Diffuse Mirrors: 3D Reconstruction from Diffuse Indirect Illumination Using Inexpensive Time-of-Flight Sensors

Felix Heide[1] Lei Xiao[1] Wolfgang Heidrich[1,3] Matthias B. Hullin[2]
1) The University of British Columbia
2) University of Bonn
3) King Abdullah University of Science and Technology

Abstract

The functional difference between a diffuse wall and a mirror is well understood: one scatters back into all directions, and the other one preserves the directionality of reflected light. The temporal structure of the light, however, is left intact by both: assuming simple surface reflection, photons that arrive first are reflected first. In this paper, we exploit this insight to recover objects outside the line of sight from second-order diffuse reflections, effectively turning walls into mirrors. We formulate the reconstruction task as a linear inverse problem on the transient response of a scene, which we acquire using an affordable setup consisting of a modulated light source and a time-of-flight image sensor. By exploiting sparsity in the reconstruction domain, we achieve resolutions in the order of a few centimeters for object shape (depth and laterally) and albedo. Our method is robust to ambient light and works for large room-sized scenes. It is drastically faster and less expensive than previous approaches using femtosecond lasers and streak cameras, and does not require any moving parts.

1. Introduction

Object reconstruction from real-world imagery is one of the central problems in computer vision, and researchers agree that the very mechanism of image formation (each pixel measuring light flux as a multidimensional plenoptic integral) is one of the main reasons why it is so challenging. To overcome the limitations of standard monocular images taken under uncontrolled illumination with respect to many vision tasks, a wide range of novel capturing approaches has emerged that extend the concept of digital imaging with structured light or new sensing techniques involving masks, filters or integral optics (light fields) [17].

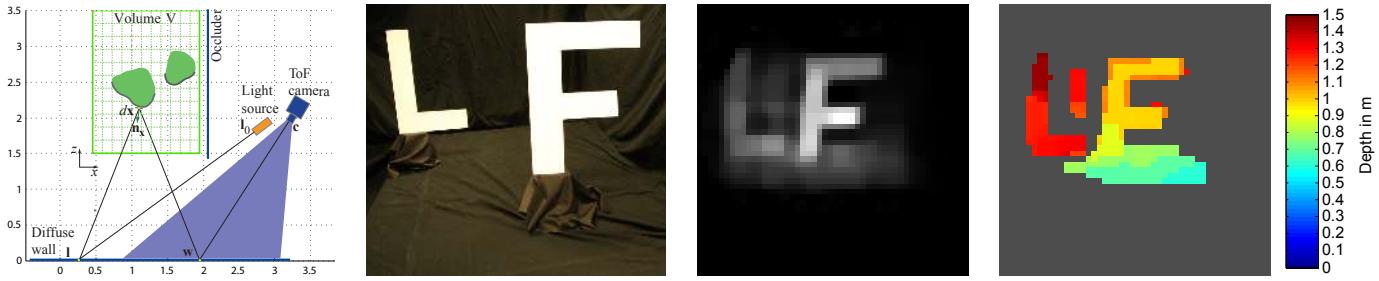
Most recently, researchers have started probing the temporal response of macroscopic scenes to non-stationary illumination, effectively resolving light contributions by the total length of the optical path [1, 11]. Experimental evidence suggests that such unmixing of light contributions will benefit many challenges in computer vision, including the use of diffuse reflectors to image objects via the time profile of reflections from ultra-short laser pulses, so-called *transient images* [20, 7]. However, reconstruction of the this data from transient images is a numerically ill-posed problem, that is sensitive to the exact parametrization of the problem as well as the priors and regularization terms that are employed. In this paper, develop a new parametrization for this inverse problem, and combine it with a novel set of sparsity inducing priors to achieve a robust reconstruction of geometry and albedo from transient images.

Another challenge in this work is that the instrumentation required to measure the transient images themselves has traditionally suffered from severe practical limitations including excessive hardware cost (hundreds of thousands of dollars), long acquisition times (hours) and the difficulty of keeping the sensitive system calibrated. In this work we address this problem by building on our recent work on using widespread CMOS time-of-flight sensors for obtaining the transient image. The inverse problems for transient image reconstruction and geometry recover can be merged into a single non-linear optimization problem that can be solved efficiently. The result is a system that is by several orders of magnitude more affordable and acquires data faster than previous solutions. In summary, we make the following contributions:

- We formulate a transient image formation model for indirect geometry reconstruction, and derive a framework for its inversion, including a novel set of sparsity enhancing priors. This framework is largely independent of the way the input transient image is acquired.
- Building on our earlier work [9], we propose an imaging setup that is budget-friendly, robust to ambient illumination, and captures the required data in only a few minutes.
- We demonstrate the effectiveness of setup and computational scheme by reconstructing both a low-contrast albedo and the geometry of hidden objects.

2. Related Work

Time-of-Flight (ToF) Sensors also known as *Photonic Mixer Devices* are image sensors where each pixel can di-



(a) Setup (top view) (b) Unknown scene (c) Reconstructed depth (volume as probability) (d) Reconstructed depth (strongest peak)

Figure 1. (a) top view schematic of our measurement scenario. All distances are in meters. A diffuse wall is illuminated by a modulated laser beam and observed by a time-of-flight camera. From diffuse reflections, we infer the geometry and albedo of objects within a bounding volume (green) that is completely occluded to both light source and camera, but visible from most locations on the wall. In this example, the shape of two letters cut from cardboard (b) becomes clearly visible in the reconstruction (c,d) (see text for details).

rect the charge from incoming photons to two or more storage sites within the pixel [19, 18, 13, 12]. This effectively allows the incident illumination to be modulated with a periodic reference signal. ToF sensors are typically used in conjunction with periodically modulated light sources of the same modulation frequency and in this setting measure the phase-dependent correlation between the illumination and the reference signal, from which the scene depth can be inferred.

Recent improvements and extensions to PMD design and operation include heterodyned modulation of light and PMD sensor to improve resolution [3, 4], multi-path and scattering suppression for depth estimation [5], as well as tomography based on time of flight information [8]. Heide et al. [9] recently showed that ToF sensors can also be used to reconstruct transient images. In our work we use this approach to reconstruct geometry that is not directly visible from the camera.

Transient Imaging a.k.a. Light-in-Flight Imaging [1] refers to a novel 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. Starting with Kirani et al.'s original work [11], there have been several proposals to use transient images to capture surface reflectance [14], or simply to visualize light transport in complex environments to gain a better understanding of optical phenomena [21]. Wu et al. [22] proposed to use transient images together with models of light/object interaction to factor the illumination into direct and indirect components.

Transient images have also been proposed as a means of reconstructing 3D geometry that is not directly visible to either the camera or the light sources ("Looking around the corner", [15, 20, 7]). In our work, we aim to perform this kind of indirect geometry reconstruction without the significant hardware complexity usually associated with transient imaging. Using standard ToF sensors, we build on Heide et al.'s work [9] to devise an image formation model, objective function, and optimization procedure to directly reconstruct geometry from indirect illumination measured with ToF cameras.

3. Image Formation Model

We make several assumptions for the image formation process in the scene (see Figure 1 (a)):

- The hidden scene is modeled as a diffuse height field, which in turn is represented as a collection of differential patches $d\mathbf{x}$ with orientation $\mathbf{n_x}$ inside a volume V.
- The camera points at a section of the diffuse wall, and is in focus.
- Light is emitted as a laser ray from position $\mathbf{l}_0$ and illuminates a single point $\mathbf{l}$ on the diffuse wall, outside the field of view of the camera. Radiometrically, we treat this point as a single, isotropic point light emitting a radiance $L_e(\mathbf{l})$.
- From $\mathbf{l}$, the light illuminates the scene, and after a single bounce returns to the diffuse wall. Patches $d\mathbf{w}$ on the wall are chosen such that there is a one-to-one correspondence between patches and camera pixels.
- Occlusion in the height field is ignored.

3.1. Stationary Light Transport

With these assumptions, and starting from the diffuse Rendering Equation [10], we can therefore model the stationary (i.e. time independent) light transport as follows.

$$L(\mathbf{l}) = L_e(\mathbf{l}) \tag{1}$$

$$L(\mathbf{x}) = L_e(\mathbf{l})\rho(\mathbf{x})\,G(\mathbf{l},\mathbf{x}) \tag{2}$$

$$L(\mathbf{w}) = \int_V L_e(\mathbf{l})\rho(\mathbf{x})\,G(\mathbf{l},\mathbf{x})\rho(\mathbf{w})\,G(\mathbf{x},\mathbf{w})\,d\mathbf{x} \tag{3}$$

with $\rho(.)$ denoting the diffuse albedo of a patch, and the *unoccluded* geometry term

$$G(\mathbf{x},\mathbf{y}) = \frac{\cos\measuredangle(\mathbf{y}-\mathbf{x},\mathbf{n_x})\cdot\cos\measuredangle(\mathbf{x}-\mathbf{y},\mathbf{n_y})}{|\mathbf{y}-\mathbf{x}|^2}. \tag{4}$$

We re-write the radiance at a wall patch (Equation 3) as

$$L(\mathbf{w}) = L_e(\mathbf{l})\rho(\mathbf{w}) \int_V g(\mathbf{x})v(\mathbf{x})\, d\mathbf{x}, \tag{5}$$

where the geometry term

$$g(\mathbf{x}) = \frac{\cos \measuredangle(\mathbf{x}-\mathbf{l}, \mathbf{n}_\mathbf{l}) \cdot \cos \measuredangle(\mathbf{x}-\mathbf{w}, \mathbf{n}_\mathbf{w})}{|\mathbf{x}-\mathbf{l}|^2 \cdot |\mathbf{x}-\mathbf{w}|^2} \tag{6}$$

is independent of both the albedo and orientation of the patch $d\mathbf{x}$, while

$$v(\mathbf{x}) = \rho(\mathbf{x}) \cdot \cos \measuredangle(\mathbf{l}-\mathbf{x}, \mathbf{n}_\mathbf{x}) \cdot \cos \measuredangle(\mathbf{w}-\mathbf{x}, \mathbf{n}_\mathbf{x}) \tag{7}$$

is a term that isolates both of these unknown quantities. We can interpret $v(\mathbf{x})$ either as a generalized albedo term or as a continuous volume occupancy that indicates whether or not a given voxel location is occupied by the surface to be reconstructed. Note that in this parametrization, the image formation process is linear in $v(\mathbf{x})$.

3.2. Transient Light Transport

The transient version of Equation 5 is obtained by adding a time coordinate t and counting only light contributions such that t is the sum of emission time t_0 and the travel time τ for a given light path from the laser $\mathbf{l}_0$ to a camera pixel $\mathbf{c}$. In our image formation model, the relevant light paths only differ in the position of the surface element $d\mathbf{x}$, i.e. $t = t_0 + \tau(\mathbf{x})$.

Recalling that we assume a one-to-one correspondence between wall patches $\mathbf{w}$ and camera pixels $\mathbf{c}$, we obtain the transient image formation model

$$L(\mathbf{c}, t) = \int_0^t L_e(\mathbf{l}, t_o)\rho(\mathbf{w}) \int_V \delta(t_0 + \tau(\mathbf{x}) - t) g(\mathbf{x}) v(\mathbf{x})\, d\mathbf{x}\, dt_0, \tag{8}$$

where the travel time $\tau(\mathbf{x})$ is given as the total path length divided by the speed of light c:

$$\tau(\mathbf{x}) = (|\mathbf{l}_0 - \mathbf{l}| + |\mathbf{l} - \mathbf{x}| + |\mathbf{x} - \mathbf{w}| + |\mathbf{w} - \mathbf{c}|)/c \tag{9}$$

We note that this transient image formation model is independent of the way the transient image has been acquired. It therefore applies to all known approaches for generating transient images, including femtosecond imaging [20] as well as correlation-based measurements with PMD sensors [9].

3.3. Discretization

The problem of reconstructing geometry from indirect light amounts to recovering the diffuse height field represented as the continuous voxel densities $v(\mathbf{x})$. To this end, we discretize the volume $v(\mathbf{x})$ from Equation 8 into a Euclidean voxel grid, and represent it as a vector $\mathbf{v}$ of size M. The transient image (radiance) is represented as a vector $\mathbf{i} \in \mathbb{R}^{NT}$, where N is the number of camera pixels/wall patches, and T is the number of time steps. The discrete version of Equation 8 is then given as

$$\mathbf{i} = \mathbf{P}\mathbf{v} \tag{10}$$

with the light transport tensor $\mathbf{P} \in \mathbb{R}^{NT \times M}$.

3.4. Transient Model with PMD Sensors

Unlike Velten et al. [20], in our implementation we do not measure the transient image directly. Instead, we build on-top of our recent work on transient imaging with low-cost sensors [9]: using a standard time-of-flight sensor with a modulated illumination source, we obtain a sequence of modulated exposure measurements $\mathbf{h}$ using different modulation frequencies and phases. The transient image can then be recovered as a linear inverse problem:

$$\mathbf{h} = \mathbf{C}\mathbf{i}, \tag{11}$$

where the correlation matrix $\mathbf{C}$ is obtained through a straightforward calibration step. Substituting Equation 10 for $\mathbf{i}$, we arrive at our full image formation model:

$$\mathbf{h} = \mathbf{C}\mathbf{P}\mathbf{v} \tag{12}$$

4. Inverse Problem

The image formation model from Equation 12 cannot be inverted directly, since both the light transport matrix $\mathbf{P}$ is poorly conditioned, as is the correlation matrix $\mathbf{C}$ (see [9]). It is therefore necessary to include additional regularization terms and solve a non-linear optimization problem. These steps are described in the following.

4.1. Objective Function

We formulate the inverse problem as the optimization problem

$$\mathbf{v}_{\text{opt}} = \underset{\mathbf{v}}{\operatorname{argmin}} \; \frac{1}{2} \|\mathbf{C}\mathbf{P}\mathbf{v} - \mathbf{h}\|_2^2 + \Gamma(\mathbf{v}), \tag{13}$$

which is regularized with three terms:

$$\Gamma(\mathbf{v}) = \lambda \sum_z \|\nabla_{x,y}\mathbf{v}_z\|_1 + \theta \|\mathbf{W}\mathbf{v}\|_1 + \omega \sum_{x,y} \text{ind}_\mathcal{C}(\mathbf{v}_{x,y}) \tag{14}$$

From left to right, the individual terms represent:

- A sparse spatial gradient distribution in the height field, implemented as the ℓ_1 penalized spatial gradients for all volume depths z.
- A sparse volume $\mathbf{v}$, justified by our assumption of height field geometry. This term is implemented as a weighted ℓ_1 norm of the volume itself. The weight matrix $\mathbf{W}$ will be obtained using an iteratively reweighted ℓ_1 scheme (IRL1, see Section 4.2).

- An explicit enforcement of the height field assumption, by constraining the volume to have at most one non-zero entry for each 2D (x, y) coordinate. We encode this prior using a projection onto an indicator set of possible depth values for each (x, y) coordinate:

$$\text{ind}_{\mathcal{C}}(\mathbf{p}) = \begin{cases} 0 & \text{if } \mathbf{p} \in \mathcal{C} \\ \infty & \text{else} \end{cases} \quad \text{with} \quad \mathcal{C} = \{\mathbf{d} \in \mathbb{R}^z | \, \text{card}\,(\mathbf{d}) = 1 \wedge \mathbf{1}^T\mathbf{d} = \mathbf{1}^T\mathbf{p}\} \tag{15}$$

We note that the second and third term of the regularizer both have the purpose of encouraging a single surface reflection along the z-dimension of the reconstruction volume. The term from Equation 15 is stronger than the ℓ_1 regularization, since it prefers exactly single-non-zero solutions (in contrast to just sparse solutions). On the other hand, it makes the overall optimization non-convex as $\mathcal{C}$ is a non-convex set. So having both terms enables us to trade the convexity of our objective function for the sparsity of our model by adjusting the weights θ, ω from Equation 14.

In order to solve the optimization problem from Equation 13, we split the regularization term into a linear operator $\mathbf{K}$ and a function $F(.)$: $\Gamma(\mathbf{v}) = F\,(\mathbf{K}\mathbf{v})$, with

$$\mathbf{K} = \left[\mathbf{D}_x^T, \mathbf{D}_y^T, \mathbf{W}\mathbb{I}^T, \mathbb{I}^T\right]^T, \tag{16}$$

where $\mathbf{D}_x, \mathbf{D}_y$ are derivative operators for the x, y dimensions for all z coordinates (stacked on-top of each other) and $\mathbb{I}$ is the identity matrix. We note that the minimum of $\Gamma(\mathbf{v})$ is obtained by independently minimizing F for each component of $\mathbf{K}\mathbf{v}$.

Having reformulated our optimization problem using $\mathbf{K}$, we have now mapped our problem to one that can solved efficiently using a variant of the alternate direction method of multipliers method (ADMM) in Algorithm 1.

Algorithm 1 Our ADMM algorithm

1: $\mathbf{v}^{k+1} := \left(\mathbf{P}^T\mathbf{C}^T\mathbf{C}\mathbf{P} + \mu\mathbb{I}\right)^{-1}\left(\mathbf{P}^T\mathbf{C}^T\mathbf{h} + \mu\mathbf{v}^k - \rho\left(\mathbf{K}^T\mathbf{K}\mathbf{v}^k - \mathbf{K}^T\mathbf{j}^k\right) + \mathbf{K}^T\lambda^k\right)$ // **v-step**

2: $\mathbf{j}^{k+1} := \mathbf{prox}_{(1/\rho)F}\left(\mathbf{K}\mathbf{v}^{k+1} - \lambda^k/\rho\right)$ // **j-step**

3: $\lambda^{k+1} := \lambda^k + \rho\left(\mathbf{K}\mathbf{v}^{k+1} - \mathbf{j}^{k+1}\right)$ // **λ-step**

A detailed derivation and description of Algorithm 1 and the proximal operator $\mathbf{prox}_{(1/\rho)F}$ can be found in the appendix and supplement.

4.2. Enhancing Volume Sparsity

To further enhance the sparsity of the convex ℓ_1-regularized part of our objective, we have placed a weight $\mathbf{W}$ on the individual components of the ℓ_1 volume penalty (second term in Eq. (14)).

This approach has been proposed by [2]. The idea is that the weights $\mathbf{W}$ capture the support of our sparse solution. This support is estimated iteratively from the last solution, which allows for improved recovery of the sparse non-negative components. As proposed in [2], we use the update rule

$$\mathbf{W}^{j+1} := \text{diag}\left(\frac{1}{|\mathbf{v}^j| + \epsilon}\right), \tag{17}$$

where the division is here point-wise. The iteration variable j from above is for an *outer iteration on top of our original iteration* from Algorithm 1.

5. Implementation and Parameter Selection

Parameters. For Algorithm 1, we use the parameters $\rho = 1.1$ and $\mu = 0.5 * 1/\left(\rho\|\mathbf{K}\|_2^2\right)$, which produced good results for all of our tested datasets. Note that $\mathbf{K}$ changes for every outer IRL1 iteration, and thus μ has to be recomputed for every iteration. We estimate $\|\mathbf{K}\|_2^2$ by running the power method for $\mathbf{K}^T\mathbf{K}$ with random initialization. We use 3 outer IRL1 iterations and an update weight of $\epsilon = 0.1$.

Implementation of the v-step. For a very high resolution sensor and reconstruction volume, storing $\mathbf{P}$ would be infeasible. In this scenario one can implement $\mathbf{P}$ as the procedural operator performing the transient light transport exactly as described in Section 3.2. The transient rendering operation parallelizes very well over each input pixel. One can implement its transpose $\mathbf{P}^T$ similarly as the dot product of each transient image for a considered voxel accumulated over the whole voxel volume. Thus again only transient rendering and some additional dot-products are required. Finally, the $\mathbf{v}$-step from Algorithm 1 can be implemented using conjugate gradient (CG). Instead of applying explicit matrix multiplication inside CG, we replace each of the products with $\mathbf{P}$ or $\mathbf{P}^T$ with the operations defined above.

We implemented this version first. However, since our sensor only has a very low resolution of 120×160, we were actually able to fully precompute and efficiently store $\mathbf{P}$ (in $<$ 8GB RAM) as a sparse matrix which speeds up the reconstruction dramatically. Note that this approach would not be possible if the sensor or reconstruction resolution were significantly higher.

Pre-factorization for Speedup. Instead of minimizing $\|\mathbf{C}\mathbf{P}\mathbf{v} - \mathbf{h}\|_2^2$ as a data term one can also pre-factor the optimization and first solve for a transient image $\mathbf{C}^{-1}\mathbf{h}$ and then use this as an observation in the changed data term $\|\mathbf{P}\mathbf{v} - \mathbf{C}^{-1}\mathbf{h}\|_2^2$. We have used the i-step from Heide et al. [9] to pre-factor the optimization and did not notice a no strong difference in reconstruction quality in comparison to using the not pre-factored version. The advantage of pre-factoring is that the method gets sped up even more since

all matrix application of $\mathbf{C}$ have been handled before and $\mathbf{C}$ itself can be inverted more efficiently than the full $\mathbf{CP}$.

6. Results

6.1. Experimental Setup

Our instrumentation comprises a modulated light source and a PMD detector, as first used for the purpose of transient imaging by Heide et al. [9].

The detector is based on a filterless version of the time-of-flight development kit CamBoard nano by PMD Technologies, and extended with an external frequency-controllable modulation source (a workaround in lack of access to the FPGA configuration for the CamBoard). We determined that for our setup an integration time of 10 milliseconds to delivers the optimal signal-to-noise ratio, which we further improve by averaging over multiple measurements (see also Section 6.3).

The light source consists of six 650 nm, 250 mW laser diodes with collimation optics and custom driving hardware to emit pulses of approximately 2-3 nanoseconds duration at variable repetition rate. The primary difference to the hardware setup by Heide et al. [9] is that in our setup, the diodes are not diffused to act as a spot light. Instead, we focus each laser diode with individual optics onto a single spot 1 on the wall (Figures 1, 2). Their overall duty cycle during capture is less than 1%, allowing operation with only the lens holders doubling as heat sinks.

Our reconstruction volume has a size of 1.5 m×1.5 m×2.0 m and is distanced 1.5 m from the flat, diffuse wall. The camera and illumination are about 2 m from the wall; please see Figure 1 (a) for the exact spatial arrangement.

6.2. Qualitative Reconstruction Results

Geometry. Our first test is to reconstruct the geometry of two letters cut out of cardboard that was painted with white color, and placed at different depths (Figure 1). We show two visualizations of the recovered depth information in the volume. In the second image from the right we treat the voxels as an occupancy probability and simply the expected value of the distribution for each pixel, i.e. the sum of distances weighted by the occupancy probability.

Since the expected value is not robust to outliers, we show in the rightmost image the depth value with the strongest peak for each (x, y) pixel. This amounts to the voxel with the highest probability of occupancy in our reconstruction. Note that in this image we threshold the volume such that all pixels with a very low albedo/occupancy probability for all depths are shown as grey.

Albedo. The next experiment 3 shows the recovery of a spatially varying albedo on a flat surface. The color-coded depth map shows the depth of the strongest density in the reconstructed volume for each pixel (x, y) as before. The left of the figure shows the albedo $v(\mathbf{x})$ sampled exactly at the depth map positions (the position of the strongest peak).

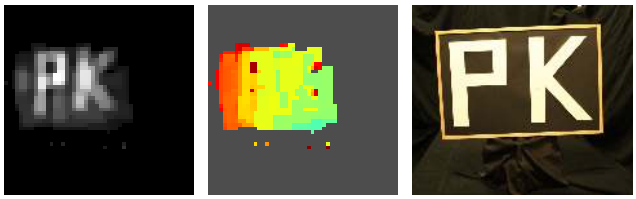


Figure 3. Albedo reconstruction example: Reconstruction of scene image with a flat surface but varying albedo (right). Left: the color-coded depth map of strongest peak along z-coordinate visualized shows an essentially flat geometry. Middle: Albedo image, reconstruction value exactly at the depth map's depth.

Albedo and Geometry. Figure 4 shows an example of variation in both geometry and albedo. In this case, the planar surface in the front could not be reconstructed in the depth map due to the low albedo limiting the reflected light.

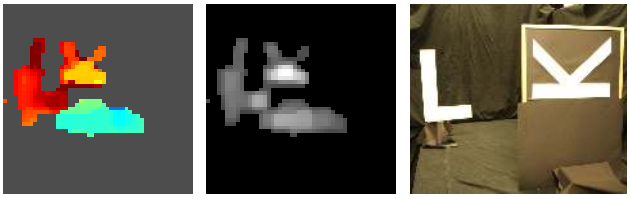


Figure 4. Simultaneous albedo and geometry reconstruction example: Reconstruction of scene image with varying albedo (letter on plane in the front) and varying depth for the letter in back (right). Albedo image, reconstruction value exactly at the depth position from the depth map (left). Color-coded depth map of strongest peak along z-coordinate visualized (middle).

Different Materials. In the supplemental material we show several examples of reconstructions with non-Lambertian surfaces. We find that Lambertian scenes result in very sparse volume reconstructions that clearly represent a height field structure. With increasingly non-Lambertian surfaces the energy is spread out more and more throughout the volume (as our model is violated).

6.3. Effects of Ambient Light and Frame Averaging

One of the advantages of Heide et al.'s method for reconstructing transient images [9], is that it is rather insensitive to ambient illumination. We tested whether this robustness also applies to our approach for reconstructing geometry (Figure 5) and albedo (Figure 6). In both cases we performed our capture once with the ceiling lights in the room switched off, and once with them switched on. We can see that there is only a minor effect on the overall sharpness and reconstruction quality in both cases.

As mentioned before, we average several measurements before reconstruction. This improves SNR, since the measured indirect reflection results in very low light levels.

Figure 2. Left: 3D model of our setup (to scale). Center: Photo of our capture setup facing the diffuse wall (light source covered with black photo board). To the left, behind an occluder, lies the reconstruction volume. Right: Close-up on the light source without cover.

Figure 7 shows different depth and albedo reconstructions, where each measurement respectively is the average of 10 or 500 individual ToF images with a specific modulation frequency and phase. We see that we still get a reasonable reconstruction by averaging only 10 images. The corresponding capture time of 4 minutes (200 minutes for averaging 500 measurements) could be significantly improved by better synchronizing the PMD camera and light source so that the camera can capture at video rates. Still, even with the current setup, our capture times compare very favorably to those reported for femtosecond laser setups [20].

6.4. Quantitative Measurements

To evaluate our reconstruction results, we compared the distance maps with manually measured scene geometry. Figure 8 shows a quantitative evaluation for the geometry reconstruction example shown above.

Lateral resolution. In order to evaluate the spatial resolution, we show an image of the measured scene geometry of the flat targets. The same discretization as for the shown depth map has been used. Having in mind that our reconstruction volume for all results in this paper had a size of $1.5\,\text{m} \times 1.5\,\text{m} \times 2.0\,\text{m}$ $(x \times y \times z)$, we see that we can achieve an (x, y) resolution of approximately $\pm 5\,\text{cm}$. The

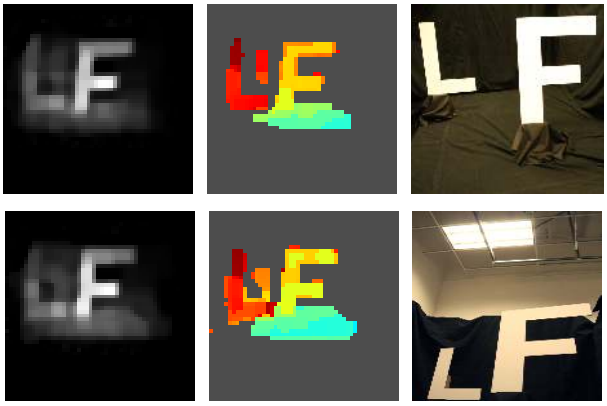


Figure 5. Effects of ambient illumination on albedo reconstruction: All lights in room off (top) and lights on(bottom). We see that we still get a reasonable reconstruction with strong ambient illumination

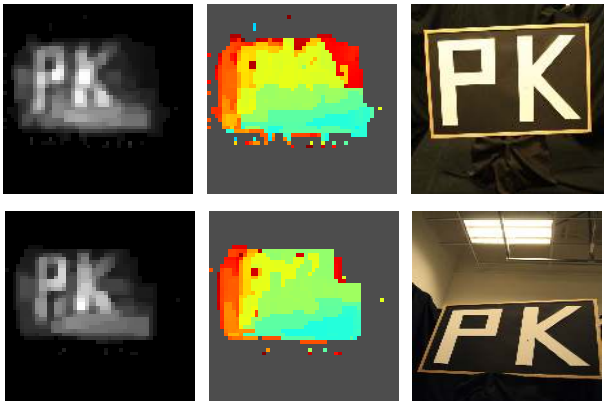


Figure 6. Effects of ambient illumination on albedo reconstruction: All lights in room off (top) and lights on(bottom). We see that we still get a reasonable reconstruction with strong ambient illumination

accuracy of the reconstruction varies with different materials. Materials that have little or no overlap in the space-time profile (e.g. mirror example in the supplement), allow for high reconstruction precision (around $\pm 2\,\text{cm}$ for the mirror example). The precision for more complex materials degraded to around $\pm 15\,\text{cm}$ tolerance. Overall the spatial resolution is limited by the low resolution of our sensor (which was only 120×160 pixels).

Note that due to our robust measurement and reconstruction procedure we are able to achieve the shown results for significantly larger scenes than previously possible in with the femtosecond laser approach demonstrated in [20]. Velten et al. report distances of up to 25 cm from object to the wall and a reconstruction volume of $(40\,\text{cm})^3$ due to the low

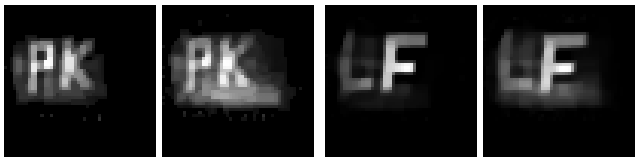


Figure 7. Effects of frame averaging on albedo (left) and geometry (right). The left image in each pair is based on averaging 500 ToF images for each measurement, while the right image in each pair uses only 10.

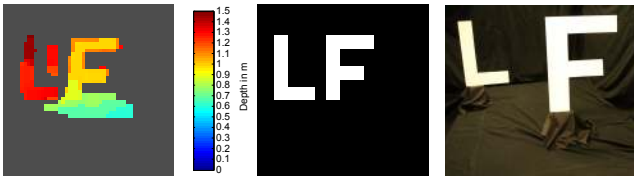


Figure 8. Quantitative evaluation: Reconstruction of scene image with letter "L" and "F" cut out of white painted cardboard (right). Color-coded depth map of strongest peak along z-coordinate visualized with color bar for depth in m (left). (x, y) ground truth geometry acquired from scene measurements (middle).

SNR for large distance bounces, whereas we demonstrate for the first time much larger room-sized environments.

Depth resolution. For the temporal resolution achieved in the above example of Fig. 8, we see from the given color bar a depth distance of approximately $0.6m$, where the measured distance was $0.75m$. For all similarly diffuse materials we reach also roughly a tolerance of $\pm 15cm$. For simple strong reflectors like the mirror we have less temporal superposition, so for the mirror example we obtain a high temporal resolution of below 5 cm error in our 2 m depth range, with more complex materials producing a precision of around ± 20 cm.

As shown above the resolution of our approach depends on the scene content. The achievable resolution should in the future scale linearly with the availability of higher resolution ToF cameras, such as the upcoming Kinect 2. We have shown that our method degrades somewhat gracefully with using different materials, although a certain scene dependence is inherent in the non-linear nature of the inverse problem we solve.

7. Conclusions

We have presented a method for reconstructing hidden geometry and low contrast albedo values from transient images of diffuse reflections. This approach involves hard inverse problems that can only be solved using additional priors such as sparsity in the geometry, and our primary contribution is to identify a linearized image formation model, regularization terms, and corresponding numerical solvers to recover geometry and albedo under this difficult scenario.

Despite these numerical challenges, we show that our method can be combined with our recent work on transient imaging using inexpensive time of flight cameras [9], which itself involves a hard inverse problem. We demonstrate that it is possible to combine these two inverse problems and solve them jointly in a single optimization method. As a result our approach has several advantages over previous methods employing femtosecond lasers and streak cameras [20]. These include a) low hardware cost, b) no moving parts and simplified calibration, c) capture times that are reduced from hours to minutes, and d) robustness under ambient illumination in large room-sized environments. We believe that, as a result, our method shows promise for applications of indirect geometry reconstruction outside the lab.

Acknowledgement. Part of this work was done while Matthias Hullin was a junior group leader at the Max Planck Center for Visual Computing and Communication (German Federal Ministry of Education and Research, 01IM10001).

A. Solving the Optimization Problem

This section describes how to derive Algorithm 1 after having introduced $\Gamma(\mathbf{v}) = F(\mathbf{K}\mathbf{v})$ in Eq. (16). For an expanded version see the supplement.

To derive our ADMM method, we rewrite the problem as

$$\mathbf{v}_{\text{opt}} = \underset{\mathbf{v}}{\text{argmin}}\ G(\mathbf{v}) + F(\mathbf{j}) \text{ s.t. } \mathbf{K}\mathbf{v} = \mathbf{j}\ , \tag{18}$$

and can then form the augmented Lagrangian

$$\begin{aligned}\mathcal{L}_\rho(\mathbf{v}, \mathbf{j}, \lambda) =& G(\mathbf{v}) + F(\mathbf{j}) + \\ & \lambda^T(\mathbf{K}\mathbf{v} - \mathbf{j}) + \frac{\rho}{2}\|\mathbf{K}\mathbf{v} - \mathbf{j}\|_2^2,\end{aligned} \tag{19}$$

where λ is a dual variable associated with the consensus constraint. ADMM now minimizes $\mathcal{L}_\rho(\mathbf{v}, \mathbf{j}, \lambda)$ w.r.t. one variable at a time while fixing the remaining variables. The dual variable is then the scaled sum of the consensus constraint error. For more details see, for example [16]. The minimization is then done iteratively, by alternatingly updating $\mathbf{v}$, $\mathbf{j}$, and the Lagrange multiplier λ. The key steps of this algorithm are as follows:

$$\begin{aligned}\mathbf{v}^{k+1} =& \underset{\mathbf{v}}{\text{argmin}}\ \mathcal{L}_\rho\left(\mathbf{v}, \mathbf{j}^k, \lambda^k\right) \\ =& \underset{\mathbf{v}}{\text{argmin}}\ \frac{1}{2}\|\mathbf{C}\mathbf{P}\mathbf{v} - \mathbf{h}\|_2^2 + (\lambda^k)^T\left(\mathbf{K}\mathbf{v} - \mathbf{j}^k\right) + \\ & \frac{\rho}{2}\left\|\mathbf{K}\mathbf{v} - \mathbf{j}^k\right\|_2^2 \\ \approx& \underset{\mathbf{v}}{\text{argmin}}\ \frac{1}{2}\|\mathbf{C}\mathbf{P}\mathbf{v} - \mathbf{h}\|_2^2 + (\lambda^k)^T\left(\mathbf{K}\mathbf{v} - \mathbf{j}^k\right) + \\ & \rho\left(\mathbf{K}^T\mathbf{K}\mathbf{v}^k - \mathbf{K}^T\mathbf{j}^k\right)^T\mathbf{v} + \frac{\mu}{2}\left\|\mathbf{v} - \mathbf{v}^k\right\|_2^2 \\ =& \left(\mathbf{P}^T\mathbf{C}^T\mathbf{C}\mathbf{P} + \mu\mathbb{I}\right)^{-1}\left(\mathbf{P}^T\mathbf{C}^T\mathbf{h} + \mu\mathbf{v}^k - \right. \\ & \left.\rho\left(\mathbf{K}^T\mathbf{K}\mathbf{v}^k - \mathbf{K}^T\mathbf{j}^k\right) + \mathbf{K}^T\lambda^k\right)\end{aligned} \tag{20}$$

Note that in the third line we have made an approximation that linearizes the quadratic term from the second line in the proximity of the previous solution $\mathbf{v}^k$. This linearization approach is known under several different names, including Linearized ADMM or inexact Uzawa method (e.g. [23, 6]). The additional parameter μ satisfies the relationship $0 < \mu \leq 1/\left(\rho\|\mathbf{K}\|_2^2\right)$.

$$\begin{aligned}\mathbf{j}^{k+1} =& \underset{\mathbf{j}}{\operatorname{argmin}}\ \mathcal{L}_\rho\left(\mathbf{v}^{k+1}, \mathbf{j}, \lambda^k\right) \\ =& \underset{\mathbf{j}}{\operatorname{argmin}}\ F\left(\mathbf{j}\right) + \frac{\rho}{2}\left\|\left(\mathbf{K}\mathbf{v}^{k+1} - \frac{\lambda^k}{\rho}\right) - \mathbf{j}\right\|_2^2 \end{aligned} \quad (21)$$

Both $F(.)$ and the least square term can be minimized independently for each component in $\mathbf{j}$. Using the slack variable $\mathbf{j}$, the minimization involving the difficult function F has now been turned into a sequence of much simpler problems in just a few variables.

To derive the specific solutions to these problems, we note that the last line in Equation 21 can be interpreted as a proximal operator [16]:

$$\mathbf{j}^{k+1} = \mathbf{prox}_{(1/\rho)F}\left(\mathbf{K}\mathbf{v}^{k+1} - \frac{\lambda^k}{\rho}\right). \quad (22)$$

Proximal operators are well-known in optimization and have been derived for many terms. For our problem, we require the proximal operators for the ℓ_1 norm and for the indicator set. These are given as

$$\begin{aligned}\mathbf{prox}_{\gamma|\cdot|}(a) =& (a-\gamma)_+ - (-a-\gamma)_+ \\ \mathbf{prox}_{\gamma\,\mathrm{ind}_{\mathcal{C}}(\cdot)}(a) =& \Pi_{\mathcal{C}}(a)\end{aligned} \quad (23)$$

The first term is the well-known point-wise shrinkage and the second is the projection on the set $\mathcal{C}$.

The final step of the ADMM algorithm is to update the Lagrange multiplier by adding the (scaled) error:

$$\lambda^{k+1} := \lambda^k + \rho\left(\mathbf{K}\mathbf{v}^{k+1} - \mathbf{j}^{k+1}\right) \quad (24)$$

References

[1] N. Abramson. Light-in-flight recording by holography. *Optics Letters*, 3(4):121–123, 1978. 1, 2

[2] E. J. Candes, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted l1 minimization. *J. Fourier Analysis and Applications*, 14(5-6):877–905, 2008. 4

[3] R. Conroy, A. Dorrington, R. Künnemeyer, and M. Cree. Range imager performance comparison in homodyne and heterodyne operating modes. In *Proc. SPIE*, volume 7239, page 723905, 2009. 2

[4] A. Dorrington, M. Cree, A. Payne, R. Conroy, and D. Carnegie. Achieving sub-millimetre precision with a solid-state full-field heterodyning range imaging camera. *Meas. Sci. and Technol.*, 18(9):2809, 2007. 2

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

[6] E. Esser, X. Zhang, and T. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imag. Sci*, 3(4):1015–1046, 2010. 7

[7] O. Gupta, T. Willwacher, A. Velten, A. Veeraraghavan, and R. Raskar. Reconstruction of hidden 3d shapes using diffuse reflections. *Opt. Express*, 20(17):19096–19108, Aug 2012. 1, 2

[8] A. Hassan, R. Künnemeyer, A. Dorrington, and A. Payne. Proof of concept of diffuse optical tomography using time-of-flight range imaging cameras. In *Proc. Electronics New Zealand Conference*, pages 115–120, 2010. 2

[9] F. Heide, M. B. Hullin, J. Gregson, and W. Heidrich. Low-budget transient imaging using photonic mixer devices. *ACM Trans. Graph. (Proc. SIGGRAPH 2013)*, 32(4):45:1–45:10, 2013. 1, 2, 3, 4, 5, 7

[10] J. T. Kajiya. The rendering equation. In *Proc. SIGGRAPH*, pages 143–150, 1986. 2

[11] A. Kirmani, T. Hutchison, J. Davis, and R. Raskar. Looking around the corner using transient imaging. In *Proc. ICCV*, pages 159–166, 2009. 1, 2

[12] R. Lange and P. Seitz. Solid-state time-of-flight range camera. *IEEE J. Quantum Electronics*, 37(3):390–397, 2001. 2

[13] R. Lange, P. Seitz, A. Biber, and S. Lauxtermann. Demodulation pixels in CCD and CMOS technologies for time-of-flight ranging. *Sensors and camera systems for scientific, industrial, and digital photography applications*, pages 177–188, 2000. 2

[14] N. Naik, S. Zhao, A. Velten, R. Raskar, and K. Bala. Single view reflectance capture using multiplexed scattering and time-of-flight imaging. *ACM Trans. Graph.*, 30(6):171, 2011. 2

[15] R. Pandharkar, A. Velten, A. Bardagjy, E. Lawson, M. Bawendi, and R. Raskar. Estimating motion and size of moving non-line-of-sight objects in cluttered environments. In *Proc. CVPR*, pages 265–272, 2011. 2

[16] N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, pages 1–96, 2013. 7, 8

[17] R. Raskar and J. Tumblin. *Computational Photography: Mastering New Techniques For Lenses, Lighting, and Sensors*. A K Peters, Limited, 2007. 1

[18] R. Schwarte. Verfahren und Vorrichtung zur Bestimmung der Phasen und/oder Amplitudeninformation einer elektromagnetischen Welle. German Patent 19704496, 1997. 2

[19] 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

[20] A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. Bawendi, and R. Raskar. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nature Communications*, 3:745, 2012. 1, 2, 3, 6, 7

[21] A. Velten, D. Wu, A. Jarabo, B. Masia, C. Barsi, C. Joshi, E. Lawson, M. Bawendi, D. Gutierrez, and R. Raskar. Femto-photography: Capturing and visualizing the propagation of light. *ACM Trans. Graph.*, 32, 2013. 2

[22] D. Wu, M. O'Toole, A. Velten, A. Agrawal, and R. Raskar. Decomposing global light transport using time of flight imaging. In *Proc. CVPR*, pages 366–373, 2012. 2

[23] X. Zhang, M. Burger, and S. Osher. A unified primal-dual algorithm framework based on bregman iteration. *J. Sci. Comp.*, 46(1):20–46, 2011. 7