Low-budget Transient Imaging using Photonic Mixer Devices

Felix Heide* Matthias B. Hullin* James Gregson Wolfgang Heidrich
The University of British Columbia

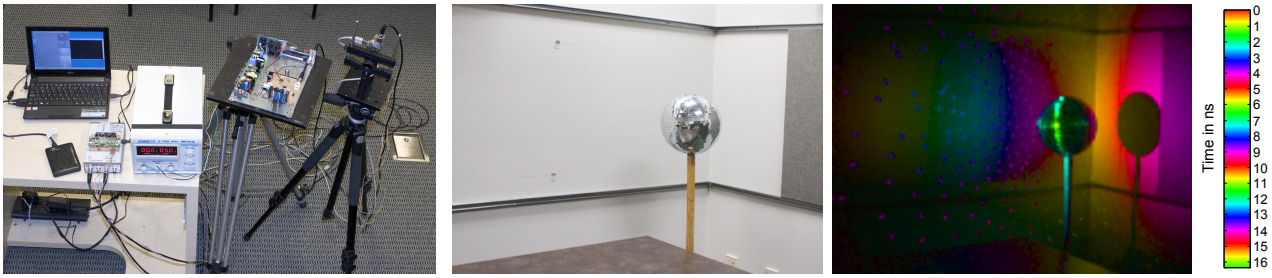

Figure 1: *Left: Our capture setup for transient images (from left: computer, signal generator, power supply, modulated light source, PMD camera). Middle: A disco ball with many mirrored facets. Right: The same sphere as seen by our transient imager when illuminated from the left, colored according to the time offset of the main intensity peak.*

Abstract

Transient imaging is an exciting a new imaging modality that can be used to understand light propagation in complex environments, and to capture and analyze scene properties such as the shape of hidden objects or the reflectance properties of surfaces.

Unfortunately, research in transient imaging has so far been hindered by the high cost of the required instrumentation, as well as the fragility and difficulty to operate and calibrate devices such as femtosecond lasers and streak cameras.

In this paper, we explore the use of photonic mixer devices (PMD), commonly used in inexpensive time-of-flight cameras, as alternative instrumentation for transient imaging. We obtain a sequence of differently modulated images with a PMD sensor, impose a model for local light/object interaction, and use an optimization procedure to infer transient images given the measurements and model. The resulting method produces transient images at a cost several orders of magnitude below existing methods, while simultaneously simplifying and speeding up the capture process.

CR Categories: I.3.3 [Computer Graphics]: Picture/Image Generation—Digitizing and scanning

Keywords: computational photography, transient imaging

Links: DL PDF WEB

1 Introduction and Related Work

Transient imaging refers to a recent imaging modality in which short pulses of light are observed "in flight" as they traverse a scene and before the light distribution achieves a global equilibrium. Specifically, a transient image is a rapid sequence of images representing the impulse response of a scene. The original idea behind transient imaging goes back to work performed in the late 70s by Abramson [1978; 1983] under the name "light-in-flight recording". Abramson created holographic recordings of scenes illuminated by picosecond lasers, from which it was possible to optically reconstruct an image of the wavefront at a specific time. While the scene complexity was limited by technical constraints of the holographic setup, other researchers already used this approach for tasks such as shape measurements (e.g. [Nilsson and Carlsson 1998]).

Recently, interest in transient imaging has been rekindled by the development of ultra-fast camera technologies [Velten et al. 2011], which allow for simplified setups compared to the holographic approach, and significantly more general scene geometries. This new imaging technology has many exciting applications. Starting with the introduction of an image formation model [Smith et al. 2008] and the pilot experiments by Kirmani et al. [2009], there have been several proposals to use transient images as a means of reconstructing 3D geometry that is not directly visible to either the camera or the light sources [Pandharkar et al. 2011; Velten et al. 2012], to capture surface reflectance [Naik et al. 2011], or simply to visualize light transport in complex environments to gain a better understanding of optical phenomena [Velten et al. 2013]. Wu et al. [2012] recently proposed to use transient images together with models of light/object interaction to factor the illumination into direct and indirect components.

Unfortunately, transient imaging currently relies on expensive custom hardware, namely a femtosecond laser as a light source, and a streak camera [Velten et al. 2011] for the image capture. Together, these components amount to hundreds of thousands of dollars worth of equipment that is bulky, extremely sensitive, difficult to operate, potentially dangerous to the eye, and slow. For example, a streak camera measures only a single scanline of a transient image in each measurement. To obtain a full transient image it is therefore necessary to mechanically scan the scene. Due to the very limited amount of light in a femtosecond pulse, averaging of multiple measurements and complicated calibration and noise suppression algorithms are required to obtain good image quality. All in all, capture times of an hour or more have been reported for a single transient image.

In our work, we seek to replace this complex setup with a modified, but simple, photonic mixer device (PMD). PMD sensors are commonly used in time-of-flight cameras, and can be obtained for

*Joint first authors

ACM Reference Format
Heide, F., Hullin, M., Gregson, J., Heidrich, W. 2013. Low-Budget Transient Imaging using Photonic Mixer Devices. ACM Trans. Graph. 32, 4, Article 45 (July 2013), 10 pages. DOI = 10.1145/2461912.2461945 http://doi.acm.org/10.1145/2461912.2461945.

0730-0301/13/07-ART45 $15.00.
DOI: http://dx.doi.org/10.1145/2461912.2461945

a few hundred dollars. This camera is used in conjunction with inexpensive laser diodes as an illuminant. The acquisition process involves taking several dozen images with different temporal modulation functions, a task that only takes about a minute. We demonstrate how models for local light interaction similar to the ones used by Wu et al. [2012] can be used to extract transient images from such a setup, thus enabling research on the applications of transient imaging on a much smaller budget.

Our specific technical contributions are as follows:

- We derive a theoretical model for the relationship between transient imaging and traditional time-of-flight imaging using photonic mixer devices (PMD, see Section 2).
- We formulate the estimation of transient images from PMD measurements as an inverse problem that is ill-conditioned but can be solved by introducing regularization terms as well as a model of local surface/light interactions, and by splitting the resulting hard problem into a sequence of simpler optimization problems (Section 3).
- We demonstrate a prototype system with a modified PMD camera that allows for flexible measurements of time-of-flight images using a range of different modulation frequencies and phases (Section 4). The system is inexpensive, portable, eye-safe, insensitive to background light, and acquires data at a much higher rate than streak cameras, enabling transient imaging even outside lab environments.

While our approach does currently not achieve quite the same temporal resolution as a streak camera setup, we believe that our contributions will significantly lower the barrier for conducting research in transient imaging in slightly larger environments, and will thus improve the practical applicability of many transient imaging methods that have already been explored.

2 PMD Image Formation Model

Photonic Mixer Devices (PMD) are time-of-flight image sensors where each pixel can direct the charge from incoming photons to two or more storage sites within the pixel [Schwarte et al. 1997; Schwarte 1997; Lange et al. 2000; Lange and Seitz 2001]. This operating principle is similar to recently proposed multi-bucket sensors [Wan et al. 2012], although PMDs can switch between the sites at much higher rates of dozens of MHz. This effectively allows PMDs to modulate the incident illumination with another signal during light integration, which, in combination with a modulated light source, can be used for time-of-flight depth imaging.

Recent improvements and extensions to PMD design and operation include heterodyne modulation of light and PMD sensor to improve resolution [Conroy et al. 2009; Dorrington et al. 2007], multi-path and scattering suppression for depth estimation [Dorrington et al. 2011], as well as tomography based on time of flight information [Hassan et al. 2010].

In our work we apply some of the same hardware strategies developed in these papers — increasing the number of modulation frequencies and phases — to acquire a large set of correlated modulation measurements. Using these measurements we derive a novel optimization strategy that allows us to infer the full transient image under global illumination and complex lighting by solving an optimization problem.

In the following we present a slightly simplified model for the imaging process of photonic mixer devices, and show how measurements obtained with PMD sensors relate to time-of-flight imaging and transient images. For detailed information operating principles of the hardware we refer the interested reader to the work by Schwarte et al. [1997; 1997].

PMD Sensor. Photonic mixer devices (PMDs, [Schwarte 1997; Schwarte et al. 1997]) are image sensors that can modulate the exposure incident at a single pixel with a periodic function f_ω during light integration, where f_ω is a zero-mean function with period $T = 2\pi/\omega$:

$$f_\omega(t + k \cdot T) = f_\omega(t); \qquad \int_0^T f_\omega(t)\, dt = 0. \tag{1}$$

Specifically, a PMD sensor measures the *modulated exposure*

$$H_{\omega,\phi} = \int_0^{NT} E(t) f_\omega(t + \phi)\, dt, \tag{2}$$

where ϕ is a programmable phase offset. f_ω is often a sinusoid, but may be any zero-mean periodic function, such as a rectangle wave.

Illumination. For time-of-flight (TOF) applications, PMDs are used in conjunction with illumination that is also modulated in intensity. Although some researchers have suggested working in a *heterodyne* setup [Dorrington et al. 2007; Conroy et al. 2009], where the light modulation frequency differs from the modulation frequency of the PMD, most current PMD applications assume a *homodyne* setup, where both frequencies are the same. In the homodyne setting, the modulation is provided by some function g_ω which may be different from f_ω, but has the same frequency and a fixed relative phase.

Furthermore, standard TOF applications assume that the modulated irradiance arriving at a sensor pixel is due only to *direct illumination* of a *single object point* by a *single point light source*, effectively meaning that only a single light path contributes to the sensor reading (see left of Figure 2). Under this assumption we obtain, again for a single pixel,

$$E_\omega(t) = E_0 + \alpha E_m g_\omega(t + \tau), \tag{3}$$

where E_0 is the DC component of the light source plus any ambient illumination, E_m is the modulation amplitude for the light source, α is an attenuation term due to surface reflectance and distance-based intensity falloff, and τ is the total *travel time* from the light source to the object point and then to the PMD pixel.

PMD Time-of-Flight Imaging. With the light and sensor modulations described above, the modulated exposure measured by the PMD sensor becomes

$$\begin{aligned} H_{\omega,\phi} = &\int_0^{NT} \left(E_0 + \alpha E_m g_\omega(t + \tau)\right) f_\omega(t + \phi)\, dt \\ =&E_0 \cdot N \underbrace{\int_0^T f_\omega(t + \phi)\, dt}_{=0} \\ &+ \alpha E_m \cdot \underbrace{N \int_0^T g_\omega(t + \tau) f_\omega(t + \phi)\, dt}_{=c_{\omega,\phi}(\tau)}. \end{aligned}$$

The correlation coefficients $c_{\omega,\phi}(\tau)$ can either be determined analytically for specific f_ω, g_ω such as sinusoids, or they can be calibrated using objects at different known distances d using the relationship $\tau = \omega/c \cdot d$, where c is the speed of light.

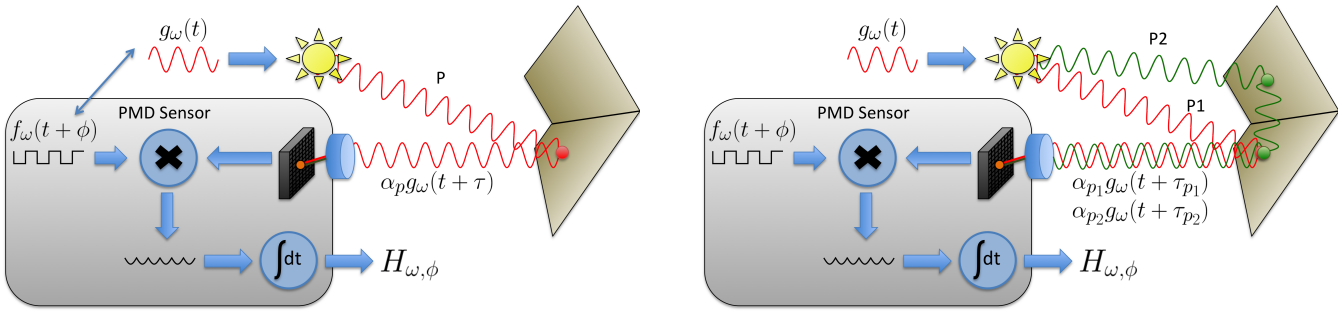

Figure 2: *Left: Operating principle of a conventional PMD time-of-flight sensor. Light from a modulated light source arrives at a pixel sensor via a single light path with time delay τ. The PMD sensor modulates the incident light with a reference signal f_ω and integrates the resulting modulated exposure to obtain a distance-dependent correlation between the two modulation functions. Right: in the presence of global illumination, the incident illumination is a superposition of light with different phase shifts. Multiple measurements at different frequencies and phases are taken to analyze such complex light interactions (see text for details).*

In normal TOF applications, two measurements $H_{\omega,0^\circ}$ and $H_{\omega,90^\circ}$ are obtained for per pixel. Using the known $c_{\omega,\phi}(\tau)$, it is then possible to solve for the pixel intensity αE_m and the distance d of the object visible at that pixel.

Multiple Modulation Frequencies and Phases. In our setting, we generalize this acquisition process to obtain measurements H_{ω_i,ϕ_i} for many different modulation frequencies and phases (ω_i, ϕ_i). This yields a collection of different, travel-time dependent correlation coefficients $c_{\omega_i,\phi_i}(\tau)$. While Dorrington et al. [2011] have recently also proposed to use more than one modulation frequency to improve depth estimates in the presence of indirect illumination, for our transient imaging system, we significantly increase the number to hundreds of different frequencies.

Global Illumination. Using a multitude of measurements over a wide frequency range, we can relax the requirement that light only arrives at the sensor via single light path, and move to a full global illumination model instead (see Figure 2, right). For the modulated exposure, this yields

$$H_{\omega_i,\phi_i} = E_m \int_0^\infty \alpha(\tau) \cdot \underbrace{N \int_0^T g_{\omega_i}(t+\tau) f_{\omega_i}(t+\phi_i)\, dt}_{=c_{\omega_i,\phi_i}(\tau)}\, d\tau, \tag{4}$$

where $\alpha(\tau)$ physically represents the integral of all contributions from different light paths p that correspond to the same travel time τ:

$$\alpha(\tau) = \int_{\mathcal{P}} \delta(|p| = \tau)\alpha_p \, dp. \tag{5}$$

Here, α_p is the light attenuation along a given ray path p connecting the light source and the sensor pixel, $\mathcal{P}$ is the space of all such ray paths, and $|p|$ is the travel time along path p.

Relationship to Transient Imaging. The sequence of $\alpha(\tau)$ for different values of τ is a *transient pixel*, which is the impulse response of a sensor element, i.e. the intensity profile of a pixel as a function of time when the scene is illuminated by a very short pulse of light. A regular grid of such transient pixels $\alpha_{x,y}(\tau)$ is known as a *transient image*

$$I(x, y, \tau) = \alpha_{x,y}(\tau).$$

A model for the formation of transient images was proposed by [Smith et al. 2008]. In recent experimental work, the authors used a femtosecond laser in conjunction with fast detectors (photodiodes, streak camera) to measure a version of $I(x, y, \tau)$ that is discretized in both space and time [Kirmani et al. 2009; Velten et al. 2013].

3 Transient Imaging with PMD Sensors

This section describes how we reconstruct a transient image from the sensor measurements H. We first formulate the corresponding optimization problem as a linear inverse problem and then present a method to solve this optimization problem.

3.1 An Optimization Problem for Transient Image Reconstruction

We wish to reconstruct a transient image I (discretized in space and time) based on a sequence H_{ω_i,ϕ_i} of modulated exposures that have been measured for each pixel in the image according to the model from Equation 4. The corresponding vectorized image (with the components stacked) is $\mathbf{i}$, and $\mathbf{h}$ denotes the vector of all modulated exposures observed at all pixels in the image. The correlation coefficients c_{ω_i,ϕ_i} can be arranged in a matrix $\mathbf{C}$ that maps a transient image to the vector of observations $\mathbf{h}$. Ignoring any measurement noise, this yields the following linear inverse problem for the recovery of $\mathbf{i}$, which is a discrete version of Equation 4:

$$\mathbf{C}\mathbf{i} = \mathbf{h} \tag{6}$$

This problem, however, is poorly conditioned, since the correlation coefficients vary slowly with distance for the range of feasible modulation frequencies and scene scales (Section 4). This makes it impossible to directly solve for $\mathbf{i}$.

To solve this challenging inverse problem, we introduce both spatial and temporal regularization terms, impose a non-linear model $m(\mathbf{u}) \approx \mathbf{i}$ for the *local* interaction of light with objects, and optimize for both the transient image $\mathbf{i}$ and the model parameters $\mathbf{u}$ in a least-squares sense:

$$\begin{aligned}(\mathbf{u}_{\text{opt}}, \mathbf{i}_{\text{opt}}) = \underset{\mathbf{u},\mathbf{i}}{\text{argmin}}\ & \frac{1}{2}\left\|\mathbf{C}\mathbf{i} - \mathbf{h}\right\|_2^2 + \\ & \lambda \sum_{\mathbf{x}} \left\|\nabla_\tau \mathbf{i}_{\mathbf{x}}\right\|_H + \theta \sum_{\tau} \left\|\nabla_{\mathbf{x}} \mathbf{i}_\tau\right\|_H + \\ & \frac{1}{2}\left\|\mathbf{C}m(\mathbf{u}) - \mathbf{h}\right\|_2^2 + \frac{\rho}{2}\left\|\mathbf{i} - m(\mathbf{u})\right\|_2^2,\end{aligned} \tag{7}$$

where the first line is a data term corresponding to the original linear inverse problem from Equation 6. The two terms in the second line respectively represent gradient regularizations along the temporal direction for each individual pixel, and along the spatial dimensions for each individual time slice. We expect the gradients for both the spatial and temporal slices to be sparse, but with occasional outliers (discontinuities). This motivates the use of robust norms in both regularizations terms. In our work we chose the Huber penalty function[1] $\|.\|_H$ since it suppresses outliers, but gives a smooth reconstruction for small gradient magnitudes.

Finally, the last line from Equation 7 contains two terms that fit the data with a model $m(\cdot)$, and tie the estimated model parameters $\mathbf{u}$ back to the reconstructed intrinsic image $\mathbf{i}$.

Thus, in Equation 7, we use three additional priors to regularize our strongly ill-conditioned problem from Equation 6. That is a) by using a model $m(\cdot)$, b) a temporal smoothness regularizer and c) a spatial smoothness regularization.

Our model is inspired by recent work by Wu et al. [2012] and applies to each pixel independently. Specifically, we model the light interaction at each local surface point as a combination of surface reflection and subsurface scattering. For surface reflection, the temporal impulse response to a short pulse of light is a Dirac peak, which we represent as a Gaussian G_σ, where σ is related to the temporal resolution of the acquisition system. For subsurface scattering, the temporal impulse response can be modeled as an exponential decay E [Wu et al. 2012]. In a complex scene with global illumination, the model $m_{\mathbf{x}}(\cdot)$ for the time profile at pixel $\mathbf{x}$ can therefore be expressed as a superposition of a number of Gaussians and exponentials

$$m_{\mathbf{x}}(\mathbf{u}) = \sum_i^{K_{\mathbf{x}}} \mathbf{g}_{\mathbf{x},i} G_\sigma(\tau - \mathbf{p}_{\mathbf{x},i}) + \mathbf{a}_{\mathbf{x},i} E(\mathbf{d}_{\mathbf{x},i}, \tau - \mathbf{p}_{\mathbf{x},i}), \quad (8)$$

where the model parameters $\mathbf{u}$ are the set of all (temporal) positions $\mathbf{p}_{\mathbf{x},i}$ for the Gaussians and exponentials, combined with their amplitudes $\mathbf{g}_{\mathbf{x},i}$ and $\mathbf{a}_{\mathbf{x},i}$, and finally the exponential decay parameters $\mathbf{d}_{\mathbf{x},i}$ which depend on the density of the scattering medium.

$$\mathbf{u} = \bigcup_{\mathbf{x},i} \{\mathbf{g}_{\mathbf{x},i}, \mathbf{p}_{\mathbf{x},i}, \mathbf{a}_{\mathbf{x},i}, \mathbf{d}_{\mathbf{x},i}\}.$$

3.2 A Splitting Algorithm for Reconstruction

With the model defined, we can now in principle minimize Equation 7 by alternating between two steps, one where we solve for $\mathbf{i}$, given fixed values of $\mathbf{u}$ from the previous iteration and vice-versa. This can be interpreted as splitting the non-linear and non-convex part of the joint optimization problem from the simple linear part, which results in two simpler subproblems are connected to each other with coupling terms (a common technique in optimization, e.g. [Beck and Teboulle 2009; Chambolle and Pock 2011]).

In the very first $\mathbf{i}$-step, all terms containing the model $m(\cdot)$ are ignored. However, the $\mathbf{u}$-subproblem is still non-convex, and so we add additional constraints to restrict the space of possible solutions for this subproblem to a region where we expect at least locally a convex function:

$$\begin{aligned}(\mathbf{u}_{\text{opt}}, \mathbf{i}_{\text{opt}}) = \underset{\mathbf{u},\mathbf{i}}{\text{argmin}}\ & \frac{1}{2}\|\mathbf{C}\mathbf{i} - \mathbf{h}\|_2^2 + \\ & \lambda \sum_{\mathbf{x}} \|\nabla_\tau \mathbf{i}_{\mathbf{x}}\|_H + \theta \sum_{\tau} \|\nabla_{\mathbf{x}} \mathbf{i}_\tau\|_H + \\ & \frac{1}{2}\|\mathbf{C}m(\mathbf{u}) - \mathbf{h}\|_2^2 + \frac{\rho}{2}\|\mathbf{i} - m(\mathbf{u})\|_2^2 + \\ & \frac{\beta}{2} \sum_{\mathbf{x}} \|\text{maxima}(\mathbf{i}_{\mathbf{x}}) - \mathbf{p}_{\mathbf{x}}\|_2^2 \\ & \text{s.t.} \quad \forall \mathbf{x}, i : K_{\mathbf{x}} = \#\text{maxima}(\mathbf{i}_{\mathbf{x}}) \ \wedge \\ & \qquad \mathbf{g}_{\mathbf{x},i}, \mathbf{a}_{\mathbf{x},i}, \mathbf{d}_{\mathbf{x},i} \in [0,1]\end{aligned} \quad (9)$$

Here $\text{maxima}(\mathbf{i}_{\mathbf{x}})$ are the positions of all local time maxima for a pixel $\mathbf{x}$, and $\#\text{maxima}(\mathbf{i}_{\mathbf{x}})$ is number of such local maxima. The two additional constraints therefore express that we prefer locations of Gaussians and exponentials close to the onset positions, that we restrict the number $K_{\mathbf{x}}$ to be the number of onsets, and constrain the numerical range of Gaussian and exponential amplitudes and decays.

We now discuss in detail how to solve the two subproblems. We initialize our method with $\mathbf{i} = 0, \mathbf{u} = 0$.

3.2.1 Solving the i-subproblem

By fixing $\mathbf{u}$ in Equation 9 and neglecting the terms constraining the location of maxima, we obtain the following optimization problem for $\mathbf{i}$:

$$\begin{aligned}\mathbf{i}_{\text{opt}} = \underset{\mathbf{i}}{\text{argmin}}\ & \frac{1}{2}\|\mathbf{C}\mathbf{i} - \mathbf{h}\|_2^2 + \\ & \lambda \sum_{\mathbf{x}} \|\nabla_\tau \mathbf{i}_{\mathbf{x}}\|_H + \theta \sum_{\tau} \|\nabla_{\mathbf{x}} \mathbf{i}_\tau\|_H + \\ & \frac{\rho}{2}\|\mathbf{i} - m(\mathbf{u})\|_2^2\end{aligned} \quad (10)$$

In this subproblem, neglecting the maxima terms is justified by the fact that these terms will be dominated by $\frac{\rho}{2}\|\mathbf{i} - m(\mathbf{u})\|_2^2$ for large ρ (recall that we only introduced the extra terms to account for the non-convexity of the $\mathbf{u}$-subproblem).

Equation 10 is a regularized linear-inverse problem. We solve it with the first-order primal-dual framework by Chambolle and Pock [2011], which considers general problems of the form

$$\mathbf{i}_{\text{opt}} = \underset{\mathbf{i}}{\text{argmin}}\ \mathbf{F}(\mathbf{K}\mathbf{i}) + \mathbf{G}(\mathbf{i}), \quad (11)$$

where the operators $\mathbf{F}, \mathbf{K}, \mathbf{G}$ are the same as in [Chambolle and Pock 2011]. For example, in an inverse problem with a TV regularizer, the first term in Equation 11 is the regularizer (that is $\mathbf{K}(\mathbf{i}) = \nabla\mathbf{i}$, $\mathbf{F}(\mathbf{j}) = \|\mathbf{j}\|_1$ for TV), while the second term is the data fitting term. Equation 10 can be expressed in this framework by defining

$$\begin{aligned}\mathbf{K}(\mathbf{i}) &= \mathbf{S}\mathbf{i} \\ \mathbf{F}(\mathbf{j}) &= \|\mathbf{j}\|_H \\ \mathbf{G}(\mathbf{i}) &= \frac{1}{2}\|\mathbf{C}\mathbf{i} - \mathbf{h}\|_2^2 + \frac{\rho}{2}\|\mathbf{i} - m(\mathbf{u})\|_2^2,\end{aligned} \quad (12)$$

where here the matrix $\mathbf{S}$ applies the temporal derivative operator to all pixel time sequences and the spatial gradient operators to all time slices (according to Eq. 10).

To solve this problem with Chambolle and Pock's algorithm, we need to derive the *proximal operators* for $\mathbf{F}^*$ and $\mathbf{G}$. These are provided in Appendix A.

[1] The Huber penalty function [Huber 1973] is an ℓ_1/ℓ_2 hybrid error measure and defined as

$$\|\mathbf{x}\|_H = \sum_i h_\epsilon(\mathbf{x}_i) \quad \text{with} \quad h_\epsilon(x) = \begin{cases} |x| - \epsilon/2 & ; |x| > \epsilon \\ x^2/(2\epsilon) & ; \text{else} \end{cases}$$

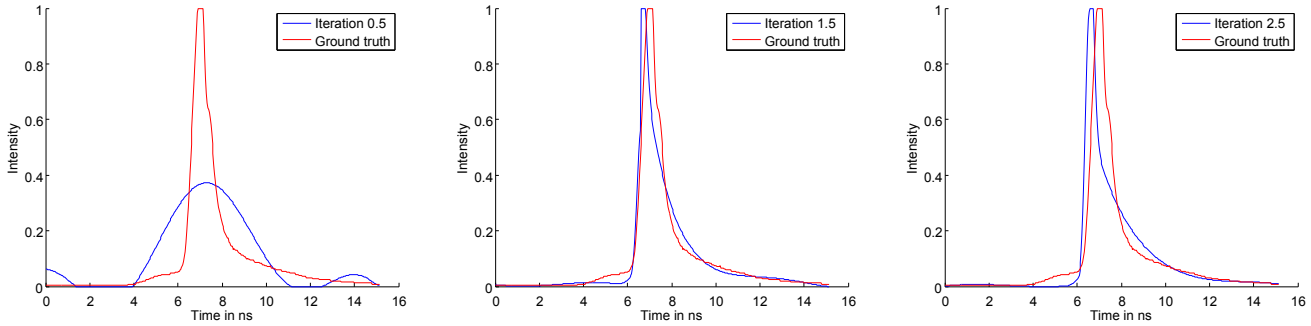

Figure 3: *A single synthetic pixel signal (over time) reconstructed with Eq. 7. Left: Solution obtained by solving for* $\mathbf{i}$ *only. Notice that the solution is smooth in temporal and spatial domain and contains the right number of modes. Center: solution after one full iteration plus one extra* $\mathbf{i}$*-step to reconstruct the transient image from the mixed Gaussian and exponential model parameters* $\mathbf{u}$*. Notice the drastically improved reconstruction fidelity. Right: an additional iteration does not improve the result significantly.*

3.2.2 Solving the u-subproblem

In the second subproblem, we fix $\mathbf{i}$ in Equation 9 and solve $\mathbf{u}$, leading to the following problem:

$$\begin{aligned}\mathbf{u}_{\text{opt}} = \underset{\mathbf{u}}{\text{argmin}} \; & \frac{1}{2}\left\|\mathbf{C}m(\mathbf{u}) - \mathbf{h}\right\|_2^2 + \frac{\rho}{2}\left\|\mathbf{i} - m(\mathbf{u})\right\|_2^2 + \\ & \frac{\beta}{2}\sum_{\mathbf{x}}\left\|\text{maxima}(\mathbf{i}_{\mathbf{x}}) - \mathbf{p}_{\mathbf{x}}\right\|_2^2 \\ \text{s.t.} \quad & \forall \mathbf{x}, i : K_{\mathbf{x}} = \#\text{maxima}(\mathbf{i}_{\mathbf{x}}) \;\wedge \\ & \mathbf{g}_{\mathbf{x},i}, \mathbf{a}_{\mathbf{x},i}, \mathbf{d}_{\mathbf{x},i} \in [0,1]\end{aligned} \tag{13}$$

This is a regularized data-fitting problem that is globally still non-convex, so there are no guarantees of obtaining a global optimum. However, by constraining the parameters, and especially the onset $\mathbf{p}$, we limit the search to regions where we expect locally convex behavior.

We first determine $K_{\mathbf{x}} = \#\text{maxima}(\mathbf{i}_{\mathbf{x}})$ by explicitly detecting the number of peaks in time for each pixel signal $\mathbf{i}_{\mathbf{x}}$. No noise suppression or further smoothing is needed since we require only a good lower bound for $K_{\mathbf{x}}$. Accurate fitting is then handled by solving Equation 13.

The resulting problem (with $K_{\mathbf{x}}$ now fixed) can be solved independently for each pixel. To account for the non-convexity we perform a two-step optimization per pixel signal: first, several steps of a global search in the constrained parameter space are done using the derivative-free Mesh Adaptive Direct Search method for non-linear constrained optimization (MADS) [Audet and Dennis Jr 2006]. This direct search method models the parameter space using a mesh. By adaptively probing the mesh and also adaptively changing the mesh refinement this method makes good progress in regions around local minimizers. We use the implementation in the MATLAB generalized pattern search function "patternsearch" in the Global Optimization Toolbox, random polling and a maximum number of 1000 iterations.

We expect the parameters found by this derivative-free global optimization to lie at least in a locally convex region. To effectively find the according minimum in this region we now use the gradient and do local constrained gradient-descent iterations using the spectral projected gradient method from the minConf package [Schmidt 2013]. The projection operator that encodes our constraints is straightforward to derive since in our case we only have simple box constraints.

We use here analytic gradient information of the unconstrained version of Eq. 13, i.e.,

$$\begin{aligned}\mathbf{u}_{\text{opt}} = \underset{\mathbf{u}}{\text{argmin}} \; & \frac{1}{2}\left\|\mathbf{C}m(\mathbf{u}) - \mathbf{h}\right\|_2^2 + \frac{\rho}{2}\left\|\mathbf{i} - m(\mathbf{u})\right\|_2^2 + \\ & \underbrace{\frac{\beta}{2}\sum_{\mathbf{x}}\left\|\text{maxima}(\mathbf{i}_{\mathbf{x}}) - \mathbf{p}_{\mathbf{x}}\right\|_2^2}_{\Phi(\mathbf{u})}\end{aligned} \tag{14}$$

For this subproblem, all the components $\mathbf{u}_{\mathbf{x}}$ for a pixel $\mathbf{x}$ are independent of the components for other pixels. So, for a single component j of $\mathbf{u}$, we obtain the following gradient component:

$$\nabla_j \Phi(\mathbf{u}) = 2\left[\frac{\partial m_{\mathbf{x}}(\mathbf{u})}{\partial \mathbf{u}_j}\right]^T \mathbf{C}^T\left(\mathbf{C}m_{\mathbf{x}}(\mathbf{u}) - \mathbf{h}\right) \tag{15}$$

Thus, with Eq. 15, we have reduced the gradient computation for $\Phi(\mathbf{u})$ to the evaluation of the partial derivatives of a pixel time sequence $m_\tau(\mathbf{u})$ with respect to its parameters (that is the Gaussian amplitude, exponential amplitude, exponential decay and the position of the Gaussian and exponential). These partial derivatives are straightforward to derive from Equation 8.

3.3 Discussion

With optimization split into two stages as described above, the $\mathbf{i}$-subproblem is fast and convex, and provides a data term plus spatial and temporal regularization, with the tendency to overly smoothen the results. The $\mathbf{u}$-subproblem, on the other hand, is a very expensive, non-convex fitting of the model parameters, that produces higher temporal resolution, but may in turn introduce some visually objectionable spatial artifacts. After the final $\mathbf{u}$-step, we always perform another $\mathbf{i}$-step to obtain a final transient image from the model parameters.

Figure 3 compares the time sequences obtained for a single pixel with ground truth data for a synthetic example (more details on the dataset can be found in Section 5). The left image shows the over-smooth reconstruction after solving just the $\mathbf{i}$ subproblem. The center and right image show reconstructions after one and two full iterations, respectively (each followed by a final $\mathbf{i}$-step). We can see that the method is converged after only a single iteration. This is consistent with behavior we observe for the time sequences of other pixels and datasets.

Figure 4 shows results for a simple scene where a wavefront travels along a diffuse wall, captured using our PMD camera setup (Sec-

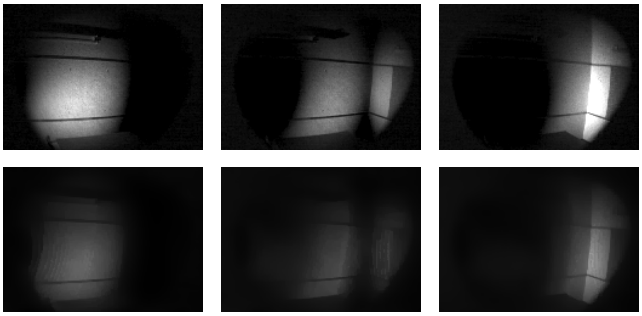

Figure 4: *Reconstructions for a simple real dataset of a wavefront traveling along a wall. Top row: image reconstruction after only solving the* **i** *subproblem. Bottom: the same images after one iteration of the full optimization problem.*

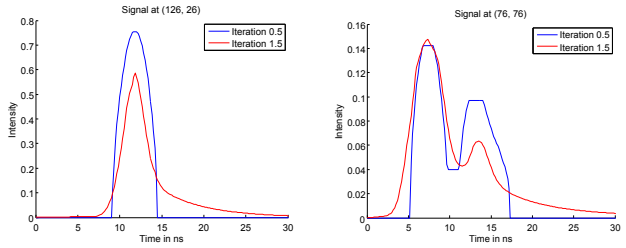

Figure 5: *Pixel time sequences for two selected pixels of Figure 4 after a full iteration and after only solving the* **i** *subproblem.*

tion 4). Using the model, pixel time profiles (Figure 5) show a similar sharpening effect as in the synthetic data set. However, as the full-frame reconstructions show, this extra sharpness comes at the cost of some spatial quantization artifacts. We refer to the accompanying video to observe the full dataset. Because of the tradeoff between temporal sharpness and spatial artifacts, it may in some cases be preferable to only run the **i** subproblem, depending on whether the goal is to obtain the most detailed measurements possible, or simply to produce visually pleasing results.

4 Prototype Hardware Setup

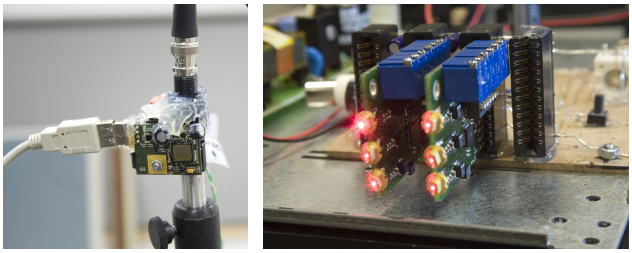

Figure 6: *The main components of our capture setup. Left: Modified PMD imager. Right: Light source with six uncollimated laser diodes.*

We now describe the physical setup we developed to acquire transient images of real scenes.

The camera is based on the PMDTechnologies CamBoard nano, an evaluation platform for the PMD PhotonICs 19k-S3 image sensor. We removed the on-board light source and the default 830 nm longpass filter. Since our technique is based on the use of a wide range of modulation frequencies, we intercepted the on-board modulation signal and replaced it with our own input from an external source, and added a trigger output that signals the start of the integration phase. We confirmed that the sensor can be operated at modulation frequencies up to 180 MHz, but managed to obtain stable results only up to 110 MHz.

Our light source is a bank of six 650 nm laser diodes with a total average output power of 2.4 W. Since the beams are not collimated, eye safety is not a concern at reasonable distances from the device. Two groups of three diodes are driven through a iC-Haus 6-channel constant current driver each (type iC-HG), with each diode occupying two driver outputs for DC bias and the high-frequency modulation. Using a fast photodiode (Thorlabs FDS010) and a 500 MHz, 2 GS/s oscilloscope, we confirm that this setup can modulate the output intensity by a full 100% up to a frequency of 180 MHz.

As our modulation source signal, we use the DDS9m/02 signal generator from Novatech Instruments, which is based on the Analog Devices AD9958 direct digital synthesizer chip. We use the digital output of the signal generator to modulate both the light source and the camera with a square wave.

A microcontroller circuit sets modulation frequencies and phases on the synthesizer board. Reacting to the trigger signal from the camera, it switches the modulation signals for light source and camera in accordance with the currently set integration time (we take an exposure series from 120 to 1920 µs in steps of one f-stop). The sensor data is read out through the API provided with the camera board.

The hardware components add up to a total cost of approximately $1500, which is well within reach for most research labs. We note that the lab signal generator alone accounts for more than half of the total budget. The Analog Devices chip that this signal generator is based on is available for only $35, so that substantial savings are possible for future revisions of our setup.

Measurement Routine. In order to complete a full measurement over a frequency range from 10 to 120 MHz in steps of 0.5 MHz, our capture system takes about 90 seconds. Note that we operate the setup at a duty cycle of less than 1%, in order to avoid overheating of the critical components (signal buffer and laser driver ICs, laser diodes, PMD sensor) that are not equipped with heatsinks or active cooling. We therefore estimate that with proper thermal management, another significant speedup will be achieved, reducing the overall acquisition time to only a few seconds.

There are several ways our methods deals with the dynamic range issues. The working principle of PMD sensors is itself is very effective in suppressing ambient illumination, the sensor provides high-bit depth readouts (14 bits) and, finally, we do take an exposure sequence as described just above. That said, if very bright light paths and very dim light paths mix in the same pixel, the reconstruction quality of the dim paths will suffer.

Calibration. In order to obtain the correlation matrix $\mathbf{C}$, we perform a calibration step. We place camera and light source as close to each other as possible, and facing a diffuse white wall, with no other objects nearby that could scatter light into the line of sight. For each frequency, we sample the distance-dependent correlation coefficients by varying the relative phase between the sensor and the light source modulation signals. This allows us to emulate different optical delays without mechanically moving parts.

The calibration measurements for frequencies from 10 to 120 MHz in 1 MHz steps and distances from 0 to 20 m in 0.1 m steps take about 6 hours to complete, with a further 30 minutes to extract the matrix from the raw sensor readings. We average the calibration

matrices over a 40×40 pixel region in order to reduce acquisition noise.

We note, however, that the calibrated matrix $\mathbf{C}$ obtained in this fashion is also valid for different geometric configurations of sensor and camera, with only a change in the physical interpretation of the resulting reconstructed transient images. This means that the calibration step is essentially a one-time operation. Figure 7 shows a color visualization of the recovered correlation coefficients, with the vertical axis corresponding to different frequencies, and the horizontal axis corresponding to different path lengths, or travel times.

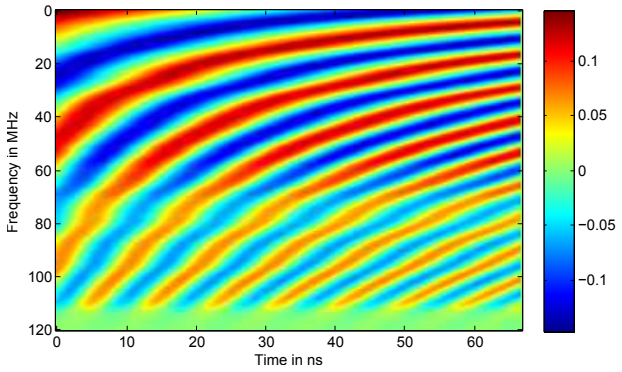

Figure 7: *A visualization of the correlation coefficients for different path lengths (horizontal axis) and modulation frequencies (vertical axis).*

5 Results

We evaluated our approach using both synthetic data for ground truth comparisons, and measurements using our custom setup.

5.1 Synthetic Examples

For our synthetic experiments, we used a transient image that Velten et al. obtained using a femtosecond laser and streak camera [Velten et al. 2013]. Since this dataset was captured for a smaller scene than the size of scenes we target with our setup, we simulated a larger scene by scaling the time dimension by a factor of 20. We then used the PMD sensor model from Equation 4 to simulate measurements of H_{ω_i,ϕ_i} for different frequencies and phases, also adding Gaussian noise with a sigma of 1%.

Figure 9 shows reconstructed frames (bottom row) in comparison with the ground truth (top). The key features are reproduced, although a certain amount of temporal smoothing is noticeable. An examination of time profiles for additional pixels confirms this analysis (Figure 10). In green, we also show a direct least-squares fit of the model to the ground truth curves. These curves demonstrate the expressiveness of the model. Even though the direct fit exhibits shows sharp discontinuities due to the use of exponential segments with a step function onset, the examples show that the key features of the time sequences can be represented using the model. Furthermore, in the actual reconstruction from PMD measurements, the $\mathbf{i}$-step smoothes out the discontinuities. These results demonstrate that our method can qualitatively reconstruct transient images, including complex scenarios with multiple discrete pulses of light arriving at a surface point, as well as broad temporal profiles resulting from indirect illumination via a diffuse surface. A full quantitative analysis of the resolution limit for our method is complicated by the non-linear nature of our optimization problem and is beyond the scope of this paper.

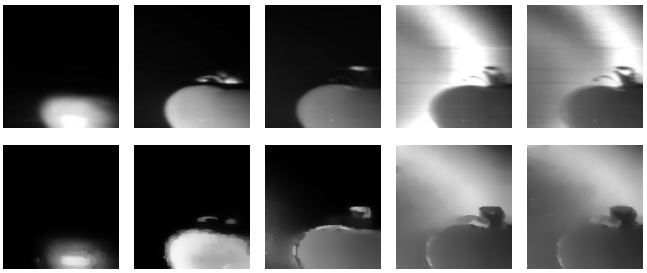

Figure 9: *A couple of selected frames from simulation with ground-truth data. Top: Original ground-truth frames. Bottom: Reconstructions using our method. Temporal profiles for a few representative pixels are shown in Figure 9. See supplemental material for the full video.*

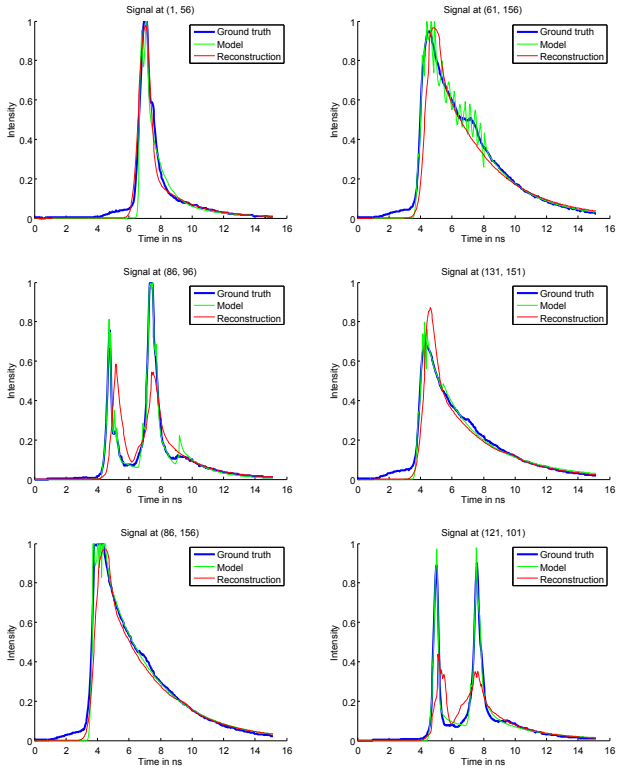

Figure 10: *A selection of time profiles for different pixels from the example in Figure 9. Each graph shows the original time profile, which we treat as ground truth (blue), a direct fit of the model to this ground truth curve (green), and finally the result of our full optimization procedure using simulated PMD measurements (red, see text).*

5.2 PMD Measurements

Using our own setup, we captured a few settings with characteristic light transport, including the simple scene from Figure 4 already discussed in Section 3. Individual time slices of three additional datasets are shown in Figure 8. The scenes and single-image visualizations of the recovered transient images are shown in Figures 1 and 11. We also encourage the reader to watch the accompanying video to see the full transient images.

The top row of Figure 8 shows a wavefront propagating though a scene with a mirrored disco ball placed in the corner of a room.

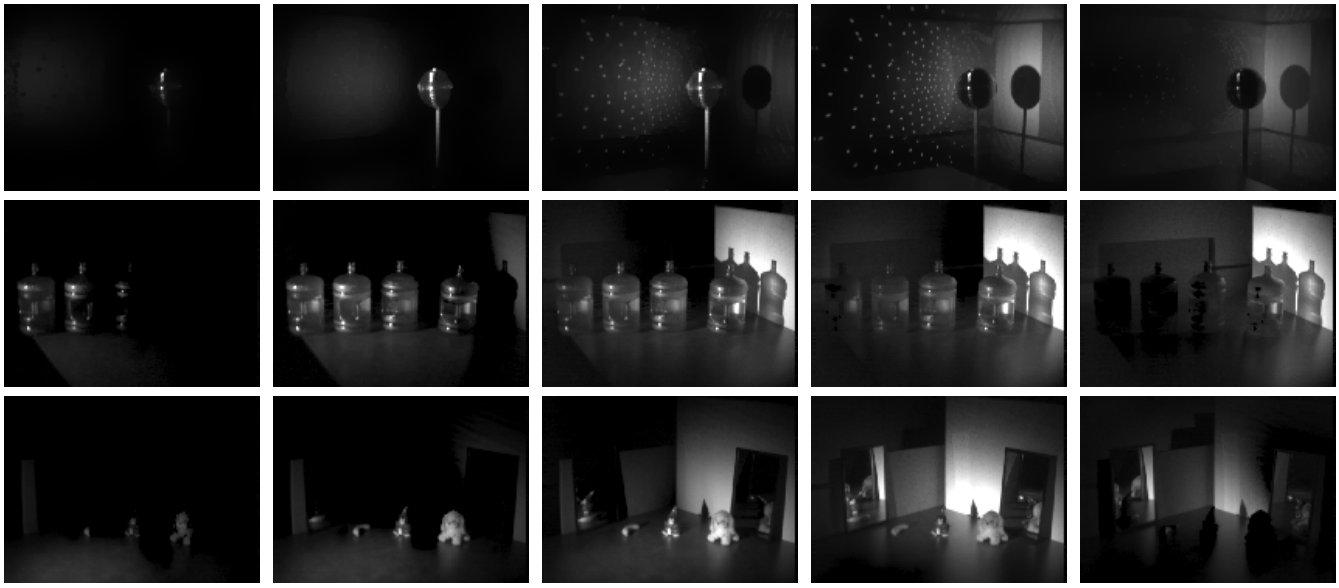

Figure 8: *Time slices from three transient images captured with our setup, and reconstructed with the method from Section 3.* **Top row:** *we see light hitting a disco ball. We observe the wavefront propagating along the walls, and the caustics from the mirrors arriving shortly after.* **Center row***: light propagates through a scene containing several bottles filled with water. We see first reflection of the surface of the bottles, followed by scattering inside the bottles, caustics, and light scattering onto the back wall.* **Bottom row:** *A scene with several objects and two mirrors. We observe the objects being illuminated first, followed by the reflections of the object in the mirrors, and finally a reflection of the right mirror in the left one. More detailed explanations of the sequences can be found in the text.*

In the first frame, the wavefront has just reached the front of the ball. In the second frame, the ball is now fully illuminated, and we see the wavefront propagating along the left wall. The third frame shows the first caustics generated by reflections in the mirror. More caustics appear for longer light paths near the top and bottom of the fourth image, and the direct illumination is now approaching the back of the corner from both sides. First indirect illumination in the floor is visible. In the last frame, the caustics have disappeared, and the indirect illumination is now lighting up the shadow of the ball.

The second row of the figure shows a scene with several bottles, filled with water and a small amount of milk to create scattering. In the first frame, the wavefront has just reached the front of the left-most bottles, and is reflecting off their surface. In the second frame, scattering effects are becoming visible in the bottles. Next, the light reaches the far wall, showing caustics of the light transport through the bottles. Indirect illumination of the back wall from light scattered in the bottles appears in the fourth frame. This light continues to illuminate the back ball even after the bottles themselves have darkened (last frame).

Finally, the bottom row of Figure 8 shows a scene with several foreground objects and two mirrors. We first see initial reflections coming off the foreground objects. As the light propagates further, the foreground objects are now fully illuminated, and the wavefront reaches the back walls, but the mirrors remain dark. In the third frame reflections of the foreground objects are starting to become visible in both the left and the right mirror. In the fourth frame, the left mirror shows a reflection of an object in the right mirror. This reflection lingers in the last frame, even after the wavefront has passed by the foreground objects.

Figure 11: *Our test scenes. All scenes are illuminated from the left, with the rainbow images encoding the temporal location of the main intensity peak for each pixel. Top: Various objects on a table, and their reflections in mirrors placed nearby. Note how the items and mirrored counterparts light up at different times. Bottom: Four bottles filled with slightly scattering water. The specular reflections reach the camera before the light scattered inside the bottles.*

6 Discussion and Conclusions

In this paper we presented a method for transient imaging with PMD sensors, a technology that is widely available in commercial time-of-flight cameras. The hardware was modified to allow for different modulation frequencies and phases for both the sensor and the illumination. Using a simple model for local light/surface interaction, we can robustly reconstruct transient images from measurements made with this hardware.

Unlike existing systems for the capture of transient images, ours does not require ultrafast temporal sensing or illumination, but

works with correlation measurements obtained over longer integration periods. As a result, we achieve transient images with hardware costing a fraction of the price of the femtosecond laser/streak camera combination described in the literature. Additional benefits of our approach include significantly reduced acquisition times, more portable hardware, and easier modes of operation.

Although our prototype uses only red laser diodes as illumination, true color measurements could be obtained easily by adding green and blue lighting to the system and measuring the corresponding channels in a time-sequential fashion.

A disadvantage of our system at the moment is the limited spatial and temporal resolution of the sensor used. The PMD PhotonICs 19k-S3 has a spatial resolution of only 160×120 pixels, although higher resolution variants of the technology are likely to become available in the near future. The technology is also inexpensive enough that multiple sensors could be tiled together for applications that require higher resolution. The time resolution of our approach is limited by the maximum modulation frequency that can be applied to the illumination and the PMD sensor. In our prototype we were able to achieve modulation frequencies of up to 110 MHz, which corresponds to a spatial wavelength of about 2.70 m. The size of the investigated scene should not be significantly below this wavelength, otherwise the correlation matrix becomes too ill-conditioned to reliably solve the problem. We therefore work with larger-scale scenes than Velten et al. [2013].

In terms of computational post processing, solving the first half of the optimization problem (the $\mathbf{i}$-subproblem) takes only a few minutes at the full sensor resolution, and can thus provide a quick preview of a smoothed reconstruction. The second step (the $\mathbf{u}$-subproblem) is much more expensive, and takes several hours per dataset. An interesting avenue for future research is to determine ways to accelerate this step or derive alternative local models that can be fit more efficiently. It would also be interesting to add a spatial component to the model itself to avoid some of the spatial quantization and banding artifacts that can occur in some results.

Our approach models the transient image as a sparse signal in a basis that corresponds to certain light-object interactions, currently surface reflection and subsurface scattering. In the future, this basis could be extended to model other types of light-object interaction with higher fidelity.

In summary, we have demonstrated the combination of an inexpensive hardware setup and an optimization problem to acquire transient images using PMD sensors. It is our hope that this approach will reduce the barrier of entry for performing research in transient imaging and its applications, including the use of transient images for reconstructing geometry from indirect illumination [Kirmani et al. 2009; Pandharkar et al. 2011; Velten et al. 2012], and analyzing reflectance properties [Naik et al. 2011; Wu et al. 2012].

Acknowledgments

We thank PMD Technologies for helpful advice on pinout and operation of their sensors. We further thank Andreas Velten and Ramesh Raskar for sharing their "tomato" dataset, Lukas Chrostowski and Owen Roberts for their valuable input on modulation of lasers, and Mark Greenstreet for lending us the fast oscilloscope.

A Proximal Operators for the i-Subproblem

Using $\mathbf{i}$ as the primary domain parameters and $\mathbf{j}$ as the dual domain parameters, the proximal operators for the optimization problem defined in Equation 12 can be derived as

$$\begin{aligned} \mathbf{j} = \mathbf{prox}_{\sigma \mathbf{F}^*}(\tilde{\mathbf{j}}) &\Leftrightarrow \mathbf{j}_i = \frac{\frac{\tilde{\mathbf{j}}_i}{1+\sigma\epsilon}}{\max\left(1, \left|\frac{\tilde{\mathbf{j}}_i}{1+\sigma\epsilon}\right|\right)} \\ \mathbf{i} = \mathbf{prox}_{\tau \mathbf{G}}(\tilde{\mathbf{i}}) &\Leftrightarrow \left[\tau \mathbf{C}^T \mathbf{C} + \mathbb{I}\right] \mathbf{i} = \tau \mathbf{C}^T \mathbf{h} + \tilde{\mathbf{i}}, \end{aligned} \tag{16}$$

where in this equation, σ and τ are parameters described in Chambolle and Pock's original paper [2011], and ϵ is the parameter of the Huber norm, which defines the transition point between the linear and the quadratic segment of that norm.

Note that the first proximal operator amounts to a simple pointwise evaluation. The second operator is a linear system that we solve to high accuracy using SVD.

References

ABRAMSON, N. 1978. Light-in-flight recording by holography. *Optics Letters 3*, 4, 121–123.

ABRAMSON, N. 1983. Light-in-flight recording: high-speed holographic motion pictures of ultrafast phenomena. *Applied optics 22*, 2, 215–232.

AUDET, C., AND DENNIS JR, J. 2006. Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on optimization 17*, 1, 188–217.

BECK, A., AND TEBOULLE, M. 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imag. Sci. 2*, 183–202.

CHAMBOLLE, A., AND POCK, T. 2011. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis. 40*, 120–145.

CONROY, R., DORRINGTON, A., KÜNNEMEYER, R., AND CREE, M. 2009. Range imager performance comparison in homodyne and heterodyne operating modes. In *Proc. SPIE*, vol. 7239, 723905.

DORRINGTON, A., CREE, M., PAYNE, A., CONROY, R., AND CARNEGIE, D. 2007. Achieving sub-millimetre precision with a solid-state full-field heterodyning range imaging camera. *Measurement Science and Technology 18*, 9, 2809.

DORRINGTON, A., GODBAZ, J., CREE, M., PAYNE, A., AND STREETER, L. 2011. Separating true range measurements from multi-path and scattering interference in commercial range cameras. In *Proc. SPIE*, vol. 7864.

HASSAN, A., KÜNNEMEYER, R., DORRINGTON, A., AND PAYNE, A. 2010. Proof of concept of diffuse optical tomography using time-of-flight range imaging cameras. In *Proc. Electronics New Zealand Conference*, 115–120.

HUBER, P. 1973. Robust regression: asymptotics, conjectures and monte carlo. *The Annals of Statistics 1*, 5, 799–821.

KIRMANI, A., HUTCHISON, T., DAVIS, J., AND RASKAR, R. 2009. Looking around the corner using transient imaging. In *Proc. ICCV*, 159–166.

LANGE, R., AND SEITZ, P. 2001. Solid-state time-of-flight range camera. *IEEE J. Quantum Electronics 37*, 3, 390–397.

LANGE, R., SEITZ, P., BIBER, A., AND LAUXTERMANN, S. 2000. Demodulation pixels in CCD and CMOS technologies for time-of-flight ranging. *Sensors and camera systems for scientific, industrial, and digital photography applications*, 177–188.

NAIK, N., ZHAO, S., VELTEN, A., RASKAR, R., AND BALA, K. 2011. Single view reflectance capture using multiplexed scattering and time-of-flight imaging. *ACM Trans. Graph. 30*, 6, 171.

NILSSON, B., AND CARLSSON, T. E. 1998. Direct three-dimensional shape measurement by digital light-in-flight holography. *Applied Optics 37*, 34, 7954–7959.

PANDHARKAR, R., VELTEN, A., BARDAGJY, A., LAWSON, E., BAWENDI, M., AND RASKAR, R. 2011. Estimating motion and size of moving non-line-of-sight objects in cluttered environments. In *Proc. CVPR*, 265–272.

SCHMIDT, M., 2013. minConf (retrieved Jan 7, 2013). `http://www.di.ens.fr/~mschmidt/Software/minConf.html`.

SCHWARTE, R., XU, Z., HEINOL, H., OLK, J., KLEIN, R., BUXBAUM, B., FISCHER, H., AND SCHULTE, J. 1997. New electro-optical mixing and correlating sensor: facilities and applications of the photonic mixer device. In *Proc. SPIE*, vol. 3100, 245–253.

SCHWARTE, R., 1997. Verfahren und Vorrichtung zur Bestimmung der Phasen und/oder Amplitudeninformation einer elektromagnetischen Welle. German Patent DE 19704496.

SMITH, A., SKORUPSKI, J., AND DAVIS, J. 2008. Transient rendering. Tech. Rep. UCSC-SOE-08-26, University of California Santa Cruz, School of Engineering.

VELTEN, A., RASKAR, R., AND BAWENDI, M. 2011. Picosecond camera for time-of-flight imaging. In *OSA Imaging Systems and Applications*.

VELTEN, A., WILLWACHER, T., GUPTA, O., VEERARAGHAVAN, A., BAWENDI, M., AND RASKAR, R. 2012. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications 3*, 745.

VELTEN, A., WU, D., JARABO, A., MASIA, B., BARSI, C., JOSHI, C., LAWSON, E., BAWENDI, M., GUTIERREZ, D., AND RASKAR, R. 2013. Femto-photography: Capturing and visualizing the propagation of light. *ACM Trans. Graph. 32*.

WAN, G., LI, X., AGRANOV, G., LEVOY, M., AND HOROWITZ, W. 2012. CMOS image sensors with multi-bucket pixels for computational photography. *IEEE J. Solid-State Circuits 47*, 4, 1031–1042.

WU, D., O'TOOLE, M., VELTEN, A., AGRAWAL, A., AND RASKAR, R. 2012. Decomposing global light transport using time of flight imaging. In *Prov. CVPR*, 366–373.