Material Classification Using Raw Time-of-Flight Measurements

Shuochen Su[3,1] Felix Heide[3,4] Robin Swanson[1] Jonathan Klein[2] Clara Callenberg[2]
Matthias Hullin[2] Wolfgang Heidrich[1,3]
[1]KAUST [2]University of Bonn [3]University of British Columbia [4]Stanford University

Abstract

We propose a material classification method using raw time-of-flight (ToF) measurements. ToF cameras capture the correlation between a reference signal and the temporal response of material to incident illumination. Such measurements encode unique signatures of the material, i.e. *the degree of subsurface scattering inside a volume. Subsequently, it offers an orthogonal domain of feature representation compared to conventional spatial and angular reflectance-based approaches. We demonstrate the effectiveness, robustness, and efficiency of our method through experiments and comparisons of real-world materials.*

1. Introduction

Material classification is a popular, yet difficult, problem in computer vision. Everyday scenes may contain a variety of visually similar, yet structurally different, materials that may be useful to identify. Autonomous robots and self driving vehicles, for example, must be aware of whether they are driving on concrete, metal, pavement, or black ice. As further advances in robotics and human computer interaction are made, the need for more accurate material classification will grow.

One aspect of materials that has seen little use in classification is the way light temporally interacts with a material. As light interacts with an object, *e.g.* via reflection and subsurface scattering, it creates a temporal point spread function (TPSF), a unique signature that can describe the physical properties of each material. Past efforts to capture and analyze this signature have relied on detailed reconstructions of this temporal point spread function in the form of transient images. These can be captured either directly using bulky and expensive equipment such as streak cameras and femtosecond lasers [25, 27], or indirectly with inexpensive time of flight cameras [6], albeit at a significant computational cost as well as a lower resolution.

[1]Each material was captured under the exact same illumination and camera settings. No adjustments were made except for some white balancing only for reproducing the images.

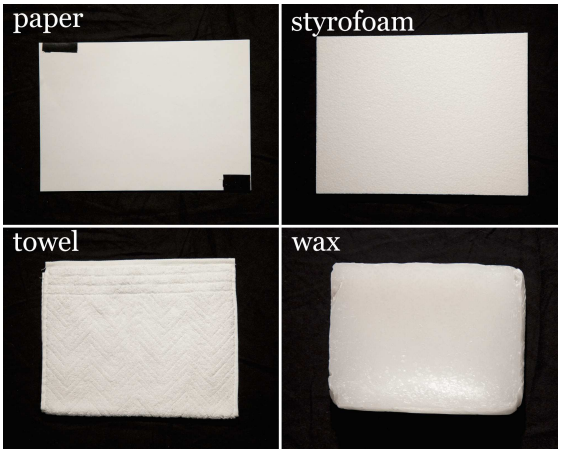


Figure 1: Visually similar[1] but structurally distinct material samples in RGB.

Our approach exploits raw measurements from ToF cameras' correlation image sensor for material feature representation. This method requires very few frequency sweeps allowing for near one-shot captures similar to coded flash methods. By completely circumventing the inverse problem which is neither easy to solve nor able to produce robust solutions [6], our features can be directly fed into a pretrained material classifier that predicts results in a timely manner. Furthermore, our method allows for per pixel labeling which enables more accurate material classification. Nevertheless, there are significant challenges inherent to this approach, including depth and noise which create ambiguities due to the correlation nature of the ToF image formation model, and the camera's limited temporal resolution [6] relative to that of a streak camera [25].

In this work, we take the first step to collect a dataset consisting of visually similar but structurally distinct materials, *i.e.* paper, styrofoam, towel, and wax as seen in Figure 1. To ensure that our classifier is robust to both distance and angle variations, we take measurements from a variety of positions. Experimental results show that classification from ToF raw measurements alone can achieve accuracies up to 81%. We also present superior results of our method

compared to those based on reflectance in real world scenario where the latter fail, *e.g*. classifying printed replicas. Together these experiments show that the representation of materials with raw ToF measurements, although at the expense of sacrificing temporal resolution, has the potential to work well on material classification tasks.

Our specific contributions are:

- We develop a method to represent materials as raw measurements from inexpensive and ubiquitous correlation image sensors which are both budget and computational friendly;
- Near single-shot, pixel wise material classification which is robust to ambient light and thus can be potentially deployed in everyday environments;
- Finally, we show that our recognition results can be further improved by including spatial information.

2. Related Work

Material classification. The robust classification of materials from optical measurements is a long-standing challenge in computer vision. Existing techniques rely on color, shading, and texture in both active and passive settings; some even use indirect information like the shape or affordance of an object. For a comprehensive overview of the state of the art, we refer the reader to a recent survey by Weinmann and Klein [26]. Here we identify the following groups of works, some of which are associated with reference databases:

- techniques based on natural RGB images and textures [1, 12, 24];
- gonioreflectometric techniques [11, 28, 19, 13] that investigate materials' response to directional illumination;
- techniques that use patterned active illumination to recover parameters of subsurface light transport [22], and finally,
- techniques that employ other aspects of materials, such as their thermal properties [18], micro-scale surface geometry obtained through mechanical contact [9], or other physical parameters like elasticity [2].

Common to all these approaches is that they fail if suitable information is not available or their capture conditions are not strictly met. Some methods, in particular ones that rely on natural RGB images, are susceptible to adversarial input and could easily be fooled by human intervention, printed photos of objects, or reflections. Furthermore, these techniques often rely on object detection to infer material information [21]. As a whole, the problem of classifying materials remains unsolved. With our method, we propose temporal analysis of subsurface scattering as a new source of information. To our knowledge, it is the first method capable of per-pixel classification without the need for structured illumination or gonioreflectometric measurements. We demonstrate that our method is capable of producing robust results under lab conditions, and that it forms a valuable complement to existing techniques.

Correlation image sensors. Correlation image sensors are a class of device that have been well explored for use in depth acquisition [20, 5] and since extended for various applications. When operated as range imagers, the quality delivered by correlation sensors suffers from multi-path interference, whose removal has therefore been the subject of extensive research [3, 4, 14]. Contrary to this line of work, our method is enabled by the insight that time-domain effects of multi-path scattering can carry valuable information about the material. To our knowledge, the only other work that explicitly makes use of this relation is a method by Naik et al. [15], in which low-parameter models are fitted to streak tube images to recover the reflectance of non-line-of-sight scenes. In a sense, Naik et al.'s method maps angular information to the time domain where it is picked up by a high-end opto-electronic imaging system. Our proposed method, in contrast, does not make use of angular or spatial distributions and works on a type of measurement that is routinely available from both configurable ToF development toolkits such as ESPROS EPC 660 and TI OPT8241-CDK-EVM, and consumer-grade hardware like the Microsoft Kinect v2 and Google's Project Tango smartphone (with multi-frequency functionality enabled) .

3. ToF Sensors and Material Properties

In this section we relate raw ToF measurements to material optical properties. Here we focus on two phenomena in particular: multiple scattering inside of a surface volume (*e.g*. between strands of a towel), and subsurface scattering which commonly occurs in wax and styrofoam. Throughout this paper we restrict materials to planar geometries and ignore macro-scale multi-bounce reflections between object surfaces. This results in the TPSF model in Equation 9, which is simpler than the mixture model found in [8].

3.1. Image Formation Model of Correlation Sensors

The image formation model of correlation sensors in homodyne mode has been derived in previous works [6, 8, 7]. Following Heide et al.'s interpretation [6], a correlation pixel measures the modulated exposure

$$b = \int_0^T E(t) f_\omega \left(t - \phi/\omega\right) dt, \tag{1}$$

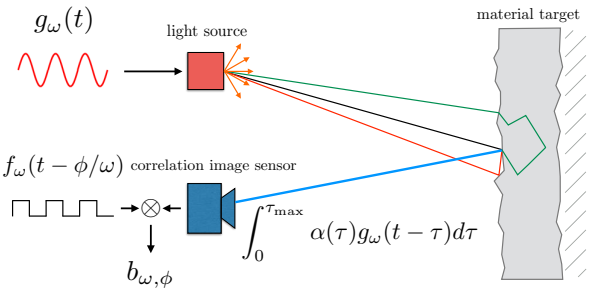


Figure 2: Temporal interaction of light and material. A correlation image sensor correlates the reference signal with a mixture of (a) direct surface reflection (black); (b) surface inter-reflection or multiple scattering inside surface volume (red); and (c) subsurface scattering (green). The camera observes a mixture of (a),(b) and (c), as indicated in blue.

where $E(t)$ is the pixel irradiance, $f_\omega(t)$ is a periodic reference function of angular frequency ω and programmable phase offset ϕ, both evaluated at time t. Typically, this reference function is zero-mean to make the imager insensitive to ambient light, *i.e.* the DC component of $E(t)$.

Materials with different reflection or scattering properties can cause multiple path contributions to be linearly combined in a single sensor pixel, as shown in Figure 2. In an active illumination setting where the light source $g_\omega(t)$ is intensity modulated at the same frequency, $E(t)$ becomes a superposition of many attenuated and phase shifted copies of $g_\omega(t)$, along all possible paths $p \in \mathcal{P}$:

$$E(t) = \int_{\mathcal{P}} \alpha_p g_\omega(t - |p|) dp \tag{2}$$

We define the temporal point spread function (TPSF) $\alpha(\tau)$ as the summed contribution of all paths of equal length $|p| = \tau$:

$$\alpha(\tau) = \int_{\mathcal{P}} \alpha_p \delta(|p| - \tau) dp. \tag{3}$$

The combined multi-path backscatter can thus be expressed as the convolution of $g_\omega(t)$ with the TPSF $\alpha(\tau)$:

$$E(t) = \int_0^{\tau_{\max}} \alpha(\tau) g_\omega(t - \tau) d\tau. \tag{4}$$

By substituting $E(t)$ in Equation 1, we obtain a correlation integral of sensor response and optical impulse response:

$$b_{\omega,\phi} = \int_0^T f_\omega(t - \phi/\omega) \int_0^{\tau_{\max}} \alpha(\tau) g_\omega(t - \tau) d\tau dt \tag{5}$$

$$= \int_0^{\tau_{\max}} \alpha(\tau) \underbrace{\int_0^T f_\omega(t - \phi/\omega) g_\omega(t - \tau) dt}\, d\tau \tag{6}$$

$$=: \int_0^{\tau_{\max}} \alpha(\tau) \cdot c(\omega, \phi/\omega + \tau) \, d\tau, \tag{7}$$

where the scene-independent functions f_ω and g_ω have been folded into a correlation function $c(\omega, \phi/\omega + \tau)$ that is only dependent on the imaging device and can be calibrated in advance (Section 4.1). Expressing the real-valued c by its Fourier series we arrive at:

$$b_{\omega,\phi} = \sum_{k=1}^{\infty} g_k \int_0^{\tau_{\max}} \alpha(\tau) \cos\left(k\omega\left(\phi/\omega + \tau\right) + \phi_k\right) d\tau, \tag{8}$$

where g_k is the amplitude and ϕ_k the phase of the k^{th} harmonic. In essence, Equation 8 shows that the correlation sensor probes a TPSF's frequency content. The change in the temporal profile $\alpha(\tau)$ will be reflected in its Fourier spectrum. This is the effect we expect to see in the measurement $b_{\omega,\phi}$ between structurally different materials.

3.2. Material Signatures From Raw Measurements

Our camera images a material target while cycling through the relative phases $\{\phi_{j=1\ldots n}\}$ and frequencies $\{\omega_{i=1\ldots m}\}$ from Equation 8, generating m measurement vectors $\mathbf{b}_{1\ldots m}$, each of which corresponds to one modulation frequency and is sampled at n different phases. We stack all these vectors together and obtain the *total* measurement matrix $\mathbf{B} = (\mathbf{b}_1 \ldots \mathbf{b}_m)$. The latent TPSF $\alpha(\tau)$ only helps with the derivation and is never reconstructed.

Both the strength and challenges of using correlation measurements as material features can be illustrated via simulation. In Figure 3, for example, we demonstrate the simulation of $\mathbf{B}$ at $\phi = 0$ and $\pi/2$ and why it is necessary to address depth ambiguities.

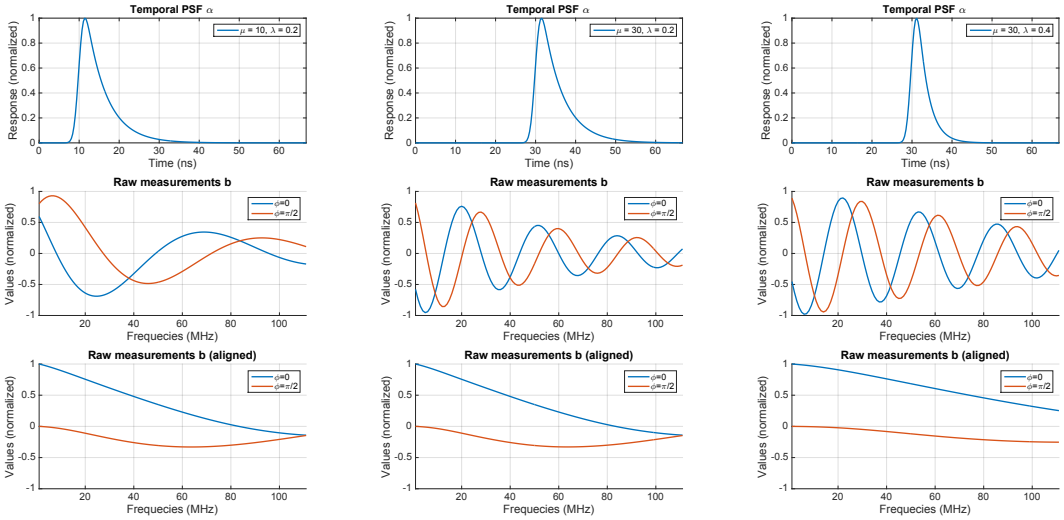
First, we approximate $\alpha(\tau)$ with an exponentially modified Gaussian model which Heide et al. [8] found to compactly represent typical TPSFs:

$$\alpha(\tau; a, \sigma, \lambda, \mu) = a \cdot \exp\left((\sigma\lambda)^2/2 - (\tau - \mu)\lambda\right) \cdot \left(1 + \operatorname{erf}\left(\frac{(\tau - \mu) - \sigma^2\lambda}{\sqrt{2}\sigma}\right)\right). \tag{9}$$

The intensity of TPSF at any given time τ is a function of amplitude a, Gaussian width σ, skew λ, and peak center μ. While λ relates to a material's scattering coefficient [27], a and μ are connected to albedo, light falloff and depth related parameters which are irrelevant to material structural properties. Similarly, σ models the temporal resolution of a correlation image sensor, which, without lack of generality, remains constant in our simulation.

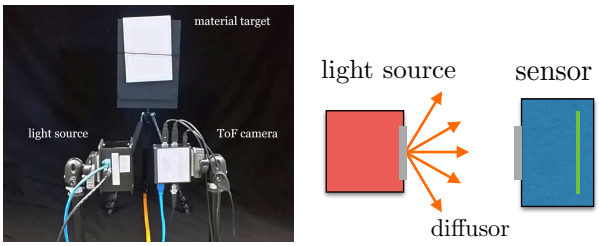
To test our concept in simulation, we assume the correlation function $c(\omega, \phi/\omega + \tau)$ (Equation 7) to be a pure sinusoid. By applying Equation 8 to the given $c(\omega, \phi/\omega + \tau)$ and $\alpha(\tau)$, we simulate measurements at several discrete frequencies ω_i from 10 to 120 MHz and two modulation delays $\phi = 0$ and $\phi = \pi/2$.

On the top row of Figure 3 we show three TPSFs with varying peak centers and skews. Specifically, Figure 3a and



(a) Material 1 at distance 1. (b) Material 1 at distance 2. (c) Material 2 at distance 2.

Figure 3: Simulation of TPSF α, measurement $\mathbf{B}$ and $\mathbf{B}^{\text{aligned}}$ at $\phi = 0$ (blue) and $\phi = \pi/2$ (red). Note that for effect the time scale of the TPSF has been exaggerated. The temporal support of the true TPSF is typically below a nanosecond.



(a) Dataset acquisition setup. (b) Noise removal setup.

Figure 4: Illustrations of experimental setup.

Figure 3b differ in μ, while Figure 3b and Figure 3c differ in λ. As can be compared in the middle row of Figure 3, $\mathbf{B}$ is affected by both the material-independent parameter μ, and material-dependent λ. To make $\mathbf{B}$ invariant to μ, i.e., to depth variations, we need to compensate for a global temporal shift that originates from translation.

To this end, we have developed a depth normalization method which is detailed in Section 4.1. An example of normalized measurements are plotted along the bottom row of Figure 3. While the depth dependent differences are eliminated in Figures 3a and 3b, the material intrinsic properties remain intact when comparing Figures 3b and 3c.

4. Methods

4.1. Data Preprocessing

Removing fixed pattern noise. Our measurements show that there exists modulation frequency dependent fixed pattern noise which necessitates per-pixel calibration for their removal. Similar to [17], and as illustrated in Figure 4b, we expose the camera sensor to diffuse light to create a noise calibration matrix. We can then divide this amplitude normalized data from future measurements to compensate for the fixed pattern noise.

Depth normalization. Next we describe how unwanted variations in amplitude and phase were removed from the input data. This serves to align measurements regardless of distance and intensity while leaving frequency-dependent effects unaffected. First, we take measurements from a set of modulation frequencies $\{\omega_{i=1\ldots m}\}$ and choose one to serve as a point of reference for each material – for the sake of convenience, let ω_1 be the reference frequency for any given material.

The following procedure summarized, is performed for all pixels, materials, and distances independently. It determines the total temporal delay of a given measurement vector by analyzing the phase shift at the fundamental of the ω_1 measurement. It then shifts all measurements by the respective phase to compensate for this delay.

1. Determine the complex amplitude of the signal at its

base frequency ω_1. To this end, we take the vector of n phase-shifted $\{\phi_{i=1\ldots n}\}$ measurements b_{ω_1,ϕ_j} and, using a discrete Fourier transform, obtain coefficients $c_{\omega_1,k}$ such that

$$b_{\omega_1,\phi_i} = \sum_{k=0}^{n/2-1} c_{\omega_1,k} \cdot e^{ik\phi_i}, \qquad c_{\omega_1,k} \in C \quad (10)$$

Note that the negative spectral coefficients follow directly from the convex conjugate and are thus omitted from our derivation. From the coefficient of the fundamental frequency, $c_{\omega_1,1}$, we obtain the desired delay τ_{ref} and amplitude factor a_{ref} by which we will compensate:

$$\tau_{\text{ref}} = \angle(c_{\omega_1,1})/\omega_1, \qquad a_{\text{ref}} = |c_{\omega_1,1}| \quad (11)$$

2. We then propagate this correction to the measured signal at all modulation frequencies $\omega_{i=1\ldots m}$ by altering their corresponding Fourier coefficients. Again, we Fourier transform the n phase samples for modulation frequency ω_i as in Equation 10.

$$b_{\omega_i,\phi_i} = \sum_{k=0}^{n/2-1} c_{\omega_i,k} \cdot e^{ik\phi_i}, \qquad c_{\omega_i,k} \in C \quad (12)$$

Next, we phase-shift the coefficients $c_{\omega_i,k}$ to compensate for the delay τ_{ref}, and normalize their amplitude with respect to a_{ref}:

$$c^{\text{aligned}}_{\omega_i,k} = c_{\omega_i,k} \cdot e^{-ik\omega_i\tau_{\text{ref}}}/a_{\text{ref}} \quad (13)$$

$$= \frac{c_{\omega_i,k}}{|c_{\omega_1,1}|} \cdot \left(\frac{c_{\omega_1,1}}{|c_{\omega_1,1}|}\right)^{-|k|\omega_i/\omega_1} \quad (14)$$

3. Finally, by substituting the new coefficients $c^{\text{aligned}}_{\omega_i,k}$ back into Equation 12 we obtain the compensated measurements $b^{\text{aligned}}_{\omega_i,\phi_j}$.

An equivalent algorithm which is more compact and straightforward to implement is provided in Algorithm 1.

4.2. Features

After preprocessing, the raw correlation measurement at each pixel is denoted by $\mathbf{B}^{\text{aligned}}$, where each element is a depth and amplitude normalized complex number $b^{\text{aligned}}_{\omega_i,\phi_j}$. This complex matrix is then vectorized into an n×m×2 dimensional feature vector for training and testing. Representing materials in such a high dimensional space poses well known challenges to classification. Overfitting could be unavoidable if our number of data points is limited. Furthermore, higher dimensional feature data requires longer training time.

Algorithm 1 Depth alignment for measurements at modulation frequency ω_i

Input: $\mathbf{b}_{\omega_1}$: vector of n phase-shifted $\{\mu_{l=1..n}\}$ measurements at base modulation frequency ω_1; $\mathbf{b}_{\omega_i}$: vector of n phase-shifted measurements at other frequency ω_i

Output: Aligned measurement: $\mathbf{b}^{\text{aligned}}_{\omega_i}$

1: $\hat{\mathbf{b}}_{\omega_1} := \text{FFT}(\mathbf{b}_{\omega_1})$
2: $\hat{\mathbf{b}}_{\omega_i} := \text{FFT}(\mathbf{b}_{\omega_i})$
3: **for** $k = 1$ to #harmonics **do**
4: $\quad \hat{b}^{\text{aligned}}_{\omega_i,k} := \frac{\hat{b}_{\omega_i,k}}{|\hat{b}_{\omega_1,1}|} \cdot \left(\frac{\hat{b}_{\omega_1,1}}{|\hat{b}_{\omega_1,1}|}\right)^{-|k|\omega_i/\omega_1}$
5: **end for**
6: $\mathbf{b}^{\text{aligned}}_{\omega_i} := \text{IFFT}(\hat{\mathbf{b}}^{\text{aligned}}_{\omega_i})$

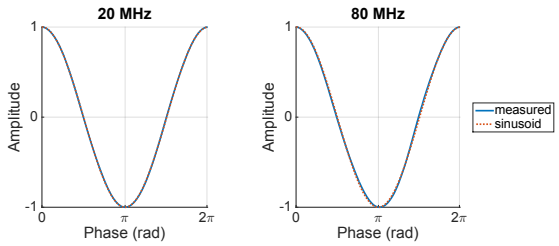


Figure 5: Our modulation signal $\phi_\omega(t)$ at 20MHz and 80MHz.

To address these two issues, we compare the classification accuracy using features in both the original space and dimensional reduced space, *e.g.* after PCA. In theory, features that share similar rows in $\mathbf{B}^{\text{aligned}}$ are highly correlated, as the only difference between the fundamentals of $b^{\text{aligned}}_{\omega_i,\phi_{j_1}}$ and $b^{\text{aligned}}_{\omega_i,\phi_{j_2}}$ is a fixed phase shift $|\phi_{j_2} - \phi_{j_1}|$. Our modulation signal $\phi_\omega(t)$ is nearly sinusoidal, see Figure 5, therefore most features in the original space may still be correlated to a certain degree. We show a comparison between the two with a real dataset and how the number of required measurements can be minimized in Section 5.2.

4.3. Learning Models

It is important to see whether the classification accuracy is benefited most from tweaking parameters for different learning models, or from the features themselves. To this end we evaluated several supervised learning methods including Nearest Neighbors, Linear SVM, RBF SVM, Decision Tree, Random Forest, AdaBoost, LDA, and QDA using both the MATLAB classificationLearner and Scikit-learn [16] implementations. The best results from every model can be found in Table 1.

Training and validation. During the training stage we performed 5-fold cross validation and reported the mean accuracy. A confusion matrix of the best performing model is given in Section 5.2.

Testing. Additionally, several test cases are reported in Section 5.3 and 5.4. These tests include special cases such as detecting material photo replicas, and scene labeling.

5. Experiments and Results

5.1. Dataset Acquisition

A prototype ToF camera system composed of a custom RF modulated light source and a demodulation camera was used to collect our dataset, similar to that used in [8]. Our light source is an array of 650 nm laser diodes equipped with a diffusor sheet to provide full-field illumination. The sensor is a PMD CamBoard nano development kit with a clear glass PMD Technologies PhotonICs 19k-S3 sensor (without NIR filter), a spatial resolution of 165×120 pixels, and a custom 2-channel modulation source with 150 MHz bandwidth that serves to generate the signals $f(t)$ and $g(t)$ (Section 3). In our experiments, we limit ourselves to the frequency range from 10 MHz to 80 MHz that is also commonly used by other ToF sensor vendors.

Data points. We collected data from 4 structurally different yet visually similar materials

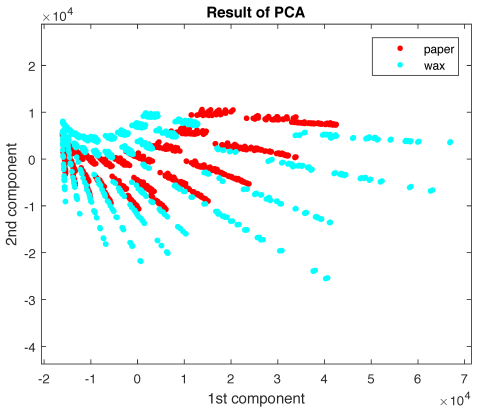
- *paper*: a stack of normal printing paper;
- *styrofoam*: a regular piece of polystyrene foam;
- *towel*: two layers of a hand towel;
- *wax*: a large block of wax.

A photo of our experimental setup can be seen in Figure 4a. To cover a wide range of distances and viewing angles, we placed the material samples at 10 distances ranging from 1 to 2 meters from the camera. The three viewing angles, *flat*, *slightly tilted* and *more tilted*, were achieved by adjusting the tripod to set positions. Under each physical and modulation setting, a total of 4 frames were captured with identical exposure times to account for noise. We then randomly sample 25 locations from the raw correlation frames as data points. In total, our dataset consists of $10 \times 3 \times 4 \times 25 = 3000$ observations for each material.

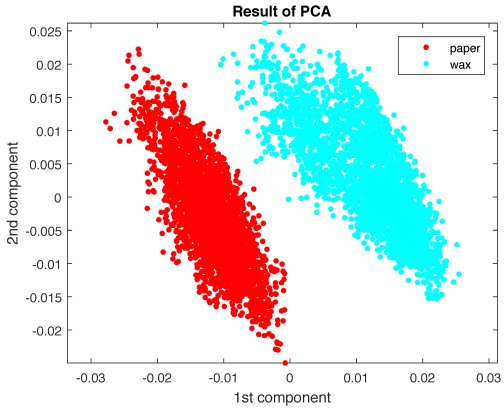
Features. Under each of the 30 physical distance and angle settings, a total of 64 frames were captured to cover a wide range of modulation frequencies ω and phases ϕ in $\mathbf{B}$, including,

- 8 frequencies: 10, 20, . . . , 80 MHz, and
- 8 equispaced delays from 0 to 2π

We then follow the data preprocessing method in Section 4.1 to normalize the depth and amplitudes, where ω_1 is chosen as 10 MHz. This leaves us with a 64-dimensional complex-valued or, equivalently, a 128-dimensional real-valued feature $\mathbf{B}^{\text{aligned}}$ at each data point. Finally, before training each feature is standardized to have zero mean and unit variance.



(a) PCA visualization of features before preprocessing.



(b) PCA visualization of features after preprocessing.

Figure 6: Effectiveness of depth normalization, visualized in the dimensional reduced [23] feature space.

Figure 6 shows a 2D projection of the wax and paper features from our dataset before and after preprocessing. It's clear that the removal of fixed pattern noise and normalization of amplitude and phase significantly improve the separability of our features.

5.2. Classification Results

As previously mentioned in Section 4.3, we now report and analyze the classification accuracies of different learning models. The mean validation accuracy for each method can be found in Table 1. We observe that while SVM with an RBF kernel generally has the greatest precision, most methods (Decision Tree, Nearest Neighbor, SVM and Random forest) perform comparably. This suggests that the power of our algorithm is a result of the features, *i.e.* the

	Original	After PCA	High freqs
Decision tree	72.2	64.3	68.4
Nearest neighbor	69.5	74.7	69.1
Linear SVM	76.9	69.8	68.6
RBF SVM	**80.9**	**77.7**	**71.5**
Random forest	79.9	75.1	70.0
AdaBoost	72.1	61.1	69.3
LDA	60.0	58.3	62.7
QDA	62.6	60.4	64.8

Table 1: Validation accuracies (%) from different learning models.

	paper	styrofoam	towel	wax
paper	**62.7**	3.9	33.5	0.0
styrofoam	6.1	**82.2**	11.7	0.0
towel	18.3	4.1	**77.6**	0.0
wax	0.0	0.1	0.0	**99.9**

Table 2: Confusion matrix (%). Labels in the left most column denote the true labels, while those in the top row correspond to predicted labels.

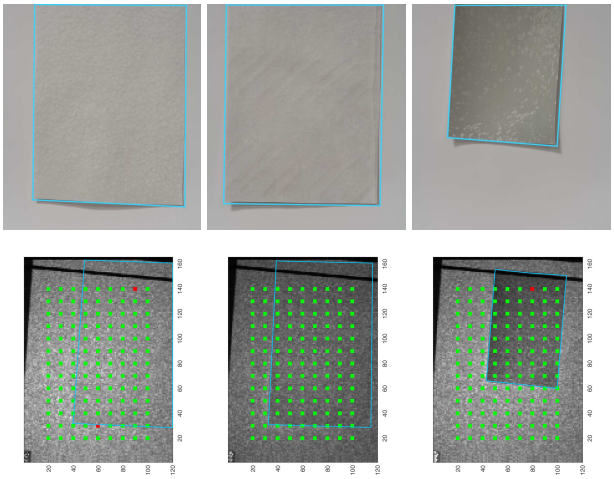
	without spatial coherence	with spatial coherence
paper	70.6	**80.0**
styrofoam	90.8	**95.8**
towel	72.0	**74.1**
wax	100.0	**100.0**

Table 3: Testing accuracies (%) with and without considering spatial coherence.

ToF raw measurements, rather than the learning model.

The confusion matrix in Table 2 shows how often each category is misclassified as another. Paper and towel, for example, are most commonly misclassified to each other. One possible explanation could be that the paper used in our experiments had stronger absorption, thus behaving similarly to the surface inter-reflectance of the towel. Wax, however, comes with a greater degree of subsurface scattering compared to the other materials which is reflected directly in its accuracy. Throughout this experiment, we fix the RBF kernel scale as $\sigma = \sqrt{P}$, where P is the number of features.

To study if we are able to further reduce the dimensionality of discriminative feature representation for each material, we performed two experiments. First, we reduce the feature dimension by performing a principal component analysis prior to training, validation, and testing. If 95% variance is kept, we are left with 5D features. These accuracies are shown in the center column of Table 1. We also empirically handpicked $\mathbf{b}$ at two higher modulation frequencies: 70MHz and 80MHz as features. These accuracies are shown in the rightmost column of Table 1. As we can see, although the highest validation accuracy is achieved by representing features in the original high dimensional space, there is a balance between the number of features and acceptable accuracy which warrants further research. Futhermore, when only the selected higher frequencies are used for measuring and predicting unseen material, the capturing time is greatly reduced from 12.3s to 3.0s.



(a) Styrofoam. (b) Towel. (c) Wax.

Figure 7: Our classifier successfully recognizes the actual material of each paper printed replica when attached to a paper stack. Top: Reference RGB images taken with a DSLR camera which could be confusing to RGB based methods and even human eyes. Bottom: classification results overlayed on top of the ToF amplitude image. Green dots indicate a correct classification (paper) and red indicates a misclassification. For clarity, the boundaries of each printed replica are highlighted in blue.

Lastly, we test our best trained classifier on a separate testing set. These test accuracies are reported in column "without spatial consistency" in Table 3. We also show that by simply introducing spatial consistency in our testing stage, up to a 10% improvement can be reached for paper. This spatial consistency is implemented by ranking all the predicted labels within a region of interest in each test frame. Then the label with highest probability is chosen as the final prediction. These results are shown in column "with spatial consistency".

5.3. Comparison with RGB Based Approaches

Reflectance-based methods can be easily fooled by replicas, *e.g.* printed pictures of actual materials, as the RGB information itself is insufficient to represent the intrinsic material properties. Furthermore, they are sensitive to even small changes in illumination. While a colored ambient light source may change the RGB appearance and there-

fore its classification using traditional methods, the material structure is unchanged.

To validate the advantages and robustness of our method over RGB approaches we devised a simple experiment. First, we photographed our raw materials, making small post-processing alterations to ensure the photos appear as similar to our lab conditions as possible. Those photos were then printed to regular printing paper, similar to those used in our earlier classification experiments, and placed on top of the paper stack used earlier before taking ToF measurements.

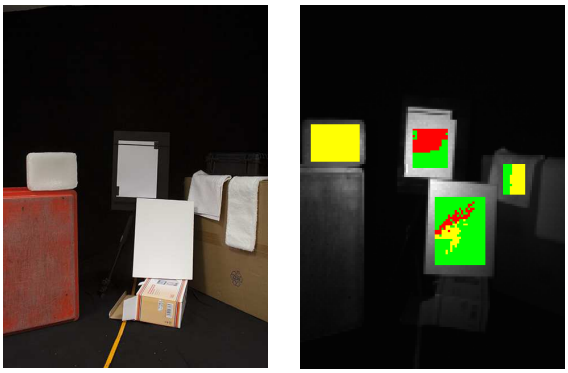
Experimental results, shown along the bottom row of Figure 7, reveal that our feature representation is invariant to RGB textures, as all paper replicas were correctly classified as the actual material: paper. Reference RGB images captured by a Canon EOS 350D DSLR camera next to the ToF camera can be seen in the top row of Figure 7. It is worth noting that this approach is limited by the fact that printed paper itself is less reflective, and therefore darker, than the actual materials.

Due to the different scope and nature of our methods, direct comparison with RGB based approaches may be unfair because they unavoidably rely on object cues. Nevertheless, we explored results from many of the best trained RGB-based Deep Neural Nets methods. When testing our photo replicas (seen previously in Figure 7, top) on a pretrained CaffeNet model based on the network architecture of Krizhevsky et al. [10] for ImageNet, the wax replica is classified as "doormat", while both towel and styrofoam are tagged with "towel" as the top predictions. When only a central region within the blue boundary of our photo replicas are fed to the network, wax, towel and styrofoam replicas are recognized as "water snake", "paper towel" and "velvet" respectively. These results are not surprising as these models only use local correlation from RGB information whereas our approach exploits completely new features for classification.

5.4. Scene Labeling

Finally, we created a scene, shown in Figure 8, where each material was placed at different distances and angles from the camera. In this scene we used an 8mm wide angle lens with the TOF camera instead of the 50mm lens used previously as it was difficult to assemble the materials into such a narrow field of view without significant occlusion.

As we can see, the entire wax region is correctly labeled at each pixel. For the most part both styrofoam and paper are correctly classified as well. Towel, on the other hand, is recognized as wax and styrofoam. One possible explanation could be that as it is placed at the edge of the frame, vignetting becomes significant and introduces additional noise to the features after preprocessing.



(a) Scene of materials. (b) Segmented and labeled.

Figure 8: Our classifier successfully labeled the segmented scene. (a) Scene of materials captured by the reference RGB camera; (b) Labeled amplitude image from our ToF camera. Red: paper; green: styrofoam; blue: towel; yellow: wax.

6. Summary

We have proposed a method for distinguishing between materials using *only* ToF raw camera measurements. While these are merely the first steps towards ToF material classification, our technique is already capable of identifying different materials which are very similar in appearance. Furthermore, through careful removal of noise and depth dependencies, our method is robust to depth, angle, and ambient light variations allowing for classification in outdoor and natural settings.

Future work. Although we are able to achieve high accuracy with our current classifiers and datasets, our method could be further refined by additional training data and a more diverse set of materials. As a valuable complement to existing techniques, we believe that our method could also be used in combination with state of the art RGB algorithms, or by incorporating traditional spatial and image priors. In the future, we would also like to relax the restriction of planar materials and investigate the robustness of our method to object shape variations.

Acknowledgements: This work was supported through the X-Rite Chair and Graduate School for Digital Material Appearance, the German Research Foundation, Grant HU 2273/2-1, the Baseline Funding of the King Abdullah University of Science and Technology, and a UBC 4 Year Fellowship.

References

[1] B. Caputo, E. Hayman, and P. Mallikarjuna. Class-specific material categorisation. In *Computer Vision, 2005. ICCV*

2005. Tenth IEEE International Conference on, volume 2, pages 1597–1604. IEEE, 2005. 2

[2] A. Davis, K. L. Bouman, J. G. Chen, M. Rubinstein, F. Durand, and W. T. Freeman. Visual vibrometry: Estimating material properties from small motions in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5335–5343, 2015. 2

[3] A. Dorrington, J. Godbaz, M. Cree, A. Payne, and L. Streeter. Separating true range measurements from multipath and scattering interference in commercial range cameras. In *Proc. SPIE*, volume 7864, 2011. 2

[4] D. Freedman, E. Krupka, Y. Smolin, I. Leichter, and M. Schmidt. SRA: Fast removal of general multipath for tof sensors. *arXiv preprint arXiv:1403.5919*, 2014. 2

[5] M. Grzegorzek, C. Theobalt, R. Koch, and A. Kolb. *Time-of-Flight and Depth Imaging. Sensors, Algorithms and Applications*, volume 8200. Springer, 2013. 2

[6] 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):45, 2013. 1, 2

[7] 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 *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 3222–3229. IEEE, 2014. 2

[8] 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):26338–26350, 2014. 2, 3, 6

[9] M. K. Johnson, F. Cole, A. Raj, and E. H. Adelson. Microgeometry capture using an elastomeric sensor. In *ACM Transactions on Graphics (TOG)*, volume 30, page 46. ACM, 2011. 2

[10] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012. 8

[11] C. Liu and J. Gu. Discriminative illumination: Per-pixel classification of raw materials based on optimal projections of spectral brdf. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(1):86–98, 2014. 2

[12] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz. Exploring features in a bayesian framework for material recognition. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 239–246. IEEE, 2010. 2

[13] M. A. Mannan, D. Das, Y. Kobayashi, and Y. Kuno. Object material classification by surface reflection analysis with a time-of-flight range sensor. In *Advances in Visual Computing*, pages 439–448. Springer, 2010. 2

[14] N. Naik, A. Kadambi, C. Rhemann, S. Izadi, R. Raskar, and S. B. Kang. A light transport model for mitigating multipath interference in time-of-flight sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 73–81, 2015. 2

[15] N. Naik, S. Zhao, A. Velten, R. Raskar, and K. Bala. Single view reflectance capture using multiplexed scattering and time-of-flight imaging. In *ACM Transactions on Graphics (TOG)*, volume 30, page 171. ACM, 2011. 2

[16] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. 5

[17] C. Peters, J. Klein, M. B. Hullin, and R. Klein. Solving trigonometric moment problems for fast transient imaging. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, 34(6), Nov. 2015. 4

[18] P. Saponaro, S. Sorensen, A. Kolagunda, and C. Kambhamettu. Material classification with thermal imagery. June 2015. 2

[19] 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 *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2191–2200. ACM, 2015. 2

[20] R. Schwarte, Z. Xu, H. Heinol, J. Olk, R. Klein, B. Buxbaum, H. Fischer, and J. Schulte. New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device. In *Proc. SPIE*, volume 3100, pages 245–253, 1997. 2

[21] G. Schwartz and K. Nishino. Automatically discovering local visual material attributes1. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3565–3573, 2015. 2

[22] J. Steimle, A. Jordt, and P. Maes. Flexpad: Highly flexible bending interactions for projected handheld displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 237–246, New York, NY, USA, 2013. ACM. 2

[23] L. J. van der Maaten, E. O. Postma, and H. J. van den Herik. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10(1-41):66–71, 2009. 6

[24] M. Varma and A. Zisserman. A statistical approach to material classification using image patch exemplars. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(11):2032–2047, 2009. 2

[25] 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):44, 2013. 1

[26] M. Weinmann and R. Klein. A short survey on optical material recognition. In *Proceedings of the Eurographics Workshop on Material Appearance Modeling*, pages 35–42. Eurographics, 2015. 2

[27] D. Wu, A. Velten, M. OToole, B. Masia, A. Agrawal, Q. Dai, and R. Raskar. Decomposing global light transport using time of flight imaging. *International Journal of Computer Vision*, 107(2):123–138, 2014. 1, 3

[28] H. Zhang, K. Dana, and K. Nishino. Reflectance hashing for material recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. 2