGB and depth images only achieve 1.66% and 1.54% mean accuracy in the face recognition task. Second, both the Mozart ToF methods and system and Mozart-manual consistently outperform other baselines in different datasets with various settings. Moreover, the Mozart ToF methods and system can approach or even surpass the performance of the manually designed Mozart-manual in different datasets, showing that the autoencoder texture generation is robust and scalable for various applications and the three loss functions are sufficient to generate high-quality Mozart ToF maps in most cases.

8.3 Performance in Dynamic Conditions

The performance of the Mozart ToF methods and system is evaluated under dynamic conditions, including different environments, illumination conditions, and distances of objects. The results are shown in Figure 17.

Different Environments

The impacts of different environments are first evaluated in human tracking, including the square, room, and corridor. First, both Mozart-manual and The Mozart methods and system have a robust performance across different environments, consistently outperforming all baselines. Radar performs extremely poorly in rooms/corridors due to significant multi-path effects. RGB images have extremely low accuracy for all settings in the dark. Depth and IR maps perform poorly in the squares where people walk in a large range (for example, 5-10 m).

Different Light Conditions

The impacts of different light conditions in face recognition are then evaluated, where Light 1, 2, 3 have sequentially increasing ambient light intensity. First, the performance of the Mozart ToF methods and system and Mozart-manual are very stable under different light conditions and outperforms all baselines except for RGB images under Light 3 (full illumination). Further, depth, IR, and radar also yield a stable performance under different light conditions as they do not rely on ambient light. However, the performance of RGB drops drastically with lower light levels since it is highly susceptible to non-ideal environmental light conditions.

Different Distances of Objects

Lastly, the impacts of different distances of objects in face recognition are evaluated. Almost all modalities perform worse in the far setting (1.5 m) because the sizes of face areas are smaller than these under the near setting (0.5 m). However, both Mozart-manual and the Mozart ToF methods and system suffer subtle performance degradation and always outperform all baselines.

8.4 Ablation Study

In this section, the effectiveness of different design components is evaluated for texture generation based on the human tracking task. Both Mozart-manual and the Mozart ToF methods and system are implemented using different component combinations. Specifically, exposing textures based on the physics model is denoted as C1, redistribution of total reflection outliers is denoted as C2, and compensation for illumination attenuation is denoted as C3.

The results are shown in Figure 16. First, the Mozart-manual with C1+C2 already shows significant accuracy improvement. Second, the results with different loss combinations

of the Mozart ToF methods and system all show significant accuracy improvement (that is, at least 20%) over IR maps. Lastly, the performance of the Mozart ToF methods and system is more stable than the Mozart-manual under different configurations, which shows the robustness of autoencoder-based texture generation that exploits the physics texture models.

5

Comparison with the Existing System and Methods

1. The Mozart ToF methods and system of the subject invention is an add-on module that can be applied to all current off-the-shelf ToF modules, including standalone ToF devices and ToF sensors embedded in mobile devices.

10

2. Compared to current off-the-shelf ToF devices like Azure Kinect and Vzense DCAM, the Mozart ToF methods and system of the subject invention provide a solution for fully understanding the detailed texture of a scene. The rich-in-texture map generated by the invention can significantly improve the performance of ToF cameras in various applications.

15

3. The Mozart ToF methods and system of the subject invention provide an effective way of sensing in the dark and outperform RGB images, mmWave Radar, depth images, and IR images by 93.4%, 88.5%, 45.8%, and 29.1%, respectively, in a typical sensing application in the dark.

The Mozart ToF methods and system of the subject invention offer following advantages:

20

1. ToF camera manufacturers can significantly boost their products' sensing performance for more applications with the invention.

2. Software developers can build applications based on data from ToF cameras, providing them with more information about a scene using only a ToF sensor.

25

3. The sensing accuracy of ToF cameras is improved and the generated rich-in-texture maps are combined with depth maps for 3D reconstruction for sensing in the dark.

30

Moreover, the Mozart ToF methods and system are low-cost, high-performance sensing technology to generate high-resolution maps using a single off-the-shelf ToF camera in low-light and dark environments. An in-depth analysis of the physics models is provided for exposing and enhancing textures. The key advantage is that the textures can be generated using highly lightweight phase manipulation functions. By exploiting the physics texture models, a deep autoencoder-based texture generation approach is adopted for automatically learning efficient representations from phase maps to generate Mozart ToF maps. The Mozart ToF methods and system are implemented on various smartphone models as well as edge platforms with mainstream ToF modules. The evaluation using three self-collected datasets in

the dark shows that the Mozart methods and system outperform existing baselines and can work in real time.

The Mozart ToF methods and system of the subject invention are tested on various platforms, including mobile phones with a ToF camera and edge computing platforms with standalone ToF cameras.

In addition, the approach is advantageous in that no labeled training data are required and it is highly scalable in different applications without manual system tuning.

The Mozart ToF methods and system can be implemented on various smartphone models and mainstream standalone ToF cameras. The results show that the Mozart ToF maps can be generated in real time on smartphones due to the extremely low overhead. Moreover, the Mozart ToF methods and system can generate high-resolution maps at about 23 frames per second on edge computing platforms and can work compatibly with the mainstream ToF cameras. The performance of the Mozart ToF methods and system is evaluated using three new datasets collected in low-light and dark conditions, including human tracking, face recognition, and gesture recognition, which involves a total of 33 subjects and contains over 1,000,000 data frames. The results show that the Mozart ToF maps outperform all baseline modalities (including RGB, IR, depth, and mmWave Radar) in dark environments. For example, in gesture recognition, the results of the Mozart ToF methods and system outperform RGB images, mmWave Radar, and depth images by 93.4%, 88.46%, and 45.76%, respectively. Moreover, compared with IR maps, the Mozart ToF methods and system deliver a performance improvement of up to 29.1% with a substantially smaller variance.

Furthermore, the Mozart ToF methods and system are advantageous in development of other potential depth-sensing systems.

(1) Privacy/Security Issues of ToF Cameras

ToF cameras were previously considered privacy-preserving. However, the potential risk of privacy leakage in ToF cameras is clearly illustrated. Therefore, an important and urgent question is how to use the ToF system in a privacy-preserving manner. Fortunately, how much the Mozart methods and system expose detailed textures through phase manipulation can be easily controlled. Therefore, a possible paradigm of using ToF systems in the future is to authorize a specific degree of visual privacy exposure according to application scenarios. On the other hand, the emergence of systems like The Mozart ToF methods and system will motivate the exploration of new privacy-preserving depth measurement principles.

(2) The New APIs of ToF Cameras

The results show the importance of opening access to raw phase components on ToF devices. ToF depth systems can provide three API layers, including a raw data layer providing phase components, a manipulation layer providing tool chains of phase manipulation for generating customized maps, and a result layer providing pre-defined manipulation formulas and auto-encoder-based texture maps. These new APIs may enable a new generation of depth sensing.

The Mozart ToF methods and system are provided to leverage off-the-shelf ToF depth camera to generate high-resolution and rich-in-texture maps for low-light and dark scenes. Extensive experiments show that the Mozart ToF methods and system significantly outperform existing sensing technologies and can work on smartphones and edge platforms in real time. The native depth maps and the Mozart ToF maps may be combined for advanced 3D sensing applications such as 3D reconstruction of dark environments.

All patents, patent applications, provisional applications, and publications referred to or cited herein are incorporated by reference in their entirety, including all figures and tables, to the extent they are not inconsistent with the explicit teachings of this specification.

It should be understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application. In addition, any elements or limitations of any invention or embodiment thereof disclosed herein can be combined with any and/or all other elements or limitations (individually or in any combination) or any other invention or embodiment thereof disclosed herein, and all such combinations are contemplated with the scope of the invention without limitation thereto.

Embodiment 1. A time-of-flight (ToF) system comprising:
an off-the-shelf indirect time-of-flight (iToF) depth camera configured to provide ToF phase components;
a computing platform; and
a data connector connecting the iToF depth camera to the computing platform;
wherein the computing platform comprises:
a driver configured to communicate with the iToF depth camera to control the iToF depth camera for data collection; and
one or more processors and co-processors having stored methods and instructions configured for phase components pre-processing, transformation, and manipulation to

effectively expose high-resolution textures of scenes under different light conditions, including scenes having details captured in dark scenarios; and

a storage medium configured to store results of phase manipulation; and/or

a monitor configured to demonstrate the results of phase manipulation in real time;

5 and/or

a downstream application configured to run on the computing platform that has a manipulated texture map as an input.

10 Embodiment 2. The ToF system of Embodiment 1, wherein the off-the-shelf iToF depth camera comprises a modulator configured to modulate emitted infrared light, an emitter configured to emit IR light, a receiver configured to receive reflected IR light by objects in the scenes, and a demodulator configured to determine phase shift between the received light and the emitted light.

15 Embodiment 3. The ToF system of Embodiment 2, wherein the phase shift is calculated by the amount of light received in two successive time windows, being defined as the phase components, or equivalent phase components are calculated to denote the phase shift.

20 Embodiment 4. The ToF system of Embodiment 1, wherein the iToF depth camera is powered by either the data connector connected to the computing platform or by a corresponding power adapter.

25 Embodiment 5. The ToF system of Embodiment 1, wherein the computing platform is powered by a power adapter.

Embodiment 6. The ToF system of Embodiment 1, wherein the driver is configured to control the iToF camera to measure distances of the scenes and output raw phase components to the processor.

30

Embodiment 7. An embedded system or a mobile system, comprising:

the ToF system of Embodiment 1; and

a mobile phone or a VR headset integrated with the ToF system of Embodiment 1.

Embodiment 8. A time-of-flight (ToF) method, comprising:
a data transformation step for obtaining input data;
a step for performing phase manipulation to generate rich-in-texture maps of scenes from phase components; and
5 an autoencoding step with a learning objective for automatically learning an efficient representation of the rich-in-texture information in the scenes from the phase components.

Embodiment 9. The method of Embodiment 8, further comprising a step of converting a depth map and an intensity map to equivalent phase components.

Embodiment 10. The method of Embodiment 8, further comprising a step with a judgement criterion for checking whether operations on the phase components generate the rich-in-texture maps.

Embodiment 11. The method of Embodiment 8, further comprising an illumination compensation step to compensate for uneven brightness distributions of entire map.

Embodiment 12. The method of Embodiment 8, further comprising an outlier redistribution step for eliminating influence of abnormal pixels with total reflection.

Embodiment 13. The method of Embodiment 8, further comprising an end-to-end autoencoder-based step to automatically learn the texture representation, comprising:

an autoencoding step having phase components as the input, and the generated rich-in-texture map as an output;

a step having three sub-objectives configured to guide the autoencoding step to effectively learn texture representations; and

a step having three tunable coefficients to balance influence weight of different optimization objectives on overall texture exposure.

Embodiment 14. A computer program product, comprising:

a non-transitory computer-executable storage device having computer readable program instructions embodied thereon that when executed by a computer cause the computer to perform time-of-flight (ToF) method for texture exposure in TOF systems, the computer-executable program instruction comprising:

a data transformation step for obtaining input data;
a step for performing phase manipulation to generate rich-in-texture maps of scenes from phase components; and
an autoencoding step with a learning objective for automatically learning an efficient
5 representation of the rich-in-texture information in the scenes from the phase components.

Embodiment 15. The computer program product of Embodiment 14, wherein the computer-executable program instruction further comprises a step of converting a depth map and an intensity map to equivalent phase components.

Embodiment 16. The computer program product of Embodiment 14, wherein the computer-executable program instruction further comprises a step with a judgement criterion for checking whether operations on the phase components generate the rich-in-texture maps.

Embodiment 17. The computer program product of Embodiment 14, wherein the computer-executable program instruction further comprises an illumination compensation step to compensate for uneven brightness distributions of entire map.

Embodiment 18. The computer program product of Embodiment 14, wherein the computer-executable program instruction further comprises an outlier redistribution step for eliminating influence of abnormal pixels with total reflection.

Embodiment 19. The computer program product of Embodiment 14, wherein the computer-executable program instruction further comprises an end-to-end autoencoder-based
25 step to automatically learn the texture representation, comprising:

an autoencoding step having phase components as the input, and the generated rich-in-texture map as an output;

a step having three sub-objectives configured to guide the autoencoding step to effectively learn texture representations; and

a step having three tunable coefficients to balance influence weight of different optimization objectives on overall texture exposure.

Embodiment 20. The computer program product of Embodiment 14, wherein the autoencoding step is configured to generate high-resolution texture maps.

REFERENCES

- [1] 2022. About Face ID advanced technology. <https://support.apple.com/en-us/HT208108>.
- [2] 2022. ARCore Documentation. <https://developers.google.com/ar/develop>.
- 5 [3] 2022. AEngine Documents. <https://developer.huawei.com/consumer/en/doc/development/graphics-Guides/introduction-0000001050130900>.
- [4] 2022. Camera2 overview. <https://developer.android.com/training/camera2>.
- [5] 2022. DepthEye Wide. <https://www.seeedstudio.com/DepthEye-Wide-ToF-Camera-with-Sony-IMX556PLR-DepthSense-p-4809.html>.
- 10 [6] 2022. MediaPipe Hands. <https://google.github.io/mediapipe/solutions/hands.html>.
- [7] 2022. Nvidia Jetson Xavier. <https://developer.nvidia.com/embedded/jetson-agx-xavier-developer-kit>.
- [8] 2022. PerfDog. <https://perfdog.qq.com/>.
- 15 [9] 2022. tegrastats Utility. https://docs.nvidia.com/drive/drive_os_5.1.6.1L/nvlib_docs/index.html#page/DRIVE_OS_Linux_SDK_Development_Guide/Utilities/util_tegrastats.html.
- [10] 2022. TI IWR6843, Single-chip 60-GHz to 64-GHz mmWave Radar. <https://www.ti.com/product/IWR6843>.
- 20 [11] 2022. Time of Flight Sensor Market. <https://www.transparencymarketresearch.com/time-of-flight-sensor-market.html>.
- [12] 2022. Vzense DCAM710. <https://www.vzense.com/RGBDToFProducts.html>.
- [13] 2022. YOLOv5. https://pytorch.org/hub/ultralytics_yolov5/.
- [14] Supreeth Achar, Joseph R Bartels, William L'Red' Whittaker, Kiriakos N Kutulakos, and Srinivasa G Narasimhan. 2017. Epipolar time-of-flight imaging. *ACM Transactions on Graphics (ToG)* 36, 4 (2017), 1–8.
- 25 [15] H Andrei, V Ion, E Diaconu, A Enescu, and I Udriou. 2019. Energy Consumption Analysis of Security Systems for a Residential Consumer. In 2019 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE). IEEE, 1–4.
- 30 [16] Seung-Hwan Baek, Noah Walsh, Ilya Chugunov, Zheng Shi, and Felix Heide. 2022. Centimeter-Wave Free-Space Neural Time-of-Flight Imaging. *ACM Transactions on Graphics (TOG)* (2022).
- [17] Emanuele Bugliarello, Ryan Cotterell, Naoaki Okazaki, and Desmond Elliott. 2021. Multimodal pretraining unmasked: A meta-analysis and a unified framework of

vision-and-language BERTs. Transactions of the Association for Computational Linguistics 9 (2021), 978–994.

- [18] Hao Chen and Youfu Li. 2018. Progressively complementarity-aware fusion network for RGB-D salient object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition. 3051–3060.
- [19] Tao Chen, Kai-Kuang Ma, and Li-Hui Chen. 1999. Tri-state median filter for image denoising. IEEE Transactions on Image processing 8, 12 (1999), 1834–1838.
- [20] Xianda Chen, Yifei Xiao, Yeming Tang, Julio Fernandez-Mendoza, and Guohong Cao. 2021. ApneaDetector: Detecting Sleep Apnea with Smartwatches. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 5, 2 (2021), 1–22.
- [21] Richard LIU Chenmeijing LIANG, Pierre CAMBOU. 2020. Status of the CMOS Image Sensor Industry 2020. https://s3.i-micronews.com/uploads/2020/11/YDR20106-Status-of-the-CMOS-Image-Sensor-Industry-2020_sample.pdf.
- [22] Ziteng Cui, Guo-Jun Qi, Lin Gu, Shaodi You, Zenghui Zhang, and Tatsuya Harada. 2021. Multitask aet with orthogonal tangent regularity for dark object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 2553–2562.
- [23] DayDayNews. 2020. The civil war for ToF technology is far from over. <https://daydaynews.cc/en/technology/683608.html>.
- [24] Jiankang Deng, Jia Guo, Evangelos Ververas, Irene Kotsia, and Stefanos Zafeiriou. 2020. Retinaface: Single-shot multi-level face localisation in the wild. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 5203–5212.
- [25] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. 2019. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).
- [26] DOMI. 2022. 3D ToF camera-DMOM2508CL. <https://www.domisensor.com/products/17-dmom2508c>.
- [27] S Burak Gokturk, Hakan Yalcin, and Cyrus Bamji. 2004. A time-of-flight depth sensor-system description, issues and solutions. In 2004 conference on computer vision and pattern recognition workshop. IEEE, 35–35.
- [28] Linjie Gu, Zhe Yang, Mithun Mukherjee, Zhigeng Pan, Mian Guo, Xiushan Liu, Rakesh Matam, and Jaime Lloret. 2021. HAWK-i: a remote and lightweight thermal

imaging-based crowd screening framework. In Proceedings of the 27th Annual International Conference on Mobile Computing and Networking. 885–887.

- [29] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. 2020. Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 1780–1789.

- [30] Tian Hao, Guoliang Xing, and Gang Zhou. 2013. isleep: Unobtrusive sleep quality monitoring using smartphones. In Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems. 1–14.

- [31] Weijia He, Maximilian Golla, Roshni Padhi, Jordan Ofek, Markus Dürmuth, Earlence Fernandes, and Blase Ur. 2018. Rethinking Access Control and Authentication for the Home Internet of Things. In 27th USENIX Security Symposium (USENIX Security 18). 255–272.

- [32] Weijia He, Valerie Zhao, Olivia Morkved, Sabeeka Siddiqui, Earlence Fernandes, Josiah Hester, and Blase Ur. 2021. SoK: Context sensing for access control in the adversarial home IoT. In 2021 IEEE European Symposium on Security and Privacy (EuroS&P). IEEE, 37–53.

- [33] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, et al. 2011. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In Proceedings of the 24th annual ACM symposium on User interface software and technology. 559–568.

- [34] Anil K Jain. 1989. Fundamentals of digital image processing. Prentice-Hall, Inc.

- [35] Ishani Janveja, Akshay Nambi, Shruthi Bannur, Sanchit Gupta, and Venkat Padmanabhan. 2020. Insight: monitoring the state of the driver in low-light using smartphones. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 4, 3 (2020), 1–29.

- [36] Evangelos Kazakos, Arsha Nagrani, Andrew Zisserman, and Dima Damen. 2019. Epic-fusion: Audio-visual temporal binding for egocentric action recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 5492–5501.

- [37] Diederik P Kingma and Max Welling. 2019. An introduction to variational autoencoders. arXiv preprint arXiv:1906.02691 (2019).

- [38] Xiangyuan Lan, Mang Ye, Shengping Zhang, and Pong Yuen. 2018. Robust collaborative discriminative learning for RGB-infrared tracking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.
- [39] Robert Lange and Peter Seitz. 2001. Solid-state time-of-flight range camera. *IEEE Journal of quantum electronics* 37, 3 (2001), 390–397.
- [40] Benjamin Langmann, Klaus Hartmann, and Otmar Loffeld. 2013. Increasing the accuracy of Time-of-Flight cameras for machine vision applications. *Computers in industry* 64, 9 (2013), 1090–1098.
- [41] Chengxi Li, Xiangyu Qu, Abhiram Gnanasambandam, Omar A Elgendy, Jiaju Ma, and Stanley H Chan. 2021. Photon-limited object detection using non-local feature matching and knowledge distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3976–3987.
- [42] Marvin Lindner and Andreas Kolb. 2009. Compensation of motion artifacts for time-of-flight cameras. In *Workshop on Dynamic 3D Imaging*. Springer, 16–27.
- [43] Xiulong Liu, Dongdong Liu, Jiuwu Zhang, Tao Gu, and Keqiu Li. 2021. RFID and camera fusion for recognition of human-object interactions. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*. 296–308.
- [44] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. 2017. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition* 61 (2017), 650–662.
- [45] David G Lowe. 1999. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, Vol. 2. Ieee, 1150–1157.
- [46] Chris Xiaoxuan Lu, Muhamad Risqi U Saputra, Peijun Zhao, Yasin Almalioglu, Pedro PB De Gusmao, Changhao Chen, Ke Sun, Niki Trigoni, and Andrew Markham. 2020. milliEgo: single-chip mmWave radar aided egomotion estimation via deep sensor fusion. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 109–122.
- [47] Li Lu, Jiadi Yu, Yingying Chen, Hongbo Liu, Yanmin Zhu, Yunfei Liu, and Minglu Li. 2018. Lippass: Lip reading-based user authentication on smartphones lever-aging acoustic signals. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*. IEEE, 1466–1474.

- [48] Jun Luo, Wenqi Ren, Tao Wang, Chongyi Li, and Xiaochun Cao. 2022. Under-Display Camera Image Enhancement via Cascaded Curve Estimation. *IEEE Transactions on Image Processing* 31 (2022), 4856–4868.
- [49] Wei Luo, Jun Li, Jian Yang, Wei Xu, and Jian Zhang. 2017. Convolutional sparse autoencoders for image classification. *IEEE transactions on neural networks and learning systems* 29, 7 (2017), 3289–3294.
- [50] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, Ian Goodfellow, and Brendan Frey. 2015. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644* (2015).
- [51] Daniel McDuff, Abdelrahman Mahmoud, Mohammad Mavadati, May Amr, Jay Turcot, and Rana el Kaliouby. 2016. AFFDEX SDK: a cross-platform real-time multi-face expression recognition toolkit. In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems*. 3723–3726.
- [52] Xiaomin Ouyang, Xian Shuai, Jiayu Zhou, Ivy Wang Shi, Zhiyuan Xie, Guoliang Xing, and Jianwei Huang. 2022. Cosmo: contrastive fusion learning with small data for multimodal human activity recognition. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*. 324–337.
- [53] 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.
- [54] 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* 19, 1 (2022), 1–32.
- [55] HyeonJung Park, Youngki Lee, and JeongGil Ko. 2021. Enabling real-time sign language translation on mobile platforms with on-board depth cameras. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–30.
- [56] Frank L Pedrotti, Leno M Pedrotti, and Leno S Pedrotti. 2017. *Introduction to optics*. Cambridge University Press.
- [57] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. 1987. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing* 39, 3 (1987), 355–368.

- [58] Kun Qian, Zhaoyuan He, and Xinyu Zhang. 2020. 3D point cloud generation with millimeter-wave radar. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 4 (2020), 1–23.
- [59] Xian Shuai, Yulin Shen, Yi Tang, Shuyao Shi, Luping Ji, and Guoliang Xing. 2021. millieye: A lightweight mmwave radar and camera fusion system for robust object detection. In *Proceedings of the International Conference on Internet-of-Things Design and Implementation*. 145–157.
- [60] Myunghoon Suk and Balakrishnan Prabhakaran. 2014. Real-time mobile facial expression recognition system-a case study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 132–137.
- [61] Ke Sun, Wei Wang, Alex X Liu, and Haipeng Dai. 2018. Depth aware finger tapping on virtual displays. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. 283–295.
- [62] 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.
- [63] Anran Wang, Jacob E Sunshine, and Shyamnath Gollakota. 2019. Contactless infant monitoring using white noise. In *The 25th Annual International Conference on Mobile Computing and Networking*. 1–16.
- [64] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.
- [65] Wikipedia. 2022. Inverse-square law. https://en.wikipedia.org/wiki/Inverse-square_law.
- [66] Zhiyuan Xie, Xiaomin Ouyang, Xiaoming Liu, and Guoliang Xing. 2021. Ultra-Depth: Exposing High-Resolution Texture from Depth Cameras. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 302–315.
- [67] Zhiyuan Xie, Xiaomin Ouyang, Li Pan, Wenrui Lu, Xiaoming Liu, and Guoliang Xing. 2022. HiToF: a ToF camera system for capturing high-resolution textures. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*. 764–765.

- [68] Lei Yang, Qiongzhen Lin, Xiangyang Li, Tianci Liu, and Yunhao Liu. 2015. See through walls with COTS RFID system!. In Proceedings of the 21st Annual International Conference on Mobile Computing and Networking. 487–499.
- [69] Wei Zhai, Yang Cao, Zheng-Jun Zha, HaiYong Xie, and Feng Wu. 2020. Deep structure-revealed network for texture recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 11010–11019.
- [70] Yunfan Zhang, Tim Scargill, Ashutosh Vaishnav, Gopika Premsankar, Mario Di Francesco, and Maria Gorlatova. 2022. InDepth: Real-time Depth Inpainting for Mobile Augmented Reality. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 6, 1 (2022), 1–25.
- [71] Jufeng Zhao, Yueting Chen, Huajun Feng, Zhihai Xu, and Qi Li. 2014. Infrared image enhancement through saliency feature analysis based on multi-scale de-composition. Infrared Physics & Technology 62 (2014), 86–93.
- [72] Yanzi Zhu, Yuanshun Yao, Ben Y Zhao, and Haitao Zheng. 2017. Object recognition and navigation using a single networking device. In Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services. 265–277.
- [73] Michael Zollhöfer, Patrick Stotko, Andreas Görlitz, Christian Theobalt, Matthias Nießner, Reinhard Klein, and Andreas Kolb. 2018. State of the art on 3D reconstruction with RGB-D cameras. In Computer graphics forum, Vol. 37. Wiley Online Library, 625–652.

CLAIMS

We claim:

1. A time-of-flight (ToF) system comprising:
 - an off-the-shelf indirect time-of-flight (iToF) depth camera configured to provide ToF phase components;
 - a computing platform; and
 - a data connector connecting the iToF depth camera to the computing platform;wherein the computing platform comprises:
 - a driver configured to communicate with the iToF depth camera to control the iToF depth camera for data collection; and
 - one or more processors and co-processors having stored methods and instructions configured for phase components pre-processing, transformation, and manipulation to effectively expose high-resolution textures of scenes under different light conditions, including scenes having details captured in dark scenarios; and
 - a storage medium configured to store results of phase manipulation; and/or
 - a monitor configured to demonstrate the results of phase manipulation in real time; and/or
 - a downstream application configured to run on the computing platform that has a manipulated texture map as an input.
2. The ToF system of claim 1, wherein the off-the-shelf iToF depth camera comprises a modulator configured to modulate emitted infrared light, an emitter configured to emit IR light, a receiver configured to receive reflected IR light by objects in the scenes, and a demodulator configured to determine phase shift between the received light and the emitted light.
3. The ToF system of claim 2, wherein the phase shift is calculated by the amount of light received in two successive time windows, being defined as the phase components, or equivalent phase components are calculated to denote the phase shift.
4. The ToF system of claim 1, wherein the iToF depth camera is powered by either the data connector connected to the computing platform or by a corresponding power adapter.

5. The ToF system of claim 1, wherein the computing platform is powered by a power adapter.

6. The ToF system of claim 1, wherein the driver is configured to control the iToF camera to measure distances of the scenes and output raw phase components to the processor.

7. An embedded system or a mobile system, comprising:

the ToF system of claim 1; and

a mobile phone or a VR headset integrated with the ToF system of claim 1.

8. A time-of-flight (ToF) method, comprising:

a data transformation step for obtaining input data;

a step for performing phase manipulation to generate rich-in-texture maps of scenes from phase components; and

an autoencoding step with a learning objective for automatically learning an efficient representation of the rich-in-texture information in the scenes from the phase components.

9. The method of claim 8, further comprising a step of converting a depth map and an intensity map to equivalent phase components.

10. The method of claim 8, further comprising a step with a judgement criterion for checking whether operations on the phase components generate the rich-in-texture maps.

11. The method of claim 8, further comprising an illumination compensation step to compensate for uneven brightness distributions of entire map.

12. The method of claim 8, further comprising an outlier redistribution step for eliminating influence of abnormal pixels with total reflection.

13. The method of claim 8, further comprising an end-to-end autoencoder-based step to automatically learn the texture representation, comprising:

an autoencoding step having phase components as the input, and the generated rich-in-texture map as an output;

a step having three sub-objectives configured to guide the autoencoding step to

effectively learn texture representations; and

a step having three tunable coefficients to balance influence weight of different optimization objectives on overall texture exposure.

14. A computer program product, comprising:

a non-transitory computer-executable storage device having computer readable program instructions embodied thereon that when executed by a computer cause the computer to perform time-of-flight (ToF) method for texture exposure in TOF systems, the computer-executable program instruction comprising:

a data transformation step for obtaining input data;

a step for performing phase manipulation to generate rich-in-texture maps of scenes from phase components; and

an autoencoding step with a learning objective for automatically learning an efficient representation of the rich-in-texture information in the scenes from the phase components.

15. The computer program product of claim 14, wherein the computer-executable program instruction further comprises a step of converting a depth map and an intensity map to equivalent phase components.

16. The computer program product of claim 14, wherein the computer-executable program instruction further comprises a step with a judgement criterion for checking whether operations on the phase components generate the rich-in-texture maps.

17. The computer program product of claim 14, wherein the computer-executable program instruction further comprises an illumination compensation step to compensate for uneven brightness distributions of entire map.

18. The computer program product of claim 14, wherein the computer-executable program instruction further comprises an outlier redistribution step for eliminating influence of abnormal pixels with total reflection.

19. The computer program product of claim 14, wherein the computer-executable program instruction further comprises an end-to-end autoencoder-based step to automatically learn the texture representation, comprising:

an autoencoding step having phase components as the input, and the generated rich-in-texture map as an output;

a step having three sub-objectives configured to guide the autoencoding step to effectively learn texture representations; and

a step having three tunable coefficients to balance influence weight of different optimization objectives on overall texture exposure.

20. The computer program product of claim 14, wherein the autoencoding step is configured to generate high-resolution texture maps.

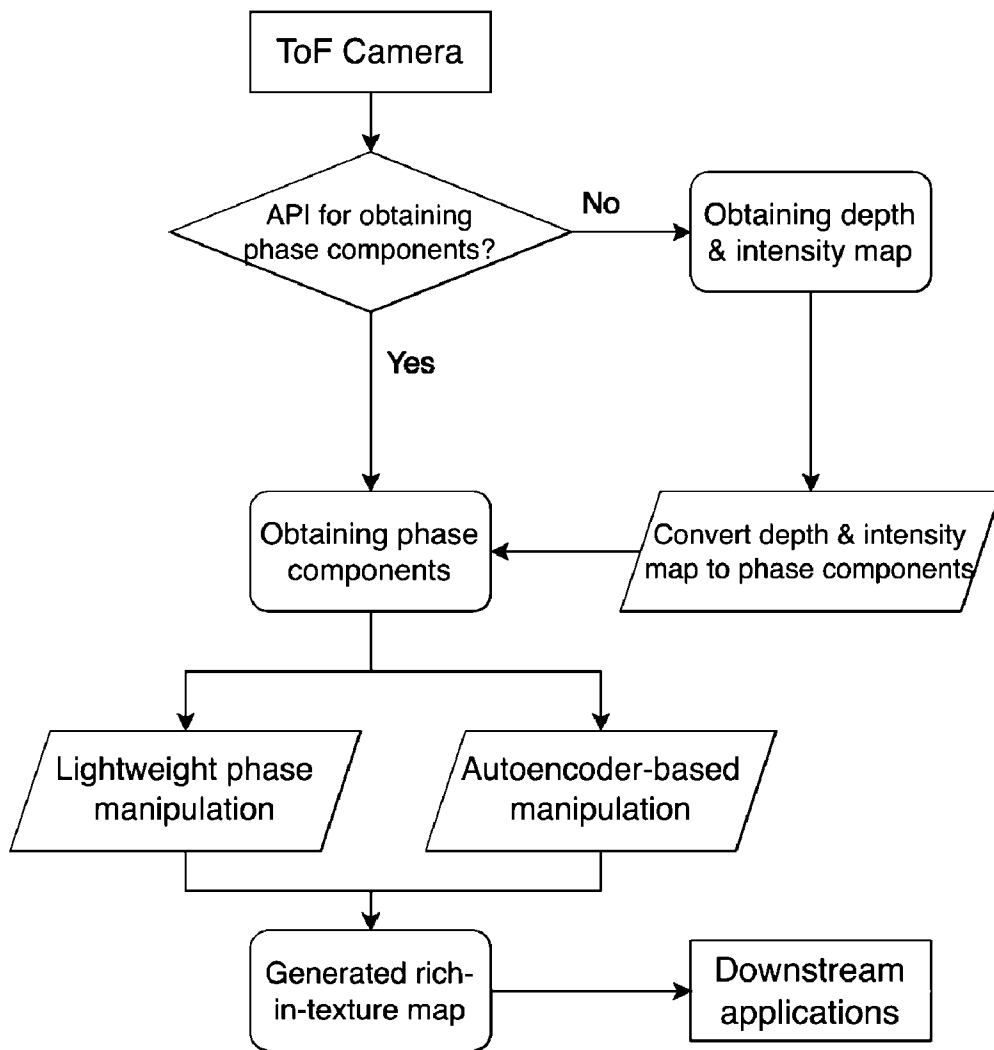


FIG. 1

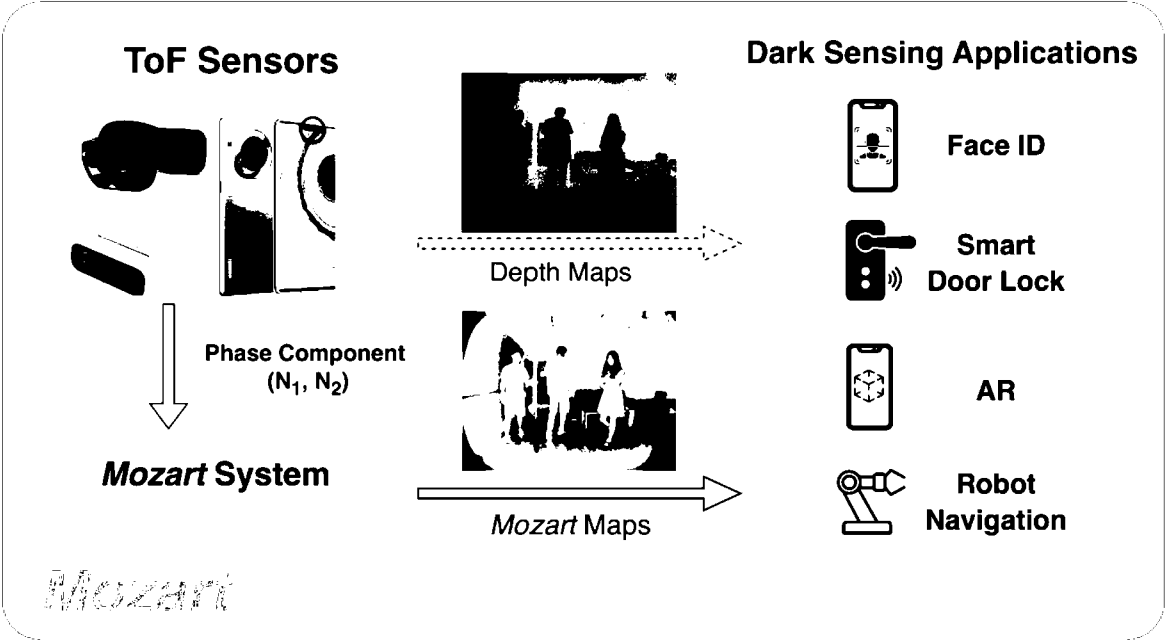
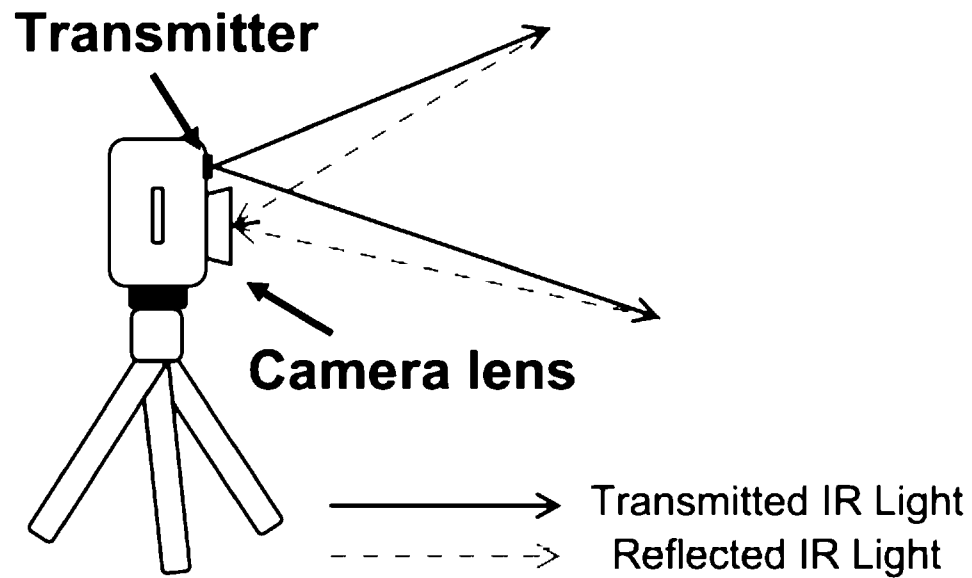
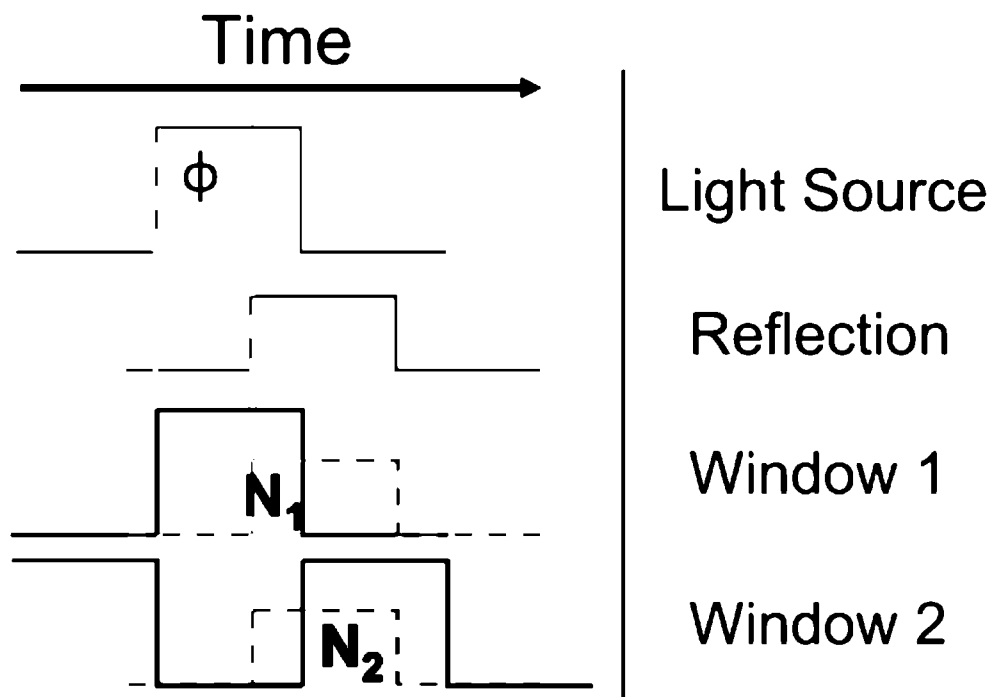


FIG. 2



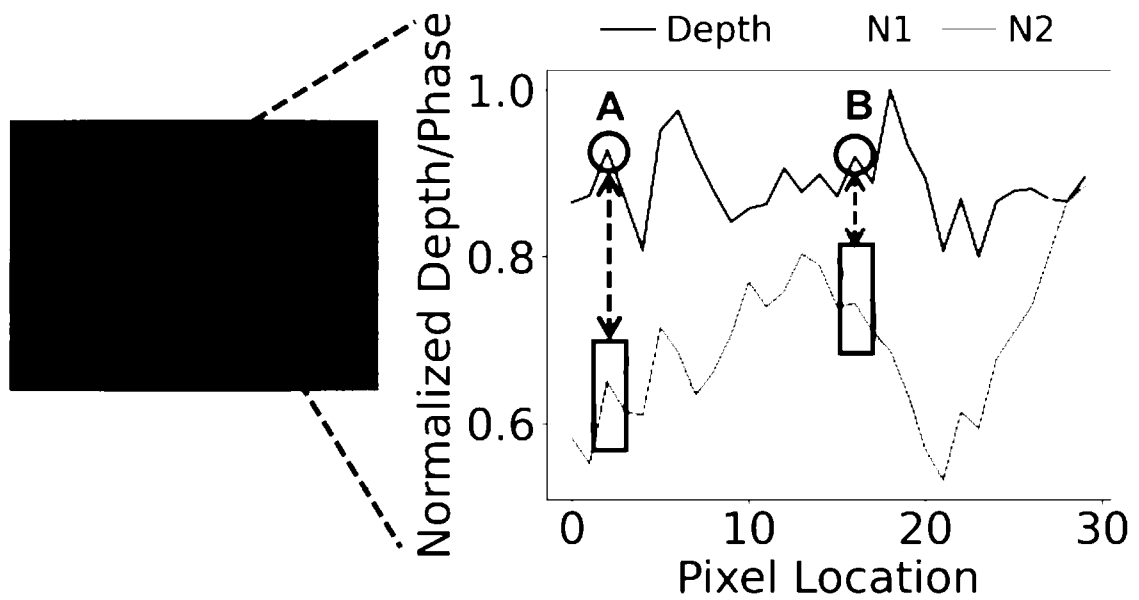
(a) Capturing depth maps of the scene with a ToF camera

FIG. 3A



(b) Principles of iToF depth cameras

FIG. 3B



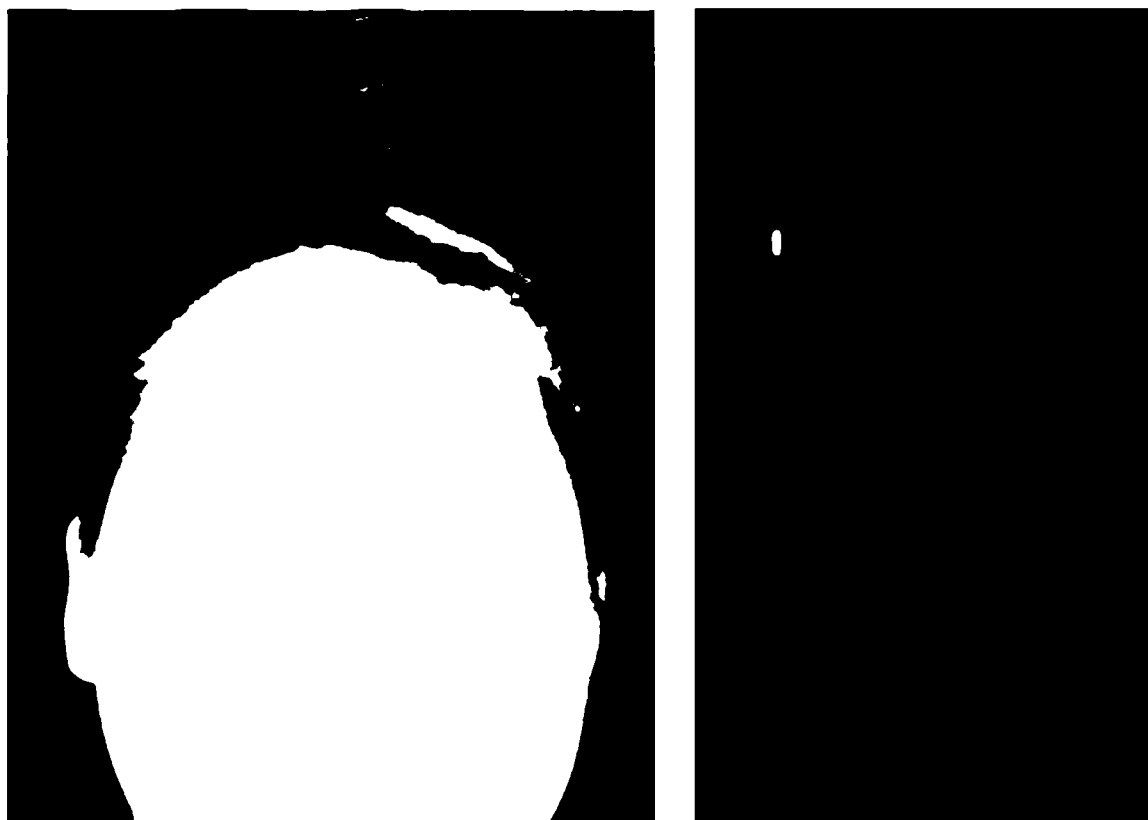
(a) The original depth map and values of phase components

FIG. 4A

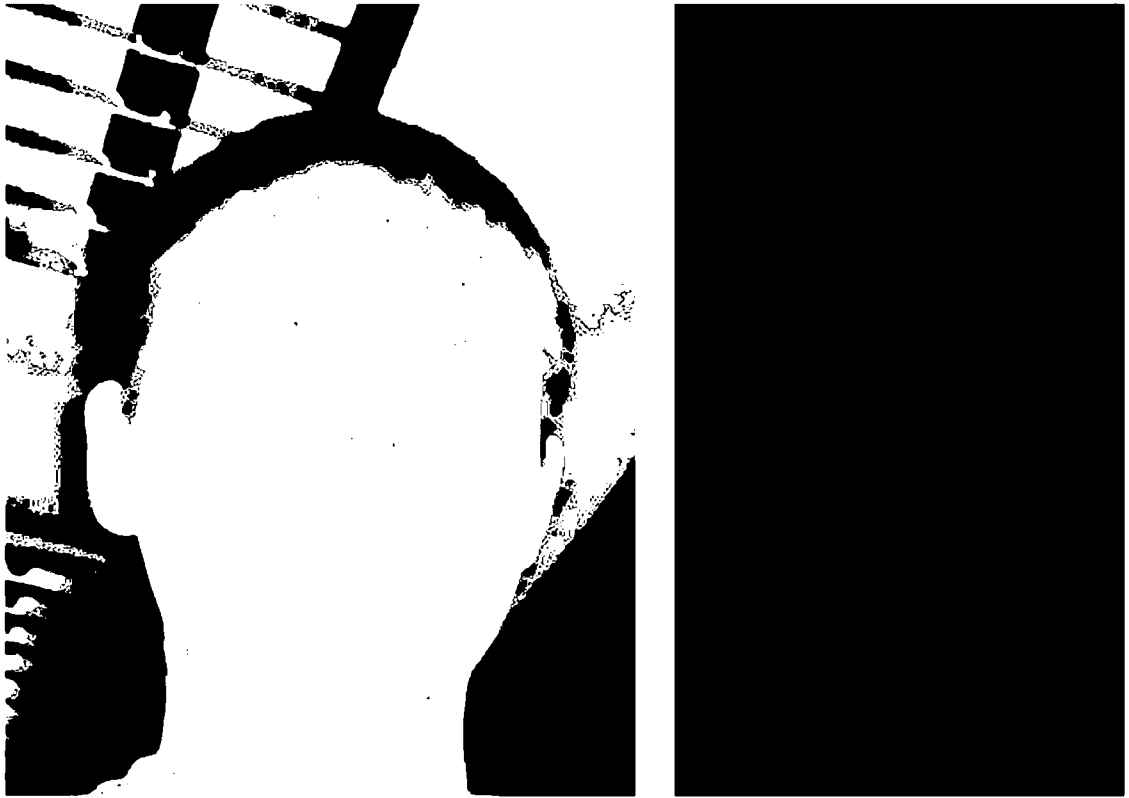


(b) The new depth map
obtained by shifting N_1 .

FIG. 4B



(a) Collected by IR cameras



(b) Collected by ToF cameras

FIG. 5B

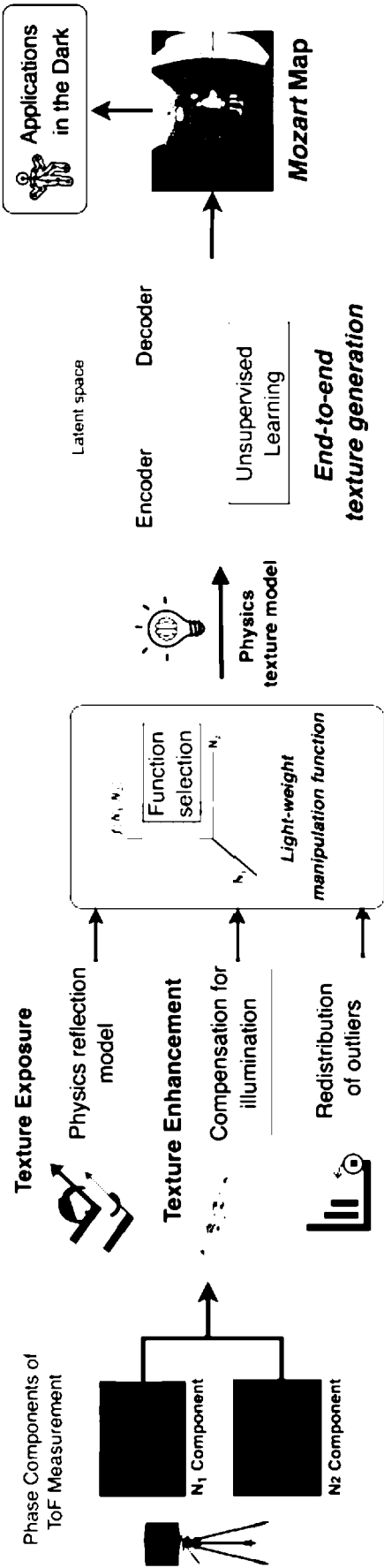
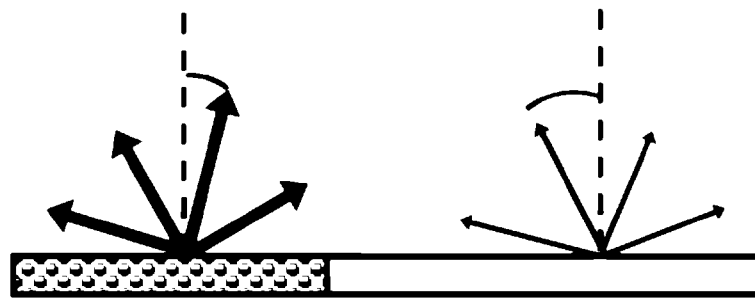
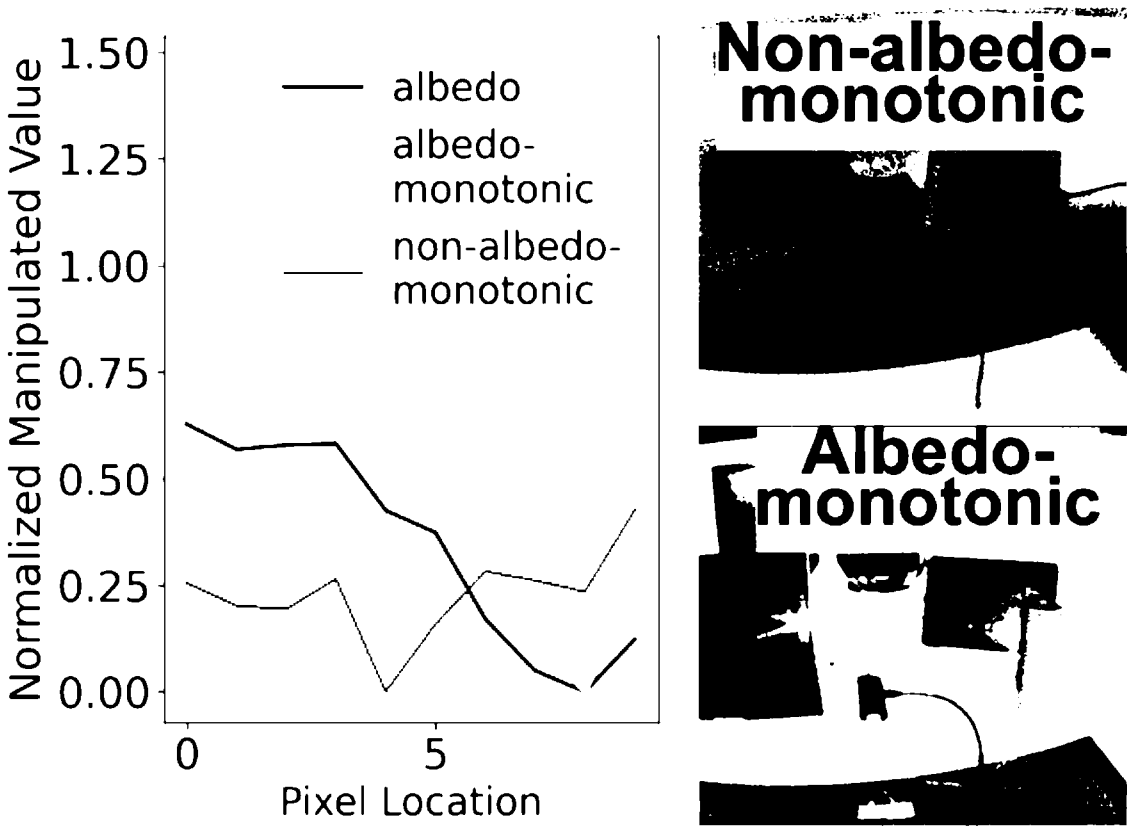


FIG. 6



(a) Lambertian reflection model.

FIG. 7A



(b) Pixel value over the map.

FIG. 7B



FIG. 8A

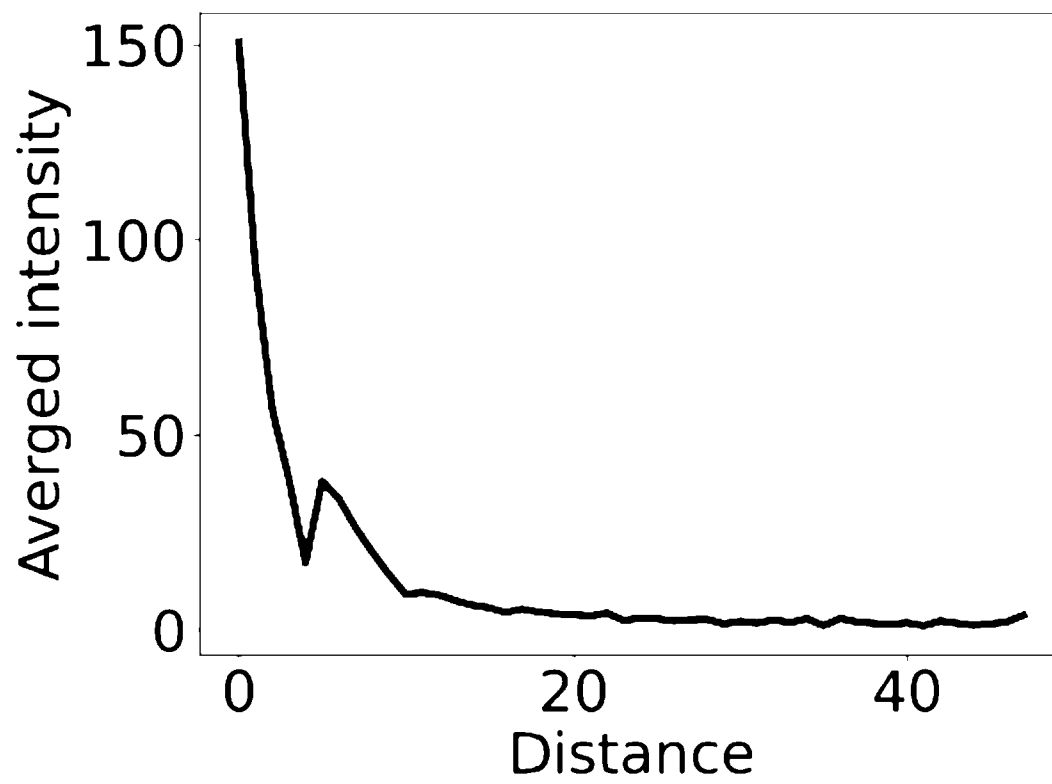


FIG. 8B

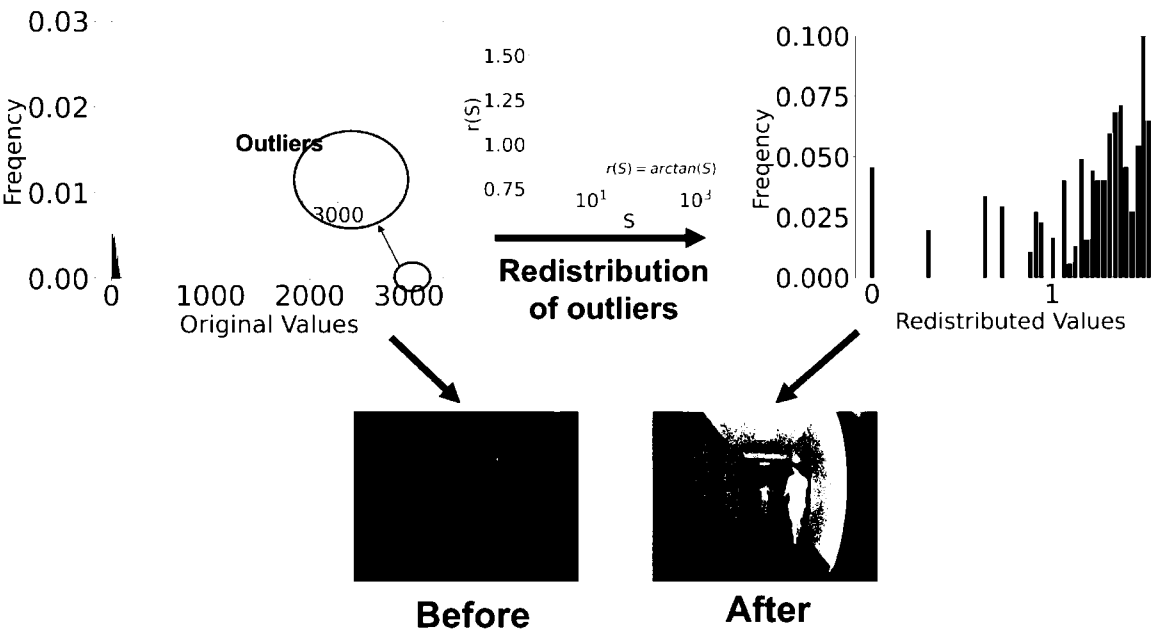
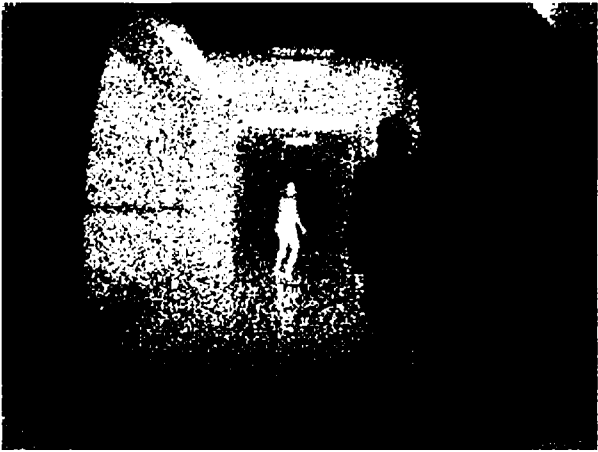


FIG. 9

5



(a) $f_0 = N_2/N_1.$

FIG. 10A

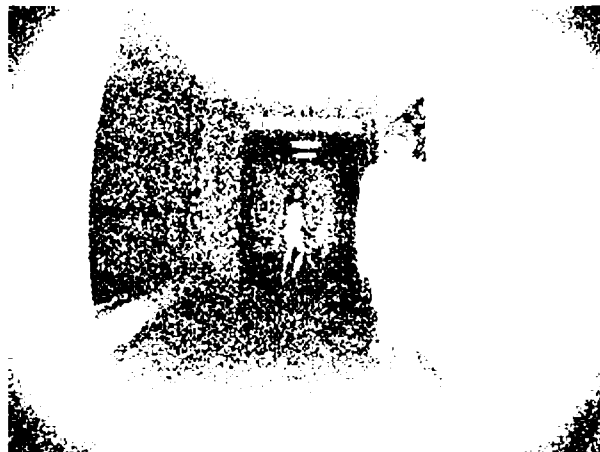
10



$$(b) f_1 = N_1 N_2 / (N_1 + N_2).$$

5

FIG. 10B



$$(c) f_3 = N_1 N_2^2.$$

FIG. 10C

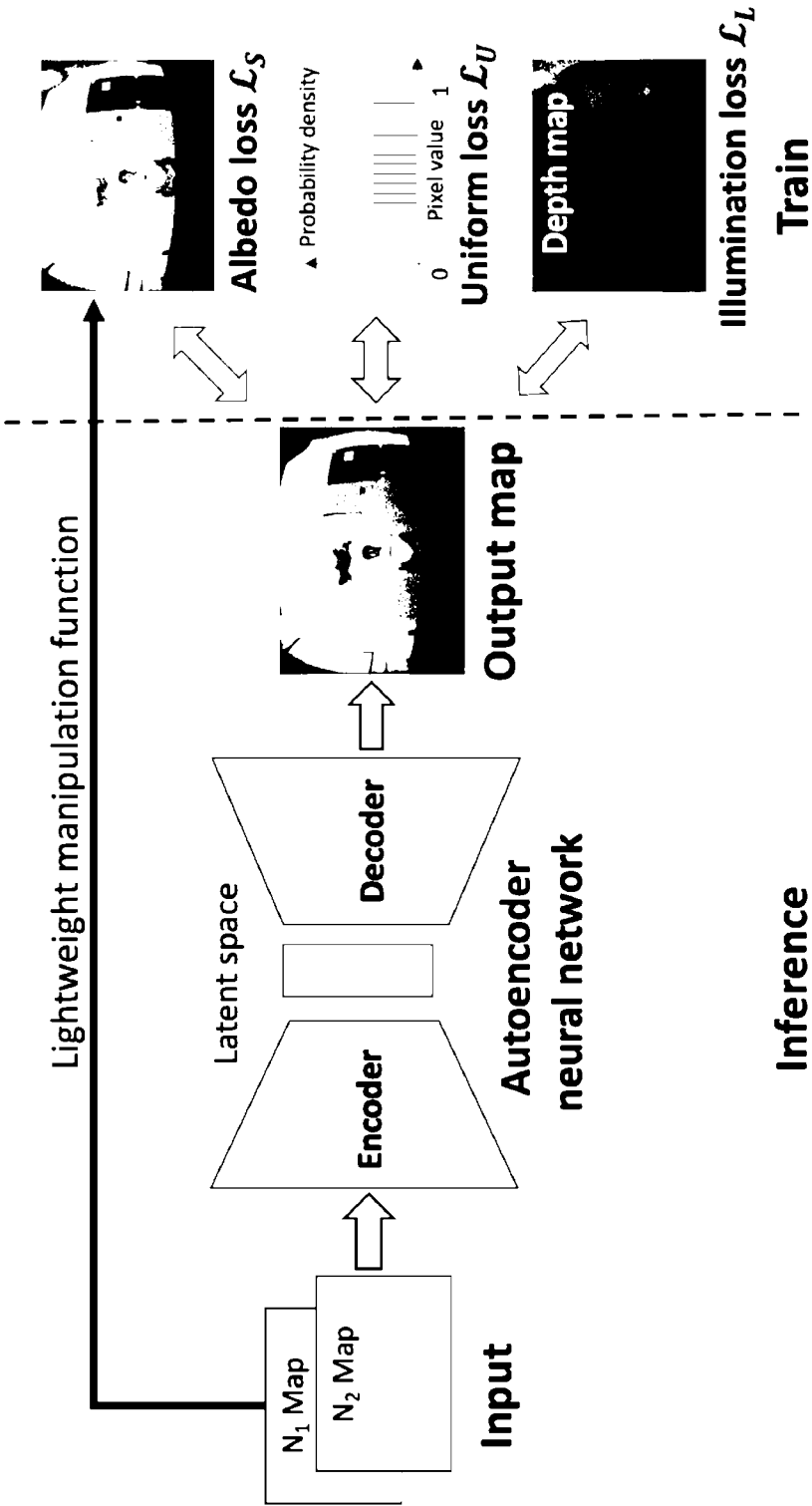
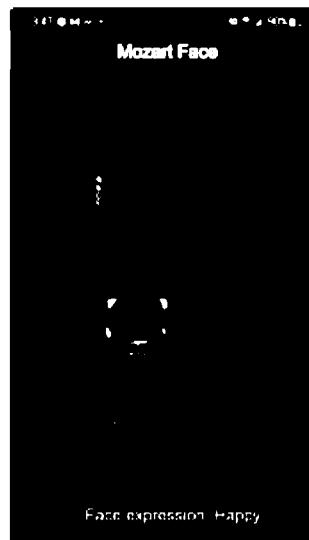
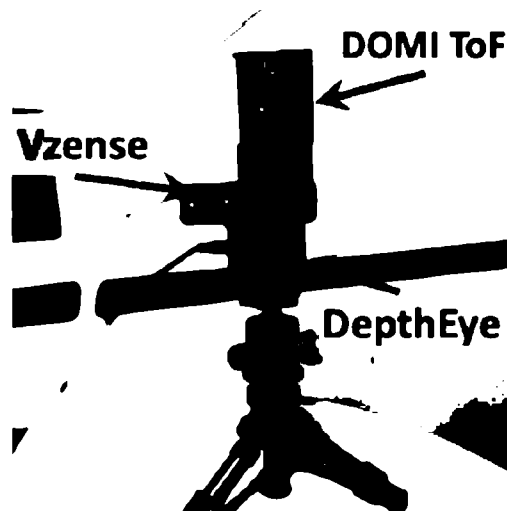


FIG. 11



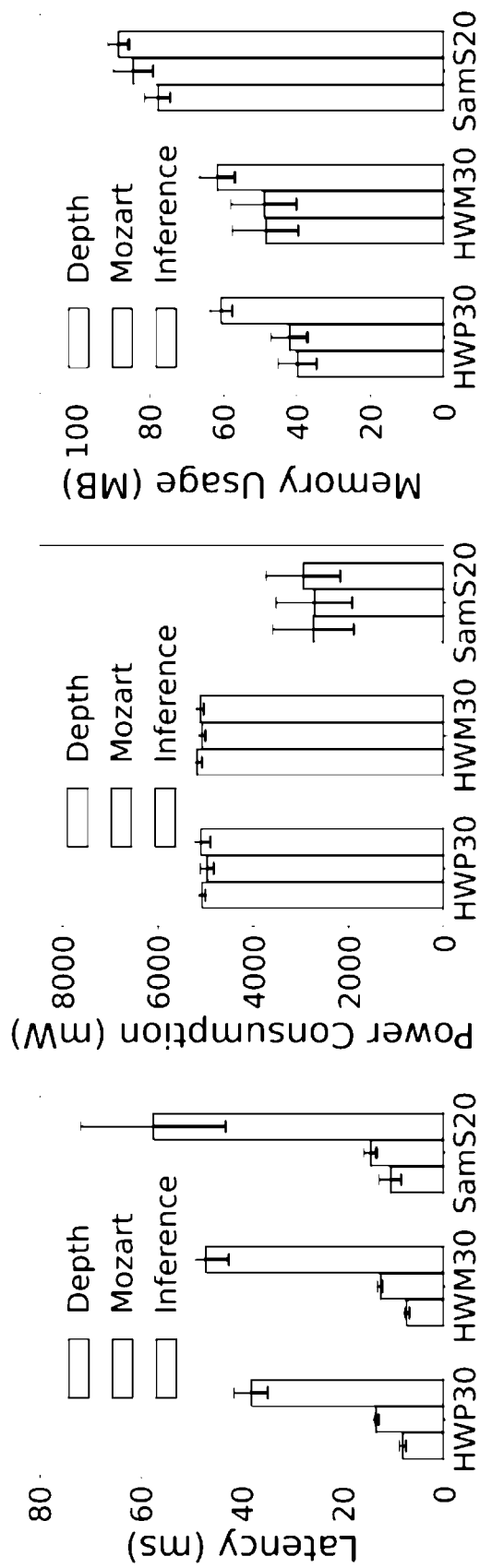
(a) Real-time Android App.

FIG. 12A



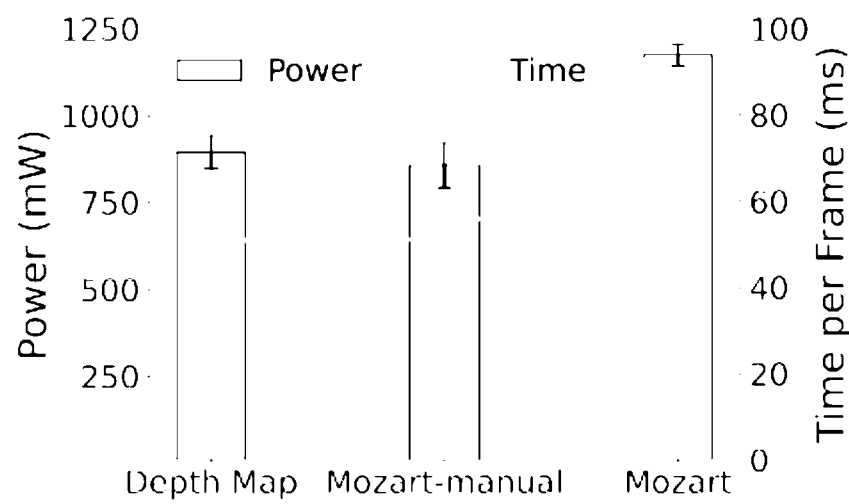
(b) Three ToF modules.

FIG. 12B



(a) Generating depth, Mozart maps, and inference with Mozart maps on smartphones.

FIG. 13A



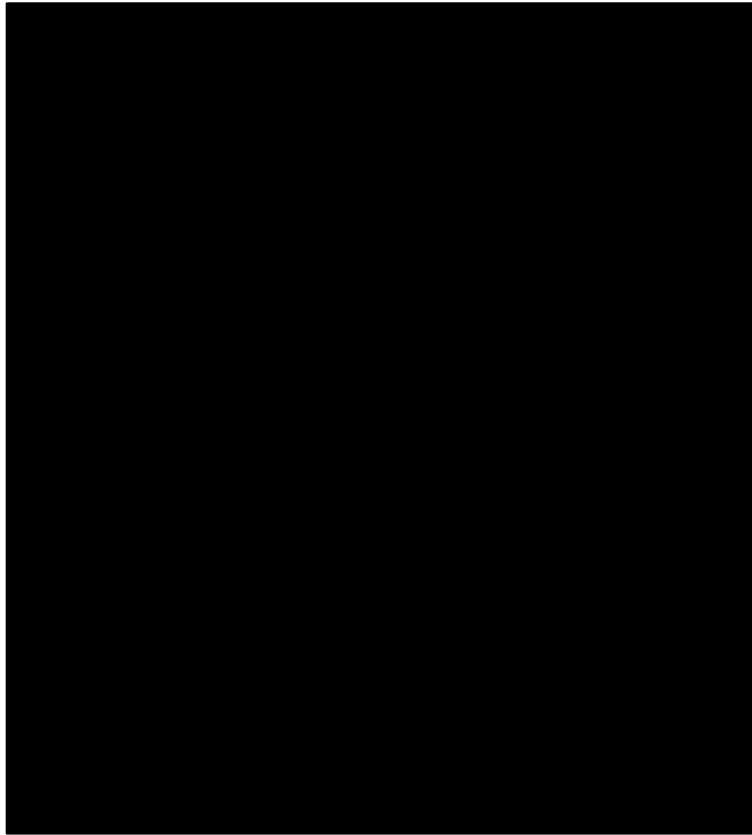
(b) Generating maps on Jetson Xavier.

FIG. 13B



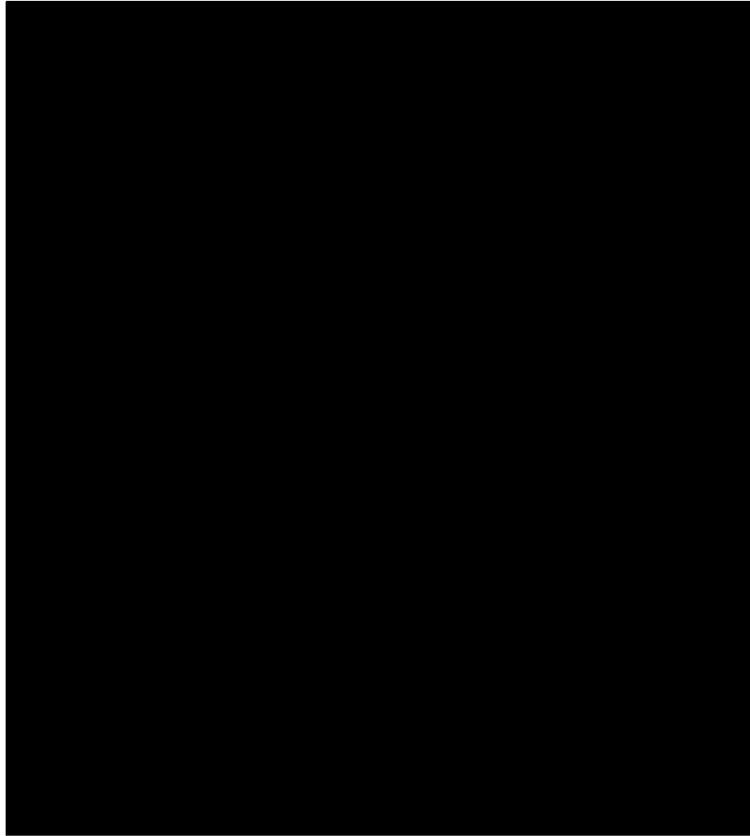
(a) Human tracking

FIG. 14A



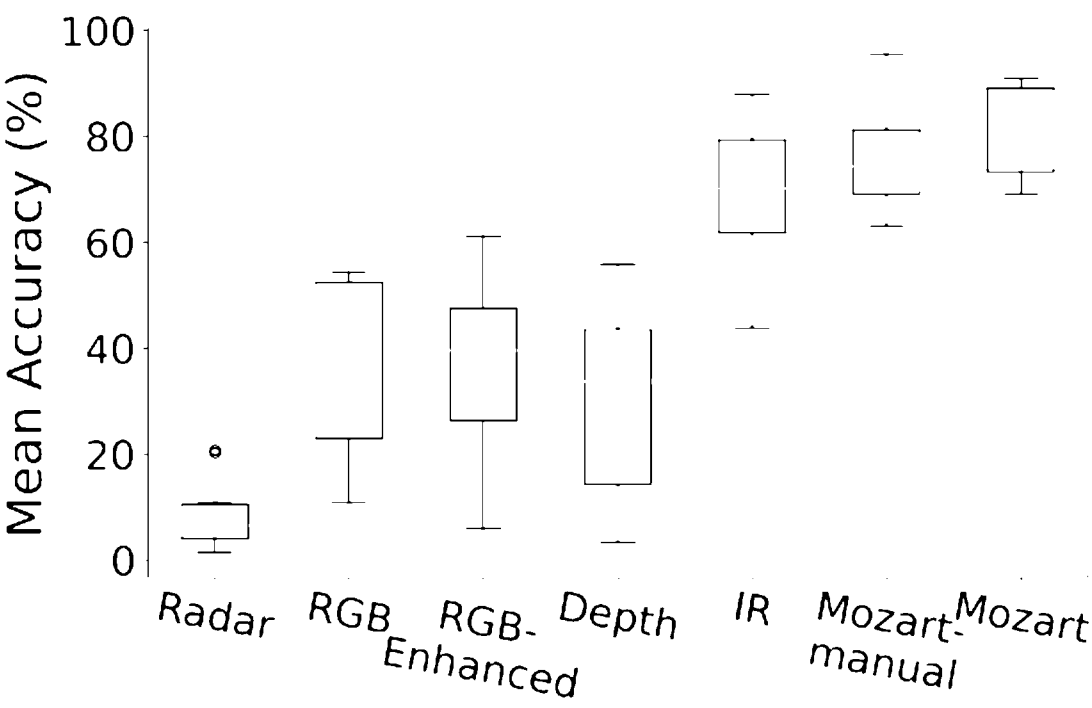
(b) Face recognition

FIG. 14B



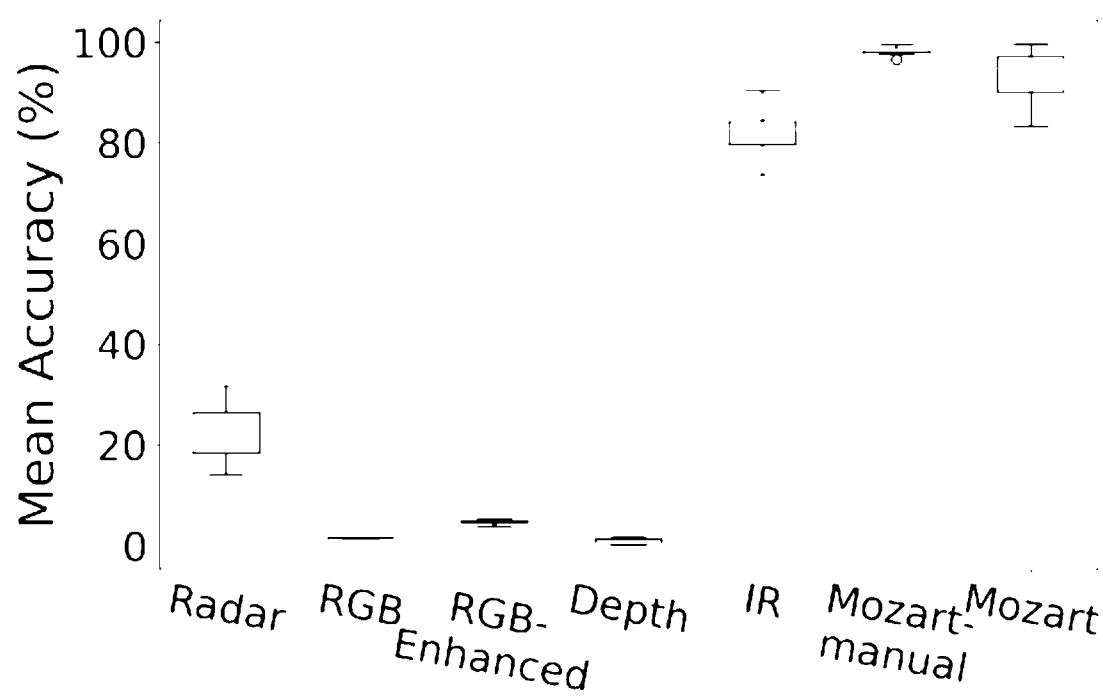
(c) Gesture recognition

FIG. 14C



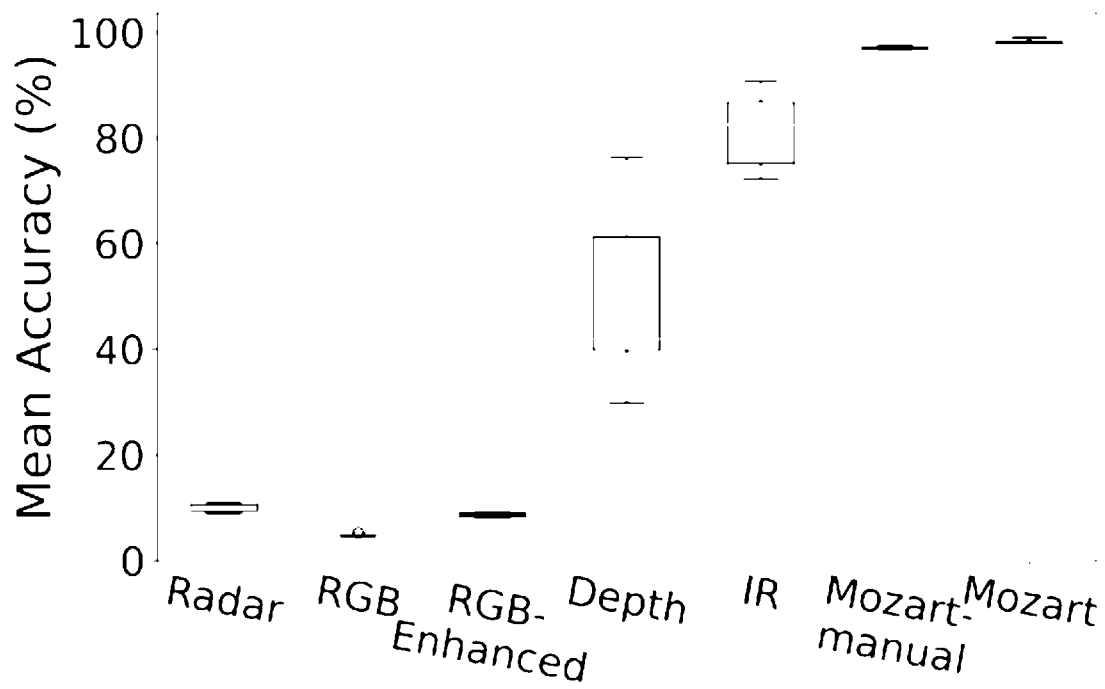
(a) Human tracking

FIG. 15A



(b) Face recognition

FIG. 15B



(c) Gesture recognition

FIG. 15C

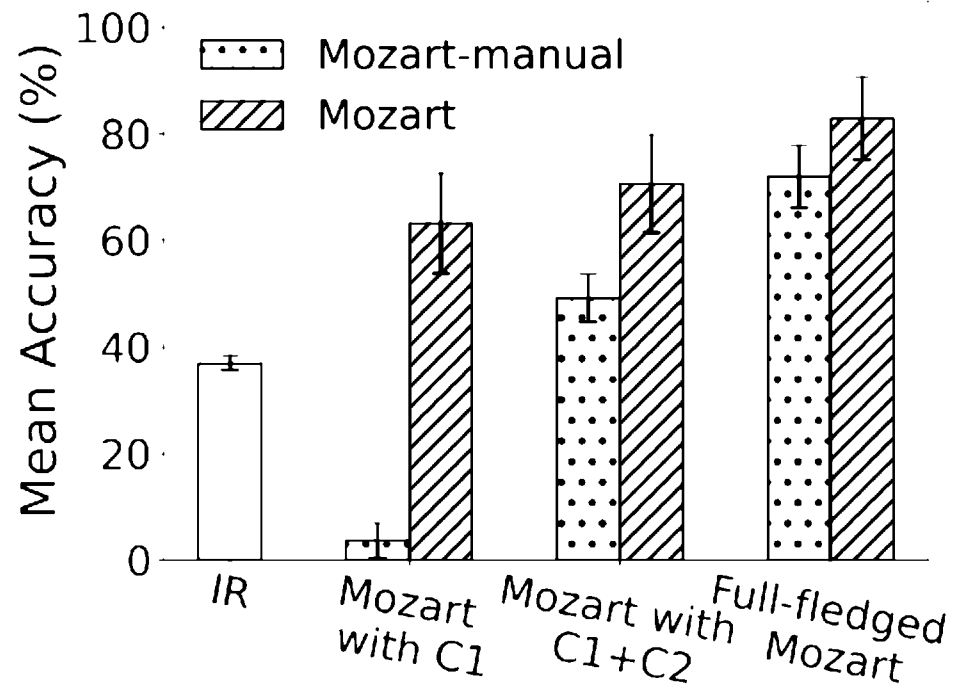


FIG. 16

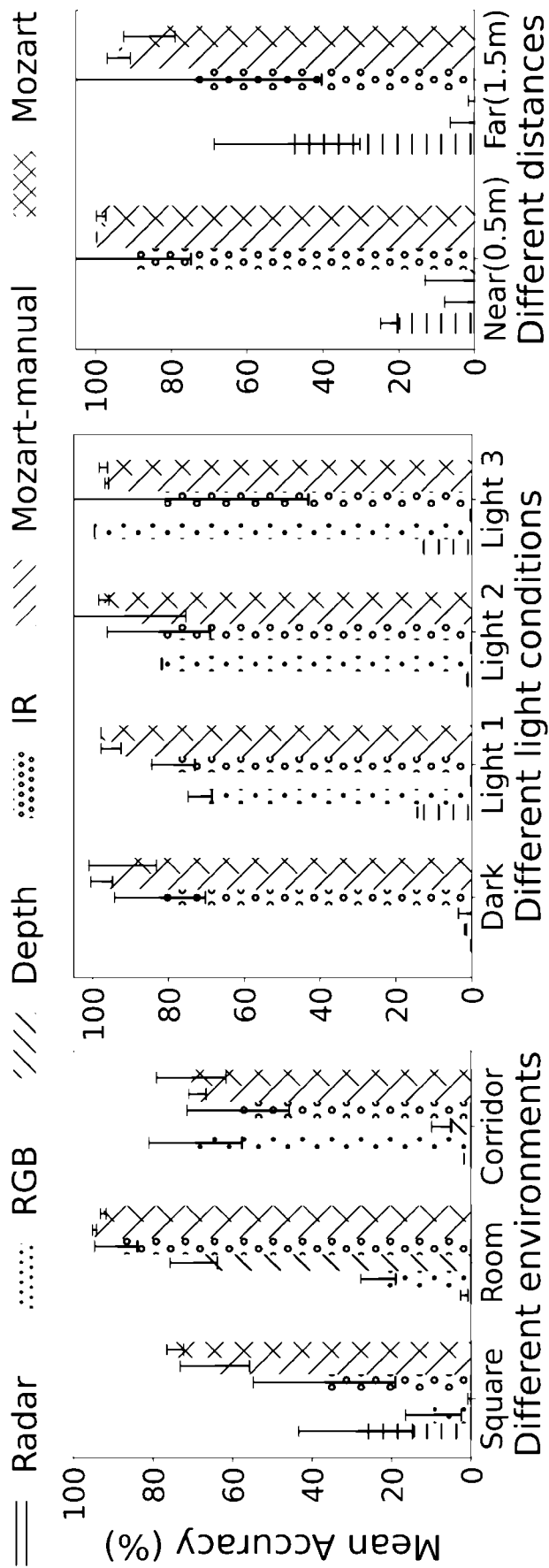


FIG. 17

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/113154

A. CLASSIFICATION OF SUBJECT MATTER

G06T15/04(2011.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC:G06T

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNTXT;ENTXT;DWPI;CNKI;IEEE: TOF、time w of w flight、depth、map、camera、phase、component、textur+、dark、light+、bright+、IR、infra+、distance、VR、learn+、monitor、illuminat+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 111402392 A (HONG KONG SHINING CLOUD TECHNOLOGY COMPANY LIMITED) 10 July 2020 (2020-07-10) description, pages 3-5	1-20
Y	CN 115249225 A (ALIBABA CLOUD COMPUTING LTD.) 28 October 2022 (2022-10-28) description, pages 6-13	1-20
A	CN 109213138 A (POWERSHION TECHNOLOGY INC.) 15 January 2019 (2019-01-15) the whole document	1-20
A	CN 112866675 A (OPPO GUANGDONG MOBILE TELECOMMUNICATIONS CO., LTD.) 28 May 2021 (2021-05-28) the whole document	1-20
A	US 2019033448 A1 (CORNELL UNIV.) 31 January 2019 (2019-01-31) the whole document	1-20
A	US 2021264626 A1 (GUANGDONG OPPO MOBILE TELECOMMUNICATIONS CORP., LTD.) 26 August 2021 (2021-08-26) the whole document	1-20

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

05 February 2024

Date of mailing of the international search report

21 February 2024

Name and mailing address of the ISA/CN

CHINA NATIONAL INTELLECTUAL PROPERTY
ADMINISTRATION
6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing
100088, China

Authorized officer

WU,Qian

Telephone No. (+86) 010-53961822

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/113154

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	111402392	A	10 July 2020	None			
CN	115249225	A	28 October 2022	None			
CN	109213138	A	15 January 2019	None			
CN	112866675	A	28 May 2021	None			
US	2019033448	A1	31 January 2019	JP	2018513967	A	31 May 2018
				JP	6722194	B2	15 July 2020
				KR	20170134502	A	06 December 2017
				KR	102456875	B1	19 October 2022
				WO	2016149438	A1	22 September 2016
				US	10605916	B2	31 March 2020
				EP	3271747	A1	24 January 2018
				EP	3271747	B1	25 May 2022
				CA	2979836	A1	22 September 2016
				CA	2979836	C	29 August 2023
				US	2020191967	A1	18 June 2020
				US	10983216	B2	20 April 2021
US	2021264626	A1	26 August 2021	US	11494925	B2	08 November 2022
				WO	2020087485	A1	07 May 2020
				EP	3876528	A1	08 September 2021
				EP	3876528	A4	23 February 2022
				CN	112889272	A	01 June 2021