Material Classification from Time-of-Flight Distortions

Kenichiro Tanaka, *Member, IEEE*, Yasuhiro Mukaigawa, *Member, IEEE*, Takuya Funatomi, *Member, IEEE*, Hiroyuki Kubo, Yasuyuki Matsushita, *Senior Member, IEEE*, and Yasushi Yagi, *Member, IEEE*

Abstract—This paper presents a material classification method using an off-the-shelf Time-of-Flight (ToF) camera. The proposed method is built upon a key observation that the depth measurement by a ToF camera is distorted for objects with certain materials, especially with translucent materials. We show that this distortion is due to the variation of time domain impulse responses across materials and also due to the measurement mechanism of the ToF cameras. Specifically, we reveal that the amount of distortion varies according to the modulation frequency of the ToF camera, the object material, and the distance between the camera and object. Our method uses the depth distortion of ToF measurements as a feature for classification and achieves material classification of a scene. Effectiveness of the proposed method is demonstrated by numerical evaluations and real-world experiments, showing its capability of material classification, even for visually indistinguishable objects.

Index Terms—Alternative sensor, subsurface scattering, time-of-flight camera, temporal point spread functions

1 Introduction

Material classification plays an important role for computer vision applications, such as semantic segmentation and object recognition. One of the major challenges in material classification is that different materials may yield very similar visual appearances. For example, artificial plastic fruits and real fruits confronting a camera produce visually similar RGB images that are difficult to distinguish. One of the plausible strategies to distinguish similar appearances is to use the optical responses of the target object such as spatial, angular, and temporal spreads of the incident light. Because different materials may have different optical responses due to their own subsurface scattering and diffuse reflection properties, it is expected that a more reliable material classification can be achieved using such optical cues together with the RGB observations.

Recently, Heide et al. [1] have developed a method that recovers transient images from observations by a low-cost Time-of-Flight (ToF) camera, which is originally designed for depth measurement. There are other related studies that use ToF cameras for recovering ultra-fast light propagation, e.g., impulse response, of the scene [2], [3], [4] with some hardware modifications and computation. Motivated by these previous approaches that exploit the temporal spread of light, we aim to classify materials using an indirect temporal cue that is obtained from an off-the-shelf ToF camera without explicitly recovering the impulse response.

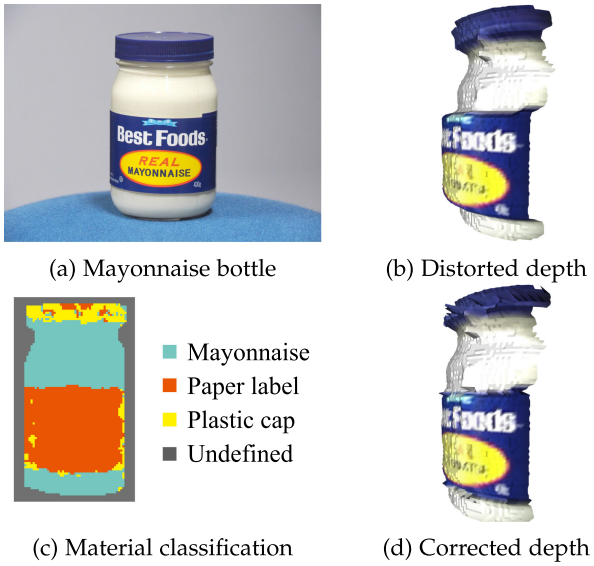
We develop a material classification method based on a key observation that the measured depth of a translucent object becomes greater than the actual depth as shown in Fig. 1b, where the depth gap between the mayonnaise glass bottle and paper label regions is obvious. We show that this depth distortion is caused by the time delay due to subsurface scattering and varies along with both the modulation frequency of ToF camera and the distance between the camera and target object. Using the depth distortions as a *feature* that conveys the material information, we propose an exemplar-based material classification method.

The chief contributions of this paper are twofold. First, we demonstrate that the material classification is tractable by an off-the-shelf ToF camera, e.g.., Xbox One Kinect. Our method uses the distorted depth measurements as an indirect temporal cue for material classification without explicitly recovering impulse responses; therefore it does not require any modifications of hardware unlike [1], [2]. Second, we show how the ToF measurements are distorted inside materials and along with depths. By moving the target object along the depth direction, rich information about the target can be obtained, which serves as important clue for achieving material classification.

This paper extends its preliminary version [5] with the following differences. Extensions have been made to (1) the analysis of the proposed feature characteristics, (2) comparison with another method, (3) robustness evaluation against the object's color, and (4) analysis of the amplitude components as well as depth components. In this manner, we

- *K. Tanaka, Y. Mukaigawa, T. Funatomi, and H. Kubo are with Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma, Nara Prefecture 630-0192, Japan. E-mail: {ktanaka, mukaigawa, funatomi, hkubo}@is.naist.jp.*
- *Y. Matsushita is with Graduate School of Information Science and Technology, Osaka University, Suita, Osaka Prefecture 565-0871, Japan. E-mail: yasumat@ist.osaka-u.ac.jp.*
- *Y. Yagi is with Institute of Scientific and Industrial Research, Osaka University, Suita, Osaka Prefecture 567-0047, Japan. E-mail: yagi@am.sanken.osaka-u.ac.jp.*

Manuscript received 2 Apr. 2018; revised 22 Aug. 2018; accepted 9 Sept. 2018. Date of publication 13 Sept. 2018; date of current version 31 Oct. 2019. (Corresponding author: Kenichiro Tanaka.) Recommended for acceptance by J. Yu. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TPAMI.2018.2869885



(a) Mayonnaise bottle (b) Distorted depth

(c) Material classification (d) Corrected depth

Fig. 1. Depth distortion of a ToF camera. (a) A mayonnaise bottle is measured by a Microsoft Kinect. (b) Measured depth in a 3D view. There is a depth gap at the glass bottle and label boundary. We use this depth *distortion* for material classification. (c) Pixel-wise material classification. (d) Application of material classification to depth correction. Depths are corrected based on the segmentation result and the distortion database. Depth gaps among materials are corrected and a faithful 3D shape is recovered.

assess the effectiveness of the proposed method from various aspects.

The rest of this paper is organized as follows. Section 2 reviews related prior works. Section 3 describes the Time-of-Flight measurement model for translucent objects and the depth distortions along with the frequency and the depth. Based on the model, we describe the proposed method for classifying the material of translucent objects in Section 4. Section 5 shows results and the accuracy of the method using real-world objects. Finally, we conclude the paper in Section 6.

2 Related Work

Non-invasive and non-contact material classification is an important research topic in computer vision and yet remains a challenging task. There are several prior works for material classification. The methods based on the visual appearance, e.g., color, shape, and/or textures of the material [6], [7], [8], [9], [10], typically only require a single RGB image; thus, the setups are easy to realize. The major problem is that this approach suffers from similar appearances of different materials, e.g., texture-less boards, resulting in a lower accuracy due to the lack of information.

A class of approaches based on the optical properties, such as BRDF [11], [12], shading [13], and spectrum [14], has a capability of distinguishing visually similar objects with higher accuracy because the optical properties convey richer information about the material. However, constructing such measurement system and building database of samples generally require extra effort in carefully controlled settings. This class includes approaches based on other physical properties, e.g., elasticity [15], and water permeation and heating/cooling process [16]. Our method falls into this class because we use a temporal response of the incident light, which implicitly measures the optical and physical properties of target objects. However, unlike these methods, our method uses an off-the-shelf ToF camera and needs only single observation at minimum, hence the cost of constructing the system is as low as the appearance-based methods.

In the context of material classification using a ToF camera, Su et al.'s method [17] is closely related. They propose a method that classifies a material from raw ToF measurements by sweeping over several modulation frequencies and phases. While the approach is shown effective, it requires special hardware customization of a ToF camera for obtaining the measurements. In contrast, our method only uses an off-the-shelf ToF camera without hardware modification. We show that the material classification can be achieved by such a simple setup by exploiting the depth-dependency of the measurements. In addition, while Su et al.'s method requires calibration for building a correlation matrix and post-processing of the data after measurement, our method does not require either of them.

For a comprehensive overview of temporal light transport, we refer the reader to the recent survey by Jarabo et al. [18]. A time domain impulse response of a scene, as known as light-in-flight and transient imaging, can be obtained using an interferometer [19], holography [20], [21], femtosecond-pulsed laser [22], [23], [24], and single photon avalanche diode sensor [25]. The time domain impulse response can be also recovered using the ToF camera, with which the cost can be significantly reduced while the temporal resolution becomes lower. Because the ToF camera is able to measure sub-nanosecond phenomena, it can be used for visualizing the light propagation of the scene by frequency and phase sweep [1], [4], [26], [27] and optical coding [2], [3], while it requires customization of a ToF camera for these purposes. These measurement methods may be applied to the task of material classification [28], although they require careful and expensive setups. On the other hand, our method bypasses the exact recovery of the time domain impulse responses and simply uses the measured depth of a ToF camera.

When a ToF camera measures a multi-path scene, the measured depth is distorted due to inter-reflections and subsurface scattering, known as the multi-path interference. Mitigating the multi-path interference and recovering the correct depth is of broad interest, and it has been studied by assuming two-bounce and simplified reflection model [29], [30], [31], [32], parametric model [33], [34], K-sparsity [35], [36], [37], consistency between ToF and stereo [38], simplified indirect reflections [39], and large-scale multi-path [40]. Instead of recovering the correct depth, we use a distorted depth as a cue for the material classification. We show that, once the material classification has been achieved, the classification result can then be used for correcting depths.

There are other scene analysis methods using ToF cameras, e.g., recovering the shape of transparent and translucent objects [41], [42], and measuring a slice of BRDF [43]. In addition, computational imaging methods using a ToF camera, such as non-line-of-sight imaging [44], [45], [46], separating direct and indirect light transport [3], [28], [47], imaging the velocity of the object [48], [49], and imaging at a specific depth [50] are proposed. Our method can also be regarded as one of the scene analysis methods because it aims at material classification of the scene.

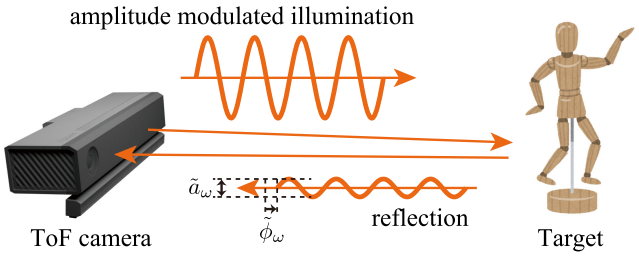


Fig. 2. ToF cameras measure the attenuated amplitude and phase delay of the amplitude modulated light, and then calculate the depth of the object from the phase delay and the speed of light.

3 Time-of-Flight Observation

To begin with, we briefly review the measurements that are obtained by a ToF camera. A correlation-based ToF camera illuminates a scene by an amplitude modulated wave $f_\omega(t)$ governed by the modulation frequency ω and measures its attenuated amplitude $\tilde{a}_\omega \in \mathbb{R}$ and phase delay $\tilde{\phi}_\omega \in \mathbb{R}$ as illustrated in Fig. 2. From the phase delay and the speed of light c, the measured depth of the scene $\tilde{d}_\omega \in \mathbb{R}$ can be obtained as

$$\tilde{d}_\omega = \frac{4\pi\omega\tilde{\phi}_\omega}{c}.$$

In general, a scene can have the *multi-path* effect due to inter-reflections and subsurface scattering, which degrades the depth estimation accuracy. The image formation process and models that involve the multi-path effect have been well understood thanks to the previous works [1], [2], [47]; hence, we briefly explain one of the models that we are going to use in this paper. Following the phasor representation [47], the amplitude and phase of the returned wave can be represented by a single complex value $\mathbf{c} \in \mathbb{C}$, called *phasor*, which is governed by the modulation frequency ω. The magnitude and argument of the phasor directly correspond to the measured amplitude and phase as

$$\begin{cases} \arg \mathbf{c}(\omega) & = -\tilde{\phi}_\omega, \\ |\mathbf{c}(\omega)| & = \tilde{a}_\omega, \end{cases}$$

where the function $\arg : \mathbb{C} \mapsto \mathbb{R}$ extracts the angle of the complex number. Note that the sign of the phase is flipped because the original definition of phasor is clock-wise rotation while the standard complex is counter-clock-wise.

When the illumination wave is a sine wave, i.e., $f_\omega(t) = \sin(2\pi\omega t)$, the observed phasor can be represented as

$$\mathbf{c}(\omega) = \int_0^\infty r(t-\tau)e^{-2\pi i\omega t}dt, \tag{1}$$

where $\tau(>0)$ is the time of flight toward the surface of the object, $r(t)$ is the impulse response, or a point spread function (PSF), of the object along with the time t, and i is the imaginary unit. The impulse response is the summation of all possible paths $\rho \in \mathcal{P}$; therefore, $r(t)$ can be written as

$$r(t) = \int_{\mathcal{P}} r_\rho \delta(|\rho| - t)d\rho,$$

where $r_\rho \in \mathbb{R}$ is the contribution of the path ρ, $|\rho| \in \mathbb{R}$ is the time traveled along the path ρ, $\delta(t)$ is the Dirac delta function, and $t = 0$ indicates the time when the impulse light hits the surface of the object. Fig. 3b illustrates a phasor representation of the multi-path ToF observation. The time domain PSF $r(t-\tau)$ is expanded onto the imaginary plane, and the phasor depicted by a red arrow is the integration of expanded PSF over the angle. Because the time domain PSF $r(t)$ is zero when $t < 0$, Eq. (1) is equivalent to the Fourier transform as

$$\begin{aligned} \mathbf{c}(\omega) &= \int_{-\infty}^\infty r(t-\tau)e^{-2\pi i\omega t}dt \\ &= \mathcal{F}[r(t-\tau)], \end{aligned}$$

where $\mathcal{F}[\cdot]$ represents the Fourier transform. This equation expresses that ToF camera measures the Fourier coefficients of the impulse response at the frequency of the light modulation.

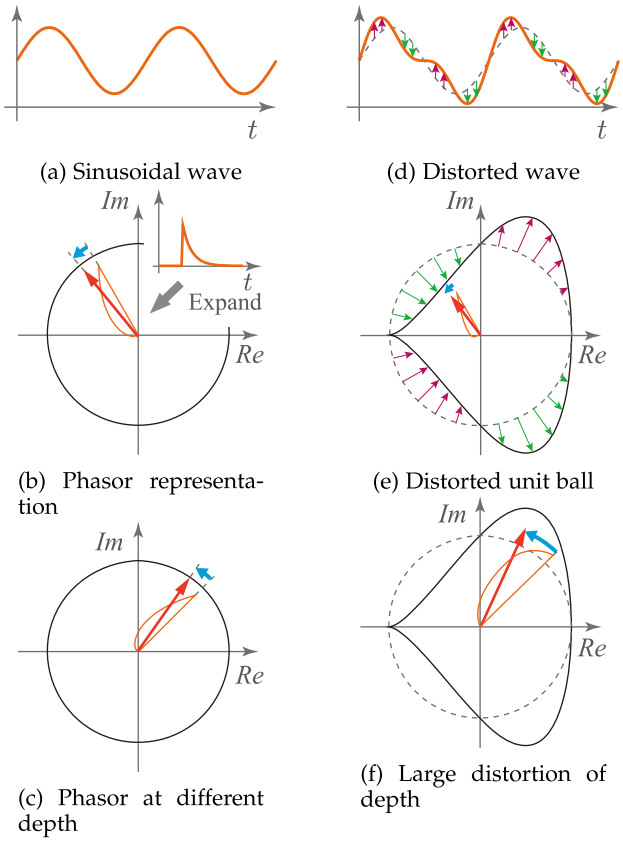


Fig. 3. Phasor representation of ToF observations. (a) Sinusoidal illumination, (b) Time domain PSF is expanded to the imaginary plane (orange). (c) When the object is placed at different depths, the observation gets rotated but phase distortion remains the same as (b). (d) Biased periodic illumination. This toy example adds 20 percent harmonics to the sinusoid for biasing. (e) The unit ball of the phasor representation is distorted due to the biased illumination. (f) The object is placed at the same depth as (c). The distortion of the phase becomes different than (e) and (c).

Frequency Dependent Depth Distortion

The principle of the ToF camera assumes that the impulse response consists of the Dirac delta function as $r(t) = \alpha\delta(t)$, where $\alpha \in \mathbb{R}$ is the amplitude decay of the modulated light. In this case, the measured depth $\tilde{d}_\omega \in \mathbb{R}$ becomes

$$\tilde{d}_\omega = \frac{c}{4\pi\omega}\underbrace{\arg\int_{-\infty}^\infty \alpha\delta(t-\tau)e^{-2\pi i\omega t}dt}_{=2\pi\omega\tau} = \frac{c\tau}{2} = d, \tag{2}$$

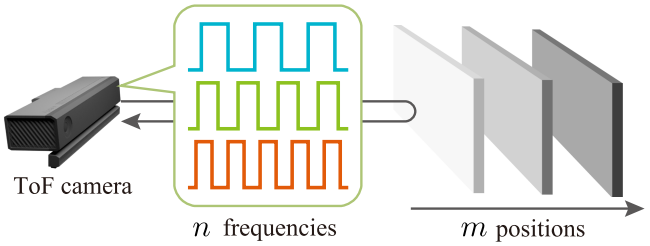


Fig. 4. Measurement setup. The target object is measured by changing the modulation frequency and its distance to the camera.

where $d = c\tau/2$ is the ground truth depth of the object. Eq. (2) represents that the accurate depth can be obtained regardless of modulation frequency ω, if the impulse response of the scene is exactly the Dirac delta function.

In reality, almost all materials except for the perfect mirror surface yield various shapes of impulse responses due to diffuse and subsurface scattering [28] that are different from the Dirac delta function. When the target object exhibits a temporally broad shape of the impulse response, the band-pass characteristic in the frequency domain becomes unique to the object. Accordingly, ToF observation $\mathbf{c}(\omega)$ can take an arbitrary value, because $\mathbf{c}(\omega)$ is a Fourier coefficient of the impulse response $r(t)$ at the frequency ω. In such a case, $\arg \mathbf{c}(\omega)$ does not necessarily represent the correct phase $2\pi\omega\tau$, as a result, the measured depth $\tilde{d}_\omega$ becomes distorted, and the distortion varies with the modulation frequency ω. This frequency-dependent depth distortion is one of our key observations, and our method exploits this property for the goal of material classification.

The shift in the time domain corresponds to the shift of phase in the Fourier domain

$$\begin{aligned}\mathcal{F}[r(t-\tau)] &= e^{-2\pi i\omega\tau}\mathcal{F}[r(t)] \\ &= e^{-2\pi i\omega\tau}\hat{r}(\omega),\end{aligned}$$

where $\hat{r}(\omega)$ is the Fourier transform of the function $r(t)$. The measured depth $\tilde{d}_\omega$ can then be described as

$$\begin{aligned}\tilde{d}_\omega &= \frac{c}{4\pi\omega}\arg\left(e^{-2\pi i\omega\tau}\hat{r}(\omega)\right) \\ &= d + \frac{c}{4\pi\omega}\arg\hat{r}(\omega).\end{aligned} \tag{3}$$

The second term $\frac{c}{4\pi\omega}\arg\hat{r}(\omega)$ is the depth distortion at frequency ω.

While a single observation of the depth distortion can be the same among different materials by chance, multiple observations using varying modulation frequencies can be used for enriching the measurement. Such multiple observations can be obtained with negligible latency because the ToF measurement is much faster than the ordinary video frame intervals, e.g., 12 ToF measurements are achieved within approximately $1/20$ seconds [4]. However in practice, it is not straightforward to measure distortions using many different frequencies by an off-the-shelf ToF camera. For example, Kinect has only three modulation frequencies, and the frequencies cannot be easily changed; hence, only three distortion measurements are practically available, which may be too few for reliable measurements of depth distortions for the purpose of material classification. To increase the information about the material in an alternative and easy way, our method employs a strategy of changing the distance between the camera and object. Now, we discuss the *depth-dependency* of the depth distortion.

Depth-Dependent Depth Distortion

When the target object is placed at a different depth $d + \Delta d$, the impulse response $r(t-\tau)$ is shifted by $\Delta\tau = 2\Delta d/c$ in the time domain. As a result, the measured depth $\tilde{d}'_\omega$ becomes

$$\begin{aligned}\tilde{d}'_\omega &= \frac{c}{4\pi\omega}\arg\left(e^{-2\pi i\omega(\tau+\Delta\tau)}\hat{r}(\omega)\right) \\ &= d + \Delta d + \frac{c}{4\pi\omega}\arg\hat{r}(\omega).\end{aligned}$$

The measured depth is just shifted by Δd, and the depth distortion $\frac{c}{4\pi\omega}\arg\hat{r}(\omega)$ remains the same as the one at the original position as in Eq. (3). Fig. 3c illustrates the depth distortion at different depth positions in a phasor representation. The blue arrow, which represents the depth distortion, is the same as that of the original position as illustrated in Fig. 3b.

So far, we have assumed that the illumination is a perfect sinusoidal wave. In practice, because a high-frequency sinusoidal wave is difficult to generate, today's ToF cameras emit non-sinusoidal periodic waves that contain high-order harmonics [51], [52]. When the illumination wave has harmonics components as illustrated in Fig. 3d, the ToF observation exhibits depth-dependency as drawn in Figs. 3e and 3f. Let us suppose that the distorted sinusoidal wave is biased as $f_\omega(t) = b_\omega(2\pi\omega t)\sin(2\pi\omega t)$, where $b_\omega(2\pi\omega t)$ is a periodic bias of the illumination wave due to harmonics. The observed phasor can then be written as

$$\mathbf{c}(\omega) = \int_0^\infty r(t-\tau)b_\omega(2\pi\omega t)e^{-2\pi i\omega t}dt.$$

The above indicates that the observation $\mathbf{c}(\omega)$ is the Fourier coefficient of $r(t-\tau)b_\omega(2\pi\omega t)$, where the impulse response $r(t)$ is distorted by the bias $b_\omega(2\pi\omega t)$. Obviously, the biased impulse response $r(t-\tau)b_\omega(2\pi\omega t)$ varies along with τ, i.e., the observation varies along with the depth.

Usually, this depth-dependent variation is unwanted; therefore, previous works attempted to eliminate it. For example, Su et al. [17] remove the depth-dependent variation using a correlation matrix. In contrast, we use the depth-dependent distortion as an important cue for material classification as it contains rich information about the target's response.

4 Material Classification

Our method uses either or both of the frequency- and depth-dependent depth distortions of ToF observations for the purpose of material classification. We assume that the depths are measured multiple times by changing the modulation frequency and/or the distance of the object as illustrated in Fig. 4. This section describes the feature and classifier that are used in our method.

4.1 Feature

For describing the use of depth distortions as features for material classification, we begin with the case where the actual depth is known and later describe a more general case where such an assumption is eliminated. When the

target object is placed at a known depth location, the depth distortion with respect to the actual depth is directly measurable. Let us suppose that the target object is measured by $n(\geq 1)$ modulation frequencies and $m(\geq 1)$ positions. The absolute depth distortion v_{ω_i,d_j} can be obtained by

$$v_{\omega_i,d_j} = d_j - \tilde{d}_{\omega_i,j},$$

where $\tilde{d}_{\omega_i,j}$ is the measured depth at the ith modulation frequency ω_i $(i \in \{1,\ldots,n\})$ and the jth position $(j \in \{1,\ldots,m\})$, and d_j is the actual depth at the jth position. By aligning these distortions, an mn-length vector $\mathbf{v} \in \mathbb{R}^{mn}$ can be formed as a *feature* vector of the object as

$$\mathbf{v} = \begin{bmatrix} v_{\omega_1,d_1} & \cdots & v_{\omega_n,d_m} \end{bmatrix}^T.$$

Because the actual depth of the target object is not generally accessible, we develop a feature that does not require the knowledge of the actual depth. Although we cannot directly obtain the depth distortion in this case, the *relative* depth distortions among multiple frequencies and/or multiple depths can be alternatively used. When multiple modulation frequencies are available, i.e., $n \geq 2$ case, the relative depth distortion v'_{ω_i,d_j} can be computed by regarding the measurement of one of the modulation frequencies, say the nth modulation frequency, as the reference measurement. The relative depth distortions can be obtained by taking differences from the reference measurement as

$$v'_{\omega_i,d_j} = v_{\omega_i,d_j} - v_{\omega_n,d_j} = \tilde{d}_{\omega_n,j} - \tilde{d}_{\omega_i,j}, \tag{4}$$

where i ranges from 1 to $n-1$. We can then setup an $m(n-1)$-length vector $\mathbf{v} \in \mathbb{R}^{m(n-1)}$ by aligning the relative depth distortions, and it can be used as a feature vector for material classification. Although the reference measurement $\tilde{d}_{\omega_n,j}$ may be distorted depending on the material, the feature vector $\mathbf{v}$ encapsulating the relative distortions conveys discriminative cues for classifying materials.

In a similar manner, for the case where a single modulation frequency and multiple depth locations is available, i.e., $n = 1$ and $m \geq 2$, the relative depth distortions among depth locations v''_{ω_1,d_j} can be obtained by regarding the measurement of the mth depth position as the reference measurement as

$$v''_{\omega_1,d_j} = v_{\omega_1,d_j} - v_{\omega_1,d_m} = \tilde{d}_{\omega_1,m} - \tilde{d}_{\omega_1,j} + \Delta d_j, \tag{5}$$

where Δd_j is the amount of movement from the base position, which should be measured.

In summary, for the feature vector generation, we need at least either a single ToF measurement plus the actual depth or multiple ToF observation changing the modulation frequency or the actual depth. A vector aligning the differences of such observations describes the feature of the material.

4.2 Classifier

We assume that we have a database of materials that consists of the feature vectors measured using predefined modulation frequencies and depth locations in a certain range beforehand. For classification, the target object is measured with the full or partial set of the predefined modulation frequencies and depth locations. Once we obtain the feature vector of the target object as a query, we use the material database to look up the closest material.

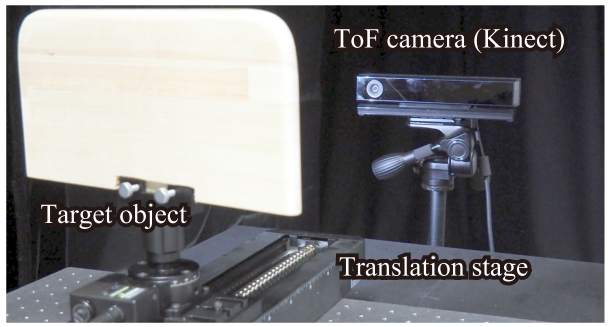


Fig. 5. Experimental setup. We use Kinect as a ToF camera, and the target object is placed on a linear translation stage.

While any arbitrary classifiers can also be alternatively used, it is desired for classifiers to have the following two properties. First, since the feature vectors tend to be high-dimensional while the number of materials and measurements in the database may be small, it is preferred that the classifier uses a model with a small number of parameters, or non-parametric like our choice. Second, a capability of handling missing elements in the feature vector is practically important, because the measurement is sometimes missing due to specular reflection on the object surface, or becomes saturated by near-distant reflectances.

For these reasons, we adopt a simple nearest neighbor classifier, which assesses the euclidean distance (ℓ_2-norm) for look-up. To deal with the missing or uninformative saturated observations, we remove such elements in the feature vector when evaluating the distance. The distance ξ_p between the feature vector $\mathbf{v}$ of the target object and the training vector $\mathbf{v}^p$ of the object p in the dataset can be computed as

$$\xi_p = \frac{1}{N}\sum_{k=1}^{nm} \begin{cases} 0 & v_k = \mathrm{N/A} \\ (v_k - v_k^p)^2 & \text{otherwise,} \end{cases} \tag{6}$$

where N is the number of valid elements, and v_k and v_k^p are kth element of vectors $\mathbf{v}$ and $\mathbf{v}^p$, respectively. Using this distance, we can classify the object by searching the nearest class $\hat{p}$ as

$$\hat{p} = \arg\min_p \xi_p.$$

Throughout the evaluation in this paper, we use this simple nearest neighbor classifier to assess the effectiveness of the depth distortion features for material classification.

5 Experiments

We evaluate the proposed method using a setup with a ToF camera and a linear translation stage shown in Fig. 5. We use Microsoft Kinect v2 for the ToF camera, which has three modulation frequencies ($n = 3$), and an OptoSigma's translation stage (SGSP46-800). Because the official Kinect API does not support an access to depth measurements of each frequency, we have slightly altered an open-source software `libfreenect2`[1] to obtain the frequency-wise depth measurements.[2]

1. https://github.com/OpenKinect/libfreenect2
2. The source code is publicly available on our website.

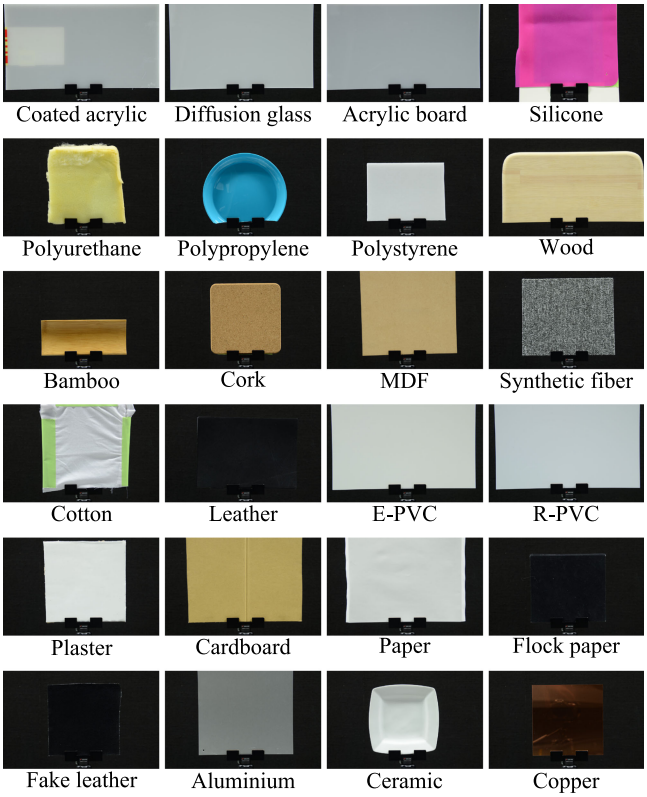


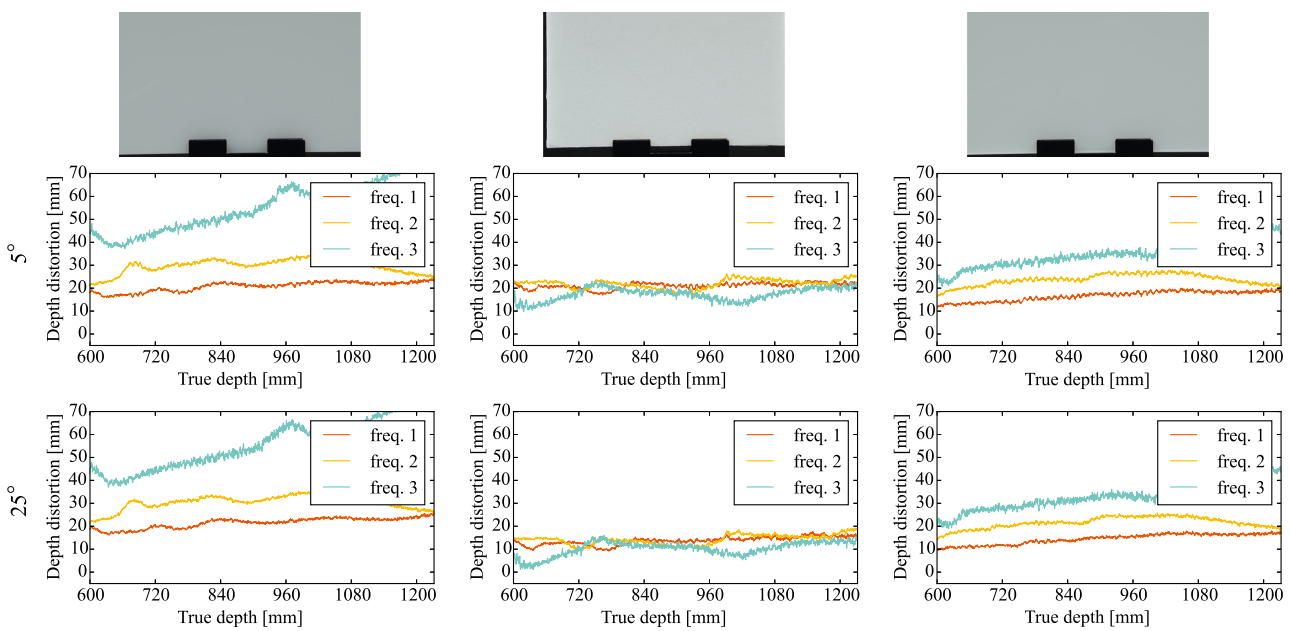
Fig. 6. All materials of our database. All images are captured by the same camera parameters, e.g., ISO, f-number, shutter speed, and focal length. The center part of each material is measured.

The translation stage and the ToF camera are roughly aligned so that approximately the same surface point is measured. Specifically, the camera pose is adjusted with changing the position of a target surface from near to far such that the center of the target is always projected to the center of the image plane. In our setting, only the center pixel is used for the evaluation. Because the target objects are homogeneous and planar, very accurate alignment is unneeded.

First, we measure the depth distortion data of 24 materials including metal, wood, plastic, fabric, and so on, as summarized in Fig. 6 and examine their differences across materials. The target object is placed on the linear translation stage and measured by changing the depth from 600 mm to 1280 mm ($m = 1700$) at varying orientations several times. The ground truth depth is obtained from the position of the translation stage, which is only used for the purpose of evaluating the depth distortions but not for the material classification. Fig. 7 shows the depth distortion of three materials; white acrylic board, polystyrene board, and opal diffusion glass. They are visually similar objects (white, planer, and no texture); hence, appearance-based methods have difficulty in distinguishing them. On the other hand, depth distortions of ToF observations show significant differences across materials, and the trend is consistent over measurement sessions indicating that the depth distortion feature is stable.

Feature Characteristics

Here we analyze the statistics of the measured depth distortions. The mean and standard deviation within each class are shown as a box plot in Fig. 8, in which the euclidean distance to the mean of its own class is assessed. The blue point represent the upper bound of classification, that corresponds to the shortest distance from the class mean to the mean of any of other classes. Most of the blue points show



(a) White acrylic board (b) White polystyrene board (c) White diffusion glass

Fig. 7. Measured depth distortions using Kinect for three objects. The ground truth depth is obtained via a linear translation stage. The top row shows photographs of the target objects. Measurements shown in the second and third rows are different in terms of the surface orientation (5 and 25 degrees). Depth distortion of each frequency varies along with the actual depth and material. Depth distortion is similar for the same material regardless of the surface orientation, but largely distinct in different materials. This frequency- and depth-dependent depth distortion is our key observation for material classification.

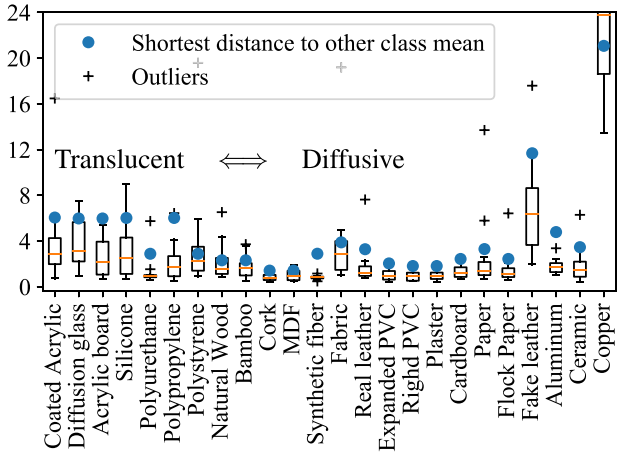


Fig. 8. Statistics of measured depth distortions. Mean and standard deviation within each class are visualized as a box plot. Plus marks and blue points are outliers and the shortest distance to other class mean, respectively.

greater distances than upper quartiles (i.e., located above the boxes), hence the depth distortion feature is discriminative. The standard deviation of translucent materials becomes larger than that of diffusive materials, which implies that the translucent materials are more sensitive to the environmental factors. However, because their upper bounds are also large, the translucent classes can also be well classified.

To further assess the depth distortion features, we show the distance matrix and feature space visualization in Fig. 9. In the distance matrix (Fig. 9 left), the average distance between every pair of materials in the feature space is visualized (darker the nearer). The materials are manually sorted based on their apparent optical properties from translucent to diffusive. It shows that materials with similar optical properties exhibit feature similarity, which are seen as dark block-diagonal components in the distance matrix. It indicates that the material's optical property information is well conveyed in the depth distortion feature. We also use the t-distributed stochastic neighbor embedding (t-SNE) [53] for the feature space visualization. In this visualization as well, it shows meaningful clusters, which supports the applicability of the depth distortion feature to material classification.

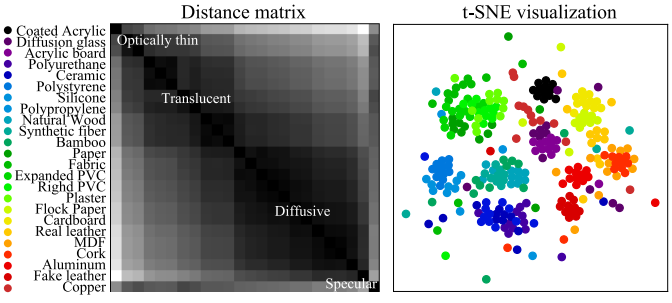


Fig. 9. Distance matrix and t-SNE visualization of measured depth distortion features. *Left:* The matrix represents the distance among classes, where the value increases from black to white. Materials with similar optical properties, e.g., optically thin, moderately translucent, and diffusive objects, show their similarity as seen in the dark block diagonal components. *Right:* t-SNE visualization. It shows well-defined groups, which indicates the discriminative power of the features.

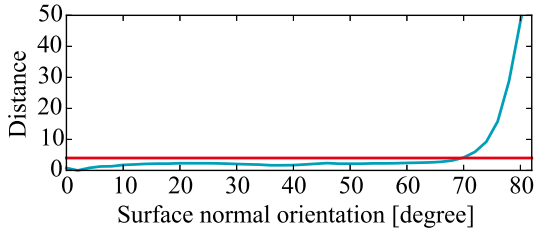


Fig. 10. Feature vector variation over surface orientation. We change the orientation of the target object, and plot the distance of features along with the orientation. The feature is stable under around 70 degrees, and shows large deviation at steep-slant orientation. Red line indicates the upper-bound distance for the correct classification.

Feature Variation w.r.t. Surface Orientation

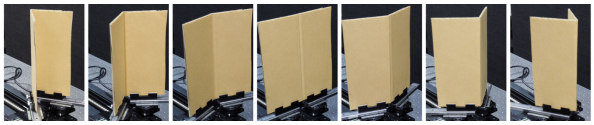
Although we measure a certain object orientation, the time-domain impulse response may vary when the surface orientation of the target object varies. To observe the effect of surface orientations on the depth distortion, we measure a planar material (material "wood" in this case) by changing its orientations. When the rotation axis of the object and the optical axis of the camera intersect, the rotational change does not vary the ground-truth depth of the object. Fig. 10 shows the variation of the shortest euclidean distance from the center of the wood class in the feature space along with the surface orientation of the target object. The red line indicates the upper-bound distance from the wood class, under which the query feature vector is correctly classified as wood. In other words, once the distance from the wood class to the query feature exceeds this upper-bound distance, it will be misclassified. The feature is stable under around 70 degrees, which indicates that the depth distortion feature is reliable for a confronting surface in practice but may break down for a steep-slanted surface, e.g., near the edges of a round shaped object.

Feature Variation w.r.t. Shape

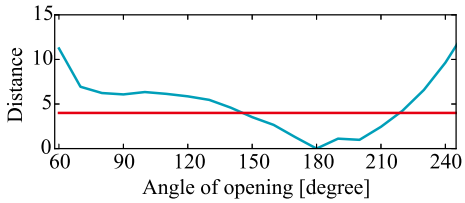
As observed above, the depth distortion feature may vary with the shape of the target object. In particular, for a concave shape where significant inter-reflections occur and for largely distorted shape where the strength of subsurface scattering varies, the effect becomes more significant. To observe the effect of shape on depth distortions, we set up a scene of a folded cardboard with varying opening angles. We measure the area around the cardboard's folding edge with changing the opening angle from the small angle (closed), via 180 degrees (flat), to large angle (protruded) as shown in Fig. 11a. When the opening angle changes, the ground-truth depth of the object is shifted accordingly. This offset is geometrically compensated in the measuring process. The euclidean distances of feature vectors between the folded and flat cardboards are plotted in the blue line in Fig. 11b. The red line represents the upper-bound of the flat cardboard class, under which the target is correctly classified as a flat cardboard. It shows the moderate robustness against the shape variations, but also indicates that the feature is affected by target's shape.

Quantitative Evaluation of Material Classification

Using this depth distortion measurements, we assess the accuracy of material classification by a nearest neighbor



(a) Scenes of the folded cardboard



(b) Variation of feature distance between flat and folded cardboards

Fig. 11. Shape dependency of the feature vector. We measure a cardboard with folding from 60 to 240 degrees. By folding cardboard less than 180 degrees, the scene exhibits strong inter-reflections. Red indicates the upper-bound distance for the correct classification, and blue indicates the distance of features along with the folding angles.

classifier. The dataset consists of 24 materials as shown in Fig. 6, each measured at 13 orientations to enable the classifier to deal with diverse orientations of the target objects. We evaluate the classification accuracy using three different features: (1) frequency-dependent distortion, (2) depth-dependent distortion, and (3) a combination of them. With the frequency-dependent distortion features ($n = 3$ and $m = 1$), the accuracy is limited only to 55.0 percent, as shown in the form of confusion matrix in Fig. 12a. This low accuracy is due to the limited availability of the number of frequency channels. In this test, measurements at 925 mm are extracted from the database and the depth measurements at frequency #3 are used as the reference. Using the depth-dependent distortion features ($n = 1$, $m = 1700$, and using Eq. (5)), the accuracy is improved to 80.3 percent as shown in Fig. 12b. For this test, the depth measurements at frequency #3 are used, and the depth at 925mm is selected as the reference. The amount of movement Δd_j is obtained from the position of the translation stage. Finally, with the combination of frequency- and depth-dependent distortion features ($n = 3$ and $m = 1700$), the accuracy is further improved to 89.9 percent. The confusion matrix is shown in Fig. 12c. While many materials are correctly classified, some

TABLE 1
Comparison with Su et al. [17]

Method	Observations	#Material	Accuracy (Ref.)
Su et al. [17]	Raw data	4	80.5%
Ours	**Depth distortion**	**24**	85.8%

The number of frequencies and depth variations are the same. Although the accuracy cannot be directly compared due to the difference in settings, they are at a similar level of accuracy.

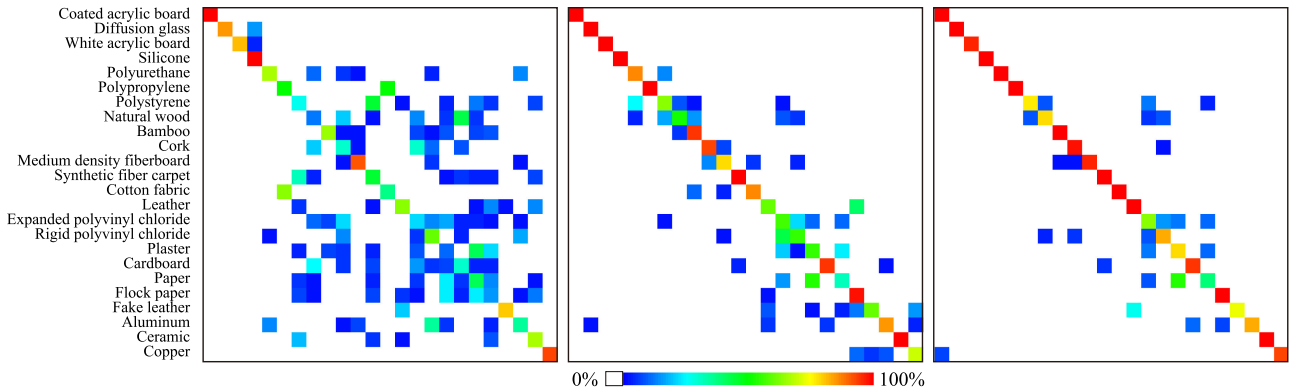
materials are difficult to classify. For example, plaster and paper, or expanded and rigid polyvinyl chlorides have similar impulse responses due to similar scattering properties, resulting in more frequent misclassification.

Comparison with Su et al. [17]

While Su et al.'s method requires hardware modification to obtain rich raw data, our method only uses the depth data that can be acquired by a commodity device with retaining the accuracy of material classification. When the experiment is conducted by reducing the number of depth variations m to 10, the accuracy of our method becomes 85.8 percent. Although it is difficult to directly compare the two methods, they are at a similar level of the accuracy; Su et al.'s paper [17] reports that the accuracy is 80.5 percent for a 4-material classification task using rich raw observations that consist of 8 frequencies and 8 phases of illumination (our case is 24-material classes using depths of 3 frequency channels). The result is summarized in Table 1.

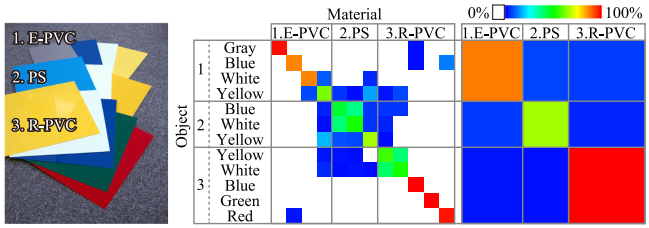
Simultaneous Classification of the Material and Color

We examine the proposed method's robustness against object's color variations. In this experiment, 12 colorful boards made of three different plastic materials shown in Fig. 13a are measured for material classification. The materials are well classified regardless of the surface color as shown in the right of Fig. 13b in the form of a confusion matrix with an accuracy of 91.7 percent. While the impulse responses may vary with the pigment, the method shows a stability against the little variations. Moreover, the classification of material-and-color can also be achieved by our



(a) Only frequency-dependent distortions (b) Only depth-dependent distortions (c) Both frequency- and depth-dependent distortions

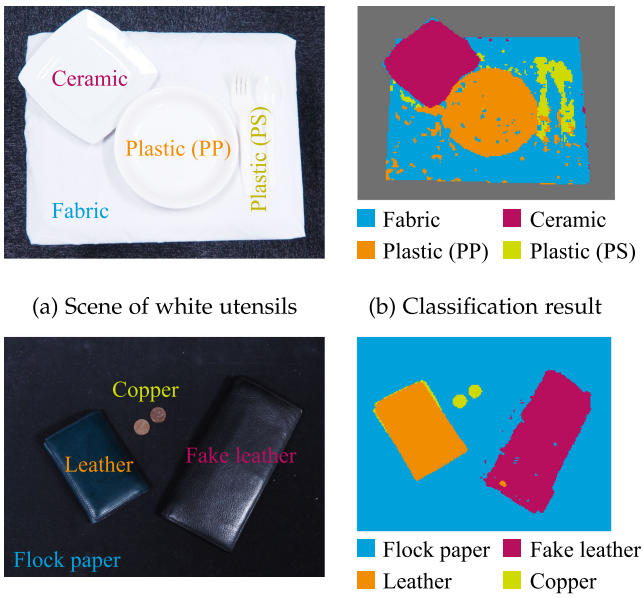
Fig. 12. Confusion matrices of material classification. Red indicates the higher value and it appears on the diagonal. (a) Only frequency-dependent distortion is used. The accuracy is 55.0 percent. (b) Only depth-dependent distortion is used. The accuracy is 80.3 percent. (c) Both of them are used. Overall accuracy is 89.9 percent.



(a) Objects (b) Confusion matrices

Fig. 13. Classification result for different color boards. The center confusion matrix shows the classification of material-and-color, which is defined as a joint classification of the base material and color pigment. The right confusion matrix shows the result of base material classification. The result shows that our method is capable of classifying objects without being confused by their colors, and also that it can further classify materials with different pigments.

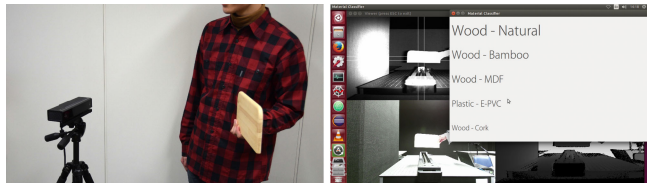
method, capturing the subtle differences in the measured impulse responses as shown in the left of Fig. 13b. The overall accuracy of the material-and-color classification is 75.3 percent, without using the RGB information for classification. While some objects made from polystyrene (PS) are confused by the color, they are still recognized as the same material. The reason for the misclassification may be that some color pigments more drastically vary the impulse response of the material. Other objects are clearly classified without being confused by their colors, e.g., blue boards made by E-PVC are not confused by the blue boards made by R-PVC. Overall, the results indicate that our method is capable of classifying objects without being confused by their colors, and also that it can be further applied classify materials with different pigments.



(a) Scene of white utensils (b) Classification result

(c) Scene of wallets (d) Classification result

Fig. 14. Material segmentation results. (a) All utensils are white hence it is difficult to classify only with an RGB image. (b) Result of our material classification. Although there are some estimation errors due to that only the frequency-dependent distortions are used for classification, it shows convincing material-based segmentation of the scene. (c) Wallets made of genuine and fake leather and copper coins are placed in the scene. (d) Material segmentation result.



(a) Measuring scene (b) Screen capture

Fig. 15. Real-time material classification. (a) A person waving a target object in front of a ToF camera. (b) A screen capture of the system. A ranking of the estimated material class is shown at interactive speed.

5.1 Applications

We now show a few applications of the proposed method; material segmentation, real-time classification system, enhanced 3D recovery that takes into account the depth distortions, and thickness classification.

Material Segmentation

Our method can be applied in a pixel-wise manner that enables material-based segmentation. Fig. 14 shows two examples of material segmentation. For the scene in Fig. 14a, all objects are whitish, and the material classes are not obvious to human eyes in the RGB image. With our method, the materials are classified for each pixel as shown in Fig. 14b. For this application, we use only frequency-dependent distortions but not depth-dependent distortions, i.e., $m = 1$, because the alignment of pixels becomes hard when the geometric relationship between the camera and scene changes. Due to that, the result appears to be a little bit noisy, but it still shows faithful classification and segmentation. For this experiment, we used a reduced database containing only four materials because the dimensionality of the feature vector is limited. Fig. 14c shows another scene, where wallets made of genuine and fake leather are placed, and they are correctly classified as shown in Fig. 14d.

Real-Time Material Classification System

We develop a near real-time material classification system,[3] which can recognize the target material class by a handheld ToF camera or by moving the target object in front of a fixed ToF camera. Fig. 15a shows the setting, where an object is waved in front of a ToF camera, and its material class is estimated at interactive speed and displayed as depicted in Fig. 15b. Because the ground-truth depth is not available, the relative depth distortion is used as described in Eq. (4); thus the number of relative frequencies is 2. Tracking is not employed in this system and a small area on the optical axis of the camera is used for classification. Using the partial matching strategy described in Eq. (6), our method yields estimates in near real-time even when observations at only a small number of depth locations m is available. By increasing the variation of depths m by moving the target object or camera, the classification accuracy is gradually improved because of the availability of richer information.

3. A video is included in a supplementary material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPAMI.2018.2869885 for demonstration.

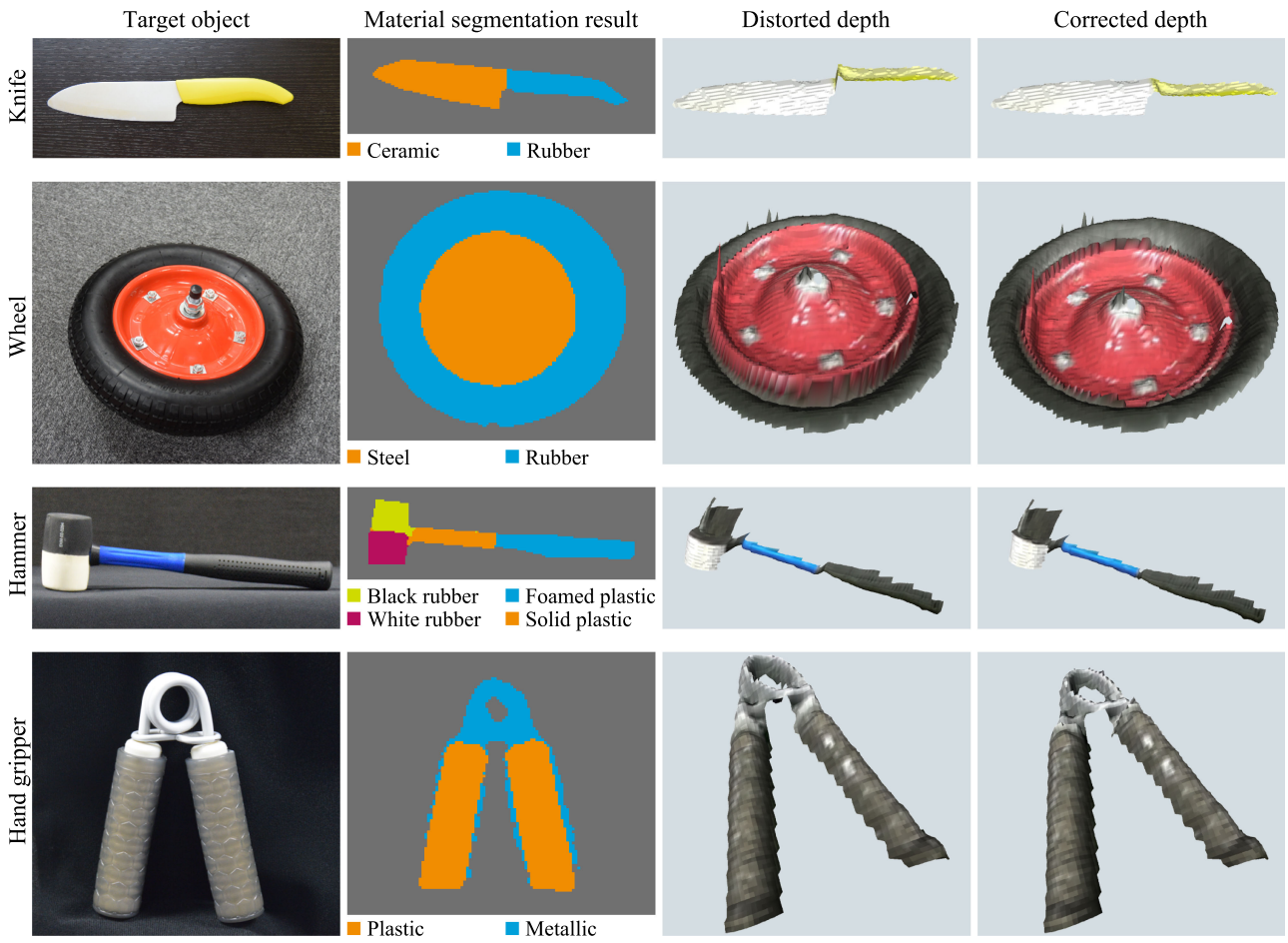


Fig. 16. Material segmentation and depth correction result for a knife, wheel, hammer, and hand gripper. Each target object consists of distinct material parts. After material segmentation, the distorted depths are corrected and the depth gaps across the different materials are reduced, yielding more convincing shape recovery for heterogeneous objects.

Depth Correction

Once materials are classified, the distorted depths can be corrected for recovering an accurate depth map using the material database that contains the samples of distortions for all materials. Some examples of depth correction are shown in Fig. 16. A knife in the figure consists of ceramic blade and rubber grip, and the measured depth of the rubber part is greatly distorted than the depth of the ceramic region due to subsurface scattering. Based on the material segmentation result, the depth can be corrected and the gap between the ceramic and rubber regions is reduced. For this experiment, we do not change the depth of the target; therefore, only the frequency-dependent distortions are used ($m = 1$). We reduced the database as same as the material classification application. Other depth correction results for wheel, hammer, and hand gripper scenes are also shown in Fig. 16. In a similar manner to the knife scene, using the material segmentation result, the measured depths are corrected as shown in the rightmost column. The depth gap among different materials are suppressed to exhibit convincing reconstruction.

Another example is shown in Fig. 1. Because mayonnaise has significant subsurface scattering, the measured depth of mayonnaise region is strongly distorted than that of the label region as shown in Fig. 1b. Fig. 1c depicts our result of material segmentation. Although some artifacts are observable because of the limited amount of measurement and slanted surface orientations, mayonnaise and the label regions are largely well separated. Using the segmentation result and depth distortion database, a faithful 3D shape of the mayonnaise bottle is recovered as shown in Fig. 1d. Compared to the original shape, the depth discontinuity

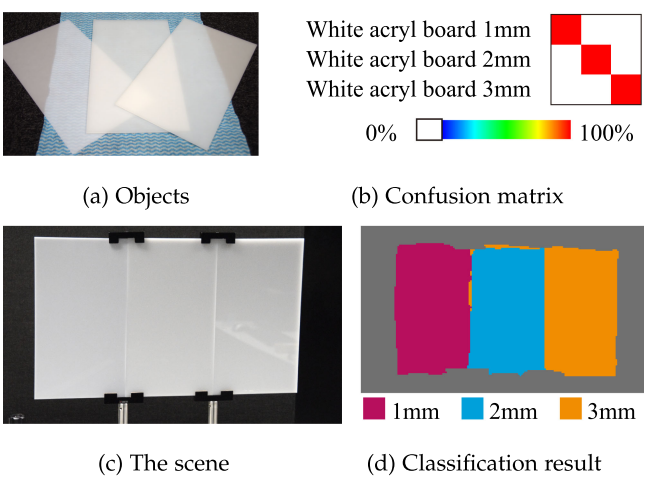


Fig. 17. Thickness classification result. Acrylic boards of different thicknesses are measured. Because they are optically thin, they have significantly different temporal response with respect to the thickness.

TABLE 2
Comparison Among Different Settings

Method	Accuracy
Chance rate	4.2%
Ours (depth distortions only)	**89.9%**
Ours (amplitude components only)	53.4%
Ours (both of them)	89.1%

Amplitude component has the capability but not have significant effectiveness than the depth distortions.

between mayonnaise and the label regions is noticeably reduced.

Thickness Classification

The proposed depth distortion feature is also useful for thickness estimation of optically thin materials. For example, white acrylic boards in Fig. 17a are optically thin, and their impulse response varies with their physical thickness. The thickness of the white acrylic board can therefore be classified by our method. In a similar manner to the quantitative evaluation (Figs. 12a, 12b, and 12c), the confusion matrix of the classification with $n = 3$ and $m = 1700$ observations is summarized in Fig. 17b, which shows accurate classification of target's thickness. The thickness classification can be conducted in a pixel-wise manner so that it enables a thickness-based image segmentation. Acrylic boards with three distinct thickness are placed side-by-side and measured as in Fig. 17c, and the thickness-based segmentation result is visualized in Fig. 17d.

6 Conclusion and Discussion

This paper describes a material classification method that uses an off-the-shelf ToF camera. It is shown that the depth measured by a ToF camera is distorted according to the time domain impulse responses, and the distortion varies along with the modulation frequency and the distance between the object and the camera. The ToF depth distortions are used as a cue for material classification instead of discarding as a measuring error. The effectiveness of the proposed method is evaluated by various real-world experiments and a few applications are shown.

Since the ToF observation is a pair of amplitude and phase components, one may consider using the amplitude component for the task of material classification, while the proposed method only uses the phase component. Indeed, the amplitude and phase are are jointly distorted in the phasor space. In the next paragraph, we discuss the potential of the use of the amplitude component for material classification.

Amplitude Component for Material Classification

Instead of relative distortion as explained in Eq. (4), relative amplitude components can be alternatively used in a similar manner

$$z_{\omega_i,d_j} = \tilde{a}_{\omega_n,j} - \tilde{a}_{\omega_i,j},$$

where $\tilde{a}_{\omega_i,j}$ is the measured amplitude at the ith modulation frequency and the jth position, and z is the relative amplitude component. Aligning these distortions, an $m(n-1)$-length vector $\mathbf{z}$ can be created and used as a feature vector.

Moreover, both the depth distortions and the relative amplitude components can be combined to obtain richer information. In that case, they have to be normalized because the domains of the depth and amplitude are different. A concatenated feature vector $\mathbf{s}$ can be set up as

$$\mathbf{s} = \begin{bmatrix} \bar{\mathbf{v}}^T & \bar{\mathbf{z}}^T \end{bmatrix}^T,$$

where $\bar{\mathbf{v}}$ and $\bar{\mathbf{z}}$ are normalized feature vectors of depth and amplitude components, respectively.

Table 2 shows the comparison of the effectiveness of depth distortion, amplitude distortion, and their combination as a feature. In this test, the object is measured while changing both the frequency and position ($n = 3$ and $m = 1700$), and the result of depth distortion is the same as before as shown in Fig. 12c. While amplitude component also is able to classify materials to a certain degree, it turned out that the accuracy was much lower than the result of depth distortion. Moreover, adding amplitude components to the depth distortion did not contribute to the improvement of accuracy. The reason may be that the amplitude is strongly affected by low signal-to-noise ratio (SNR). From this test, we consider that while the amplitude component may have a potential, but for a practical use a further study is needed.

Limitations

Our material classification exploits differences of time domain impulse responses that depend on materials. In other words, we assume that the impulse response is the same for the same materials. However, it may not be always true because difference in shape, color, and geometry may cause variations in impulse responses. As we assessed in the paper, there is a limitation in material classification accuracy due to these differences. Related to this problem, optically thin object's impulse response also varies depending on the thickness of the target object. On one hand, it serves as a cue for classifying thickness of the target object as we have seen, but on the other hand, it also suggests that a database with varying thicknesses is needed for correctly classifying materials of an object that may have an arbitrary thickness. This is one of the current limitations of our method. Using realistic simulation such as Jarabo et al.'s renderer [54], a very large database that includes all the materials and variations may be obtained, but it still appears non-straightforward.

Another limitation is that the shape of illumination wave may be device-dependent; therefore, the measured impulse responses may have a dependency on ToF cameras. At this point, a database could be built for each ToF camera product, but eliminating the device-dependency appears to be an important future work.

Acknowledgments

We thank all anonymous reviewers who gave us various insightful and constructive comments. This work is partly supported by JSPS KAKEN JP18H03265, JP18K19822, and JST CREST JPMJCR1764.

References

[1] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich, "Low-budget transient imaging using photonic mixer devices," *ACM Trans. Graph.*, vol. 32, no. 4, 2013, Art. no. 1.

[2] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, and R. Raskar, "Coded time of flight cameras: Sparse deconvolution to address multipath interference and recover time profiles," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 1–10, 2013.

[3] M. O'Toole, F. Heide, L. Xiao, M. B. Hullin, W. Heidrich, and K. N. Kutulakos, "Temporal frequency probing for 5D transient analysis of global light transport," *ACM Trans. Graph.*, vol. 33, no. 4, pp. 1–11, 2014.

[4] C. Peters, J. Klein, M. B. Hullin, and R. Klein, "Solving trigonometric moment problems for fast transient imaging," *ACM Trans. Graph.*, vol. 34, no. 6, pp. 1–11, 2015.

[5] K. Tanaka, Y. Mukaigawa, T. Funatomi, H. Kubo, Y. Matsushita, and Y. Yagi, "Material classification using frequency- and depth-dependent time-of-flight distortion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2740–2749.

[6] B. Caputo, E. Hayman, and P. Mallikarjuna, "Class-specific material categorisation," in *Proc. IEEE Conf. Comput. Vis.*, 2005, pp. 1597–1604.

[7] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2032–2047, Nov. 2009.

[8] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz, "Exploring features in a Bayesian framework for material recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 239–246.

[9] G. Schwartz and K. Nishino, "Visual material traits: Recognizing per-pixel material context," in *Proc. Color Photometry Comput. Vis. (Workshop Held Conjunction ICCV)*, 2013, pp. 883–890.

[10] G. Schwartz and K. Nishino, "Automatically discovering local visual material attributes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3565–3573.

[11] H. Zhang, K. Dana, and K. Nishino, "Reflectance hashing for material recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3071–3080.

[12] C. Liu and J. Gu, "Discriminative illumination: Per-pixel classification of raw materials based on optimal projections of spectral BRDF," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 86–98, Jan. 2014.

[13] M. A. Mannan, D. Das, Y. Kobayashi, and Y. Kuno, "Object material classification by surface reflection analysis with a time-of-flight range sensor," in *Proc. Int. Symp. Visual Comput.*, 2010, pp. 439–448.

[14] M. Sato, S. Yoshida, A. Olwal, B. Shi, A. Hiyama, T. Tanikawa, M. Hirose, and R. Raskar, "SpecTrans: Versatile material classification for interaction with textureless, specular and transparent surfaces," in *Proc. ACM Conf. Human Factors Comput. Syst.*, 2015, pp. 2191–2200.

[15] A. Davis, K. L. Bouman, J. G. Chen, M. Rubinstein, F. Durand, and W. T. Freeman, "Visual vibrometry: Estimating material properties from small motion in video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 5335–5343.

[16] P. Saponaro, S. Sorensen, A. Kolagunda, and C. Kambhamettu, "Material classification with thermal imagery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 4649–4656.

[17] S. Su, F. Heide, R. Swanson, J. Klein, C. Callenberg, M. Hullin, and W. Heidrich, "Material classification using raw time-of-flight measurements," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3503–3511.

[18] A. Jarabo, B. Masia, J. Marco, and D. Gutierrez, "Recent advances in transient imaging: A computer graphics and vision perspective," *Visual Inform.*, vol. 1, no. 1, pp. 1–16, 2017.

[19] I. Gkioulekas, A. Levin, F. Durand, and T. Zickler, "Micron-scale light transport decomposition using interferometry," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 37:1–37:14, 2015.

[20] N. Abramson, "Light-in-flight recording by holography," *Opt. Lett.*, vol. 3, no. 4, 1978, Art. no. 121.

[21] T. Kakue, K. Tosa, J. Yuasa, T. Tahara, Y. Awatsuji, K. Nishio, S. Ura, and T. Kubota, "Digital light-in-flight recording by holography by use of a femtosecond pulsed laser," *IEEE J. Sel. Topics Quantum Electron.*, vol. 18, no. 1, pp. 479–485, Jan./Feb. 2012.

[22] A. Velten, D. Wu, A. Jarabo, B. Masia, C. Barsi, C. Joshi, E. Lawson, M. Bawendi, D. Gutierrez, and R. Raskar, "Femto-photography: Capturing and visualizing the propagation of light," *ACM Trans. Graph.*, vol. 32, no. 4, 2013, Art. no. 1.

[23] A. Kirmani, T. Hutchison, J. Davis, and R. Raskar, "Looking around the corner using ultrafast transient imaging," *Int. J. Comput. Vis.*, vol. 95, no. 1, pp. 13–28, 2011.

[24] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," *Nature Commun.*, vol. 3, 2012, Art. no. 745.

[25] M. O'Toole, F. Heide, D. Lindell, K. Zang, S. Diamond, and G. Wetzstein, "Reconstructing transient images from single-photon sensors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2289–2297.

[26] J. Lin, Y. Liu, M. B. Hullin, and Q. Dai, "Fourier analysis on transient imaging with a multifrequency time-of-flight camera," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3230–3237.

[27] K. Kitano, T. Okamoto, K. Tanaka, T. Aoto, H. Kubo, T. Funatomi, and Y. Mukaigawa, "Recovering temporal PSF using ToF camera with delayed light emission," *IPSJ Trans. Comput. Vis. Appl.*, vol. 9, Jun. 2017, Art. no. 15.

[28] D. Wu, A. Velten, M. O'Toole, B. Masia, A. Agrawal, Q. Dai, and R. Raskar, "Decomposing global light transport using time of flight imaging," *Int. J. Comput. Vis.*, vol. 107, no. 2, pp. 123–138, 2014.

[29] S. Fuchs, "Multipath interference compensation in time-of-flight camera images," in *Proc. Int. Conf. Pattern Recognit.*, 2010, pp. 3583–3586.

[30] A. A. Dorrington, J. P. Godbaz, M. J. Cree, A. D. Payne, and L. V. Streeter, "Separating true range measurements from multi-path and scattering interference in commercial range cameras," in *Proc. SPIE 7864, Three-Dimensional Imaging Interaction Meas.*, 2011, pp. 1–10.

[31] J. P. Godbaz, M. J. Cree, and A. A. Dorrington, "Closed-form inverses for the mixed pixel/multipath interference problem in AMCW lider," in *Proc. SPIE 8296, Comput. Imaging X*, 2012, pp. 1–15.

[32] D. Jimenez, D. Pizarro, M. Mazo, and S. Palazuelos, "Modelling and correction of multipath interference in time of flight cameras," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 893–900.

[33] F. Heide, L. Xiao, A. Kolb, M. B. Hullin, and W. Heidrich, "Imaging in scattering media using correlation image sensors and sparse convolutional coding," *Opt. Express*, vol. 22, no. 21, pp. 26338–26350, 2014.

[34] A. Kirmani, A. Benedetti, and P. A. Chou, "SPUMIC: Simultaneous phase unwrapping and multipath interference cancellation in time-of-flight cameras using spectral methods," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2013, pp. 1–6.

[35] A. Bhandari, M. Feigin, S. Izadi, C. Rhemann, M. Schmidt, and R. Raskar, "Resolving multipath interference in kinect: An inverse problem approach," in *Proc. IEEE SENSORS*, 2014, pp. 614–617.

[36] D. Freedman, E. Krupka, Y. Smolin, I. Leichter, and M. Schmidt, "SRA: Fast removal of general multipath for ToF sensors," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 1–15.

[37] H. Qiao, J. Lin, Y. Liu, M. B. Hullin, and Q. Dai, "Resolving transient time profile in ToF imaging via log-sum sparse regularization," *Opt. Lett.*, vol. 40, no. 6, pp. 918–921, 2015.

[38] S. Lee and H. Shim, "Skewed stereo time-of-flight camera for translucent object imaging," *Image Vis. Comput.*, vol. 43, no. C, pp. 27–38, 2015.

[39] N. Naik, A. Kadambi, C. Rhemann, S. Izadi, R. Raskar, and S. BingKang, "A light transport model for mitigating multipath interference in time-of-flight sensors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 73–81.

[40] A. Kadambi, J. Schiel, and R. Raskar, "Macroscopic interferometry: Rethinking depth estimation with frequency-domain time-of-flight," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 893–902.

[41] H. Shim and S. Lee, "Recovering translucent object using a single time-of-flight depth camera," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 5, pp. 841–854, May 2016.

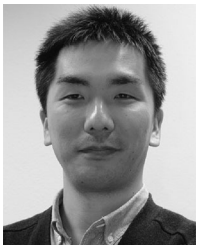
[42] K. Tanaka, Y. Mukaigawa, H. Kubo, Y. Matsushita, and Y. Yagi, "Recovering transparent shape from time-of-flight distortion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 4387–4395.

[43] N. Naik, S. Zhao, A. Velten, R. Raskar, and K. Bala, "Single view reflectance capture using multiplexed scattering and time-of-flight imaging," *ACM Trans. Graph.*, vol. 30, no. 6, 2011, Art. no. 1.

[44] F. Heide, L. Xiao, W. Heidrich, and M. B. Hullin, "Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3222–3229.

[45] A. Kadambi, H. Zhao, B. Shi, and R. Raskar, "Occluded imaging with time-of-flight sensors," *ACM Trans. Graph.*, vol. 35, no. 2, pp. 1–12, 2016.

[46] C.-Y. Tsai, K. N. Kutulakos, S. G. Narasimhan, and A. C. Sankaranarayanan, "The geometry of first-returning photons for non-line-of-sight imaging," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2336–2344.
[47] M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin, "Phasor imaging: A generalization of correlation-based time-of-flight imaging," *ACM Trans. Graph.*, vol. 34, no. 5, 2015, Art. no. 156.
[48] F. Heide, W. Heidrich, M. Hullin, and G. Wetzstein, "Doppler time-of-flight imaging," *ACM Trans. Graph.*, vol. 34, no. 4, pp. 36:1–36:11, 2015.
[49] S. Shrestha, F. Heide, W. Heidrich, and G. Wetzstein, "Computational imaging with multi-camera time-of-flight systems," *ACM Trans. Graph.*, vol. 35, 2016, Art. no. 33.
[50] R. Tadano, A. KumarPediredla, and A. Veeraraghavan, "Depth selective camera: A direct, on-chip, programmable technique for depth selectivity in photography," in *Proc. IEEE Conf. Comput. Vis.*, 2015, pp. 3595–3603.
[51] Z. Xu, T. Perry, and G. Hills, "Method and system for multi-phase dynamic calibration of three-dimensional (3D) sensors in a time-of-flight system," U.S. Patent US 8587771 B2, 2013.
[52] J. P. Godbaz, A. Dorrington, and M. J. Cree, "Understanding and ameliorating mixed pixels and multipath interference in AMCW lidar," in *TOF Range-Imaging Cameras.*, Berlin, Germany: Springer, 2013, pp. 91–116.
[53] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *The J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[54] A. Jarabo, J. Marco, A. Muñoz, R. Buisan, W. Jarosz, and D. Gutierrez, "A framework for transient rendering," *ACM Trans. Graph.*, vol. 33, no. 6, 2014, Art. no. 177.



Kenichiro Tanaka received the MS and PhD degrees in CS from Osaka University, in 2014 and 2017, respectively. In April 2017, he joined Nara Institute of Science and Technology (NAIST) as an assistant professor. His research interests include physically-based computer vision and computational photography, imaging, and illumination. He is a member of the IEEE and CVF.

Yasuhiro Mukaigawa received the ME and PhD degrees from the University of Tsukuba, in 1994 and 1997, respectively. He became a research associate at Okayama University, in 1997, an assistant professor at University of Tsukuba, in 2003, an associate professor at Osaka University, in 2004, and a professor at Nara Institute of Science and Technology (NAIST), in 2014. His current research interests include photometric analysis and computational photography. He is a member of the IEEE.

Takuya Funatomi received the BS degree in engineering, MS and PhD degrees in informatics from the Graduate School of Informatics, Kyoto University, Japan, in 2002, 2004, and 2007, respectively. He is currently an associate professor at Information Science Department, Nara Institute of Science and Technology (NAIST), Japan since 2015. He was an assistant professor at Kyoto University, Japan from 2007 to 2015, and a visiting assistant professor at Stanford University, in 2014. He has been working on 3D shape reconstruction of human as a deformable object and image processing techniques to observe human activity such as cooking. His research interests include computer vision, computer graphics, and pattern recognition. He is a member of the ACM and IEEE, the IEEE Computer Society and Communication Society, and Institute of Electronics, Information and Communication Engineers of Japan.

Hiroyuki Kubo received the MS and PhD degrees from Waseda University, in 2008 and 2012, respectively. He is an assistant professor at Nara Institute of Science and Technology (NAIST), Japan since 2014. His research interests include computer graphics and computer animation. He is a member of the ACM.

Yasuyuki Matsushita received the BS, MS and PhD degrees in EECS from the University of Tokyo, in 1998, 2000, and 2003, respectively. From April 2003 to March 2015, he was with Visual Computing group at Microsoft Research Asia. In April 2015, he joined Osaka University as a professor. His research area includes computer vision, machine learning and optimization. He is on the editorial board of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* (TPAMI), the *International Journal of Computer Vision* (IJCV), the *IPSJ Journal of Computer Vision and Applications* (CVA), the *The Visual Computer Journal*, and the *Encyclopedia of Computer Vision*. He served/is serving as a program co-chair of PSIVT 2010, 3DIMPVT 2011, ACCV 2012, ICCV 2017, and a general co-chair for ACCV 2014. He is a senior member of the IEEE.

Yasushi Yagi received the PhD degree from Osaka University, in 1991. He is the executive vice president of Osaka University. In 1985, he joined the Product Development Laboratory, Mitsubishi Electric Corporation, where he worked on robotics and inspections. He became a research associate in 1990, a lecturer in 1993, an associate professor in 1996, and a professor in 2003 at Osaka University. He was also director of the Institute of Scientific and Industrial Research, Osaka university from 2012 to 2015. International conferences for which he has served as chair include: FG1998 (Financial chair), OMINVIS2003 (Organizing chair), ROBIO2006 (Program co-chair), ACCV2007 (Program chair), PSVIT2009 (Financial chair), ICRA2009 (Technical Visit chair), ACCV2009 (General chair), ACPR2011 (Program co-chair) and ACPR2013 (General chair). He has also served as the editor of the IEEE ICRA Conference Editorial Board (2007–2011). He is an editorial member of IJCV and the editor-in-chief of the *IPSJ Transactions on Computer Vision & Applications*. He was awarded ACM VRST2003 Honorable Mention Award, IEEE ROBIO2006 Finalist of T.J. Tan Best Paper in Robotics, IEEE ICRA2008 finalist for Best Vision Paper, MIRU2008 Nagao Award, and PSIVT2010 Best Paper Award. His research interests are computer vision, medical engineering and robotics. He is a fellow of IPSJ and a member of IEICE, RSJ, and IEEE.