

# Measuring Translucent Objects using Spatially and Temporally Modulated Light

KENICHIRO TANAKA<sup>1,2,a)</sup> YASUHIRO MUKAIGAWA<sup>2</sup> HIROYUKI KUBO<sup>2</sup>  
TAKUYA FUNATOMI<sup>2</sup> YASUYUKI MATSUSHITA<sup>1</sup> YASUSHI YAGI<sup>1</sup>

**Abstract:** Measurement of translucent objects is an important technology with broad potential applications. In the computer vision field, computational photography approaches that combine optical design and computational algorithms to obtain informative images have been actively developed to enhance and restore images. However, it remains difficult to recover heavily degraded information within translucent objects because light penetrates and scatters inside the object. In this thesis, we propose a measurement framework using either spatially or temporally modulated light. Using a projector to realize spatial modulation and a Time-of-Flight (ToF) camera to realize temporal modulation, we reveal how the modulated light conveys the information of spatial and temporal spread of light, respectively. Based on these relationships, we recover clear images stored inside translucent objects, accurate 3-d shape, and material. We demonstrate the effectiveness of our method in real-world experiments.

## 1. Introduction

Measurement of transparent and translucent objects is an important technology that may have applications in fields, including autonomous robotics, industrial inspection, underwater science, art, and historical research. However, such measurements remain challenging in computer vision applications because of the complicated optical phenomena involved, such as subsurface scattering and refraction. For example, when a translucent object is illuminated by a laser pointer, the laser beam becomes blurred and degraded as a result of the light penetrating and scattering inside the object. To overcome these problems in measuring transparent and translucent objects, their complicated optical responses, referred to as ‘light transport’, must be compensated for.

Properly dealing with the light transport of a scene can help a computer to correctly understand the scene. For example, clear images can be recovered in foggy weather if the scattering properties of the fog are identified. Translucency of the scene is also important for recognizing the scene, as well as removing its effects. Taking into account the magnitude of subsurface scattering, the material of the target object can be recognized because translucency conveys the optical properties of the material.

We aim to recover invisible information from images degraded by complicated light transport such as scattering. Because scattering is too complex to be fully measured and analyzed, we do not explicitly recover either the scattering parameters or the point spread functions (PSF). Instead, we propose model-based approaches for visualizing invisible information that indirectly

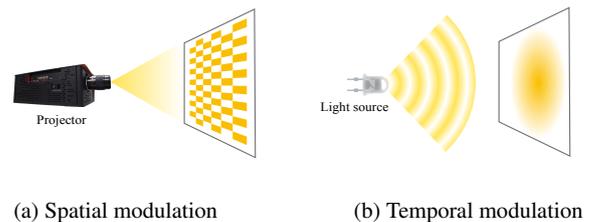


Fig. 1: Amplitude-modulated illumination. (a) High-frequency illumination, in which the illumination is spatially modulated by a projector. Illumination is rapidly switched on and off with the illuminated position. A typical wavelength is sub-millimeter. (b) Illumination is temporally modulated at several tens of megahertz, using a ToF camera.

observes spatial and temporal PSFs.

We adopt either spatially or temporally amplitude-modulated illumination as shown in Fig. 1. Spatially modulated illumination can be realized using a projector, which can control the amplitude of light by projecting a high-frequency pattern. Temporal modulation can be realized by turning the light source on and off at high speed, which is achieved using a Time-of-Flight(ToF) camera. Observation under an ordinary illumination such as a light bulb degrades the information of the scene because scattering behaves as a low-pass filter. In contrast, observation under amplitude-modulated light maintains the frequency response of the scene. Using various frequencies of the amplitude-modulated illumination, many frequency responses can be obtained; thus, invisible information can be recovered via computation.

We propose a framework that simultaneously deals with spatially and temporally modulated illumination at different frequencies, as shown in Fig. 2. We illuminate the scene by spatially or

<sup>1</sup> Osaka University, Suita, Osaka 567-0047, Japan

<sup>2</sup> NARA Institute of Science and Technology (NAIST), Ikoma, Nara 630-0192, Japan

<sup>a)</sup> ktanaka@is.naist.jp

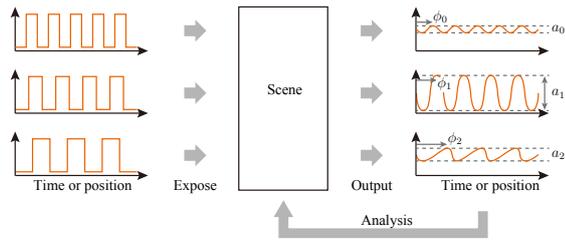


Fig. 2: An overview of our framework. We illuminate the scene using either spatially or temporally amplitude-modulated light at different frequencies and observe the response as amplitude and phase shift of the returned wave. From these observations, we recover invisible information from the scene via computation based on the distortion model of the returned wave.

temporally amplitude-modulated light and observe the amplitude and phase shift of the returned wave. We repeat observations at various illumination frequencies. From these observations, we analyze the scene to recover invisible information based on the model about the PSFs and modulated illumination.

In this paper, we extract two works from the Ph.D. thesis [65]. In Section 3, a spatial analysis of translucent objects is discussed, and the temporal analysis is summarized in Section 4.

## 2. Related Work

We summarize active illumination works that illuminate the scene by amplitude-modulated light. They are strongly related to our work because our aim is to analyze spatial and temporal light transport using amplitude-modulated light. Related to our specific implementations, we summarize the work of separating layered translucent objects and material classification, respectively.

### Spatially Coded Illumination

Our method can be grouped in a class of active sensing techniques that use high-frequency pattern projection. These techniques are based on the relationship between the pattern projection and the spatial spread of light. The original Nayar *et al.* [47]’s method separate direct and global components by projecting multiple high-frequency patterns, where direct components include specular and diffuse reflection and global components include inter-reflection, volume and subsurface scattering. Extending this method, Lamond *et al.* [36] separate specular and diffuse refraction, Mulaigawa *et al.* [42], [43] separate single and multiple scattering, Tanaka *et al.* [68] separate transmissive and scattered lights, and Tanaka *et al.* [66] separate upper and inner layers. We summarize these method in terms of what is exactly the direct components in the viewpoint of optical design in this thesis. In addition, Achar *et al.* [2], [3] overcome the problem of motion and defocus of high-frequency illumination by total variation optimization and changing projector’s focus while projection. Reddy *et al.* [54] separate light transport into direct, near-indirect, and far-indirect rays by frequency-domain modeling and analysis. There are other pattern projection techniques to decompose light transport. Gupta *et al.* [20] acquire scene depths with direct-global separation by modeling both projector’s defocus and global light transport. O’Toole *et al.* [48], [50], [51] illuminate

a scene by a pattern while masking the camera by the complementary pattern to spatially probing the light transport of the target scene. Our methods is also based upon Nayar’s method [47] and we use a relationship among direct components, the size of scattering blur, and the pitch of the projection pattern to separate depth layers. We use the pattern projection not only for separating scattering effects, but also for recovering appearance of multiple inner slices.

Our work is also related to imaging through scattering/occlusion methods in computational photography. For imaging through scattering media, Narasimhan *et al.* [46]’s and Gu *et al.* [18]’s methods sharpen images of a target scene in muddy liquid by precisely modeling single scattering. Their methods work well for those scenes that do not exhibit multiple scattering. Differently from these works, our method recovers images inside an optically thick translucent object, where significant multiple scattering is observed and the optical thickness of the target is unknown.

### Temporally Coded Illumination

A time domain impulse response of the scene, as known as light-in-flight and transient imaging, can be obtained using an interferometer [15], holography [1], [32], and femtosecond-pulsed laser [35], [71], [72]. The time domain impulse response can be also recovered using the ToF camera. 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 [23], [38], [52] and optical coding [30], [49], while it requires customization of a ToF camera. These measurement methods may be able to be applied to the task of material classification, 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 [10], [13], [17], [28], parametric model [25], [34],  $K$ -sparsity and optimization [6], [12], [53], stereo ToF cameras [37], using external projector [44], and frequency sweep [29]. Instead of recovering the correct depth, we use a distorted depth as a cue for the material classification.

There are other scene analysis methods using ToF cameras, *e.g.*, recovering the shape of transparent and translucent objects [60], [67], and measuring a slice of BRDF [45]. In addition, computational imaging methods using the ToF camera, such as imaging around the corner [24], [31], separating direct and indirect light transport [19], [49], [74], imaging the velocity of the object [22], [62], and imaging at a specific depth [64] are proposed. Our method can also be considered one of the scene analysis methods as it aims at material classification of the scene.

### Separating Layers from Superposed Images

There have been independent developments of technologies for

imaging internal structures of target objects for special purposes. In art analysis, several techniques have been developed for imaging hidden layers of paintings. Infrared reflectography [14] and X-ray transmission radiography [9] have been used for visualizing internal layers of paintings, although the surface texture cannot be separated. X-ray fluorescence technique [9] uses spectroscopic information measured over a couple of days and estimates the metallic atom distribution for determining colored appearance of inner layers. Tera-hertz imaging [4], [55] is another technique that can see inner layers of paints and books. In the medical imaging and its related areas, optical coherence tomography [11], [26] techniques are widely used, especially for visualizing retina. These techniques enable to observe inside translucent objects based on interference of coherent light at the cost of expensive wavelength-order optics and mechanics. In contrast, our method uses a commodity camera and projector for recovering slices inside translucent objects, which allows low-cost implementation.

Time domain coding for analyzing light transport is another approach to recover images that are not directly measurable. Heide *et al.* [25] sweep the modulation frequency and phase of their customized Time-of-Flight (ToF) camera to recover the light propagation inside scattering medium. Kadambi *et al.* [30] build a coded-illumination ToF camera with a deconvolution technique and use it for recovering a sharp image by observing through a diffuser. O'Toole *et al.* [49] combine spatial probing and ToF imaging to separate direct and indirect light-in-flight images. Tadano *et al.* [64] propose an imaging system that is capable to select a target depth using a coded ToF camera. While these methods effectively recover light transport, they require carefully tailored ToF cameras. Contrary, our method uses a simple projector-camera system and spatial pattern coding to analyze light transport inside translucent objects.

### Material Estimation

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 [5], [7], [39], [58], [59], [70], [73], [77], 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 [40], [76], shading [41], and spectrum [57], 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 [8], and water permutation and heating/cooling process [56]. 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 ob-

jects. 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 [63] 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.

### 3. Analysis of Spatially Spread Lights and Multi-frequency Illumination

Translucent objects have complex appearance. It is a superposition of light rays emitted from inner slices at every depths, blurred by subsurface scattering. Because seeing internal appearance of objects is of broad interest in medical and art analyses and industry inspection, various imaging techniques have been developed in the past. In particular, since the translucency effect becomes significant for many materials in near infrared (NIR) wavelengths, infrared photography is used as one of common techniques for achieving this goal. For example, it is used for observing inner layers of oil paintings that tell us the drawing technique, growth, history, and/or authenticity of old age painters.

One of the major challenges in observing inner layers of translucent objects is to separate inner appearances with properly dealing with scattering. To overcome this problem, we develop a *multi-frequency illumination* method, which can recover sharp appearance of inner slices at a desired depth with explicitly removing scattering blur. Compared with conventional techniques that aim at a similar goal, our method is faster and safer than the X-ray fluorescence technique [9], and sharper results can be obtained differently from infrared reflectography [14].

Our method exploits the fact that the spread of light due to scattering has dependency on the depth of inner layer where light rays are emitted. By modeling the light spreads as depth-dependent point spread functions (PSFs) and utilizing their depth-dependent low-pass characteristics, we develop a method for recovering inner layer appearances from a set of images taken under a variant of high-frequency illumination [47]. Specifically, our method uses a spatial pattern projection with varying the pattern pitch – we call this *multi-frequency illumination*. Our multi-frequency illumination method allows us to separate direct (high-frequency) and global (low-frequency) components as in [47], yet at various frequency levels that define high- and low-frequencies. Our method uses the direct component observations for recovering the appearance of inner slices, which are related to the direct components via depth-dependent PSFs.

The key contributions of this section are twofold. First, we describe the relationship between depth inside a translucent object and its PSF by a physically motivated scattering model. Second, using the relationship, we develop a method for recovering the appearance of inner slices using varying pitch pattern projection.

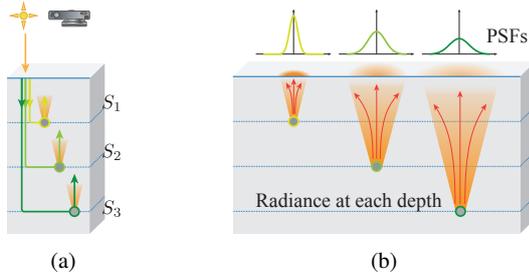


Fig. 3: Illustration of the image formation model for translucent objects. (a) Recorded intensity is the summation of all layer's appearance. (b) Spatial spread of light varies with depth.

We implement a coaxial projector-camera setup for measuring real-world objects, and show the recovery results using oil painting.

### 3.1 Appearance of Translucent Objects

When an image of a translucent layered object is recorded, the observed intensity can be modeled as a summation of the appearance of multiple depth layers as illustrated in Fig. 3(a). Let  $S_d$  be the appearance slice of the layer at depth  $d$ . The observed intensity  $L_o(q)$  at camera pixel  $q$  can be expressed by

$$L_o(q) = \sum_d S_d(q) \quad (1)$$

in a discrete form.

The appearance slice  $S_d$  is generally blurry due to the scattering effect inside the medium. The spread of radiance at a scene point inside a translucent object varies depending on its depth from the object surface [61]. In general, the spatial spread of light can be expressed using PSFs. Let us consider light rays emitted (or returned) from a specific depth inside a translucent object. When the depth is shallower, the PSF becomes sharper. On the other hand, it gradually spreads as the depth  $d$  becomes deeper inside the medium as illustrated in Fig. 3(b). In this manner, there is a close relationship between the PSF and depth. By denoting  $h_d$  as a PSF at depth  $d$ , the appearance slice  $S_d$  at depth  $d$  can be expressed as

$$S_d(q) = (R_d * h_d)(q), \quad (2)$$

where  $R_d$  is the sharp slice that we are interested in estimating, which we call a radiance slice, and  $*$  denotes a convolution operator.

Since the appearance of the translucent object under normal illumination is a superposition of radiance of multiple layers as Eq. (1), the observation can be re-written as

$$L_o(q) = \sum_d (R_d * h_d)(q). \quad (3)$$

Our goal is to recover radiance slices  $R_d$  from the composite observation  $L_o$ . Before introducing the solution method, we describe a model of depth-dependent PSFs  $h_d$ .

### 3.2 Proposed Method

We are interested in recovering radiance slices  $R_d$  from the

mixed observation  $L_o$ . To achieve this goal, we develop a *multi-frequency illumination* measurement method, which is built upon the high-frequency illumination (HFI) method proposed by Nayar *et al.* [47]. To begin with, we briefly review the original HFI method.

#### High-frequency illumination method [47]

The HFI method separates direct and global components by projecting small pitch checker patterns. When the phase of the projection pattern changes slightly, the direct component  $D(q)$  varies accordingly, but the global component  $G(q)$  remains stable. Based on this observation, their method computes direct and global components using the maximum  $L_{\max}(q)$  and minimum  $L_{\min}(q)$  intensities that are obtained by shifting the projector pattern as

$$\begin{cases} D(q) = L_{\max}(q) - L_{\min}(q), \\ G(q) = 2L_{\min}(q). \end{cases} \quad (4)$$

The direct component  $D(q)$  contains high-frequency components, while the global component  $G(q)$  contains only lower frequency components than the frequency of projection pattern. Therefore, the HFI method can be viewed as a separation technique for high- and low-frequency components.

#### Pattern pitch of HFI

In our case, when a translucent object is measured under HFI with pattern pitch  $p$ , we can obtain direct component  $D(p, q)$  and global component  $G(p, q)$  as

$$\begin{cases} D(p, q) = \sum_d D_d(p, q) \\ G(p, q) = \sum_d G_d(p, q) \\ S_d(q) = D_d(p, q) + G_d(p, q), \end{cases} \quad (5)$$

where  $D_d(p, q)$  and  $G_d(p, q)$  are the direct (high-frequency) and global (low-frequency) components at depth  $d$ , respectively, and the sum of direct and global components for each depth becomes the radiance slice. As mentioned in [47], the pattern pitch  $p$  must be sufficiently smaller than the scene texture for a faithful separation.

#### 3.2.1 Multi-frequency Illumination

By measuring the target object using multi-frequency patterns, multiple of corresponding direct components are obtained. Unfortunately, increasing the number of measurements does not make the problem easier as it also increases the number of variables to solve for. To make the problem tractable, we assume that the texture of direct components does not vary drastically when the pattern frequency is high enough and the pitch variation ( $p_v - p_u$ ) is sufficiently small. These direct components  $D_d$  at a certain depth  $d$  are supposed to have a similar texture with the original radiance; therefore, we can expect the following relationship:

$$D_d(p, q) \approx \alpha(h_d, p)R_d(q), \quad (6)$$

where  $\alpha(h_d, p)$  is the relative brightness of  $D_d(p, q)$  to  $R_d(q)$ . We call  $\alpha(h_d, p)$  the *direct component ratio* that represents the ratio of direct component's mean intensity to the radiance  $R_d(q)$ 's mean intensity. Hence, Eq. (5) can be rewritten as

$$D(p, q) = \sum_d \alpha(h_d, p)R_d(q). \quad (7)$$

These assumptions are based on the fact that the diffuse reflection and subsurface scattering can be regarded as the same physical phenomena [21], [27]; the light scatters on or beneath the surface and eventually bounces off of the material in random directions. Direct components  $D_d$  represent total intensities of lights from all the points inside the object whose distance from the incident point on the surface is smaller than the pattern pitch  $p$ . Hence the pattern pitch  $p$  controls the *scale* of the separation of scattered lights in the direct-global separation scheme, and thus controls the intensity of direct components. Furthermore, because  $p$  is sufficiently smaller than the scene texture, the texture of direct components and the original texture are largely similar. Based on these observations, we obtain the original texture at different brightnesses by changing  $p$ .

With these assumptions, a set of direct component images  $D(p, q)$  taken under the multi-frequency illumination of  $m$  pitch variations ( $p = p_1, p_2, \dots, p_m$ ) can be written in a matrix form as

$$D(q) = AR(q), \quad (8)$$

where

$$D(q) = \begin{bmatrix} D(p_1, q) & D(p_2, q) & \cdots & D(p_m, q) \end{bmatrix}^T, \\ A = \begin{bmatrix} \alpha(h_{d_1}, p_1) & \cdots & \alpha(h_{d_n}, p_1) \\ \vdots & \ddots & \vdots \\ \alpha(h_{d_1}, p_m) & \cdots & \alpha(h_{d_n}, p_m) \end{bmatrix}, \\ R(q) = \begin{bmatrix} R_{d_1}(q) & R_{d_2}(q) & \cdots & R_{d_n}(q) \end{bmatrix}^T.$$

Here,  $D \in \mathbb{R}^m$  is a vector of direct components measured under  $m$  variations of the pattern pitches at pixel  $q$ ,  $R \in \mathbb{R}^n$  is a vector of  $n$  layers of radiance slices, and  $A \in \mathbb{R}^{m \times n}$  is a matrix containing direct component ratios computed from the projected pattern pitch and the depth-dependent PSF.

When the number of projected patterns  $m$  is no less than the number of depth layers  $n$  ( $m \geq n$ ) and  $\text{rank}(A) = n$ , the radiance slices  $R$  can be obtained by a norm approximation of the residual vector, *i.e.*,  $D(q) - AR(q)$ . For example, with a least-squares approximation, the radiance slices  $R(q)$  can be determined using the pseudo-inverse  $A^+$  as

$$R(q) = A^+ D(q). \quad (9)$$

### 3.2.2 Estimation of Informative Slices

Once we know the target depths (or PSFs) to inspect, we can set up a matrix  $A$ , and thus can recover slices corresponding to the depths using Eq. (9). However, such a prior knowledge is difficult to obtain before measurement; therefore, automatically selecting a good set of depths becomes important for recovering *informative* slices. For example, if an arbitrary depth is chosen, it has a chance to correspond to the middle of distinct texture layers. To recover informative slices, we use a two-step approach. The first step is the estimation of a set of informative depths via optimization. This is equivalent to selecting a small number of useful PSFs from many other possible PSFs. The second step is the recovery of slices at the informative depths determined by the earlier step.

### Step 1: Estimation of informative depths

Frequently, there are only a small number of informative slices inside translucent objects. Hence we can regard such radiance slices exist sparsely along depth. Our method uses this sparsity to determine the informative slices by solving a  $l_1$  regularized problem (as known as the lasso [69]) with a non-negative constraint about  $R$ :

$$\hat{R}(q) = \underset{R(q)}{\operatorname{argmin}} \|AR(q) - D(q)\|_2^2 + \lambda \|R(q)\|_1 \quad (10) \\ \text{subject to } R \succeq 0.$$

We can regard the depth  $d (= \frac{T}{\sigma})$ , where  $\hat{R}_d(q)$  has a non-zero value, is informative while others are not. Equation (10) becomes a quadratic programming (QP) problem and thus can be efficiently solved in a polynomial time. We solve the optimization in a per-pixel manner. Solving a similar problem for the entire image at a time instead of computing in a pixel-wise manner is also a viable option; however, we have observed that they do not make much difference because of the following Step 2. Therefore, for efficient parallelization, we choose the per-pixel implementation.

### Step 2: Informative slice recovery

This step determines informative depth slices of the whole image by consolidating all the pixel-wise selections.

The informative depths  $\hat{d}$  are local maximas of the sum of  $l_0$  norm of  $\hat{R}_d(q)$  for all pixels. The evaluation function  $f(d)$  is defined as

$$f(d) = \sum_q \left\| \hat{R}_d(q) \right\|_0, \quad (11)$$

and we find all local maximas of  $f(d)$  along  $d$  in the range of interest. Once the depths of interest are selected, we can set up a small matrix  $\hat{A}$ . Finally, the appearances of informative slices are recovered using the matrix  $\hat{A}$  using the least-squares approximation as

$$R(q) = \hat{A}^+ D(q) \quad (12)$$

in the same manner to Eq. (9).

### 3.3 Experiments

We develop a coaxial projector-camera setup for realizing the measurement setup as shown in Fig. 4. The coaxial setup has a favorable property; correspondence between projector and camera pixels becomes invariant with respect to depths. Unlike non-coaxial settings, with which a illumination ray inside the translucent object forms a line in the image coordinates [68], the coaxial setting allows us to easily separate the direct rays. We use a LightCommander projector, which is a DMD projector development kit by Texas Instruments, and use near infrared (NIR) light for measurements. The lenses of both camera and projector are set equivalent (Ai Micro-Nikkor 105mm f/2.8S) for making the alignment easy. In the experiment, we use 18 variations of checker patterns (3px to 20px with 1px interval), and shift the pattern for one-third of square size for each pattern.

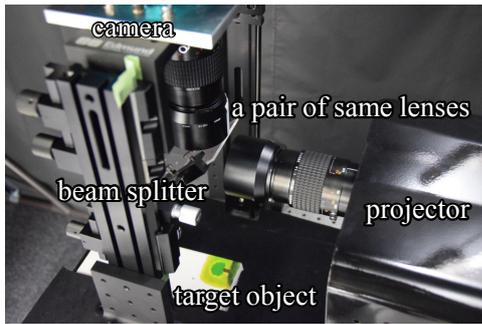


Fig. 4: Measurement setup. The coaxial system allows us to maintain the correspondences between projector and camera pixels.

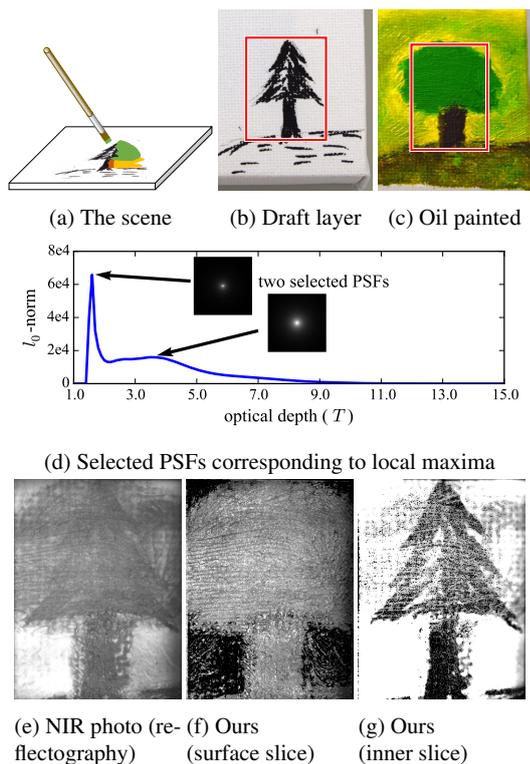


Fig. 5: Experimental result of oil painting using the baseline method. (a) Target scene. We draw a colored round tree on top of the draft of spiny tree. (b) Inner layer (draft) of the painting. (c) Painted scene. Red rectangle region is measured. (e) Normal photo using infrared light. (d) Selected PSFs. There are two peaks in the plot, hence two corresponding PSFs are selected to recover. (f, g) Results of our method. Layer of surface texture and hidden drawing, respectively. Range of the intensities is adjusted for visualization.

### Experimental results

First, we use an oil painting as a target scene as shown in Fig. 5(a), which has draft and surface pigment layers as depicted in Figs. 5(b) and 5(c). By taking a standard photograph under the near infrared light, we can only vaguely observe the draft layer as shown in Fig. 5(e). Since it is the superposition of draft and surface pigment layers, it naturally results in a blurry image. By applying our method to this scene, two PSFs are estimated as depicted in Fig. 5(d) and two slices are recovered as in Figs. 5(f)

and 5(g). The upper surface layer corresponds to the surface pigment layer. Because the yellow pigment is almost transparent in the infrared wavelength, the corresponding painting regions become dark in the surface slice. The lower layer shows the inner layer, where the texture of the tree is clearly observed.

### Shadow detection from recovered inner layers

The recovered appearance of inner layers suffers from shadows caused by their upper layers. If the upper layer's transmittance is low, it casts shadow to the inner layer. Here we describe a simple technique for dealing with shadows by post-processing. To simplify the discussion, we assume that there are only two layers in the scene, *i.e.*, upper and inner layers.

The shadowed regions have two common properties: (1) The shape of shadows observed in the inner slice is similar to the texture of upper slice. (2) The shadowed region becomes darker due to low irradiance. Based on these observations, we define a shadow likelihood measure  $P_s$  for identifying shadowed regions.

The similarity  $S$  of the texture shapes in the recovered slices  $R_1$  and  $R_2$  can be obtained using the absolute value of their cross correlation as

$$S(q) = |C_{w,c}(R_1, R_2)|, \quad (13)$$

where  $C_{w,c}$  is the cross-correlation within a small window  $w$  centered at pixel location  $q$ , and values in  $R_1$  and  $R_2$  are normalized in the range of  $[0, 1]$ . Using this similarity, the shadow likelihood  $P_s$  is defined using the darkness of the shadowed region in  $R_2$  as

$$P_s(q) = S(q)(1 - R_2(q)). \quad (14)$$

It yields a likelihood score for each pixel being in shadow, and by a simple thresholding, a shadowed region can be determined.

Once the shadow regions are identified, we can use a shadow removal technique. For example, the lost information within the shadow region can be filled in by arbitrary image inpainting methods. In this paper, we use a patch-match based image inpainting method [33].

Finally, we show a recovery and shadow removal result for a three-layer scene. Figure 6(a) shows the target scene, where three printed papers are overlaid. By applying our baseline method, three slices with shadows are recovered. With shadow detection and inpainting, we obtain the recovery of three layers as shown in Figs. 6(b)-6(d). For the shadow detection of the third slice, we compute the correlation with the original second slice which contains the shadows casted by the top slice. The lowest layer only retains a limited amount of high-frequency components hence its recovery becomes slightly blurry. However, it shows the different characteristics among layers, which indicates the separability of the layers.

### 3.4 Discussions

This section described a method for recovering inner slices of translucent objects based on multi-frequency pattern projection. The proposed method is built upon the observation that the PSF inside translucent objects varies according to the depth of slices. Based on that, we have shown that inner radiance slices can be recovered by estimating PSFs using varying pitches of projection

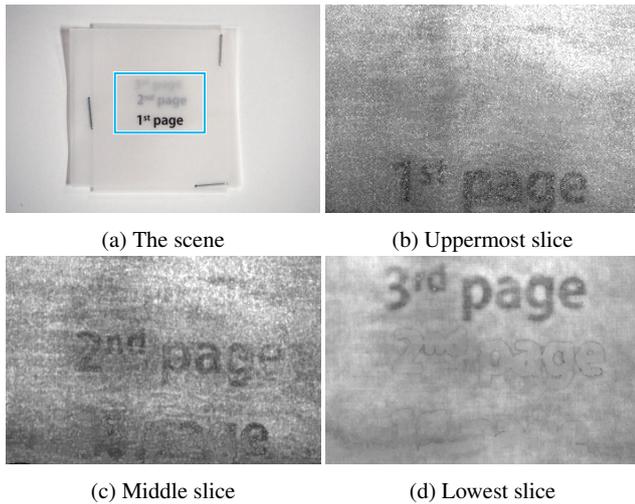


Fig. 6: The result of seeing through pages. (a) The target object. Three printed thin papers are superposed. (b) Recovered uppermost slice. Only the first page can be seen. (c) Recovered second page. Shadows from the upper layer is inpainted. (d) Recovered third page. Shadows from the first and second layers are inpainted.

patterns. We also developed a method for automatically selecting informative slices via a sparse representation, *i.e.*, determining sparse coefficients that corresponds to radiance slices. We further extended the method for dealing with inhomogeneous translucent objects based on a combination of pixel-wise appearance recovery and gradient-based image repairing. The effectiveness of the proposed method are shown by several experiments on simulation and real-world translucent objects.

#### 4. Analysis of Temporally Spread Light and Time-of-Flight Distortion

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.* [23] 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 [30], [49], [52] 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

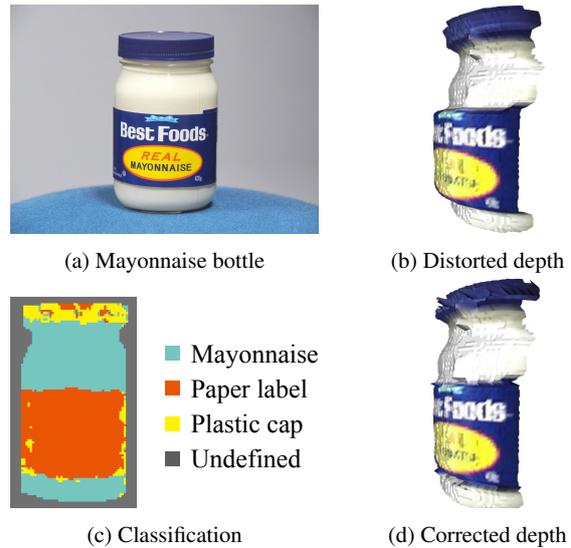


Fig. 7: Depth distortion of a ToF camera. (a) A mayonnaise bottle is measured by a Kinect. (b) Measured depth in a 3D view. There is a gap in depth between the mayonnaise and label regions. We use this depth distortion for material classification. (c) Material segmentation result. The material label 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.

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. 7(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 section 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 [23], [30]. 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.

##### 4.1 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 the 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 ob-

tained.

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 [19], [23], [30]; hence, we briefly explain one of the models that we are going to use in this paper. Following a phasor representation [19], amplitude and phase of the returned wave can be represented by a single complex value  $c \in \mathbb{C}$ , called phasor, governed by the modulation frequency  $\omega$ . The measured amplitude  $\tilde{a}_\omega \in \mathbb{R}$  and depth  $\tilde{d}_\omega \in \mathbb{R}$  of the ToF camera are obtained as

$$\begin{cases} \tilde{d}_\omega &= \frac{c}{4\pi\omega} \arg c(\omega), \\ \tilde{a}_\omega &= |c(\omega)|, \end{cases} \quad (15)$$

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

$$c(\omega) = \int_0^\infty r(t - \tau) e^{-2\pi i \omega t} dt, \quad (16)$$

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

$$r(t) = \int_{\mathcal{P}} r_\rho \delta(|\rho| - t) d\rho, \quad (17)$$

where  $r_\rho$  is the contribution of the path  $\rho$ ,  $|\rho|$  is the time travelled along the path  $\rho$ ,  $\delta(t)$  is the Dirac delta function, and  $t = 0$  is the time when the impulse light hits the surface of the object. Figure 8(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. (16) expresses that ToF camera measures the Fourier coefficients of the impulse response at the frequency of the light modulation.

#### 4.1.1 Frequency Dependent Depth Distortion

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

$$\tilde{d}_\omega = \frac{c}{4\pi\omega} \arg \underbrace{\int_0^\infty \beta\delta(t - \tau) e^{-2\pi i \omega t} dt}_{=2\pi\omega\tau} = \frac{c\tau}{2} = d, \quad (18)$$

where  $d = \frac{c\tau}{2}$  is the ground truth depth of the object. Equation (18) represents that the accurate depth can be obtained regardless of modulation frequency  $\omega$ , 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 [74]. When the target object exhibits a temporally broad shape of the impulse response, band-pass characteristic in the frequency domain becomes unique to the object.

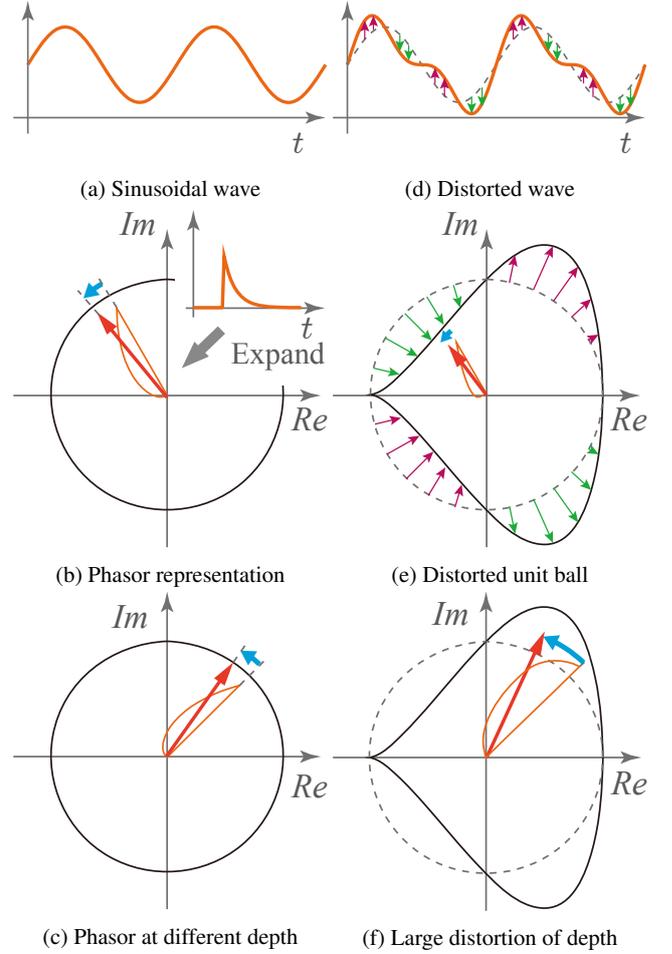


Fig. 8: Phasor representation of ToF observations. (a) Sinusoidal illumination, (b) Time domain PSF is expanded to the imaginary plane (orange). (c) When the object is placed at different depths, the observation gets rotated but phase distortion remains the same as (b). (d) Biased periodic illumination. 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).

Accordingly, ToF observation  $c(\omega)$  can take an arbitrary value, because  $c(\omega)$  is a Fourier coefficient of the impulse response  $r(t)$  at the frequency  $\omega$ . In such a case,  $\arg c(\omega)$  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  $\omega$ . 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  $\mathcal{F}[\cdot]$  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

$$\begin{aligned}\tilde{d}_\omega &= \frac{c}{4\pi\omega} \arg \left( e^{-2\pi i\omega\tau} \hat{r}(\omega) \right) \\ &= d + \frac{c}{4\pi\omega} \arg \hat{r}(\omega).\end{aligned}\quad (19)$$

The second term  $\frac{c}{4\pi\omega} \arg \hat{r}(\omega)$  is the depth distortion at frequency  $\omega$ . In Fig. 8(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 [52].

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.

#### 4.1.2 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

$$\begin{aligned}\tilde{d}'_\omega &= \frac{c}{4\pi\omega} \arg \left( e^{-2\pi i\omega(\tau + \Delta\tau)} \hat{r}(\omega) \right) \\ &= d + \Delta d + \frac{c}{4\pi\omega} \arg \hat{r}(\omega).\end{aligned}\quad (20)$$

The measured depth is just shifted by  $\Delta d$ , 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. (19). Figure 8(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. 8(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 [16], [75]. When the illumination wave has harmonics components as shown in Fig. 8(d), the ToF observation exhibits depth-dependency as illustrated in Figs. 8(e) and 8(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

$$c(\omega) = \int_0^\infty r(t - \tau) b_\omega(2\pi\omega t) e^{-2\pi i\omega t} dt. \quad (21)$$

The above indicates that the observation  $c(\omega)$  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  $\tau$ , *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.* [63] 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.

## 4.2 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  $a_{\omega_i, d_j}$  can be obtained by

$$a_{\omega_i, d_j} = d_j - \tilde{d}_{\omega_i, j}, \quad (22)$$

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

$$\mathbf{v} = [a_{\omega_1, d_1} \quad \dots \quad a_{\omega_n, d_m}]^T. \quad (23)$$

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  $a'_{\omega_i, d_j}$  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

$$a'_{\omega_i, d_j} = a_{\omega_i, d_j} - a_{\omega_n, d_j} = \tilde{d}_{\omega_n, j} - \tilde{d}_{\omega_i, j}, \quad (24)$$

where  $i$  ranges from 1 to  $n - 1$ . We can then setup a  $m(n - 1)$ -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  $a''_{\omega_1, d_j}$  can be obtained by regarding the measurement of the  $m$ -th depth position as the reference measurement as

$$a''_{\omega_1, d_j} = a_{\omega_1, d_j} - a_{\omega_1, d_m} = \tilde{d}_{\omega_1, m} - \tilde{d}_{\omega_1, j} + \Delta d_j, \quad (25)$$

where  $\Delta d_j$  is the amount of movement from the base position, which should be measured.

### 4.2.1 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 ( $\ell_2$  norm). To deal with the missing or uninformative saturated observations, we remove such elements in the feature vector when evaluating the distance. The distance  $\xi_p$  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

$$\xi_p = \frac{1}{N} \sum_{k=0}^{nm} \begin{cases} 0 & v_k = \text{N/A} \\ (v_k - v_k^p)^2 & \text{otherwise,} \end{cases} \quad (26)$$

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

$$\hat{p} = \underset{p}{\operatorname{argmin}} \xi_p.$$

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.

### 4.3 Experiments

We evaluate the proposed method by a ToF camera and a linear translation stage system as shown in Fig. 9. 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`<sup>\*1</sup> to obtain such data.

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 10 shows the depth distortion of

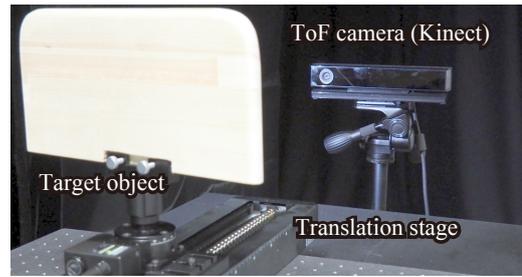


Fig. 9: Experimental setup. We use Kinect as a ToF camera, and the target object is placed on a linear translation stage.

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.

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 as shown in Fig. 11, 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. (25)), the accuracy is improved to 81.6%. Finally, with both of frequency- and depth-dependent distortions ( $n = 3$  and  $m = 2600$ ), the accuracy is further improved to 90.5%. The confusion matrices are shown in Fig. 12. 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 13 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

<sup>\*1</sup> <https://github.com/OpenKinect/libfreenect2>

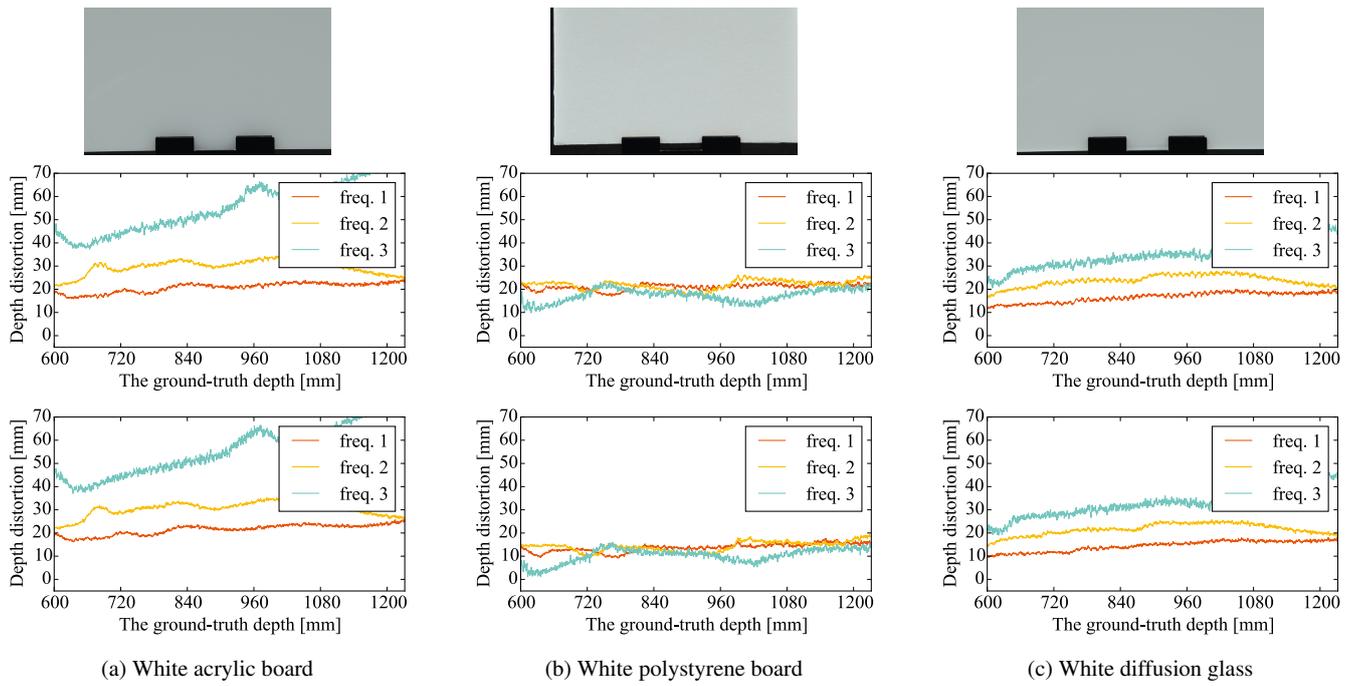


Fig. 10: Measured depth distortions using Kinect for three objects. The ground truth depth is obtained via a linear translation stage. The top row shows photographs of the target objects. Measurements of the second and third rows are different in terms of surface orientation. Depth distortion of each 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.

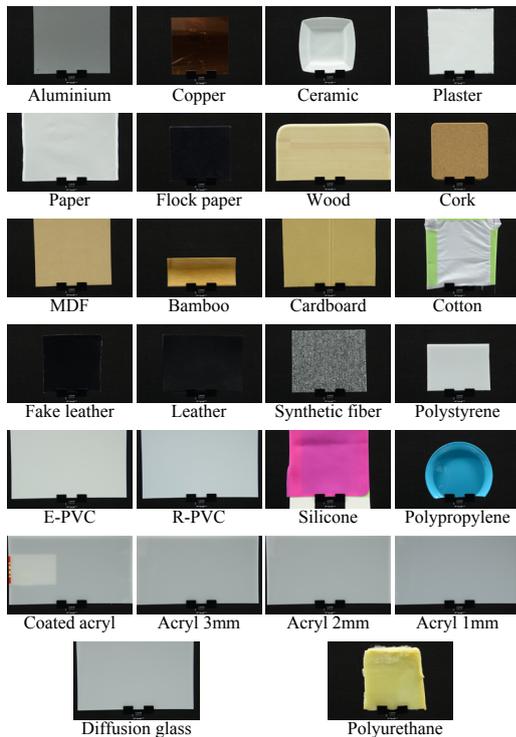


Fig. 11: All materials of our database. All images are captured by the same camera parameters *e.g.* ISO, f-number, shutter speed, and focal length.

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. 14(a). The distances of feature vectors between the folded and flat cardboards are plotted in the blue line in Fig. 14(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 15 shows a couple of example of material segmentation. For the scene shown in Fig. 15(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. 15(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 15(c) shows another scene where wallets made of genuine and fake leather are placed, and they are correctly classified as shown in Fig. 15(d).

### Depth correction

Once materials are classified, the distorted depths can be corrected for recovering an accurate 3D shape using the material

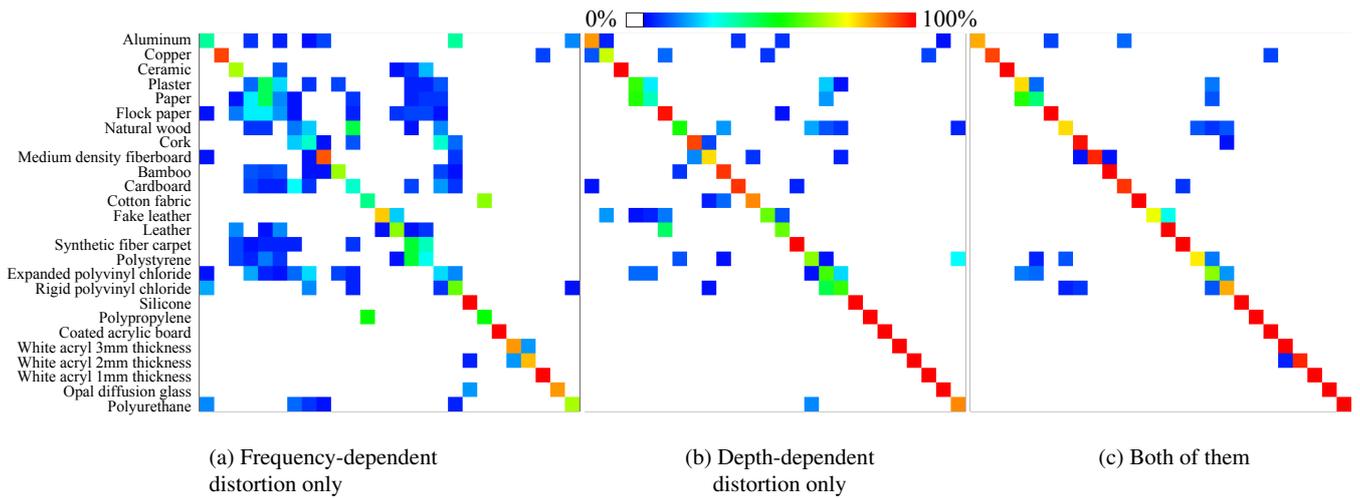


Fig. 12: Confusion matrix. Red indicates the higher value and it appears on the diagonal. (a) Classified result only using frequency-dependent distortion. Accuracy is 57.4%. (b) Classified result only using depth-dependent distortion. Accuracy is 81.6%. (c) Classified result using both of them. Overall accuracy is 90.5%.

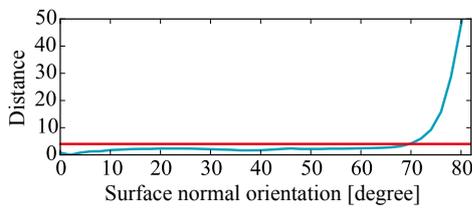
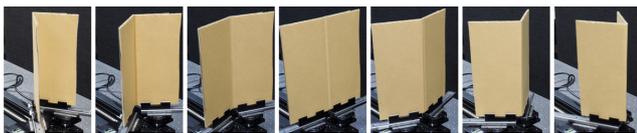
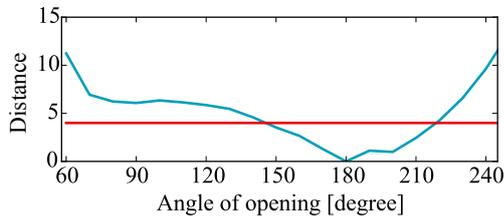


Fig. 13: Feature vector variation over surface orientation. We change the orientation of the target object, and plot the distance of features along with the orientation. The feature is stable under around 70 degrees, and shows large deviation at steep-slant orientation. Red line indicates the upper-bound distance for the correct classification.



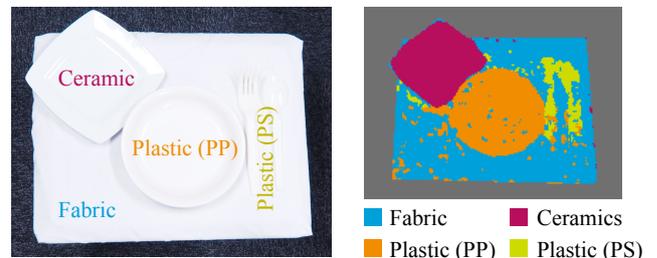
(a) Scenes of the folded cardboard



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

Fig. 14: 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.

database that contains the samples of distortions for all materials. An example of the depth correction is shown in Fig. 7(a). Because mayonnaise has significant subsurface scattering, the measured depth of mayonnaise region is strongly distorted than that

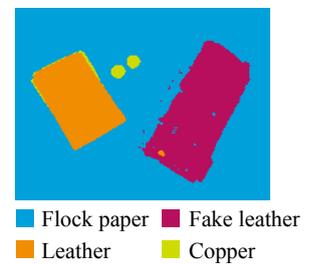


(a) The scene of white utensils

(b) Classification result



(c) The scene of wallets



(d) Classification result

Fig. 15: 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.

of the label as shown in Fig. 7(b). Figure 7(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 of the mayonnaise bot-

tle is recovered as shown in Fig. 7(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. (26), 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.

#### 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. 12. 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.

#### 4.4 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 complex shapes 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.

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.

## 5. Conclusion

We established model-based approaches for measuring translucent and transparent objects using spatially and temporally modulated light. Amplitude-modulated illumination techniques convey rich information of the PSFs, and appropriate image processing methods can recover informative images from degraded observations. We developed high-frequency illumination and image processing methods based on the model of the spatial spread of light. We also developed a method that processes ToF distortions based on the model of the temporal spread of light.

Our aim was to avoid redundant measurement of PSFs. Instead we used the frequency responses of PSFs under amplitude-modulated illumination at different frequencies. Specifically, we analyzed the relationship between the spatial PSFs and the frequencies of projection patterns, and proposed a model of frequency- and depth-dependent ToF distortion for material classification. Based on calculations using the relationship between the PSF and the frequency responses, we recovered invisible information from the scene.

### 5.1 Limitations and Future Work

Our techniques are designed for each specific purpose, limiting the eligible target scenes. This is because our optical designs and computational algorithms are based on the specialized light transport models for the target applications. Different measurement and computation techniques based on the model of target's light transport must be developed, to visualize different information of the target object.

The exact cause of ToF distortions remains unclear. In the limited condition, ToF distortions are due to the material's translucency. ToF distortions may also be affected by the shape, motion, density, temperature, and other object properties. Clarifying the effects of these properties and establishing a framework for all ToF distortions is an exciting topic for future research.

Proposed techniques include dealing with the behavior of light transport separately. Concurrently analyzing the spatio-temporal spread of light will provide more detailed and accurate analysis of the scene. Moreover, angular spread, spectral spread, and coherence spread, including polarization and interference, should also be included in future analysis. Developing a framework to incorporate all optical cues is another interesting topic for future research.

### References

- [1] Abramson, N.: Light-in-Flight Recording by Holography, *Optics Letters*, Vol. 3, No. 4, pp. 121–123 (1978).
- [2] Achar, S. and Narasimhan, S. G.: Multi Focus Structured Light for Recovering Scene Shape and Global Illumination, *Proc. European Conference on Computer Vision (ECCV)*, pp. 205–219 (2014).
- [3] Achar, S., Nuske, S. T. and Narasimhan, S. G.: Compensating for Motion During Direct-Global Separation, *Proc. International Conference on Computer Vision (ICCV)*, pp. 1481–1488 (2013).
- [4] Adam, A. J. L., Planken, P. C. M., Meloni, S. and Dik, J.: Terahertz Imaging of Hidden Paint Layers on Canvas, *Optics Express*, Vol. 17, No. 5, pp. 3407–3416 (2009).
- [5] Bell, S., Upchurch, P., Snively, N. and Bala, K.: Material Recognition in the Wild with the Materials in Context Database, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3479–3487 (2015).

- [6] Bhandari, A., Feigin, M., Izadi, S., Rhemann, C., Schmidt, M. and Raskar, R.: Resolving Multipath Interference in Kinect: an Inverse Problem Approach, *IEEE SENSORS*, IEEE, pp. 614–617 (2014).
- [7] Caputo, B., Hayman, E. and Mallikarjuna, P.: Class-Specific Material Categorisation, *Proc. International Conference on Computer Vision (ICCV)*, IEEE, pp. 1597–1604 (2005).
- [8] Davis, A., Bouman, K. L., Chen, J. G., Rubinstein, M., Durand, F. and Freeman, W. T.: Visual Vibrometry: Estimating Material Properties from Small Motion in Video, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5335–5343 (2015).
- [9] Dik, J., Janssens, K., Van der Snickt, G., van der Loeff, L., Rickers, K. and Cotte, M.: Visualization of a Lost Painting by Vincent Van Gogh using Synchrotron Radiation Based X-Ray Fluorescence Elemental Mapping., *Analytical chemistry*, Vol. 80, No. 16, pp. 6436–42 (2008).
- [10] Dorrington, A. A., Godbaz, J. P., Cree, M. J., Payne, A. D. and Streeter, L. V.: Separating True Range Measurements from Multipath and Scattering Interference in Commercial Range Cameras, *SPIE 7864, Three-Dimensional Imaging, Interaction, and Measurement*, p. 786404 (2011).
- [11] Fercher, A. F.: Optical Coherence Tomography., *Journal of biomedical optics*, Vol. 1, No. 2, pp. 157–73 (1996).
- [12] Freedman, D., Krupka, E., Smolin, Y., Leichter, I. and Schmidt, M.: SRA: Fast Removal of General Multipath for ToF Sensors, *Proc. European Conference on Computer Vision (ECCV)*, pp. 234–249 (2014).
- [13] Fuchs, S.: Multipath Interference Compensation in Time-of-Flight Camera Images, *Proc. IAPR International Conference on Pattern Recognition*, IEEE, pp. 3583–3586 (2010).
- [14] Gavrilov, D., Maeva, E., Grube, O., Maldague, X. and Maev, R.: Infrared Methods in Noninvasive Inspection of Artwork, *International Conference on NDT of Art*, pp. 1–5 (2008).
- [15] Gkioulekas, I., Levin, A., Durand, F. and Zickler, T.: Micron-Scale Light Transport Decomposition using Interferometry, *ACM Transactions on Graphics (ToG)*, Vol. 34, No. 4, pp. 37:1–37:14 (2015).
- [16] Godbaz, J. P., Dorrington, A. and Cree, M. J.: Understanding and Ameliorating Mixed Pixels and Multipath Interference in AMCW Lidar, *TOF Range-Imaging Cameras.*, pp. 91–116 (2013).
- [17] Godbaz, J. P., Cree, M. J. and Dorrington, A. A.: Closed-Form Inverses for the Mixed Pixel/Multipath Interference Problem in AMCW Lidar, *SPIE 8296, Computational Imaging X*, p. 909778 (2012).
- [18] Gu, J., Nayar, S. K., Grinspun, E., Belhumeur, P. and Ramamoorthi, R.: Compressive Structured Light for Recovering Inhomogeneous Participating Media, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 35, No. 3, pp. 555–567 (2013).
- [19] Gupta, M., Nayar, S. K., Hullin, M. B. and Martin, J.: Phasor Imaging: a Generalization of Correlation-Based Time-of-Flight Imaging, *ACM Transactions on Graphics (ToG)*, Vol. 34, No. 5, pp. 156:1–156:18 (2015).
- [20] Gupta, M., Tian, Y., Narasimhan, S. G. and Zhang, L.: A Combined Theory of Defocused Illumination and Global Light Transport, *International Journal of Computer Vision (IJCV)*, Vol. 98, No. 2, pp. 146–167 (2011).
- [21] Hanrahan, P. and Krueger, W.: Reflection from Layered Surfaces Due to Subsurface Scattering, *Proc. ACM SIGGRAPH*, ACM Press, pp. 165–174 (1993).
- [22] Heide, F., Heidrich, W., Hullin, M. and Wetzstein, G.: Doppler Time-of-Flight Imaging, *ACM Transactions on Graphics (ToG)*, Vol. 34, No. 4, pp. 36:1–36:11 (2015).
- [23] Heide, F., Hullin, M. B., Gregson, J. and Heidrich, W.: Low-Budget Transient Imaging using Photonic Mixer Devices, *ACM Transactions on Graphics (ToG)*, Vol. 32, No. 4, pp. 45:1–45:10 (2013).
- [24] Heide, F., Xiao, L., Heidrich, W. and Hullin, M. B.: Diffuse Mirrors: 3D Reconstruction from Diffuse Indirect Illumination using Inexpensive Time-of-Flight Sensors, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3222–3229 (2014).
- [25] Heide, F., Xiao, L., Kolb, A., Hullin, M. B. and Heidrich, W.: Imaging in Scattering Media using Correlation Image Sensors and Sparse Convolutional Coding., *Optics express*, Vol. 22, No. 21, pp. 26338–26350 (2014).
- [26] Huang, D., Swanson, E., Lin, C., Schuman, J., Stinson, W., Chang, W., Hee, M., Flotte, T., Gregory, K., Puliafito, C. and Et, A.: Optical Coherence Tomography, *Science*, Vol. 254, No. 5035, pp. 1178–1181 (1991).
- [27] Jensen, H. W., Marschner, S. R., Levoy, M. and Hanrahan, P.: A Practical Model for Subsurface Light Transport, *Proc. ACM SIGGRAPH*, ACM Press, pp. 511–518 (2001).
- [28] Jimenez, D., Pizarro, D., Mazo, M. and Palazuelos, S.: Modelling and Correction of Multipath Interference in Time of Flight Cameras, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 893–900 (2012).
- [29] Kadambi, A., Schiel, J. and Raskar, R.: Macroscopic Interferometry: Rethinking Depth Estimation with Frequency-Domain Time-of-Flight, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 893–902 (2016).
- [30] Kadambi, A., Whyte, R., Bhandari, A., Streeter, L., Barsi, C., Dorrington, A. and Raskar, R.: Coded Time of Flight Cameras: Sparse Deconvolution to Address Multipath Interference and Recover Time Profiles, *ACM Transactions on Graphics (ToG)*, Vol. 32, No. 6, pp. 167:1–167:10 (2013).
- [31] Kadambi, A., Zhao, H., Shi, B. and Raskar, R.: Occluded Imaging with Time-of-Flight Sensors, *ACM Transactions on Graphics (ToG)*, Vol. 35, No. 2, pp. 15:1–15:12 (2016).
- [32] Kakue, T., Tosa, K., Yuasa, J., Tahara, T., Awatsuji, Y., Nishio, K., Ura, S. and Kubota, T.: Digital Light-in-Flight Recording by Holography by Use of a Femtosecond Pulsed Laser, *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 18, No. 1, pp. 479–485 (2012).
- [33] Kawai, N. and Yokoya, N.: Image inpainting Considering Symmetric Patterns, *Proc. IAPR International Conference on Pattern Recognition*, pp. 2744–2747 (2012).
- [34] Kirmani, A., Benedetti, A. and Chou, P. A.: SPUMIC: Simultaneous Phase Unwrapping and Multipath Interference Cancellation in Time-of-Flight Cameras using Spectral Methods, *IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, pp. 1–6 (2013).
- [35] Kirmani, A., Hutchison, T., Davis, J. and Raskar, R.: Looking Around the Corner using Ultrafast Transient Imaging, *International Journal of Computer Vision (IJCV)*, Vol. 95, No. 1, pp. 13–28 (2011).
- [36] Lamond, B., Peers, P. and Debevec, P.: Fast Image-Based Separation of Diffuse and Specular Reflections, *Proc. ACM SIGGRAPH Sketches*, p. 74 (2007).
- [37] Lee, S. and Shim, H.: Skewed Stereo Time-of-Flight Camera for Translucent Object Imaging, *Image and Vision Computing*, Vol. 43, No. C, pp. 27–38 (2015).
- [38] Lin, J., Liu, Y., Hullin, M. B. and Dai, Q.: Fourier Analysis on Transient Imaging with a Multifrequency Time-of-Flight Camera, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 3230–3237 (2014).
- [39] Liu, C., Sharan, L., Adelson, E. H. and Rosenholtz, R.: Exploring Features in a Bayesian Framework for Material Recognition, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 239–246 (2010).
- [40] Liu, C. and Gu, J.: Discriminative Illumination: Per-Pixel Classification of Raw Materials Based on Optimal Projections of Spectral BRDF, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 36, No. 1, pp. 86–98 (2014).
- [41] Mannan, M. A., Das, D., Kobayashi, Y. and Kuno, Y.: Object Material Classification by Surface Reflection Analysis with a Time-of-Flight Range Sensor, *Advances in Visual Computing*, Springer Berlin Heidelberg, pp. 439–448 (2010).
- [42] Mukaigawa, Y., Raskar, R. and Yagi, Y.: Analysis of Light Transport in Scattering Media, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 153–160 (2010).
- [43] Mukaigawa, Y., Raskar, R. and Yagi, Y.: Analysis of Scattering Light Transport in Translucent Media, *IPSJ Transaction on Computer Vision and Applications (CVA)*, Vol. 3, pp. 122–133 (2011).
- [44] Naik, N., Kadambi, A., Rhemann, C., Izadi, S., Raskar, R. and Bing Kang, S.: A Light Transport Model for Mitigating Multipath Interference in Time-of-Flight Sensors, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 73–81 (2015).
- [45] Naik, N., Zhao, S., Velten, A., Raskar, R. and Bala, K.: Single View Reflectance Capture using Multiplexed Scattering and Time-of-Flight Imaging, *ACM Transactions on Graphics (ToG)*, Vol. 30, No. 6, pp. 171:1–171:10 (2011).
- [46] Narasimhan, S. G., Nayar, S. K. and Koppal, S.: Structured Light in Scattering Media, *Proc. International Conference on Computer Vision (ICCV)*, IEEE, pp. 420–427 (2005).
- [47] Nayar, S. K., Krishnan, G., Grossberg, M. D. and Raskar, R.: Fast Separation of Direct and Global Components of a Scene using High Frequency Illumination, *ACM Transactions on Graphics (ToG)*, Vol. 25, No. 3, pp. 935–944 (2006).
- [48] O’Toole, M., Achar, S., Narasimhan, S. G. and Kutulakos, K. N.: Homogeneous Codes for Energy-Efficient Illumination and Imaging, *ACM Transactions on Graphics (ToG)*, Vol. 34, No. 4, pp. 35:1–35:13 (2015).
- [49] O’Toole, M., Heide, F., Xiao, L., Hullin, M. B., Heidrich, W. and Kutulakos, K. N.: Temporal Frequency Probing for 5D Transient Analysis of Global Light Transport, *ACM Transactions on Graphics (ToG)*, Vol. 33, No. 4, pp. 87:1–87:11 (2014).
- [50] O’Toole, M., Mather, J. and Kutulakos, K. N.: 3D Shape and Indirect Appearance by Structured Light Transport, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 3246–3253 (2014).
- [51] O’Toole, M., Raskar, R. and Kutulakos, K. N.: Primal-Dual Coding to

- Probe Light Transport, *ACM Transactions on Graphics (ToG)*, Vol. 31, No. 4, pp. 39:1–39:11 (2012).
- [52] Peters, C., Klein, J., Hullin, M. B. and Klein, R.: Solving Trigonometric Moment Problems for Fast Transient Imaging, *Proc. ACM SIG-GRAPH Asia*, Vol. 34, No. 6, pp. 220:1–220:11 (2015).
- [53] Qiao, H., Lin, J., Liu, Y., Hullin, M. B. and Dai, Q.: Resolving Transient Time Profile in ToF Imaging via Log-Sum Sparse Regularization, *Optics letters*, Vol. 40, No. 6, pp. 918–21 (2015).
- [54] Reddy, D., Ramamoorthi, R. and Curless, B.: Frequency-Space Decomposition and Acquisition of Light Transport under Spatially Varying Illumination, *Proc. European Conference on Computer Vision (ECCV)*, pp. 596–610 (2012).
- [55] Sanchez, A. R., Heshmat, B., Aghasi, A., Zhang, M., Naqvi, S., Romberg, J. and Raskar, R.: Terahertz Time-gated Spectroscopic Imaging for Content Extraction through Layered Structures, *Nature Communications*, Vol. 7, No. 12665, pp. 1–7 (2016).
- [56] Saponaro, P., Sorensen, S., Kolagunda, A. and Kambhampettu, C.: Material Classification with Thermal Imagery, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4649–4656 (2015).
- [57] Sato, M., Yoshida, S., Olwal, A., Shi, B., Hiyama, A., Tanikawa, T., Hirose, M. and Raskar, R.: SpecTrans: Versatile Material Classification for Interaction with Textureless, Specular and Transparent Surfaces, *Proc. ACM Conference on Human Factors in Computing Systems (CHI)*, ACM Press, pp. 2191–2200 (2015).
- [58] Schwartz, G. and Nishino, K.: Visual Material Traits: Recognizing Per-Pixel Material Context, *Proc. Color and Photometry in Computer vision workshop in conjunction with the IEEE International Conference on Computer Vision*, pp. 883–890 (2013).
- [59] Schwartz, G. and Nishino, K.: Automatically Discovering Local Visual Material Attributes, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3565–3573 (2015).
- [60] Shim, H. and Lee, S.: Recovering Translucent Object using a Single Time-of-Flight Depth Camera, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 26, No. 5, pp. 841–854 (2015).
- [61] Shimizu, K., Tochio, K. and Kato, Y.: Improvement of Transcutaneous Fluorescent Images with a Depth-Dependent Point-Spread Function, *Applied Optics*, Vol. 44, No. 11, pp. 2154–2161 (2005).
- [62] Shrestha, S., Heide, F., Heidrich, W. and Wetzstein, G.: Computational Imaging with Multi-Camera Time-of-Flight Systems, *ACM Transactions on Graphics (ToG)*, Vol. 35, No. 4, pp. 33:1–33:11 (2016).
- [63] Su, S., Heide, F., Swanson, R., Klein, J., Callenberg, C., Hullin, M. and Heidrich, W.: Material Classification using Raw Time-of-Flight Measurements, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3503–3511 (2016).
- [64] Tadano, R., Kumar Pediredla, A. and Veeraraghavan, A.: Depth Selective Camera: a Direct, On-Chip, Programmable Technique for Depth Selectivity in Photography, *Proc. International Conference on Computer Vision (ICCV)*, pp. 3595–3603 (2015).
- [65] Tanaka, K.: Model-based Analysis of Translucent Objects using Spatially and Temporally Modulated Light, *Ph. D. Thesis* (2017).
- [66] Tanaka, K., Mukaigawa, Y., Kubo, H., Matsushita, Y. and Yagi, Y.: Recovering Inner Slices of Translucent Objects by Multi-Frequency Illumination, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5464–5472 (2015).
- [67] Tanaka, K., Mukaigawa, Y., Kubo, H., Matsushita, Y. and Yagi, Y.: Recovering Transparent Shape from Time-of-Flight Distortion, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4387–4395 (2016).
- [68] Tanaka, K., Mukaigawa, Y., Matsushita, Y. and Yagi, Y.: Descattering of Transmissive Observation using Parallel High-Frequency Illumination, *Proc. International Conference on Computational Photography (ICCP)*, IEEE, pp. 96–103 (2013).
- [69] Tibshirani, R.: Regression Shrinkage and Selection via the Lasso, *Journal of the Royal Statistical Society, Series B*, Vol. 58, pp. 267–288 (1994).
- [70] Varma, M. and Zisserman, A.: A Statistical Approach to Material Classification using Image Patch Exemplars, *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, Vol. 31, No. 11, pp. 2032–2047 (2009).
- [71] Velten, A., Raskar, R., Wu, D., Jarabo, A., Masia, B., Barsi, C., Joshi, C., Lawson, E., Bawendi, M. and Gutierrez, D.: Femto-Photography: Capturing and Visualizing the Propagation of Light, *ACM Transactions on Graphics (ToG)*, Vol. 32, No. 4, pp. 44:1–44:8 (2013).
- [72] Velten, A., Willwacher, T., Gupta, O., Veeraraghavan, A., Bawendi, M. G. and Raskar, R.: Recovering Three-Dimensional Shape Around a Corner using Ultrafast Time-of-Flight Imaging, *Nature communications*, Vol. 3, pp. 745:1–745:8 (2012).
- [73] Wang, T.-C., Zhu, J.-Y., Hiroaki, E., Chandraker, M., Efros, A. A. and Ramamoorthi, R.: A 4D Light-Field Dataset and CNN Architectures for Material Recognition, *Proc. European Conference on Computer Vision (ECCV)*, Springer International Publishing, pp. 121–138 (2016).
- [74] Wu, D., Velten, A., O’Toole, M., Masia, B., Agrawal, A., Dai, Q. and Raskar, R.: Decomposing Global Light Transport using Time of Flight Imaging, *International Journal of Computer Vision (IJCV)*, Vol. 107, No. 2, pp. 123–138 (2014).
- [75] Xu, Z., Perry, T. and Hills, G.: Method and system for multi-phase dynamic calibration of three-dimensional (3D) sensors in a time-of-flight system, *US Patent 8587771 B2* (2013).
- [76] Zhang, H., Dana, K. and Nishino, K.: Reflectance Hashing for Material Recognition, *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3071–3080 (2015).
- [77] Zheng, H., Fang, L., Ji, M., Strese, M., Oezer, Y. and Steinbach, E.: Deep Learning for Surface Material Classification using Haptic and Visual Information, *IEEE Transactions on Multimedia*, Vol. 18, No. 12, pp. 2407–2416 (2016).