UltraDepth: Exposing High-Resolution Texture from Depth Cameras

Zhiyuan Xie
The Chinese University of Hong Kong
Hong Kong SAR, China
xavier_ie@link.cuhk.edu.hk

Xiaomin Ouyang
The Chinese University of Hong Kong
Hong Kong SAR, China
xmouyang@link.cuhk.edu.hk

Xiaoming Liu
Michigan State University
East Lansing, MI, USA
liuxm@cse.msu.edu

Guoliang Xing*
The Chinese University of Hong Kong
Hong Kong SAR, China
glxing@cuhk.edu.hk

ABSTRACT

Time-of-flight (ToF) depth cameras have been increasingly adopted in various real-world applications, e.g., used with RGB cameras for advanced computer vision tasks like 3-D mapping or deployed alone in privacy-sensitive applications such as sleep monitoring. In this paper, we propose *UltraDepth*, the first system that can expose high-resolution texture from depth maps captured by off-the-shelf ToF cameras, simply by introducing a distorting IR source. The exposed texture information can significantly augment depth-based applications. Moreover, such a capability can be used to launch privacy attacks, which poses a major concern due to the prominence of ToF cameras. To design *UltraDepth*, we present an in-depth analysis on the impact of the distorting IR light on the distance measurement. We further show that, the reflection properties (reflectivity and incidence angle) of the objects will be encoded in the distorted depth map and hence can be leveraged to reveal texture of objects in *UltraDepth*. We then propose two practical implementations of *UltraDepth*, i.e., reflection-based and external IR-based implementations. Our extensive real-world experiments show that, the depth maps output by *UltraDepth* achieve 89.06%, 99.33%, 81.25% mean accuracy in object detection, face recognition and character recognition, respectively, which offers over 10× improvement over the ordinary depth maps and even approaches the performance of RGB and IR images in a number of scenarios. The findings of this work provide key insights for new research on depth-related computer vision and security of depth sensing devices.

CCS CONCEPTS

• **Computer systems organization → Sensors and actuators**; • **Security and privacy → Hardware attacks and countermeasures**; • **Computing methodologies → Computer vision**.

*Corresponding author

SenSys '21, November 15–17, 2021, Coimbra, Portugal

ACM ISBN 978-1-4503-9097-2/21/11…$15.00
https://doi.org/10.1145/3485730.3485927

KEYWORDS

ToF depth camera, Texture exposure, Depth data processing, Attack and defense

ACM Reference Format:
Zhiyuan Xie, Xiaomin Ouyang, Xiaoming Liu, and Guoliang Xing. 2021. UltraDepth: Exposing High-Resolution Texture from Depth Cameras. In *The 19th ACM Conference on Embedded Networked Sensor Systems (SenSys '21), November 15–17, 2021, Coimbra, Portugal.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3485730.3485927

1 INTRODUCTION

Recently, the Time-of-Flight (ToF) depth camera has attracted significant attention in the research community as well as in the industry. The overall Time-of-Flight sensor market is expected to grow to USD 6.9 billion by 2025 [37]. The ToF cameras are often used together with RGB cameras for advanced computer vision tasks such as scene labeling [56], 3-D mapping [61] and object detection [17]. Besides, as the depth maps only contain distance information without revealing details of the scene such as personal identities, ToF depth cameras are increasingly deployed for privacy-sensitive applications such as fall detection [32], sleep monitoring [21], and surveillance systems [57]. Moreover, a number of vendors [1–3] provide depth-only modules, which not only support a range of depth-only based studies [26, 33, 62] but also lead to privacy-preserving customer products [4, 5, 44].

In this paper, we propose *UltraDepth*, the first system that can expose high-resolution texture from the depth maps captured by off-the-shelf ToF cameras, simply by using a distorting IR source. Figure 1 shows several illustrative examples of the ordinary depth maps and the depth maps output by *UltraDepth*. It's obvious that the latter can expose rich texture of various scenes. As a result, *UltraDepth* can augment the performance of ToF cameras in various perception tasks like face recognition and object detection. Without requiring extra RGB cameras, *UltraDepth* can save cost and reduce form factor for real-world applications. Moreover, the high resolution texture exposed by *UltraDepth* can be used to launch privacy attacks, which poses a major concern due to the prominence of ToF cameras.

The design of *UltraDepth* is based on the key idea that the distance measurement of indirect Time-of-Flight (iToF) cameras will be distorted in the presence of an additional IR source, and the resulted distortion varies in different areas of the scene, which exposes the texture of the objects in the scene. To design *UltraDepth*,

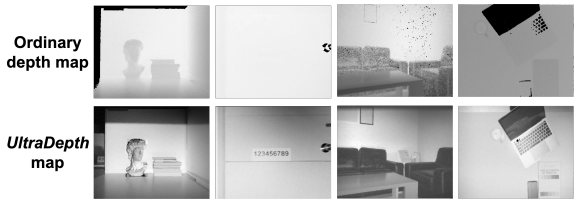


Figure 1: Illustrative examples of the ordinary depth maps and the depth maps output by *UltraDepth*. The latter is able to expose rich texture of the scene.

we first present an in-depth analysis on the impact of the distorting IR light on the distance measurement based on the working principle of iToF camera. We further show that under the influence of the distorting IR source, the reflection properties (reflectivity and incidence angle) of the objects will be encoded in the distorted depth map and hence can be leveraged to reveal texture of objects in *UltraDepth*. We then propose two implementations of *UltraDepth*, namely reflection-based and external IR-based implementations, based on different ways of constructing the distorting IR source. Specifically, the reflection-based *UltraDepth* generates a distorting IR source through the light reflected from the ambient objects (e.g., cover board, wall or furniture) near the emitter of the iToF camera. Therefore, the reflection-based *UltraDepth* needs merely a simple cover board reflecting part of the emitted IR light. The external IR-based implementation utilizes a designated external IR source (e.g., another device with the same model of the original ToF camera) to interfere with the received light from the measured area.

We validate the effectiveness of texture exposure of *UltraDepth* in extensive real-world experiments with different perception tasks, daily scenarios and settings on several practical factors. The results show that, the depth maps output by *UltraDepth* achieve 89.06%, 99.33%, 81.25% mean accuracy in object detection, face recognition and character recognition, respectively, which offers over 10× improvement compared with the ordinary depth maps. Since *UltraDepth* is insensitive to visible light, its object recognition accuracy can surpass RGB camera in the scenes with poor illumination. Moreover, we show that the reflection-based *UltraDepth* can be easily realized in real-world applications with the help of ambient objects such as a wall or furniture. Finally, to address the potential privacy attacks enabled by *UltraDepth*, we briefly discuss possible defense mechanisms for two different implementations of *UltraDepth*.

Our key contributions are summarized as follows:

- To the best of our knowledge, *UltraDepth* is the first ToF depth-based system that can expose detailed texture information from the depth maps output by off-the-shelf iToF depth cameras.
- We are the first to provide an in-depth analysis on the principle of exposing high-resolution texture from depth cameras.
- Based on our theoretical analysis, we propose two practical implementations of *UltraDepth*, i.e., reflection-based and external IR-based *UltraDepth*, which can be easily realized for real-world applications.
- We conduct comprehensive experiments to validate the effectiveness of texture exposure of *UltraDepth* in various perception tasks, e.g., object detection, face recognition and character recognition.
- We discuss possible defense solutions for privacy attacks that can be launched by two *UltraDepth* implementations.

2 RELATED WORK

Applications of Depth Cameras. Recently, ToF depth cameras have been increasingly adopted in various real-world applications. They are often used together with RGB cameras for computer vision tasks such as scene labeling [49, 56], 3-D mapping [16, 61], object detection [17, 52], semantic instance segmentation [25] etc. Depth cameras are also fused with wearable inertial sensors for motion analysis such as biomechanical gait analysis [8] and gesture recognition [10]. Moreover, with more and more vendors providing depth-only modules, depth-only cameras (instead of both depth/RGB cameras) are increasingly deployed for privacy-sensitive applications [14, 28, 36]. For example, in [28], ToF cameras are used in a smart room for privacy-preserving people tracking. In [36], an algorithm is proposed to use only the depth information from a ceiling-mounted ToF camera to detect people. In [14], the depth cameras are used for human posture recognition at homes.

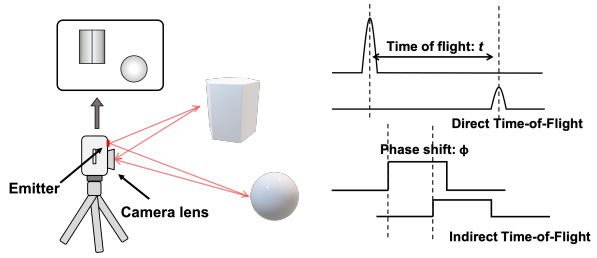
Augmenting ToF Cameras. In view of the prevalence of ToF cameras, a family of techniques have been proposed to improve the sensing performance of ToF depth cameras. Specifically, [18, 19, 42, 46] focus on improving theoretical or empirical noise models of ToF cameras for better distance measurement. The energy-efficient epipolar imaging approach proposed in [6] improves the robustness of depth measurement in several extreme scenarios, e.g., in presence of strong outdoor sunlight, interference from other ToF cameras, and severe camera shaking. Moreover, [27, 30, 31, 41] model and compensate for the internal scattering of ToF cameras where the distance measurements of far objects are severely affected by the near objects. Different from previous work that aims at improving the robustness of distance measurement of ToF cameras, our work focuses on exposing rich texture of the captured scene from the depth map, which broadens the applications of depth cameras, e.g., for face recognition or object detection.

Privacy Attacks on LiDAR and Radar. Similar to ToF depth camera, LiDAR and radar are widely accepted as privacy-preserving imaging sensors since they only show point cloud of the scene. However, recent studies have demonstrated risks of privacy leakage in systems based on LiDAR and radar. Specifically, [63] finds that the millimeter wave radar is capable of identifying individuals with the aid of a deep recurrent network. [50] demonstrates the risk of eavesdropping private conversations through LiDAR sensors. However, to the best of our knowledge, our work is the first to show that off-the-shelf iToF camera can be used to expose rich texture information of depth maps and hence reveal sensitive information in real-world scenarios.

3 BACKGROUND

In this section, we introduce the technical background, including the difference between direct ToF cameras and indirect ToF cameras, and the principle of depth measurement in iToF cameras via IR light.

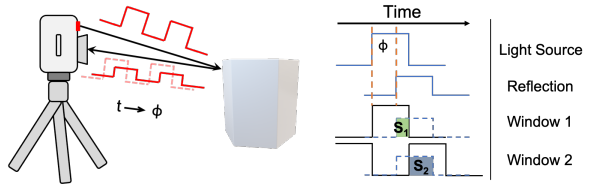
A simplified system diagram of how a Time-of-Flight (ToF) camera works is depicted in Figure 2(a). A ToF depth camera emits IR light, illuminates the scene to be captured and receives the IR light reflected by the objects in the scene. The distance measurement is derived based on the fact that the round trip time-of-flight (t) of the IR signal between the scene and the camera is strictly proportional to the distance. We have $t = 2d/c$, where d is the distance of



(a) Capturing depth maps with a ToF camera (b) dToF and iToF

Figure 2: (a) The IR light is first emitted by the ToF camera, then reflected by the scene, and finally received by the camera. The time-of-flight of IR light during this process is measured to calculate the distance of the scene. (b) Up: direct Time-of-Flight (dToF) measures the time delay. Down: indirect Time-of-flight (iToF) measures the phase shift.

the scene and c is the speed of light. Thus, the design goal of ToF camera is to measure the time-of-flight as accurately as possible. Off-the-shelf ToF cameras can be divided into two types based on how the time-of-flight is measured: direcet Time-of-Flight (dToF) and indirect Time-of-Flight (iToF). Figure 2(b) illustrates the basic idea of these two techniques: dToF measures the time delay directly using high-precision Single-Photon Avalanche Diode (SPAD) while iToF measures the phase shift between the emitted signal and received signal to calculate the time delay indirectly. Compared with dToF, iToF is more suitable for 3D imaging applications due to its low cost and high resolution [22]. Currently, most of the ToF modules on mobile devices (especially Android smartphones) on the market adopt the iToF technology [9, 11].



(a) (b)

Figure 3: Measurement of phase shift in iToF. The camera shutter is in the same phase as the emitted light pulse. The energies of received IR light in two consecutive windows S_1 and S_2 are used to measure the phase shift of the received light signal.

Specifically, the iToF camera has a shutter with the same phase as the emitted light pulses and uses the phase shift of the returned light to calculate the time-of-flight. Figure 3 illustrates the general principle of calculating the phase shift in an iToF camera. The laser source (typically a vertical-cavity surface-emitting laser, VCSEL) emits pulses of light continuously and periodically. A square wave rather than a sine wave is commonly used because it can be easily realized using digital circuits [23, 35]. The pulse width of the square wave (T) determines the range of measurement, which can be configured by the user. Then there will be a phase shift between the emitted and received light due to the time-of-flight, as indicated by ϕ in Figure 3. To measure the phase shift, the camera shutter

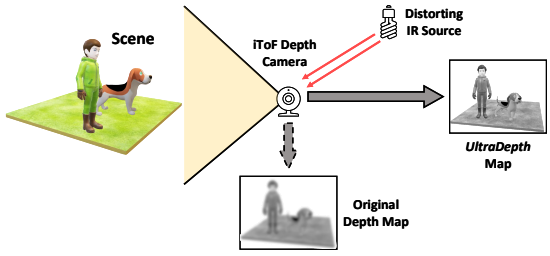


Figure 4: *UltraDepth* is designed to enable exposing high-resolution texture from the depth maps captured by the an iToF depth camera.

opens and closes periodically with the same frequency and the same phase as the emitted laser pulses. Then the camera will read out the energy of received light in two successive windows (denoted as S_1 and S_2), as shown in window 1 and window 2 of Figure 3. Therefore, the phase shift ϕ can be calculated using the energy ratio of received light in the window 1 and window 2:

$$\phi = \frac{S_2}{S_1 + S_2} \times \pi.$$

Then the time-of-flight t can be calculated by $t = \frac{\phi}{\pi} T$.

Besides the basic designs, mainstream off-the-shelf iToF cameras also adopt some advanced techniques to mitigate the influence of ambient light on the distance measurement. For example, pulse-based iToF cameras detect the ambient light in non-pulse time [51] while the continuous-wave iToF cameras take multiple samples (using more than two windows) per measurement and calculates the phase shift using the subtractions of energy samples to reduce the energy offset caused by ambient light during the process of each distance measurement [24]. However, these techniques can only reduce the interference when the ambient light irradiated within the measurement area is constant and continuous.

4 APPLICATIONS

In this paper, we propose a novel system *UltraDepth*, which can expose high-resolution texture information from the depth maps captured by iToF depth cameras. We now first briefly introduce the typical setup of *UltraDepth* and then present typical applications.

Figure 4 depicts the basic setup of *UltraDepth*. Assume there is an iToF depth camera that irradiates IR light on the scene and records the depth map, where the depth map only captures the distance information of the scene. *UltraDepth* aims to extract rich and detailed texture information from the captured depth map. To achieve this goal, *UltraDepth* first adopts the proposed methods in this paper (described in Section 6) to add a distorting IR source and impose continuous and indelible interference that can manipulate the distance measurements in the depth maps of the iToF camera. Then the detailed texture information of objects in the scene will be encoded in the distorted depth maps, turning the blurry distance maps into gray-scale like images. *UltraDepth* is thus able to expose rich texture information and leverage them for various applications.

Augmenting Depth Performance. The exposed texture information in the depth maps through *UltraDepth* can be used to augment various applications of ToF depth camera. There are scenarios such as surveillance [57] and fall detection [7] where only a

Figure 5: The key idea of texture exposure is to distort distance measurement by adding a distorting IR source

depth camera (instead of both depth/RGB cameras) is present. In such cases, we can combine the original depth map (with distance information) and the distorted depth map (with texture information) captured through *UltraDepth* to improve the performance of tasks such as object detection and 3D constructions, eliminating the need of an extra RGB camera. Our experiments show that *UltraDepth* can accurately detect faces and daily objects. Besides, the distance map and texture map in *UltraDepth* are from the same ToF camera. Therefore, they do not need to be calibrated to achieve the alignment between the two maps, which is a compute-intensive procedure needed for common RGB-D cameras due to the different radial distortion and the rotation and translation between the depth camera and the RGB camera [59].

Privacy Attacks. *UltraDepth* can be used to attack user's privacy by revealing rich texture information from ToF depth cameras. Such attacks can be launched for a number of real-world applications if an adversary can get access to the *UltraDepth* output. Specifically, an indoor attack scenario may occur when the iToF depth cameras are used in smart home applications, such as sleep monitoring [21, 38] or fall detection [7, 60], which are previously considered capable of preserving privacy and anonymity. Due to the high resolution of exposed texture information, the attacks can cause severe privacy leakage of users, especially in privacy-sensitive spots, like bedroom or bathroom. The indoor layout revealed by the attacked depth camera will also render the users vulnerable to physical attacks such as robbery. Moreover, such attacks are also possible in public areas, e.g. building entrances and elevator cars, when the iToF cameras are mounted for privacy-sensitive applications such as people counting [40]. Similarly, the adversary can obtain a video stream that exposes clear texture information in the scene, which can reveal not only personal identities but also text contents on the papers or smartphones people hold, especially when the iToF cameras are ceiling-mounted.

5 DESIGN PRINCIPLE OF *UltraDepth*

5.1 Key Idea

As introduced in Section 3, in iToF depth cameras, the phase shift of the received light signal is measured using the energy ratio of received light in two successive windows. Therefore, as Figure 5 shows, the key idea of *UltraDepth* is to introduce a distorting IR source to change the energy ratio of received signal, which will distort the distance measurement of the iToF camera. Moreover, the impact of this distorting IR source varies in different areas of the scene (e.g., digits and background on the credit card in Figure 5), effectively increasing the difference between the perceived (distorted) depth measurements. Therefore, the detailed texture information of the scene will be exposed in the distorted depth map.

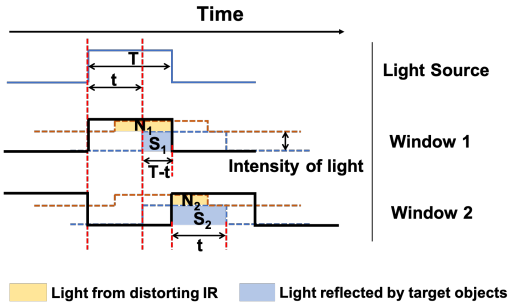


Figure 6: Impact of the distorting IR source on distance measurement. The additional IR light (N_1, N_2) from the distorting IR source will affect the measured phase shift through changing the energy ratio of the received light.

Next, we will introduce how the distorting IR source can affect the distance measurement of iToF depth cameras in Section 5.2, and how to effectively expose rich texture information from the distorted depth map of the scene in Section 5.3.

5.2 Impact of Distorting IR Source on Distance Measurements

In our context, a distorting IR source for the distance measurement of a specific object can come from the changeable ambient light, the reflected light from ambient objects and even an external IR light source. In this section, we will present how a distorting IR source affects the distance measurement of an iToF depth camera.

Figure 6 shows the received light signal on the iToF camera with a distorting IR source, where the yellow rectangles denote the additional light from the distorting IR source. Compared with the one depicted in Figure 3, the additional light from the distorting IR source (N_1 and N_2) will change the energy ratio of the received light and further affect the measured phase shift and distance. We then use the model shown in Figure 6 to give a quantitative analysis on the impact of distance measurement from the distorting IR.

Suppose S_1 and S_2 denote the energy of received light from the targeting object in window 1 and window 2, respectively; N_1 and N_2 denote the energy of received light from the distorting IR source in window 1 and window 2, respectively. T is the width of the emitted light pulse and t is the time-of-flight. Then the real distance of the object is $d = \frac{S_2}{S_1+S_2} \times \frac{cT}{2}$, where c is the speed of light. However, the measured distance after adding the distorting IR source is:

$$\widetilde{d} = \frac{S_2 + N_2}{S_1 + S_2 + N_1 + N_2} \times \frac{cT}{2}.$$

In Figure 6, the received intensity of light reflected by the objects can be expressed as E/d^2, which is inversely proportional to the square of the distance d [39]. Here E is the received intensity from the object at a unit distance, which is determined by the emission power of the iToF camera and reflection properties (e.g. reflectivity, incidence angle of light) of the object. Then as the energy of received light is equal to the accumulation of light intensity over time, S_1 and S_2 can be expressed as: $S_2 = E/d^2 \cdot t$ and $S_1 = E/d^2 \cdot (T - t)$. By substituting S_1 and S_2 with above two equations (where $t = 2d/c$), and with rearrangement, the relationship between the measured

distance $\widetilde{d}$ and the real distance d can be presented as:

$$\widetilde{d} = \left[1 - \frac{[(N_1 + N_2)d - N_2 D]d}{ET + (N_1 + N_2)d^2}\right] \times d, \quad (1)$$

where $D = \frac{cT}{2}$ is the iToF's range of measurement.

Eqn.(1) shows that for a specific object (E is fixed) and a fixed range of measurement (D), the measured distance with distorting IR is determined by the energy distribution of the distorting IR light in the two windows (N_1 and N_2) as well as the real distance d. We note that when $d > \frac{N_2}{N_1+N_2}D$, the measured distance will be smaller than the real one, and vice versa. We now discuss the impact of distance measurement with different N_1 and N_2 based on Eqn.(1):

$N_1 = N_2$. This usually occurs when the distorting IR is from the ambient light, where the energies of the additional light in all the windows are equal. In this case, when the real distance d is smaller than $\frac{D}{2}$ (half of the range of measurement), the measured distance $\widetilde{d}$ will be larger than the ground truth, and vice versa. Fortunately, as mentioned in Section 3, the interference from ambient light is easy to be eliminated by detecting the ambient light in non-pulse time [51] or subtracting the ambient light with more windows [24].

$N_1 \gg N_2$ and $N_1 \ll N_2$. When the distorting IR source is set so that $N_1 \gg N_2$, $d > \frac{N_2}{N_1+N_2}D$ will always hold, and the measured distance $\widetilde{d}$ will always be smaller than the real value. While when $N_1 \ll N_2$, $d < \frac{N_2}{N_1+N_2}D$ is true most of time, it leads to a larger distance measurement.

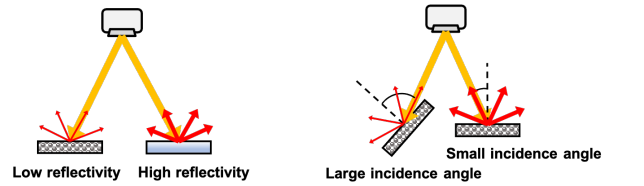
5.3 Exposing Textures in Depth Maps

As shown in Section 5.2, the distance measurement of the iToF depth camera will be affected by the interfering IR source, resulting in a distorted depth map. In this section, we will describe how the distorted depth map can expose rich texture information of the scene. The key idea is that the impact of the distorting IR source varies among different areas of the scene due to their different reflectivities and incidence angles for the light, which is implied by the variable E in Eqn.(1). Therefore, the distorted depth map (composed of a 2-D array of measured distance $\widetilde{d}$) is actually encoded with the properties (reflectivities and incidence angles) of the irradiated scene, which exposes more textures than the original depth map. In other words, by manipulating the distance measurements of different areas of a scene, we can effectively increase the "granularity" of depth maps, exposing texture details of the scene. To this end, we will present how the textures make differences in response to the distorting IR source.

In the original iToF camera system, the depth map is only related to distance of the irradiated object as the phase shift only depends on the time-of-flight of the received signal while having nothing to do with the specific intensity of received light signal. However, according to the Lambertian reflection model [39, 43], the intensity of IR signal reflected by the scene at distance d can be calculated by:

$$E_d = E_0 \alpha \cos\theta / (8d^2), \quad (2)$$

where E_0 is a constant determined by the settings of the depth camera, α is the reflectivity of the object at the IR wavelength of the ToF camera and θ is the angle of incidence.



(a) Objects with different reflectivities (b) Objects with different incidence angles

Figure 7: Two special cases of different *albedos* for objects at the same distance. Both high reflectivity and small incidence angle will make the *albedo* larger, resulting in a stronger intensity for reflected light (indicated by the thick arrow in (a) and (b).

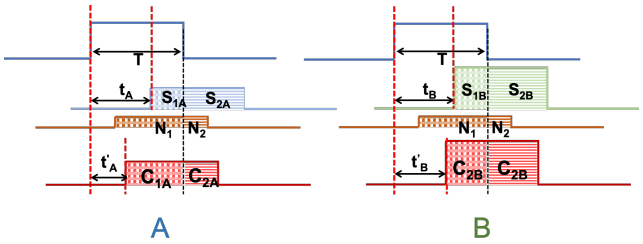


Figure 8: A and B have the same distance (energy ratio in two windows) but different *albedos* (intensities of received light). They can be differentiated in the distance measurement when adding a distorting IR source in our design.

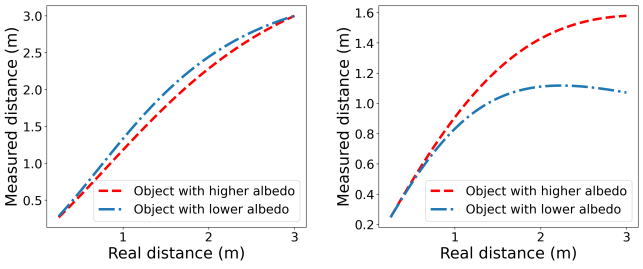
Therefore, we can see that even for two points with the same distance, the intensity of received light can still be different due to various reflectivities (α) and the incidence angles (θ). In this paper, we define a new variable "*albedo*" [1] $\beta = \alpha \cos\theta$, to quantify the two factors from the object itself that have impact on the intensity of the received light. Figure 7 gives two special cases of different *albedos* for objects with the same distance. In Figure 7(a), the two objects have unequal *albedos* due to their different reflectivities, while for the two objects in Figure 7(b), their *albedos* differ as they have different incidence angles. In reality, the impact of the two factors usually appears at the same time.

We now show how the intensity of the received light signal affects the distance measurement when a distorting IR source is present. As Figure 8 shows, two points **A** and **B** in the scene have the same distance but different *albedos*. When there is no distorting IR (i.e. in the original ToF camera systems), we have $t_A = \frac{S_{2A}}{S_{1A}+S_{2A}} = \frac{S_{2B}}{S_{1B}+S_{2B}} = t_B$, i.e., the measured distances are the same for **A** and **B**. In this case, we can not differentiate these two points even they have different *albedos*. However, when there is a distorting IR, we have $t'_A = \frac{C_{2A}}{C_{1A}+C_{2A}} \neq \frac{C_{2B}}{C_{1B}+C_{2B}} = t'_B$, where $C_{ij} = S_{ij} + N_i, i = 1 \text{ or } 2, j = A \text{ or } B$. In this case, the two points will have different distance measurements and can be differentiated in the depth maps. In summary, the distorting IR can introduce the impact of the *albedo* into the distance measurement. As a result, the depth map is distorted according to various *albedos* in different areas of objects, thereby exposing textures. Without the distorting IR

[1] We note that this is a slight abuse of term albedo, which is defined as a measure of the diffuse reflection of solar radiation in optics.

source, the original depth camera would only capture the distance of objects while oblivious to the impact of other factors including reflectivities and incidence angles of the objects, which actually are already embedded in the intensity of received IR light.

We further quantify the impact of different *albedos* on the distance measurement of the distorted depth map. Recalling Eqn.(1), for objects with the same distance d, if the distorting IR is fixed, i.e. N_1, N_2 are unchanged, the measured distance will be solely determined by E, where $E = E_0\alpha\cos\theta = E_0\beta$ and E_0 is a constant that is related to the camera's emitter. In this case, the relationship between the measured distance and the real distance will be determined by the *albedo* β, which is shown in Figure 9. We summarize three key factors and their impacts on the performance of *UltraDepth* from Figure 9: (1) The object's albedo. For both cases, the object with lower *albedo* is less resistant to interference, i.e., its distance measurements are distorted more severely than the object with higher *albedo*. (2) The distorting IR light. The setting $N_1 \gg N_2$ for the distorting IR source will be a better choice to differentiate objects and expose more texture information. When $N_1 \ll N_2$, the impact of different *albedos* on the distance measurement is less significant. However, when $N_1 \gg N_2$, the two objects with different *albedos* will have significantly different distance measurements, especially when they are both located at a larger distance (e.g., > 1m). (3) The distance of object. When $N_1 \gg N_2$ (which is the setting of two *UltraDepth* implementations in Section 6), the objects at a larger distance from the camera more likely expose texture details since the intensity of light reflected off the objects is weaker than that at a smaller distance, which makes them more vulnerable to the distorting IR light.



(a) $N_1 \ll N_2$ (b) $N_1 \gg N_2$

Figure 9: The impact of different *albedos* on the distance measurement, with two fixed distorting IR sources. (a) when $N_1 \ll N_2$, (b) when $N_1 \gg N_2$.

6 SYSTEM IMPLEMENTATION

In this section, we will discuss two types of practical implementations of *UltraDepth* that can expose texture information in the depth map, i.e., reflection-based and external IR-based *UltraDepth* implementations, which differ in how to construct a distorting IR source introduced in Section 5. Specifically, the reflection-based implementation attempts to generate a distorting IR source through the light reflected from the ambient objects near the emitter, while the second method directly utilizes a designated external IR source to interfere with the received light from the measured area. In both *UltraDepth* implementations, the intensity of received distorting IR light (namely N_1, N_2) is comparable to that of the IR light reflected off the objects, making the depth measurement of objects

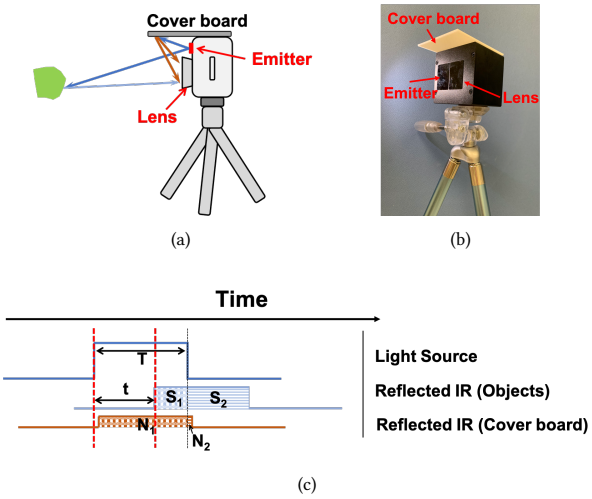


Figure 10: Schematic and physical diagram of the reflection-based *UltraDepth* implementation. The cover board near the ToF camera will reflect part of the emitted IR light to distort the distance measurement.

easier to be distorted (according to analysis in Section 5). Therefore, high-resolution texture information can be exposed in the two real-world implementations. Furthermore, we propose a region-based linear transformation method for *UltraDepth* to refine the texture information in the depth map.

6.1 Reflection-based Implementation

To impose an effective distortion on the depth map, the distorting IR must have the same wavelength and modulation frequency as the ToF depth camera. In this way, it can continuously manipulate each frame of captured depth maps and thus expose more texture information. Then the most straightforward and convenient way to construct a distorting IR light that meets the requirement is to utilize the IR light emitted by the depth camera itself.

Based on this idea, we propose a reflection-based implementation of *UltraDepth*, where a cover board is placed near the emitter to reflect part of the IR light emitted by the camera, which constitutes a distorting IR source. Figure 10(a) shows a schematic diagram for the reflection-based implementation, where the emitter emits the IR light (depicted as the blue arrows) to illuminate the scene while part of the emitted IR is blocked and reflected by the cover board (brown arrows). A fraction of the reflected IR light will fall into the camera lens, thereby introducing stable and indelible interference to the distance measurement of the scene. Compared with the measured objects in the scene, the cover board is much closer to the emitter and camera lens so that the time-of-flight for the distorting IR light will be extremely small. Therefore, as Figure 10(c) shows, the received IR reflected from the cover board in window 1 (N_1) will be much greater than that in window 2 (N_2), i.e., $N_1 \gg N_2$. In this case, as indicated in Section 5.2, the depth measurements will always be smaller than the real value in reflection-based *UltraDepth* implementation. Suppose the reflections on the scene and on the cover board follow the Lambertian reflection as described by Eqn.(2),

we can rewrite the Eqn.(1) as:

$$\widetilde{d} = \left(1 - \frac{d(d - d_0)}{\beta d_0^2/\beta_0 + d^2}\right) \times d, \tag{3}$$

where β and β_0 are the *albedos* of an object in the scene and of the cover board respectively, d is the real distance of the object and d_0 is the distance of the point on the cover board that reflects the emitted IR light. Eqn.(3) describes how a single point on the cover board affects the depth measurement of the ToF camera, which provides the guidance for setting cover board in the reflection-based *UltraDepth*. For example, in Eqn.(3), increasing β_0 and reducing d_0 will both enlarge the distortion between measured distance $\widetilde{d}$ and the real value d. Therefore, we can choose the cover board with higher reflectivity or put the cover board closer to the emitter to make it easier for texture exposure in reflection-based *UltraDepth*.

Figure 10(b) shows a prototype of the reflection-based *UltraDepth* implementation, where a 3D printed cover board is placed on top of the ToF camera module. In practice, the reflection-based implementation can be easily realized since there are no special requirements on the cover board's material and placement, as long as a light path is established between the ToF camera and the cover board. Moreover, such reflection-based implementation is difficult to detect without physical inspection of the camera device. In particular, one may accidentally cover part of the ToF camera, resulting in an equivalent reflection as the cover board. For example, if the ToF camera is placed close to a wall, the wall will serve as a cover board and reflect part of the emitted IR light to distort the ToF camera's distance measurements. In this case, the user will have a high possibility of accidentally exposing his/her privacy through the ToF depth camera. In Section 7.1, we present a feasibility study that includes a number of such scenarios.

6.2 External IR-based Implementation

Different from reflection-based *UltraDepth* that utilizes the IR from the camera as the distorting IR source, *UltraDepth* can also be implemented by using another device to construct a proper distorting IR source. Similar to the reflection-based implementation, the external IR source in this method also needs to have the same wavelength and modulation frequency as the ToF camera's IR light. Instead of developing a customized VCSEL IR source that is very labor-intensive, we choose to use a ToF camera with the same model as the original ToF camera, which naturally meets the requirements. In this section, we describe the *UltraDepth* implementation using two VZense Dcam 710 ToF cameras [58], where one serves as the original ToF camera and the other serves as the distorting IR source, although the same design can be easily adapted for other iToF models. Next, we will first introduce the IR emission pattern of the VZense ToF camera in Section 6.2.1, and then present the sniffing and spoofing procedure of the external IR-based *UltraDepth* implementation in Section 6.2.2 and Section 6.2.3.

6.2.1 The IR emission pattern of the iToF cameras. The effective distortion occurs only when both the distorting ToF camera and the original ToF camera emit IR light simultaneously. However, the ToF camera does not emit IR light all the time for the purpose of energy saving and eye safety. Therefore, the main task of external IR-based implementation is to align the IR emission times of the two ToF cameras. To this end, we need to study the IR emission pattern of the ToF camera. Firstly, the VZense ToF camera emits IR light with a wavelength of 940 nm and a modulation frequency at 100 MHz. Moreover, as is shown in Figure 11, the emission of IR light has a period of around 33 ms (30 periods per second), of which the ToF camera will emit IR light during the first 8 ms, and switch to the idle state for the remaining 25 ms. Furthermore, for any two depth cameras, the lengths of emission periods will be slightly different due to the hardware bias. Take the two depth cameras in our implementation as examples, the emission period of camera 1 is shorter than that of camera 2 by around 20 μs, which will result in an accumulative time shift between the emission times of the two cameras even if they start to emit IR light at the same time.

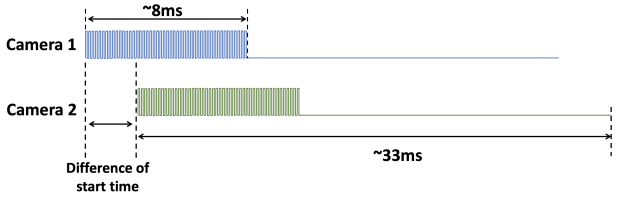


Figure 11: The emission patterns of two ToF cameras. The emission period of camera 1 is shorter than camera 2 by 20 μs. Our goal is to dynamically set the launch time of the distorting ToF camera to be aligned with the original camera.

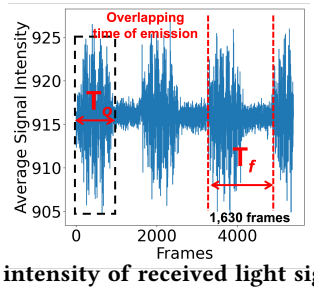
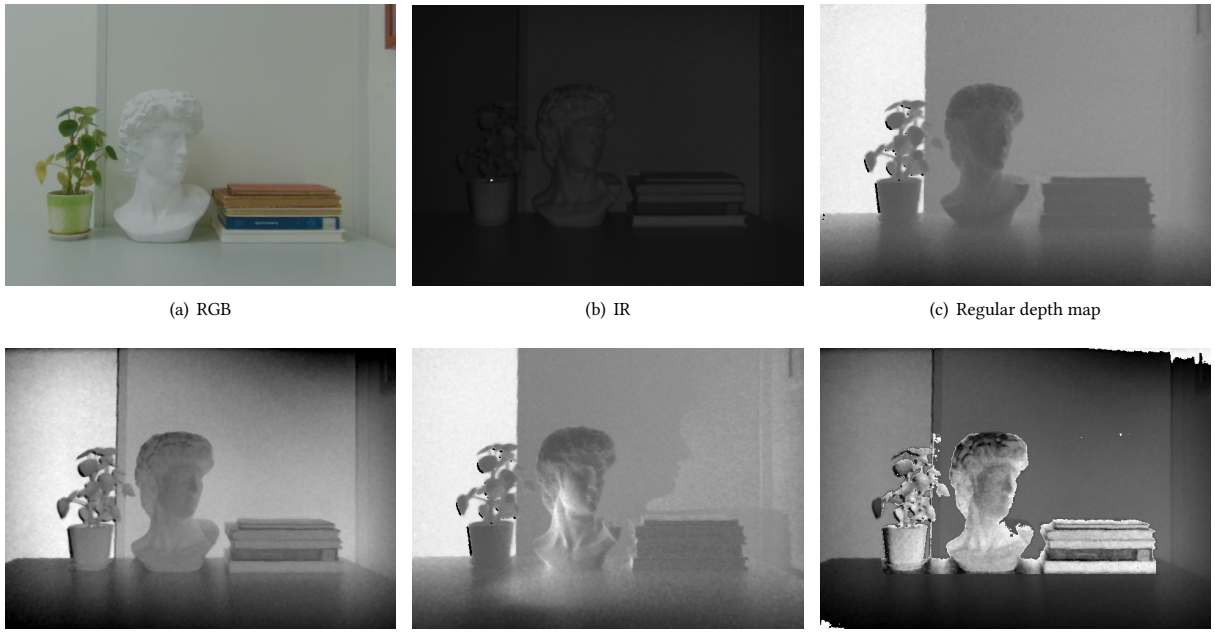


Figure 12: The intensity of received light signal on the distorting ToF camera. The signal intensity will drastically fluctuate when the IR emissions of the two cameras overlap (T_o) and the interference will appear periodically (with a period T_f). T_o and T_f can be used to align the two ToF cameras.

6.2.2 Sniffing. To align the emission time of the distorting ToF camera with that of the original one, we first need to know the emission pattern of the original ToF camera, and then control the IR emission of the distorting ToF camera to align with it. Therefore, the external IR-based *UltraDepth* includes a sniffing module to detect the emission pattern (period and launch time of each emission) of the original camera.

The design of sniffing is based on the observation that the IR light emitted by the original camera will be received by the distorting camera and hence influence the received light intensity of the distorting ToF camera. As Figure 12 shows, the average signal intensity received by the distorting camera will drastically fluctuate when the IR emissions of the two cameras overlap. Moreover, the interference will appear periodically due to the accumulated launch time shift and the periodical light emission of the two ToF cameras.

Based on the above observation, we propose to detect and predict the current time shift of starting a new emission between the



(a) RGB (b) IR (c) Regular depth map

(d) Reflection-based *UltraDepth* output (e) External IR-based *UltraDepth* output (f) *UltraDepth* output with region-based linear transformation

Figure 13: Comparison of images captured by a RGB camera, a IR camera and an ordinary depth camera and *UltraDepth* with different settings. The images output by our system *UltraDepth* (d,e,f) will reveal rich texture information. The depth map with region-based linear transformation (f) can even expose textures for distant objects.

two ToF cameras from the average signal intensity received by the distorting camera. Basically, the distorting ToF camera will continuously record the average intensity of the received IR light and analyze the fluctuation of the intensity. When the amplitude of the fluctuation is larger than a threshold, the distorting ToF camera will deem that the two camera's IR emissions are overlapping. In particular, the central time point t_0 of the strong fluctuation happens to be the time when the two cameras' emission times completely overlap. Moreover, from the period of fluctuation (T_f), we can calculate the difference of emission period for the distorting camera and the original camera as $t_d = 33/(30 \cdot T_f)$ ms.

6.2.3 Spoofing. Based on the overlapping time T_o and the period of fluctuation T_f, the distorting ToF camera will dynamically adjust its start time of IR emission to be aligned with that of the original one during the spoofing stage.

First, according to the overlapping time T_o, the distorting ToF camera is able to know the moment when the emissions of the two cameras are exactly overlapped (denoted as t_0, actually happens at the middle point of the overlapping time). Second, the difference of emission period between the attacker and the victim can be calculated as $t_d = 33/(30 \cdot T_f)$ ms, where T_f is the period of fluctuation detected during the sniffing stage. Assuming that the period of the distorting ToF camera is shorter than the original camera, then at $t_0 + 8/(30 \cdot t_d)$ s, the overlap will disappear since the distorting ToF camera has drifted 8 ms earlier than the original camera. Then the distorting ToF camera should wait 8 ms to start a new IR emission to align with the emission of the original camera again. Therefore, once the distorting ToF camera knows the aligned moment t_0 with the original camera, it only needs to wait 8 ms every $8/(30 \cdot t_d)$ s in the future to induce a continuous interference to the emission of the original camera. This scheme also works for the case that the period of the distorting ToF camera is longer than the original one.

6.3 Converting to Gray-scale Images

To make full use of the texture encoded in the depth map output by *UltraDepth*, we propose a region-based linear transformation to convert a depth map to a gray-scale image.

Generally, a simple linear transformation from the distorted depth maps output by *UltraDepth* to the gray-scale images can reveal lots of detailed textures. However, in some extreme cases where the range of measurement is large or the distance of objects of interest is concentrated within a small range, the direct linear transformation for the distance on the entire depth map will lose significant amounts of information. Specifically, for a depth map with a range of distance measurement in [150, 3,000] mm, a simple linear transform to the gray-scale images (with the range 0-255) can only differentiate two points of the depth map with a minimum granularity of 11 mm. To address this issue, we propose a region-based linear transformation method to ensure that detailed texture information is recovered as much as possible. Specifically, we first segment the scene into different regions based on the distance information in the depth map. A key observation is that the distance measurements are usually continuous but have drastic transitions at the boundaries of different regions or objects in the depth map, which enables fast and accurate region segmentation by simply detecting the drastic transitions [34, 55]. After that, regions of interest (ROI) are selected and linear transformation are performed on each separate region independently, while the remaining areas

are transformed to the gray-scale. Finally, we obtain a gray-scale image with augmented details in regions of interest.

Figure 13 shows images of a David statue from RGB camera, IR camera, normal depth camera and the distorted depth camera in *UltraDepth*. We observe that the ordinary RGB images show the richest details. While the normal depth maps can preserve privacy well because they only measure the objects' distance and cannot capture the texture information. However, *UltraDepth* will reveal high-resolution texture from distance measurement (shown in Figure 13(d), 13(e), 13(f)). Moreover, the depth map output by full-fledged *UltraDepth* (shown in Figure 13(f)) even reveals more texture information than the IR image as our region-based linear transformation technique can preserve the detailed information of distant objects in the depth maps.

7 EXPERIMENTAL EVALUATION

We evaluate *UltraDepth* on two models of iToF depth cameras, namely VZense Dcam 710 [58] which adopts the ToF technique from Analog Devices (ADI) and DepthEye Wide [45] which is developed based on IMX556PLR CMOS from Sony. In the following experiments, we use a reflection-based *UltraDepth* that has a 58 mm by 70 mm cover board, unless otherwise indicated.

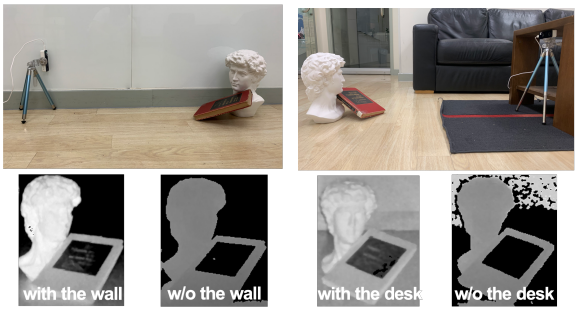
We first show the feasibility of the two *UltraDepth* implementations in daily life scenarios in Section 7.1. We then evaluate the effectiveness of exposing high-resolution texture of *UltraDepth* in three perception tasks, including object detection, character recognition and face recognition.[2] Furthermore, we evaluate the impact of several key factors on the performance of *UltraDepth* in Section 7.5. All the experiments are conducted in indoor environments since most ToF cameras are designed for indoor scenarios, and do not work well outdoors due to the excessive interference from sunlight.

7.1 A Feasibility Study

We first evaluate the feasibility of implementing the reflection-based and external IR-based *UltraDepth* in daily life scenarios. Our evaluation is focused on two aspects. First, we show how easy the reflection-based *UltraDepth* can be realized. As discussed in 6.1, to obtain a reflection path near the emitter, one may either place a cover board on the ToF camera or block part of the emitted light using an object. In particular, we focus on the later scenario because it can easily cause accidental revelation of sensitive information and hence has a major implication on privacy breach. Second, we illustrate how easy an additional IR source can be introduced in a typical set-up. The external IR-based *UltraDepth* can expose textures at a longer distance without modifying the interfered module, thereby is an important supplement to the reflection-based implementation.

We first present two cases of reflection-based *UltraDepth* implementation in Figure 14, where a ToF depth camera is placed near a wall or under a desk to establish a reflection path (equivalent to the effect of a cover board introduced in Section 6.1). As shown in Figure 14(a) and Figure 14(b), compared with the depth maps captured without the interference (w/o the wall or desk), detailed texture information of the scene is exposed from the depth maps in the presence of the wall or the desk. The results in two common

[2] All the data collection involving human subjects was approved by IRB of the authors' institution.



(a) Depth camera near a wall (b) Depth camera under a desk

Figure 14: Two cases of reflection-based *UltraDepth* implementation in daily scenarios, where rich texture of the objects is easy to be exposed with the help of ambient objects such as a wall and a desk.

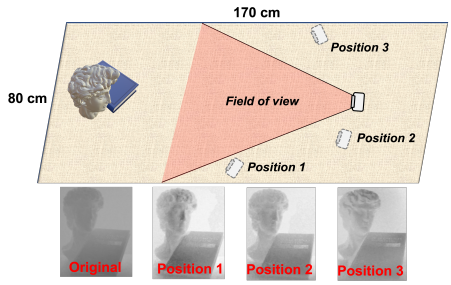


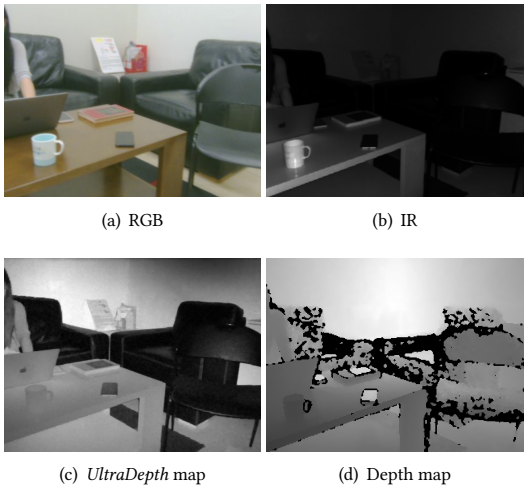
Figure 15: Experiments of external IR-based *UltraDepth* implementation with different positions and orientations of the external IR source.

daily scenarios show that the reflection-based *UltraDepth* is easy to be realized in real-world applications with the help of ambient objects such as a wall and furniture.

To study the feasibility of external IR-based *UltraDepth* implementation, as shown in Figure 15, experiments are conducted when the external IR source (a ToF depth camera of the same model as the original one) is mounted with different orientations and at different positions. The corresponding depth maps are also shown in Figure 15. Firstly, compared with the original depth map that shows little texture of the objects, the depth maps captured when the external IR source is placed at the three positions all show certain levels of texture information. The result confirms that, as long as the distorting IR light illuminates the target objects, it can be reflected and distort the distance measurement of the ToF depth camera. Moreover, another observation is that the effectiveness of the external IR source decreases with the increase of the distance and the decrease of overlap of field-of-view between the interfered camera and the external IR source due to the lower illumination on the objects. Specifically, external IR source at *Position 1* has a better performance than that at *Position 2* due to its short distance, while *Position 2* has a better performance than *Position 3* due to a larger area overlap of field-of-view between the two cameras.

7.2 Object Detection

We first evaluate the performance of *UltraDepth* in object detection tasks. In this experiment, we set up a scene with 10 selected



(a) RGB (b) IR

(c) *UltraDepth* map (d) Depth map

Figure 16: The four types of captured frames in the object detection task.

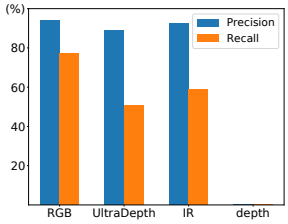
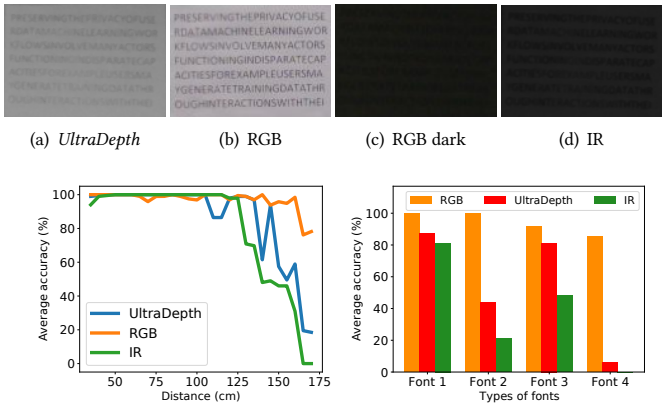


Figure 17: The performance of object detection using four types of captured frames. The *UltraDepth* maps have a significant performance improvement in object detection compared with the ordinary depth maps.

daily objects (chair, book, person, laptop, cell phone, sofa, cup, keyboard, mouse, umbrella). A RGB camera, an IR camera, an ordinary depth camera and a depth camera running *UltraDepth* are packed together to record the same scene from the same position simultaneously. The resolutions of RGB images, IR images, depth maps, and *UltraDepth* maps are all set to be 640×480. Figure 16 shows samples of the captured images/maps in the object detection task, where lots of noises and black spots are shown in the ordinary depth map. We totally collect 1,632 frames for RGB images, IR images, regular depth maps, and *UltraDepth* maps respectively with one set of objects under similar environmental conditions.

For a fair comparison, we first convert the four types of images/maps to 8-bit gray-scale images, and then we apply a widely used object detector YOLOv3 [48] on these images. An object is deemed to be detected successfully if the intersection over union (IoU) of the predicted bounding box and the ground truth is over 0.5, and the confidence of the correct label for the detected object is greater than 0.3. Figure 17 summarizes the precision and recall of object detection for the four types of images. It shows that the precision and recall for ordinary depth maps are both lower than 1%, which indicates the limitation of using ordinary depth maps for object detection. The *UltraDepth* maps, however, have a significant improvement in object detection compared with the ordinary depth maps, which is very close to the performance of RGB and IR



(a) *UltraDepth* (b) RGB (c) RGB dark (d) IR

(e) Mean accuracy at different distances (f) Mean accuracy for different fonts

Figure 18: (a)-(d): The *UltraDepth* map, RGB image, RGB image in the dark, and IR image of a paper with printed characters, which are captured at a distance of 0.8 m. (e): Mean accuracy of character recognition at different distances. (f): Mean accuracy of character recognition for different fonts.

images. Therefore, the experimental results show that the exposed texture in the *UltraDepth* maps can be leveraged to augment the performance of depth camera in applications like object detection, when only a depth camera (instead of both depth/RGB) is present.

7.3 Character Recognition

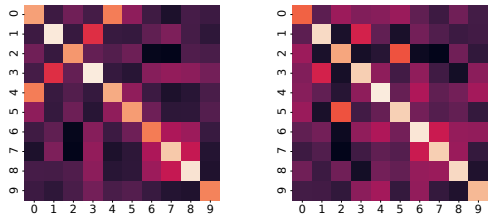
We now evaluate the effectiveness of *UltraDepth* in exposing text information. In the experiment, we concurrently capture *UltraDepth* output maps, RGB images, and IR images of a paper with printed characters or numbers. The resolution of both *UltraDepth* maps and IR images is 640×480, while the RGB images have a resolution of 640×360. The examples of the captured images are shown in Figure 18(a) to 18(d), from which we observe that the *UltraDepth* maps can well expose recognizable text information on the paper.

Specifically, a widely used optical character recognition (OCR) engine named Tesseract [54] is adopted to recognize characters in the collected images. We first evaluate the mean accuracy of character recognition when the cameras are mounted at different distances from the paper. As Figure 18(e) shows, the mean accuracy of character recognition decreases with the increase of distance for all images. Although the *UltraDepth* maps and IR images have the same resolution, the *UltraDepth* maps perform better in long-distance scenarios as the IR images captured at a long-distance are too dark due to the poor light intensity. However, *UltraDepth* measures the distance using the time-of-flight of IR light, hence is less vulnerable to the poor intensity of the received light. As expected, the RGB images achieve the best performance. However, in dark scenarios, the RGB images will fail to expose text information on the paper (Figure 18(c)) while the *UltraDepth* can still work well. We notice that the accuracy of *UltraDepth* drops significantly around 1.5 m, which is mainly due to the limited pixels for the characters. Similarly, RGB and IR images also suffer noticeable accuracy drops here. We further evaluate the performance of recognizing characters of different fonts (Arial, Chancery, Arial Bold, Bradley Hand) using the three kinds of images/maps. The results in Figure 18(f)

also show similar trends, where the mean recognition accuracy of *UltraDepth* (e.g., 81.25% for font 3) is always higher than the IR images (e.g., 47.92% for font 3) but not as good as RGB images (e.g., 91.67% for font 3).

7.4 Face Recognition

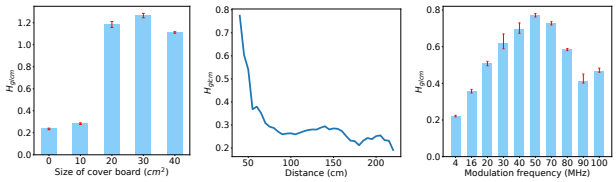
To evaluate the effectiveness of *UltraDepth* for face recognition, we recruit 10 volunteers and ask them to sit in front of two depth cameras (one runs *UltraDepth*, the other as a reference) and a RGB camera at a distance around 1 m. The three cameras are mounted to record the face images simultaneously at the same frame rate. The volunteers are instructed to randomly rotate their head and make different facial expressions during 2 minutes. We totally collect 16,033 frames for RGB images, regular depth maps, and *UltraDepth* maps of 10 volunteers, respectively. We convert all these images into gray-scale images and run a pre-trained *RetinaFace* detector [13] on them. It turns out no face can be detected for regular depth maps, which is expected, while the face detection rate for RGB images and for *UltraDepth* output maps are 100% and 99.93%, respectively, which means only 11 frames out of 16,033 *UltraDepth* maps fail to expose facial information, although this face detector is originally designed for RGB images.



(a) *UltraDepth* maps (b) RGB images

Figure 19: Heatmaps showing the inter-class and intra-class cosine similarity of the feature vectors of 10 volunteers' faces for *UltraDepth* maps and RGB images, where a lighter color denotes a higher similarity. The exposed texture in the *UltraDepth* maps is rich enough for face recognition.

We further evaluate the classification performance of the detected faces, which directly violates the personal anonymity that the depth cameras are commonly believed to preserve. We resize the detected faces areas to a unified resolution (120×120) and then the resized images are feed to a pre-trained ArcFace [12] model based on ResNet34 to generate 512-dimension feature vectors. Figure 19 shows the inter-class and intra-class cosine similarity of the 10 volunteers' face feature vectors from *UltraDepth* maps and RGB images, where the (i, j) element indicates the inter-class cosine similarities between the i-th and j-th volunteers. It's shown that the feature vectors from the same volunteer have strong cluster structures and those from different volunteers are less related, which means that the volunteers' faces can be classified easily using the *UltraDepth* maps and RGB images. We then build a simple neural network which contains only one hidden layer and takes the 512-dimensional vectors as input to classify these faces. Only 10% of the samples are used for training and the remaining samples are used for testing. Under this setting, *UltraDepth* achieves an accuracy of classification at 99.334%, which is close to the accuracy of RGB images (99.757%). Therefore, the exposed texture in the output of *UltraDepth* is rich enough to achieve good performance in face recognition even under dynamic settings.



(a) Size of cover board (b) Distance (c) Modulation frequency

Figure 20: The texture information metric H_{glcm} of captured depth maps under different settings of several factors, including the size of the cover board, the distance between the objects and the ToF camera, and the modulation frequency of the iToF camera.

7.5 The Impact of Key Factors

Now we study the impact of several key factors in *UltraDepth* design, including the size of the cover board, the distance between the objects and the ToF camera, and the modulation frequency of the iToF camera. Among them, the size of the cover board is only related to the reflection-based *UltraDepth* implementation, and the other two factors will affect both two implementations.

To quantitatively evaluate the level of exposed texture information, we propose a metric H_{glcm} to measure the amount of texture information revealed in an image as follows:

$$H_{glcm} = -\sum_{i=0}^{h-1}\sum_{j=0}^{w-1} p(i,j)\log_b p(i,j).$$

Here $p(i, j)$ is the (i, j) element of the gray level co-occurrence matrix (GLCM) [53], which is a commonly used metric to characterize the texture of an image by calculating how often pairs of pixel that have specific values with a specified spatial relationship occur in an image; w and h are the width and height of the image respectively. Therefore, the metric H_{glcm} quantifies the shannon entropy of the the co-occurrence matrix (GLCM), which is shown to represent the amount of texture in a image [20, 47] and therefore can be used to evaluate the effectiveness of texture exposure in *UltraDepth*.

In this section, all the experiments are conducted by taking the images of a David statue. For a fair comparison, we make the background of the David statue free of texture.

- **Size of cover board.** To study the impact of the size of cover board in reflection-based *UltraDepth* implementation, we mount four different cover boards on the ToF depth camera and collect depth maps of the same object at a fixed distance. Figure 20(a) shows the mean value of texture information metric H_{glcm}, where size 0 means no cover board is installed and size 4 is the largest cover board. When the size of cover board is the smallest (size 1), the reflection path is not established, hence has no impact on the depth map. When the size of cover board increases, the H_{glcm} increases drastically, as more reflected IR lights from the cover board are distorting the distance measurement and

cause more detailed texture exposed in the depth map. Finally, the largest cover board does not lead to the biggest H_{glcm}, i.e. the most significant texture exposure, since the cover board is so large that part of the depth map is blocked.

- **Distance from the camera.** Figure 20(b) shows H_{glcm} tends to decrease with the increase of distance from the camera, which is consistent with the intuition. Note that the drastic accuracy drop is mainly due to the decreasing number of pixels for the David statue as it moves away from the camera. However, when the distance is large enough, the advantage of *UltraDepth* for distant objects overwhelms the impact of reduced object pixel values, which results in a short rise of H_{glcm} ([100 cm, 150cm]).
- **Modulation frequency.** To study the impact of different modulation frequencies of iToF cameras, we conduct a set of experiments using the DepthEye ToF camera where all other settings are fixed except for the modulation frequency. As shown in Figure 20(c), H_{glcm} increases first and then decreases when the modulation frequency increases from 4 MHz to 100 MHz. The key reason for the first growth of H_{glcm} is that the modulation frequency is inversely proportional to the ToF camera's range of measurement. When the frequency is extremely small, the range of measurement is so large that the distance measurements distorted by *UltraDepth* fail to differentiate the detailed texture. However, when the modulation frequency is above 70MHz, more noises will be introduced in the captured depth map since the distance will approach the measurement limit, which causes the performance degradation of *UltraDepth*.

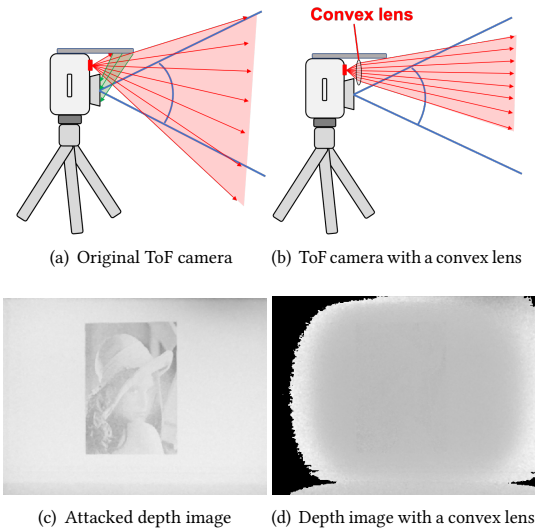
8 DEFENSE

As one important application of *UltraDepth* is to attack users' privacy, we now discuss possible solutions to defend against the attacks under two different implementations of *UltraDepth*.

Defense Against Reflection-based Attacks. The basic intuition behind the defense against the reflection-based implementation is to reduce the IR light reflected by the cover board. Based on this idea, we propose to narrow the field of view (FoV) of the IR emitter as small as the FoV of the camera lens, and position the emitter be as close as the camera lens. As a result, it will be hard to reflect the IR using a cover board without blocking the camera's field of view. Since blockage of part of the camera's FoV can be easily spotted, this method improves users' awareness of potential attacks and can help reduce the risk of privacy leakage.

To illustrate the feasibility of this defense, we place a dedicated convex lens in front of the ToF's emitter to narrow the FoV of emitted IR. As Figure 21(b) shows, compared with normal ToF camera (Figure 21(a)), the ToF camera with a convex lens can effectively prevent the same cover board from reflecting the emitted IR. The corresponding depth maps captured under the two settings are shown in Figure 21(c) and Figure 21(d), from which we can observe that the ToF camera with a convex lens can effectively defend against attacks using reflection-based *UltraDepth*.

Defense Against External IR-based Attacks. As discussed in Section 6.2, in order to distort the distance measurement of iToF cameras, the external IR source needs to employ the same modulation frequency. Therefore, our idea of defending against external IR-based *UltraDepth* is to use time-varying modulation



(a) Original ToF camera (b) ToF camera with a convex lens

(c) Attacked depth image (d) Depth image with a convex lens

Figure 21: Defend against attacks in reflection-based *UltraDepth* implementation.

frequencies in the target iToF camera, which is supported on most off-the-shelf iToF cameras [15, 29]. For instance, an iToF camera may randomly or periodically change the modulation frequency, making it difficult for the adversary to sniff the emission pattern of IR light and derive the adopted modulation frequency.

Nevertheless, we note that the defense methods for two implementations of *UltraDepth* have limitations. Specifically, the defense method against reflection-based attacks will limit the sensing scope of interfered depth camera, and the defense method against external IR-based attacks will affect the accuracy of depth sensing. We will leave addressing these limitations for our future work.

9 CONCLUSION

In this paper, we propose *UltraDepth*, the first system that can expose high-resolution texture from the depth maps captured by the off-the-shelf iToF depth cameras, simply by using a distorting IR source. To design *UltraDepth*, we first present an in-depth analysis on the impact of the distorting IR light on the distance measurement and how the texture of objects is encoded and exposed in the distorted depth maps. We then propose two practical implementations of *UltraDepth*, i.e., reflection-based and external IR-based *UltraDepth* implementations, which differ in how to construct a distorting IR source. We validate the effectiveness and feasibility of texture exposure of *UltraDepth* in extensive real-world experiments. The results show that, the depth maps output by *UltraDepth* achieve 89.06%, 99.33%, 81.25% mean accuracy in object detection, face recognition and character recognition, respectively. The findings of this work provide insights for new research on depth-related computer vision and security/privacy of depth sensing devices.

ACKNOWLEDGEMENT

This work is supported in part by Research Grants Council (RGC) of Hong Kong under General Research Fund #14203420, and National Natural Science Foundation of China under Grant No. 62032021.

REFERENCES

[1] 2021. 3D IMAGING WITH ADI TIME OF FLIGHT TECHNOLOGY. https://www.analog.com/en/applications/technology/3d-time-of-flight.html.

[2] 2021. 3D Sensing TOF (Time of Flight) Product Solution. https://www.gigabyte.com/Solutions/3D-Depth-Sensing/3d-sensing-product-solution.

[3] 2021. Depth Sensors: Precision & Personal Privacy. https://www.terabee.com/depth-sensors-precision-personal-privacy/.

[4] 2021. Helios2: The next generation of time-of-flight. https://thinklucid.com/helios-time-of-flight-tof-camera/.

[5] 2021. Vzense DCAM500 ToF Camera User Manual. https://991ef858-2cfe-44ad-9f3b-5cd69ed0861f.filesusr.com/ugd/9c9dda_d442dc06c23e45c9944689b29932f7f6.pdf.

[6] 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.

[7] Zhen-Peng Bian, Junhui Hou, Lap-Pui Chau, and Nadia Magnenat-Thalmann. 2014. Fall detection based on body part tracking using a depth camera. *IEEE journal of biomedical and health informatics* 19, 2 (2014), 430–439.

[8] Vishwanath Bijalwan, Vijay Bhaskar Semwal, and TK Mandal. 2021. Fusion of Multi-sensor based Biomechanical Gait Analysis using Vision and Wearable Sensor. *IEEE Sensors Journal* (2021).

[9] 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.

[10] Neha Dawar, Sarah Ostadabbas, and Nasser Kehtarnavaz. 2018. Data augmentation in deep learning-based fusion of depth and inertial sensing for action recognition. *IEEE Sensors Letters* 3, 1 (2018), 1–4.

[11] DayDayNews. 2020. The civil war for ToF technology is far from over. https://daydaynews.cc/en/technology/683608.html.

[12] 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)*.

[13] Jiankang Deng, Jia Guo, Yuxiang Zhou, Jinke Yu, Irene Kotsia, and Stefanos Zafeiriou. 2019. Retinaface: Single-stage dense face localisation in the wild. *arXiv preprint arXiv:1905.00641* (2019).

[14] Giovanni Diraco, Alessandro Leone, and Pietro Siciliano. 2013. Human posture recognition with a time-of-flight 3D sensor for in-home applications. *Expert Systems with Applications* 40, 2 (2013), 744–751.

[15] David Droeschel, Dirk Holz, and Sven Behnke. 2010. Multi-frequency phase unwrapping for time-of-flight cameras. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 1463–1469.

[16] Felix Endres, Jürgen Hess, Jürgen Sturm, Daniel Cremers, and Wolfram Burgard. 2013. 3-D mapping with an RGB-D camera. *IEEE transactions on robotics* 30, 1 (2013), 177–187.

[17] Deng-Ping Fan, Zheng Lin, Zhao Zhang, Menglong Zhu, and Ming-Ming Cheng. 2020. Rethinking RGB-D salient object detection: Models, data sets, and large-scale benchmarks. *IEEE Transactions on neural networks and learning systems* (2020).

[18] Mario Frank, Matthias Plaue, Holger Rapp, Ullrich Köthe, Bernd Jähne, and Fred A Hamprecht. 2009. Theoretical and experimental error analysis of continuous-wave time-of-flight range cameras. *Optical Engineering* 48, 1 (2009), 013602.

[19] Peter Fürsattel, Simon Placht, Michael Balda, Christian Schaller, Hannes Hofmann, Andreas Maier, and Christian Riess. 2015. A comparative error analysis of current time-of-flight sensors. *IEEE Transactions on Computational Imaging* 2, 1 (2015), 27–41.

[20] Peichao Gao, Zhilin Li, and Hong Zhang. 2018. Thermodynamics-based evaluation of various improved Shannon entropies for configurational information of gray-level images. *Entropy* 20, 1 (2018), 19.

[21] Timo Grimm, Manuel Martinez, Andreas Benz, and Rainer Stiefelhagen. 2016. Sleep position classification from a depth camera using bed aligned maps. In *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 319–324.

[22] Felipe Gutierrez-Barragan, Huaijin Chen, Mohit Gupta, Andreas Velten, and Jinwei Gu. 2021. iToF2dToF: A Robust and Flexible Representation for Data-Driven Time-of-Flight Imaging. *arXiv preprint arXiv:2103.07087* (2021).

[23] Miles Hansard, Seungkyu Lee, Ouk Choi, and Radu Patrice Horaud. 2012. *Time-of-flight cameras: principles, methods and applications*. Springer Science & Business Media.

[24] Radu Horaud, Miles Hansard, Georgios Evangelidis, and Clément Ménier. 2016. An overview of depth cameras and range scanners based on time-of-flight technologies. *Machine vision and applications* 27, 7 (2016), 1005–1020.

[25] Ji Hou, Angela Dai, and Matthias Nießner. 2019. 3d-sis: 3d semantic instance segmentation of rgb-d scans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4421–4430.

[26] Pengpeng Hu, Edmond Shu-lim Ho, and Adrian Munteanu. 2021. 3DBodyNet: Fast Reconstruction of 3D Animatable Human Body Shape from a Single Commodity Depth Camera. *IEEE Transactions on Multimedia* (2021).

[27] Sonam Jamtsho and Derek D Lichti. 2010. Modelling scattering distortion in 3D range camera. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38, 5 (2010), 299–304.

[28] Li Jia and Richard J. Radke. 2014. Using Time-of-Flight Measurements for Privacy-Preserving Tracking in a Smart Room. *IEEE Transactions on Industrial Informatics* 10, 1 (2014), 689–696. https://doi.org/10.1109/TII.2013.2251892

[29] Adrian PP Jongenelen, Donald G Bailey, Andrew D Payne, Adrian A Dorrington, and Dale A Carnegie. 2011. Analysis of errors in tof range imaging with dual-frequency modulation. *IEEE transactions on instrumentation and measurement* 60, 5 (2011), 1861–1868.

[30] Wilfried Karel, Sajid Ghuffar, and Norbert Pfeifer. 2012. Modelling and compensating internal light scattering in time of flight range cameras. *The photogrammetric record* 27, 138 (2012), 155–174.

[31] Tom Kavli, Trine Kirkhus, Jens T Thielemann, and Borys Jagielski. 2008. Modelling and compensating measurement errors caused by scattering in time-of-flight cameras. In *Two-and Three-Dimensional Methods for Inspection and Metrology VI*, Vol. 7066. International Society for Optics and Photonics, 706604.

[32] Michal Kepski and Bogdan Kwolek. 2014. Fall detection using ceiling-mounted 3d depth camera. In *2014 International conference on computer vision theory and applications (VISAPP)*, Vol. 2. IEEE, 640–647.

[33] Xiangbo Kong, Zelin Meng, Lin Meng, and Hiroyuki Tomiyama. 2018. A privacy protected fall detection IoT system for elderly persons using depth camera. In *2018 International Conference on Advanced Mechatronic Systems (ICAMechS)*. IEEE, 31–35.

[34] Chao Li, Zheheng Zhao, and Xiaohu Guo. 2018. Articulatedfusion: Real-time reconstruction of motion, geometry and segmentation using a single depth camera. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 317–332.

[35] Larry Li. 2014. Time-of-flight camera–an introduction. *Technical white paper* SLOA190B (2014).

[36] Carlos A Luna, Cristina Losada-Gutierrez, David Fuentes-Jimenez, Alvaro Fernandez-Rincon, Manuel Mazo, and Javier Macias-Guarasa. 2017. Robust people detection using depth information from an overhead time-of-flight camera. *Expert Systems with Applications* 71 (2017), 240–256.

[37] MARKETSANDMARKETS. 2020. Global Time-of-flight (ToF) Sensor Market 2021-2025. https://https://www.marketsandmarkets.com/Market-Reports/time-of-flight-sensor-market-264466295.html.

[38] Manuel Martinez and Rainer Stiefelhagen. 2017. Breathing rate monitoring during sleep from a depth camera under real-life conditions. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 1168–1176.

[39] Antonio Medina, Francisco Gayá, and Francisco Del Pozo. 2006. Compact laser radar and three-dimensional camera. *JOSA A* 23, 4 (2006), 800–805.

[40] Niluthpol Chowdhury Mithun, Sirajum Munir, Karen Guo, and Charles Shelton. 2018. ODDS: real-time object detection using depth sensors on embedded GPUs. In *2018 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 230–241.

[41] James Mure-Dubois and Heinz Hügli. 2007. Real-time scattering compensation for time-of-flight camera. In *International Conference on Computer Vision Systems: Proceedings (2007)*.

[42] Chuong V Nguyen, Shahram Izadi, and David Lovell. 2012. Modeling kinect sensor noise for improved 3d reconstruction and tracking. In *2012 second international conference on 3D imaging, modeling, processing, visualization & transmission*. IEEE, 524–530.

[43] Frank L Pedrotti, Leno M Pedrotti, and Leno S Pedrotti. 2017. *Introduction to optics*. Cambridge University Press.

[44] PointcloudAI. 2021. DepthEye Pro-VGA Depth camera. http://pointcloud.ai/products.

[45] PointcloudAI. 2021. DepthEye Pro-VGA Depth camera. http://pointcloud.ai/products.

[46] Holger Rapp, Mario Frank, Fred A Hamprecht, and B Jahne. 2008. A theoretical and experimental investigation of the systematic errors and statistical uncertainties of time-of-flight-cameras. *International Journal of Intelligent Systems Technologies and Applications* 5, 3-4 (2008), 402–413.

[47] QR Razlighi and N Kehtarnavaz. 2009. A comparison study of image spatial entropy. In *Visual Communications and Image Processing 2009*, Vol. 7257. International Society for Optics and Photonics, 72571X.

[48] Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).

[49] Xiaofeng Ren, Liefeng Bo, and Dieter Fox. 2012. Rgb-(d) scene labeling: Features and algorithms. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2759–2766.

[50] Sriram Sami, Yimin Dai, Sean Rui Xiang Tan, Nirupam Roy, and Jun Han. 2020. Spying with your robot vacuum cleaner: eavesdropping via lidar sensors. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*. 354–367.

[51] Hamed Sarbolandi, Markus Plack, and Andreas Kolb. 2018. Pulse based time-of-flight range sensing. *Sensors* 18, 6 (2018), 1679.

[52] Max Schwarz, Anton Milan, Arul Selvam Periyasamy, and Sven Behnke. 2018. RGB-D object detection and semantic segmentation for autonomous manipulation in clutter. *The International Journal of Robotics Research* 37, 4-5 (2018), 437–451.

[53] Bino Sebastian V, A Unnikrishnan, and Kannan Balakrishnan. 2012. Gray level co-occurrence matrices: generalisation and some new features. *arXiv preprint arXiv:1205.4831* (2012).

[54] Ray Smith. 2007. An overview of the Tesseract OCR engine. In *Ninth international conference on document analysis and recognition (ICDAR 2007)*, Vol. 2. IEEE, 629–633.

[55] Shuran Song, Fisher Yu, Andy Zeng, Angel X Chang, Manolis Savva, and Thomas Funkhouser. 2017. Semantic scene completion from a single depth image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1746–1754.

[56] Xinhang Song, Shuqiang Jiang, Luis Herranz, and Chengpeng Chen. 2018. Learning effective RGB-D representations for scene recognition. *IEEE Transactions on Image Processing* 28, 2 (2018), 980–993.

[57] Ting-En Tseng, An-Sheng Liu, Po-Hao Hsiao, Cheng-Ming Huang, and Li-Chen Fu. 2014. Real-time people detection and tracking for indoor surveillance using multiple top-view depth cameras. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 4077–4082.

[58] Vzense Technology. 2021. VZense Product Dcam710. https://www.vzense.com/products.

[59] Cha Zhang and Zhengyou Zhang. 2014. Calibration between depth and color sensors for commodity depth cameras. In *Computer vision and machine learning with RGB-D sensors*. Springer, 47–64.

[60] Zhong Zhang, Weihua Liu, Vangelis Metsis, and Vassilis Athitsos. 2012. A viewpoint-independent statistical method for fall detection. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE, 3626–3630.

[61] Cheng Zhao, Li Sun, Pulak Purkait, Tom Duckett, and Rustam Stolkin. 2018. Dense rgb-d semantic mapping with pixel-voxel neural network. *Sensors* 18, 9 (2018), 3099.

[62] Feng Zhao, Zhiguo Cao, Yang Xiao, Jing Mao, and Junsong Yuan. 2018. Real-time detection of fall from bed using a single depth camera. *IEEE Transactions on Automation Science and Engineering* 16, 3 (2018), 1018–1032.

[63] Peijun Zhao, Chris Xiaoxuan Lu, Jianan Wang, Changhao Chen, Wei Wang, Niki Trigoni, and Andrew Markham. 2019. mid: Tracking and identifying people with millimeter wave radar. In *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 33–40.