Material Classification using Frequency- and Depth-Dependent Time-of-Flight Distortion

Kenichiro Tanaka[1]   Yasuhiro Mukaigawa[1]   Takuya Funatomi[1]
Hiroyuki Kubo[1]   Yasuyuki Matsushita[2]   Yasushi Yagi [2]
[1] Nara Institute of Science and Technology (NAIST)   [2] Osaka Univerisity

{ktanaka,mukaigawa,funatomi,hkubo}@is.naist.jp, {yasumat@ist,yagi@sanken}.osaka-u.ac.jp

Abstract

This paper presents a material classification method using an off-the-shelf Time-of-Flight (ToF) camera. We use a key observation that the depth measurement by a ToF camera is distorted in objects with certain materials, especially with translucent materials. We show that this distortion is caused by the variations of time domain impulse responses across materials and also by the measurement mechanism of the existing ToF cameras. Specifically, we reveal that the amount of distortion varies according to the modulation frequency of the ToF camera, the material of the object, and the distance between the camera and object. Our method uses the depth distortion of ToF measurements as features and achieves material classification of a scene. Effectiveness of the proposed method is demonstrated by numerical evaluation and real-world experiments, showing its capability of even classifying visually similar objects.

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 appearance. For example, artificial plastic fruits and real fruits confronting a camera produce visually similar RGB images that are difficult to distinguish. One of the possible strategies to distinguish similar appearance is to use the optical responses of the target object such as spatial, angular, and temporal spread 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 on top of the RGB observations.

Recently, Heide *et al*. [13] have developed a method that recovers transient images from observations by a low-cost

(a) Mayonnaise bottle   (b) Distorted depth

■ Mayonnaise
■ Paper label
■ Plastic cap
■ Undefined

(c) Material classification   (d) Corrected depth
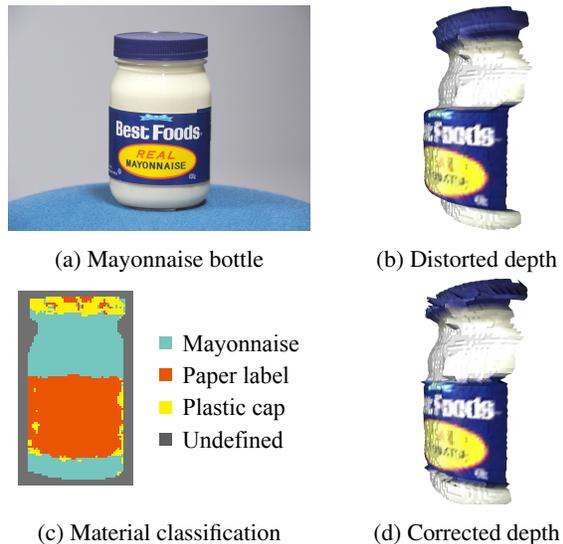
Figure 1: Depth distortion of a ToF camera. (a) A mayonnaise bottle is measured by a Kinect. (b) Measured depth in a 3D view. There is a gap in depth between the mayonnaise and label regions. We use this depth distortion for material classification. (c) Material segmentation result. The material label is assigned for each pixel. (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.

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 [20, 32, 33] 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 from an off-the-shelf ToF camera without explicitly recovering 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. 1(b), where the depth gap between the mayonnaise 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 target and the camera. Using the depth distortions as a *feature* of the material, 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 [13, 20]. Second, we show how 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 and it serves as important clue for realizing material classification.

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 estimation. The methods based on the visual appearance, *e.g.*, color, shape, and/or textures of the material [3, 44, 28, 37, 38], typically only require a single RGB image; thus, the setups are easy to realize. The main 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.

The class of approaches based on the optical properties, such as BRDF [49, 27], shading [29], and spectrum [36], has a capability of distinguishing visually similar objects in higher accuracy because the optical properties convey richer information about the material. However, constructing such measurement systems and building database of samples generally require carefully controlled settings. This class includes approaches based on other physical properties, *e.g.*, elasticity [4], and water permeation and heating/cooling process [35]. 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. Unlike these approaches, our method uses an off-the-shelf ToF camera and needs only single observation at least, 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 [41] 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 customization of a ToF camera for obtaining the measurements. In contrast, our method only uses an off-the-shelf ToF camera. 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 the comprehensible overview of temporal light transport, we refer the reader to the recent survey by Jarabo *et al.* [17]. A time domain impulse response of the scene, as known as light-in-flight and transient imaging, can be obtained using an interferometer [8], holography [1, 22], and femtosecond-pulsed laser [46, 24, 45]. The time domain impulse response can be also recovered using the ToF camera, where the cost and temporal resolution drastically decrease. Because the ToF camera is a device for measuring sub-nano second phenomena, it can be used for visualizing the light propagation of the scene by frequency sweep [13, 26, 33] and optical coding [20, 32], while it requires customization of a ToF camera. These measurement methods may be able to be applied to the task of material classification [47], although they require careful and expensive setups. On the other hand, our method bypasses the exact recovery of the time domain impulse response 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 [7, 5, 9, 18], parametric model [15, 23], K-sparity and optimization [2, 6, 34], stereo ToF cameras [25], using external projector [30], and frequency sweep [19]. 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 be used for correcting depths.

There are other scene analysis methods using ToF cameras, *e.g.*, recovering the shape of transparent and translucent objects [39, 43], and measuring a slice of BRDF [31]. In addition, computational imaging methods using a ToF camera, such as imaging around the corner [14, 21], separating direct and indirect light transport [47, 32, 11], imaging the velocity of the object [12, 40], and imaging at a specific depth [42] are proposed. Our method can also be considered one of the scene analysis methods as it aims at material classification of the scene.

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 $f_\omega(t)$ and measures its attenuated amplitude and phase delay. From the phase delay and the speed of light, the depth of the scene can be obtained.

In general, a scene can have the "multi-path" effect due to inter-reflections and subsurface scattering, which degrade the depth estimation accuracy. Image formation models regarding the multi-path effect have been well understood thanks to the previous works [13, 20, 11]; hence, we briefly explain one of the models that we are going to use in this paper. Following a phasor representation [11], amplitude and phase of the returned wave can be represented by a single complex value $\mathbf{c} \in \mathbb{C}$, called phasor, governed by the modulation frequency ω. The measured amplitude $\tilde{a}_\omega \in \mathbb{R}$ and depth $\tilde{d}_\omega \in \mathbb{R}$ of the ToF camera are obtained as

$$\begin{cases} \tilde{d}_\omega & = \frac{c}{4\pi\omega} \arg \mathbf{c}(\omega), \\ \tilde{a}_\omega & = |\mathbf{c}(\omega)|, \end{cases} \quad (1)$$

where the $\arg$ operator returns the angle of a complex phasor, and c is the speed of light.

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

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

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

$$r(t) = \int_\mathcal{P} r_\rho \delta(|\rho| - t)d\rho, \quad (3)$$

where r_ρ is the contribution of the path ρ, $|\rho|$ is the time travelled 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. Figure 2(b) 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 negative domain of $r(t)$ is zero, Eq. (2) expresses that ToF camera measures the Fourier coefficients of the impulse response at the frequency of the light modulation.

Frequency dependent depth distortion The principle of the ToF camera assumes that the impulse response forms

(a) Sinusoidal wave   (d) Distorted wave

(b) Phasor representation   (e) Distorted unit ball

(c) Phasor at different depth   (f) Large distortion of depth
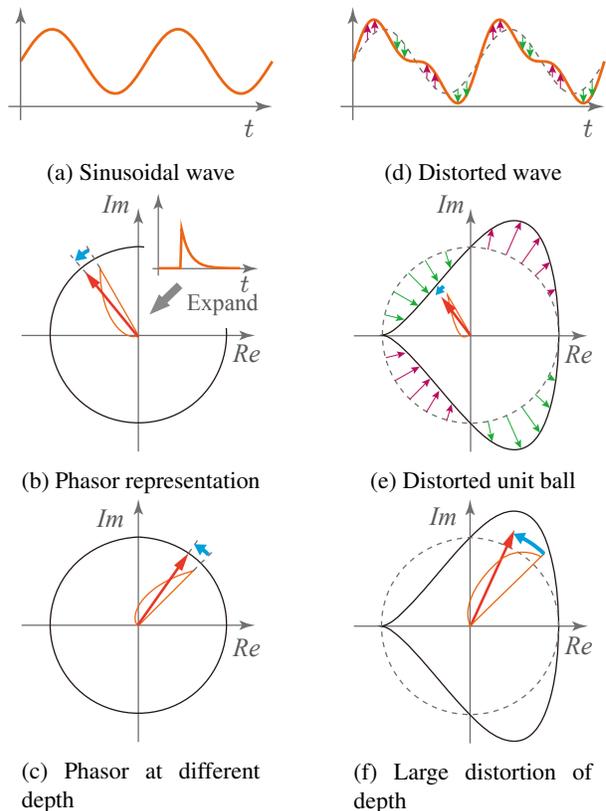
Figure 2: Phasor representation of ToF observations. (a) Sinusoidal illumination, (b) Time domain PSF is expanded to the imaginary plane (orange). (c) When the object is placed at different depths, the observation gets rotated but phase distortion remains the same as (b). (d) Biased periodic illumination. This toy example adds 20% 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).

Dirac delta function as $r(t) = \alpha\delta(t)$, where α is the amplitude decay of modulated light. In this case, the measured depth $\tilde{d}_\omega$ becomes

$$\tilde{d}_\omega = \frac{c}{4\pi\omega} \underbrace{\arg \int_0^\infty \alpha\delta(t-\tau)e^{-2\pi i\omega t}dt}_{=2\pi\omega\tau} = \frac{c\tau}{2} = d, \quad (4)$$

where $d = \frac{c\tau}{2}$ is the ground truth depth of the object. Equation (4) 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.

In reality, almost all materials except for the perfect mirror surface yield various shapes of impulse responses due to diffuse and subsurface scattering [47]. When the target object exhibits a temporally broad shape of the impulse re-

sponse, 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 $2\pi\omega\tau$, and as a result, the measured depth $\tilde{d}_\omega$ is 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:

$$\mathcal{F}\left[r(t-\tau)\right] = e^{-2\pi i \omega \tau} \mathcal{F}\left[r(t)\right]$$
$$= e^{-2\pi i \omega \tau} \hat{r}(\omega),$$

where $\mathcal{F}[\cdot]$ is the Fourier transform and $\hat{r}(\omega)$ is the Fourier transform of the function $r(t)$. Measured depth $\tilde{d}_\omega$ can then be represented as

$$\tilde{d}_\omega = \frac{c}{4\pi\omega} \arg\left(e^{-2\pi i \omega \tau} \hat{r}\left(\omega\right)\right)$$
$$= d + \frac{c}{4\pi\omega} \arg \hat{r}(\omega). \tag{5}$$

The second term $\frac{c}{4\pi\omega} \arg \hat{r}(\omega)$ is the depth distortion at frequency ω. In Fig. 2(b), the depth distortion is illustrated as a blue arrow.

While a single observation of 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 [33].

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 developing a reliable material classification system. 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 $d + \Delta d$, $r(t - \tau)$ is shifted by $\Delta\tau = \frac{2\Delta d}{c}$ in the time domain. As a result, the measured depth $\tilde{d}'_\omega$ becomes

$$\tilde{d}'_\omega = \frac{c}{4\pi\omega} \arg\left(e^{-2\pi i \omega(\tau+\Delta\tau)} \hat{r}\left(\omega\right)\right)$$
$$= d + \Delta d + \frac{c}{4\pi\omega} \arg \hat{r}(\omega). \tag{6}$$

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. (5). Figure 2(c) illustrates the depth distortion at a different depth 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. 2(b).

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 [48, 10]. When the illumination wave has harmonics components as shown in Fig. 2(d), the ToF observation exhibits depth-dependency as illustrated in Figs. 2(e) and 2(f). 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 is then written as

$$\mathbf{c}(\omega) = \int_0^\infty r(t-\tau) b_\omega(2\pi\omega t) e^{-2\pi i \omega t} dt. \tag{7}$$

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.* [41] 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. For describing how to use the depth distortions 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

$$v_{\omega_i, d_j} = d_j - \tilde{d}_{\omega_i, j}, \tag{8}$$

where $\tilde{d}_{\omega_i, j}$ is the measured depth at the i-th modulation frequency ω_i $(i \in \{1, \cdots, n\})$ and the j-th position $(j \in$

$\{1, \cdots, m\}$), and d_j is the actual depth at the j-th position. By aligning these distortions, a mn-length vector $\mathbf{v}$ can be formed as a *feature* vector of the object as

$$\mathbf{v} = \begin{bmatrix} v_{\omega_1, d_1} & \cdots & v_{\omega_n, d_m} \end{bmatrix}^T. \tag{9}$$

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 n-th modulation frequency, as the reference measurement. The relative depth distortions can be obtained by taking differences from the reference measurement as

$$v'_{\omega_i, d_j} = v_{\omega_i, d_j} - v_{\omega_n, d_j} = \tilde{d}_{\omega_n, j} - \tilde{d}_{\omega_i, j}, \tag{10}$$

where i ranges from 1 to $n - 1$. We can then setup an $m(n - 1)$-length vector $\mathbf{v}$ 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 m-th depth position as the reference measurement as

$$v''_{\omega_1, d_j} = v_{\omega_1, d_j} - v_{\omega_1, d_m} = \tilde{d}_{\omega_1, m} - \tilde{d}_{\omega_1, j} + \Delta d_j, \tag{11}$$

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

4.1. 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 as exemplar to look up the closest material.

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 tends to be high-dimensional while the number of materials in the database may be small, it is preferred 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 with near-distance reflectance.

For these reasons, we adopt a simple nearest neighbor classifier, which assesses the Euclidean distance (ℓ_2 norm). 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

$$\xi_p = \frac{1}{N} \sum_{k=0}^{nm} \begin{cases} 0 & v_k = \text{N/A} \\ (v_k - v_k^p)^2 & \text{otherwise}, \end{cases} \tag{12}$$

where N is the number of valid elements, and v_k and v_k^p are k-th 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

$$\hat{p} = \operatorname*{argmin}_p \xi_p.$$

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 by a ToF camera and a linear translation stage system as shown in Fig. 3. We use Microsoft Kinect v2 for a ToF camera, which has three modulation frequencies ($n = 3$), and a OptoSigma's translation stage (SGSP46-800). As the official Kinect API does not support an access to depth measurements of each frequency, we have slightly altered an open-source software `libfreenect2` to obtain such data[1].

First, we measure the depth distortion data for many materials and examine their differences across materials. The target object is placed on a linear translation stage changing the depth from 600 mm to 1250 mm ($m = 2600$), and is measured several times with changing the orientation of the object. The ground truth depth is obtained from the position of the translation stage, which is only used this test. Figure 4 shows the depth distortion of three materials; white acrylic board, polystyrene board, and opal diffusion glass. They are visually similar object (white, planer, and no texture) hence appearance based methods have difficulty to distinguish these objects. On the other hand, depth distortions of ToF observations show significant difference across materials and retain consistency over measurement sessions.

---

[1]The source code is publicly available on our website.
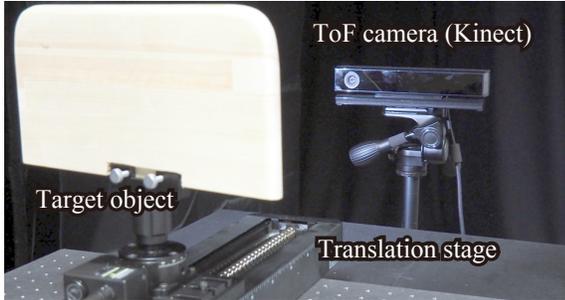
Figure 3: Experimental setup. We use Kinect as a ToF camera, and the target object is placed on a linear translation stage.

Using this depth distortion data, we assess the accuracy of material classification by the nearest neighbor classifier. The dataset consists of 26 materials including metal, wood, plastic, fabric, and so on, with 13 orientations for each material to enable the classifier to deal with target objects with arbitrary surface orientations. We evaluate the classification accuracy using three different features: Frequency-dependent distortion, depth-dependent distortion, and both of them. Using the feature with only frequency-dependent distortion ($n = 3$ and $m = 1$), the accuracy is 57.4%. This low accuracy is due to the limited availability of the number of frequency channels. Using only depth-dependent distortions ($n = 1$, $m = 2600$, and using Eq. (11)), the accuracy is improved to 81.6%[2]. Finally, with both of frequency- and depth-dependent distortions ($n = 3$ and $m = 2600$), the accuracy is further improved to 90.5%. The confusion matrix is shown in Fig. 5. While many materials are correctly classified, some materials are miss-classified. For example, plaster and paper, or expanded and rigid polyvinyl chlorides have similar impulse responses due to similar scattering properties; therefore they are sometimes miss-classified.

Feature variations w.r.t surface orientation　When the surface orientation of the target object varies, the time-domain impulse response may also vary. To confirm the effect of surface orientations, we measure a wooden board by changing the orientation and assess the variation of feature vectors with respect to varying orientations. Figure 6 shows the variation of the nearest distance from 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 goes beyond this upper-bound distance, it will be misclassified. The feature is stable under around 70 degrees, which indicate that the depth distortion feature is reliable for the

confronting surface in practice but may break down for a steep-slanted surface, *e.g.*, near the edges of a round-shape object.

Feature variations w.r.t. shape　When the shape of the target object varies, the time domain impulse response may also vary, especially for a concave shape where significant inter-reflections occur. To confirm the effect of the shape of the object, we set up a scene of folded cardboard that can change the opening angle. We measure the folding edge area of the cardboard with changing the opening angle from the small angle (closed) via 180 degrees (flat) to large angle (protruded) as shown in Fig. 7(a). The distances of feature vectors between the folded and flat cardboards are plotted in the blue line in Fig. 7(b). The red line represents the upper-bound of the flat cardboard class, under which the target is regarded a flat cardboard, and a moderate robustness against the shape variation is shown.

Material segmentation　Our method can be applied in a pixel-wise manner to achieve material-based segmentation. Figure 8 shows a couple of example of material segmentation. For the scene shown in Fig. 8(a), all objects in the scene are white and the material classes are not obvious in the RGB image. With our method, the material is classified for each pixel as shown in Fig. 8(b). For this application, we use only frequency-dependent variations without the depth-dependent ones, *i.e.*, $m = 1$, because the alignment of the pixels may become hard when the geometric relationship between the camera and scene varies. As a result, the result appears to be a little bit noisy, but it still shows faithful classification performance. For this experiment, we used a reduced database containing only four materials as the dimensionality of the feature vector is limited. Figure 8(c) shows another scene where wallets made of genuine and fake leather are placed, and they are correctly classified as shown in Fig. 8(d).

Depth correction　Once materials are classified, the distorted depths can be corrected for recovering an accurate 3D shape using the material database that contains the samples of distortions for all materials. An example of the depth correction 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 as shown in Fig. 1(b). Figure 1(c) shows our result of material segmentation. Again, we do not change the depth of the target; therefore, only frequency-dependent variation is used ($m = 1$) with a limited database. Although some artifacts are observable because of the limited amount of measurement and steep surface orientations, mayonnaise and the label regions are largely well separated. Using the segmentation result and depth distortion database, a faithful 3D shape

---

[2]Details and confusion matrices are shown in the supplementary.
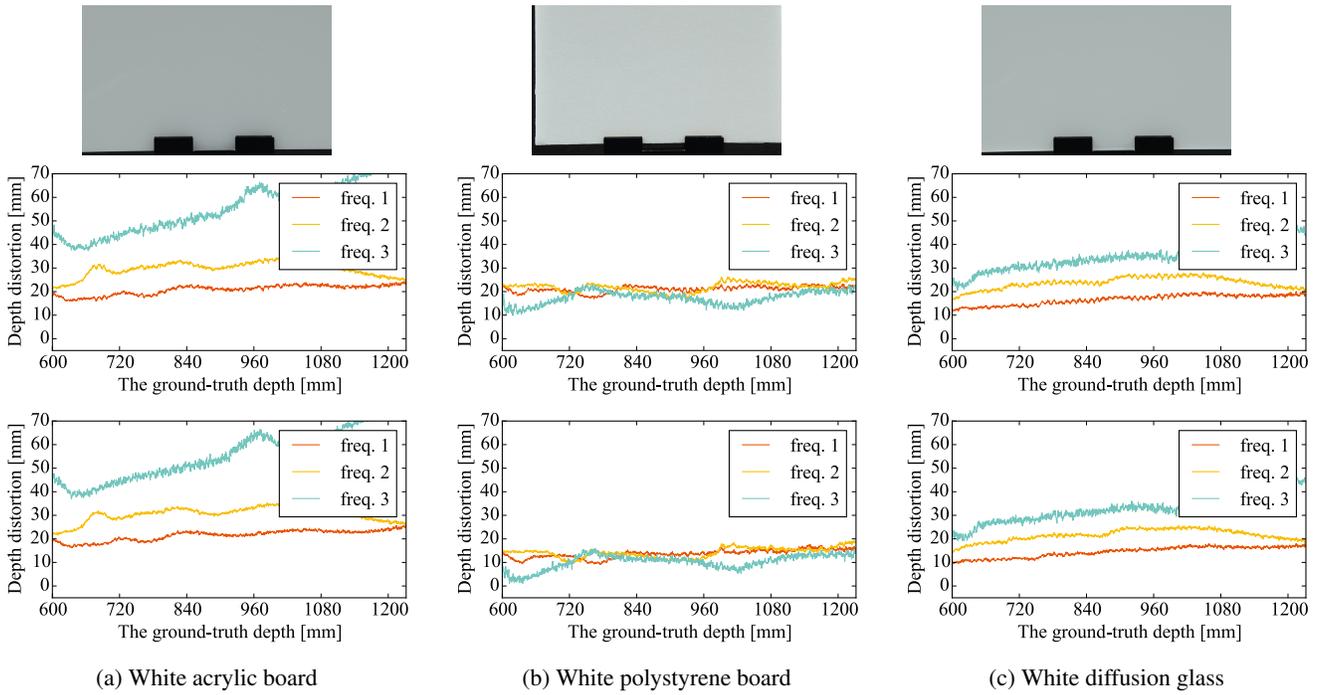
Figure 4: Measured depth distortions using Kinect for three objects. The ground truth depth is obtained via a linear translation stage. The top row shows photographs of the target objects. Measurements of the second and third rows are different in terms of surface orientation. Depth distortion of each frequency varies along with the actual depth and material. Depth distortion is similar for the same material regardless of the surface orientation, but largely different in different materials. This frequency- and depth-dependent depth distortion is our key observation for material classification.
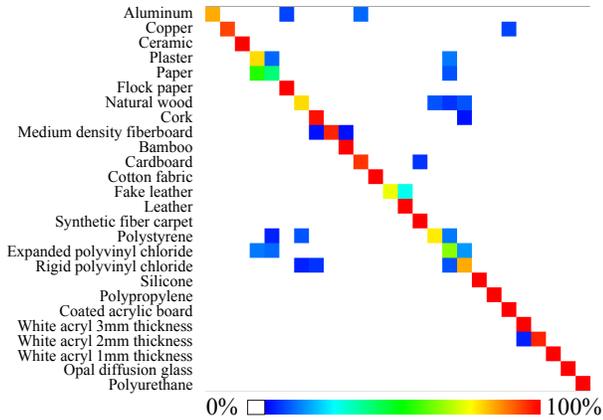
Figure 5: Confusion matrix. Red indicates the higher value and it appears on the diagonal. Overall accuracy is 90.5%.

Figure 6: 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.

of the mayonnaise bottle is recovered as shown in Fig. 1(d). Compared to the original shape, the depth discontinuity between mayonnaise and the label regions is significantly reduced.

Real-time classification system We develop a near real-time material classification system, which can recognize the target material category by a hand-held ToF camera. Using

the partial matching described in Eq. (12), our method outputs the result in near real-time even when observations at only a small number of depth locations m is available. By increasing the variation of depths by moving a target object or the camera (increasing m), the classification accuracy is gradually improved because richer information can be fed to the classifier.[3]

---

[3]A video is included in a supplementary material. Please refer to the video for this demonstration.

(a) Scenes of the folded cardboard
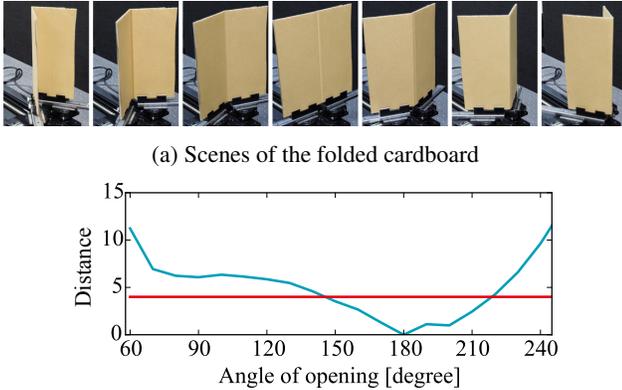
(b) Variation of feature distance between flat and folded cardboards

Figure 7: 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.

(a) The scene of white utensils

(b) Classification result

(c) The scene of wallets

(d) Classification result
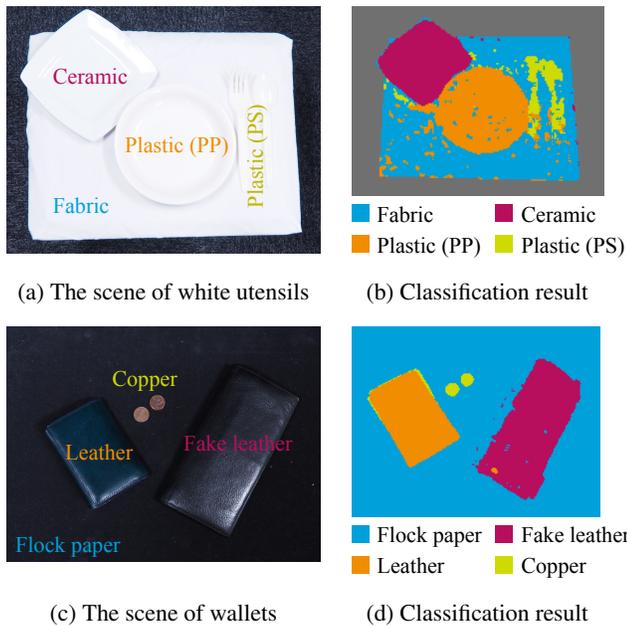
Figure 8: Material segmentation results. (a) All utensils are white hence it is difficult to classify only with an RGB image. (b) The result of our material classification. Although there are some estimation error because of the pixel-wise process and only one depth variation, the scene is much interpretable than the RGB image. (c) Wallets made of genuine and fake leather and copper coins are placed in the scene. (d) Material segmentation result.

Thickness classification Depth distortion is also useful for thickness estimation of the optically thin material. For

example, white acrylic boards are optically thin so that the impulse response varies along with its physical thickness. The thickness of the white acrylic board can therefore be classified as shown at the near bottom part of Fig. 5. Currently, our method is limited to classification of different thicknesses, but we are interested in turning the problem into a regression problem for estimating the thickness.

6. Discussions

We have developed a material classification using an off-the-shelf ToF camera. We show that the measured depth using a ToF camera is distorted according to the time domain impulse response of materials, and the distortion varies along with the modulation frequency and the distance between the object and the camera. We use the ToF depth distortion as a cue for material classification, and developed a classification method.

Our method is based on a difference of time domain impulse response among materials, hence we assume the impulse response is the same for the same material. However, it may not be always true because the shape, color, and geometry may cause varying impulse response. We have assessed the variation of the designed feature over varying shape and surface orientation and have shown the robustness of the developed feature up to a limitation on the variations. Related to this problem, optically thin object's impulse response also varies along with the object's thickness. On one hand, this allows us to classify thickness of the target object, but on the other hand, it indicates that a database with varying thicknesses is needed for correctly classifying materials of an object that may have arbitrary thickness. This is one of the current limitations of our method. Using simulation such as Jarabo *et al.*'s renderer [16], a very large database that includes all the materials and variations could be obtained. At this point, it looks non-straightforward to prepare materials' physical properties and measure the illumination bias b_ω using an oscilloscope; however, it is a new potential direction.

Another limitation is that the depth distortion measures, especially the depth-dependent distortion, is camera-dependent because the bias of illumination wave may be different across different devices. The development of the inter-device feature or transferring the database for a different camera is an important future work.

The amplitude of ToF observation also varies over different frequencies and depths, hence it can be also used for analyzing the scene. We did not use this cue in this paper, but we are interested in investigating this respect for further improving the classification accuracy.

References

[1] N. Abramson. Light-in-Flight Recording by Holography. *Optics Letters*, 3(4):121, 1978. 2

[2] A. Bhandari, M. Feigin, S. Izadi, C. Rhemann, M. Schmidt, and R. Raskar. Resolving Multipath Interference in Kinect: an Inverse Problem Approach. In *IEEE SENSORS*, pages 614–617. IEEE, 2014. 2

[3] B. Caputo, E. Hayman, and P. Mallikarjuna. Class-Specific Material Categorisation. In *Proc. International Conference on Computer Vision (ICCV)*, pages 1597–1604. IEEE, 2005. 2

[4] 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 Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5335–5343, 2015. 2

[5] 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 *SPIE 7864, Three-Dimensional Imaging, Interaction, and Measurement*, 2011. 2

[6] D. Freedman, E. Krupka, Y. Smolin, I. Leichter, and M. Schmidt. Sra: Fast Removal of General Multipath for ToF Sensors. In *Proc. European Conference on Computer Vision (ECCV)*, pages 1–15, 2014. 2

[7] S. Fuchs. Multipath Interference Compensation in Time-of-Flight Camera Images. In *International Conference on Pattern Recognition*, pages 3583–3586. IEEE, 2010. 2

[8] I. Gkioulekas, A. Levin, F. Durand, and T. Zickler. Micron-Scale Light Transport Decomposition using Interferometry. *ACM Transactions on Graphics (ToG)*, 34(4):37:1–37:14, 2015. 2

[9] J. P. Godbaz, M. J. Cree, and A. A. Dorrington. Closed-Form Inverses for the Mixed Pixel/Multipath Interference Problem in AMCW Lider. In *SPIE 8296, Computational Imaging X*, 2012. 2

[10] 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.*, pages 91–116, 2013. 4

[11] M. Gupta, S. K. Nayar, M. B. Hullin, and J. Martin. Phasor Imaging: a Generalization of Correlation-Based Time-of-Flight Imaging. *ACM Transactions on Graphics (ToG)*, 34(5), 2015. 2, 3

[12] F. Heide, W. Heidrich, M. Hullin, and G. Wetzstein. Doppler Time-of-Flight Imaging. *ACM Transactions on Graphics (ToG)*, 34(4):36:1–36:11, 2015. 2

[13] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich. Low-Budget Transient Imaging using Photonic Mixer Devices. *ACM Transactions on Graphics (ToG)*, 32(4):1, 2013. 1, 2, 3

[14] 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 Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 2

[15] F. Heide, L. Xiao, A. Kolb, M. B. Hullin, and W. Heidrich. Imaging in Scattering Media using Correlation Image Sensors and Sparse Convolutional Coding. *Optics express*, 22(21):26338–50, 2014. 2

[16] A. Jarabo, J. Marco, A. Muñoz, R. Buisan, W. Jarosz, and D. Gutierrez. A Framework for Transient Rendering. *ACM Transactions on Graphics (ToG)*, 33(6), 2014. 8

[17] A. Jarabo, B. Masia, J. Marco, and D. Gutierrez. Recent Advances in Transient Imaging: A Computer Graphics and Vision Perspective. *Visual Informatics*, 1(1):1–16, 2017. 2

[18] D. Jimenez, D. Pizarro, M. Mazo, and S. Palazuelos. Modelling and Correction of Multipath Interference in Time of Flight Cameras. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 893–900. IEEE, 2012. 2

[19] A. Kadambi, J. Schiel, and R. Raskar. Macroscopic Interferometry: Rethinking Depth Estimation with Frequency-Domain Time-Of-Flight. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 893–902, 2016. 2

[20] 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 Transactions on Graphics (ToG)*, 32(6):1–10, 2013. 1, 2, 3

[21] A. Kadambi, H. Zhao, B. Shi, and R. Raskar. Occluded Imaging with Time-of-Flight Sensors. *ACM Transactions on Graphics (ToG)*, 35(2):1–12, 2016. 2

[22] 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 Journal of Selected Topics in Quantum Electronics*, 18(1):479–485, 2012. 2

[23] 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 *IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2013. 2

[24] A. Kirmani, T. Hutchison, J. Davis, and R. Raskar. Looking Around the Corner using Ultrafast Transient Imaging. *International Journal of Computer Vision (IJCV)*, 95(1):13–28, 2011. 2

[25] S. Lee and H. Shim. Skewed Stereo Time-of-Flight Camera for Translucent Object Imaging. *Image and Vision Computing*, 43(C):27–38, 2015. 2

[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 Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3230–3237. IEEE, 2014. 2

[27] C. Liu and J. Gu. Discriminative Illumination: Per-Pixel Classification of Raw Materials Based on Optimal Projections of Spectral Brdf. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 36(1):86–98, 2014. 2

[28] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz. Exploring Features in a Bayesian Framework for Material Recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 239–246. IEEE, 2010. 2

[29] 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 *Advances in Visual Computing*, pages 439–448. Springer Berlin Heidelberg, 2010. 2

[30] N. Naik, A. Kadambi, C. Rhemann, S. Izadi, R. Raskar, and S. Bing Kang. A Light Transport Model for Mitigating Multipath Interference in Time-of-Flight Sensors. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 73–81, 2015. 2

[31] N. Naik, S. Zhao, A. Velten, R. Raskar, and K. Bala. Single View Reflectance Capture using Multiplexed Scattering and Time-of-Flight Imaging. *ACM Transactions on Graphics (ToG)*, 30(6):1, 2011. 2

[32] 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 Transactions on Graphics (ToG)*, 33(4):1–11, 2014. 1, 2

[33] C. Peters, J. Klein, M. B. Hullin, and R. Klein. Solving Trigonometric Moment Problems for Fast Transient Imaging. *Proc. SIGGRAPH Asia*, 34(6):1–11, 2015. 1, 2, 4

[34] H. Qiao, J. Lin, Y. Liu, M. B. Hullin, and Q. Dai. Resolving Transient Time Profile in ToF Imaging Via Log-Sum Sparse Regularization. *Optics letters*, 40(6):918–21, 2015. 2

[35] P. Saponaro, S. Sorensen, A. Kolagunda, and C. Kambhamettu. Material Classification with Thermal Imagery. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4649–4656, 2015. 2

[36] 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 Conference on Human Factors in Computing Systems (CHI)*, pages 2191–2200. ACM Press, 2015. 2

[37] G. Schwartz and K. Nishino. Visual Material Traits: Recognizing Per-Pixel Material Context. *Proceedings of Color and Photometry in Computer Vision (Workshop held in conjunction with ICCV)*, pages 883–890, 2013. 2

[38] G. Schwartz and K. Nishino. Automatically Discovering Local Visual Material Attributes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3565–3573, 2015. 2

[39] H. Shim and S. Lee. Recovering Translucent Object using a Single Time-of-Flight Depth Camera. *IEEE Transactions on Circuits and Systems for Video Technology*, 26:841–854, 2015. 2

[40] S. Shrestha, F. Heide, W. Heidrich, and G. Wetzstein. Computational Imaging with Multi-Camera Time-of-Flight Systems. In *Proc. SIGGRAPH*, 2016. 2

[41] 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 Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3503–3511, 2016. 2, 4

[42] R. Tadano, A. Kumar Pediredla, and A. Veeraraghavan. Depth Selective Camera: a Direct, On-Chip, Programmable Technique for Depth Selectivity in Photography. In *Proc. International Conference on Computer Vision (ICCV)*, pages 3595–3603, 2015. 2

[43] K. Tanaka, Y. Mukaigawa, H. Kubo, Y. Matsushita, and Y. Yagi. Recovering Transparent Shape from Time-of-Flight Distortion. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4387–4395, 2016. 2

[44] M. Varma and A. Zisserman. A Statistical Approach to Material Classification using Image Patch Exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 31(11):2032–2047, 2009. 2

[45] 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 communications*, 3(745), 2012. 2

[46] 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 Transactions on Graphics (ToG)*, 32(4):1, 2013. 2

[47] 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. *International Journal of Computer Vision (IJCV)*, 107(2):123–138, 2014. 2, 3

[48] Z. Xu, T. Perry, and G. Hills. Method and system for multiphase dynamic calibration of three-dimensional (3d) sensors in a time-of-flight system. *US Patent*, US 8587771 B2, 2013. 4

[49] H. Zhang, K. Dana, and K. Nishino. Reflectance Hashing for Material Recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3071–3080, 2015. 2