Research Article | Vol. 25, No. 10 | 15 May 2017 | OPTICS EXPRESS 11574

Optics EXPRESS

Fast back-projection for non-line of sight reconstruction

VICTOR ARELLANO, DIEGO GUTIERREZ, AND ADRIAN JARABO*

Universidad de Zaragoza - I3A, Zaragoza 50018, Spain
ajarabo@unizar.es

Abstract: Recent works have demonstrated non-line of sight (NLOS) reconstruction by using the time-resolved signal from multiply scattered light. These works combine ultrafast imaging systems with computation, which back-projects the recorded space-time signal to build a probabilistic map of the hidden geometry. Unfortunately, this computation is slow, becoming a bottleneck as the imaging technology improves. In this work, we propose a new back-projection technique for NLOS reconstruction, which is up to *a thousand times* faster than previous work, with almost no quality loss. We base on the observation that the hidden geometry probability map can be built as the intersection of the three-bounce space-time manifolds defined by the light illuminating the hidden geometry and the visible point receiving the scattered light from such hidden geometry. This allows us to pose the reconstruction of the hidden geometry as the voxelization of these space-time manifolds, which has lower theoretic complexity and is easily implementable in the GPU. We demonstrate the efficiency and quality of our technique compared against previous methods in both captured and synthetic data.

OCIS codes: (110.1758) Computational imaging; (100.3190) Inverse problems.

References and links

1. A. Jarabo, B. Masia, J. Marco, and D. Gutierrez, "Recent advances in transient imaging: A computer graphics and vision perspective," Vis. Inform. **1**, https://arxiv.org/abs/1611.00939 (2017).
2. A. Bhandari and R. Raskar, "Signal processing for time-of-flight imaging sensors," IEEE Signal Process. Mag. **33**(5), 45–58 (2016).
3. A. Velten, T. Willwacher, O. Gupta, A. Veeraraghavan, M. G. Bawendi, and R. Raskar, "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," Nat. Commun. **3**, 745 (2012).
4. P. Y. Han, G. C. Cho, and X.-C. Zhang, "Time-domain transillumination of biological tissues with terahertz pulses," Opt. Lett. **25**(4), 242–244 (2000).
5. D. Raviv, C. Barsi, N. Naik, M. Feigin, and R. Raskar, "Pose estimation using time-resolved inversion of diffuse light," Opt. Express **22**(17), 20164–20176 (2014).
6. F. Heide, L. Xiao, A. Kolb, M. B. Hullin, and W. Heidrich, "Imaging in scattering media using correlation image sensors and sparse convolutional coding," Opt. Express **22**(21), 26338–26350 (2014).
7. 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**(4), 44 (2013).
8. F. Heide, L. Xiao, W. Heidrich, and M. B. Hullin, "Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors," in IEEE Computer Vision and Pattern Recognition, (IEEE, 2014) pp. 3222–3229.
9. M. Buttafava, J. Zeman, A. Tosi, K. Eliceiri, and A. Velten, "Non-line-of-sight imaging using a time-gated single photon avalanche diode," Opt. Express **23**(16), 20997–21011 (2015).
10. M. Laurenzis and A. Velten, "Nonline-of-sight laser gated viewing of scattered photons," Opt. Eng. **53**(2), 023102 (2014).
11. J. Klein, C. Peters, J. Martín, M. Laurenzis, and M. B. Hullin, "Tracking objects outside the line of sight using 2D intensity images," Sci. Rep. **6**, 32491 (2016).
12. M. B. Hullin, "Computational imaging of light in flight," in SPIE/COS Photonics Asia, (2014).
13. 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 (2012).
14. M. Schwarz and H.-P. Seidel, "Fast parallel surface and solid voxelization on GPUs," ACM Trans. Graph. **29**(6), 179 (2010).
15. E. Eisemann and X. Décoret, "Fast scene voxelization and applications," in "Proceedings of the 2006 symposium on Interactive 3D graphics and games," (ACM, 2006), pp. 71–78.
16. J. Pantaleoni, "Voxelpipe: a programmable pipeline for 3D voxelization," in "Proceedings of the ACM SIGGRAPH Symposium on High Performance Graphics," (ACM, 2011), pp. 99–106.
17. A. Jarabo, J. Marco, A. Muñoz, R. Buisan, W. Jarosz, and D. Gutierrez, "A framework for transient rendering," ACM Trans. Graph. **33**(6), 177 (2014).
18. L. Zhang, W. Chen, D. S. Ebert, and Q. Peng, "Conservative voxelization," Vis. Comput. **23**(9), 783–792 (2007).

#290206 https://doi.org/10.1364/OE.25.011574
Journal © 2017 Received 14 Mar 2017; revised 7 Apr 2017; accepted 7 Apr 2017; published 9 May 2017

Optics EXPRESS

1. Introduction

One of the core applications of time-resolved imaging (see e.g. [1, 2] for recent surveys on the field) is the capability to robustly capture depth from a scene, by being able to track the time of arrival of photons. Geometry reconstruction techniques have significantly benefited from this, but in the last years the applicability of time-resolved imaging has gone beyond directly visible geometry, to include non-line of sight (NLOS) imaging [3]. This technique allows reconstructing occluded objects by analyzing multiple-scattered light, even in the presence of turbid media [4–6].

The first prototype of this technology was demonstrated with a femtosecond laser and a streak camera [7]. Further work explored alternative hardware setups, including correlation-based time-of-flight cameras [8], single photon avalanche diodes (SPAD) [9], laser-gated sensors [10], or even common DSLR cameras [11]. These setups are cheaper and more portable, although at the cost of sacrificing time resolution.

Despite all these available hardware options, the *reconstruction* step is still a bottleneck, limiting the applicability of this technology in the wild. Different approaches have been proposed for reconstruction, either by solving a non-convex optimization on three-dimensional geometry [11, 12] or a depth map [8], or by back-projecting the space-time captured image on a voxelized geometry representation [3, 13]. In both cases, the large amount of data being processed, together with the complexity of the reconstruction algorithms, impose computation times in the order of several hours.

Back-projection reconstruction builds a 3D probability map based on the recorded time-resolved image, encoding the time of arrival of photons reflected by the hidden geometry. It exploits the correlation between the time of arrival of a photon and the total distance traveled, back-projecting this distance to the reconstruction volume. This approach has several advantages over the alternative methods: First, it avoids the need to solve a non-convex optimization problem, which may fall in local minima. Second, the memory cost is significantly lower. And third, it is relatively robust to noisy captures, which is particularly important when aiming for low capturing times. However, current back-projection methods are very inefficient and redundant, which results in a poor scalability with the resolution of the voxelized scene. This imposes a hard limit of the quality of the reconstruction, regardless of the used imaging technology.

In this work we propose a new back-projection reconstruction method that yields a speed-up factor of three orders of magnitude over previous NLOS reconstruction approaches, thus addressing the main pending issue limiting the applicability of recent approaches. We build on the key observation that the manifold of probably points that might reflect light towards an observed point at a certain time form an ellipsoid with poles at the light source and the observed point, and with radii the time of flight of photons (Fig. 1). This allows us to build the probability map as the intersection of the ellipsoids of multiple space-time measurements. Posing the problem in this ellipsoid space allows to avoid redundancy, which reduces the theoretical complexity of the algorithm to just the number of space-time measurements.

More importantly, formulating the problem this way is equivalent to performing voxelization of the ellipsoids geometry, which is very suitable for modern GPUs.

This combination of lower computational complexity and an efficient hardware-accelerated implementation allows us to increase efficiency further: our techniques allows computing the probability maps of large datasets in the order of seconds with a minimum increase in error, while on the other hand allows significantly higher-resolution reconstructions with a negligible added cost. We demonstrate these capabilities by evaluating our method using both synthetic and real captured data, and comparing with existing approaches. In addition, we make our code publicly available at http://giga.cps.unizar.es/~ajarabo/pubs/nlosbackprojectionOExp17/code/.

Optics EXPRESS

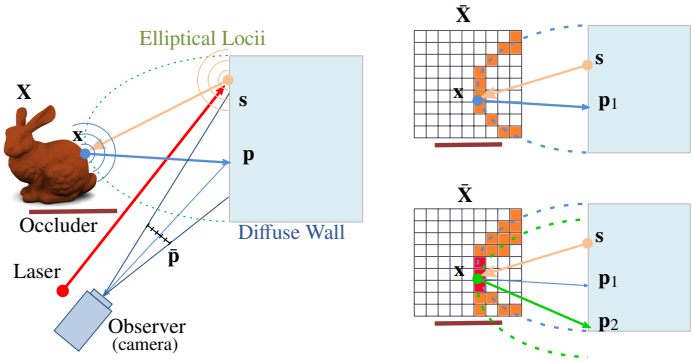

Fig. 1. Overview of our method. Left: Illustration of our reconstruction setup. A laser pulse is emitted towards a diffuse wall, creating a virtual point light **s** illuminating the occluded scene. The reflection of the occluded geometry travels back to the diffuse wall, which is imaged by the camera. The total propagation time from a hidden surface point **x** forms an ellipsoid with focal points at **s** and **p**. Right: The intersection of several of these ellipsoids defines a probability map for the occluded geometry, from which the reconstruction is performed with a speed-up factor of three orders of magnitude over previous approaches.

2. Back-projection for NLOS reconstruction

The goal of NLOS reconstruction methods is to recover an unknown hidden scene $\mathbf{X} \in \mathbb{R}^3$ from the measurements on visible known geometry $\mathbf{P} \in \mathbb{R}^3$. Such $\mathbf{P}$ is defined by a bijective map $\psi : \mathbb{R}^2 \to \mathbb{R}^3$ from pixel values $\bar{\mathbf{P}}$ imaged on the sensor. The scene is illuminated by a set of laser shots directed towards the visible geometry, hitting at positions $\mathbf{S}$; an image is captured for each shot. Figure 1 shows an example of our setup, where $\mathbf{s} \in \mathbf{S}$ is a virtual point light source created by the laser's reflection, and $\mathbf{p} \in \mathbf{P}$ is the projection of pixel $\bar{\mathbf{p}}$ in the known geometry $\mathbf{P}$ as $\psi(\bar{\mathbf{p}}) = \mathbf{p}$, with inverse mapping $\psi^{-1}(\mathbf{p}) = \bar{\mathbf{p}}$. It can be seen how the total propagation time from a hidden surface point $\mathbf{x}$ forms an ellipsoid with focal points at $\mathbf{s}$ and $\mathbf{p}$.

In our particular context, the captured signal I is a time-resolved image indexed by the spatial and the temporal domains, measuring the light's time of arrival on each pixel. Physically, the signal formation model is

$$I(\bar{\mathbf{p}}, t) = \int_{\mathbf{X}} L_o(\mathbf{x} \to \mathbf{p}, t - \tau(\mathbf{x} \to \mathbf{p}))\, G(\mathbf{x} \to \mathbf{p}) V(\mathbf{x} \to \mathbf{p}) d\mathbf{x}, \tag{1}$$

where $I(\bar{\mathbf{p}}, t)$ is the captured signal for pixel $\bar{\mathbf{p}}$ at time instant t, $L_o(\mathbf{x} \to \mathbf{p}, t - \tau(\mathbf{x} \to \mathbf{p}))$ is the reflected radiance at $\mathbf{x}$ towards $\mathbf{p}$, and $\tau(\mathbf{x} \to \mathbf{p})) = \|\mathbf{x} - \mathbf{p}\|\, c^{-1}$, with c the speed of light in the medium. G and V are respectively the geometric attenuation and binary visibility function between $\mathbf{x}$ and $\mathbf{p}$.

Assuming that all recorded radiance at $\mathbf{x}$ is due to third bounce reflection, we can remove the visibility term V. We also consider that all captured illumination comes from the virtual point light at $\mathbf{s}$. With that in place, we transform Eq. (1) to

$$I_{\mathbf{s}}(\bar{\mathbf{p}}, t) = \int_{\mathbf{X}} L_o(\mathbf{s} \to \mathbf{x}, t_o)\, f(\mathbf{s} \to \mathbf{x} \to \mathbf{p})\, G(\mathbf{s} \to \mathbf{x} \to \mathbf{p}) d\mathbf{x}, \tag{2}$$

where f represents the BRDF of the material at $\mathbf{x}$, $G(\mathbf{s} \to \mathbf{x} \to \mathbf{p}) = G(\mathbf{s} \to \mathbf{x}) G(\mathbf{x} \to \mathbf{p})$ is the geometric attenuation of the full light path, and $t_o = t - \tau(\mathbf{s} \to \mathbf{x} \to \mathbf{p}) - t_s - t_p$. Note that Eqs. (1) and (2) assume that the time coordinate starts when $\mathbf{s}$ emits light, and that the sensor is placed at $\bar{\mathbf{p}}$. t_s and t_p are the times of flight from the laser to the virtual light $\mathbf{s}$, and from the camera to the imaged surface point $\mathbf{p}$, respectively. However, since these additional terms do not affect the

Optics EXPRESS

shape of the ellipsoids, we omit them in the rest of the paper for clarity (although we do take them into account in our calculations).

2.1. NLOS by back-projection

Taking into account the signal formation model (2), we would like to recover the unknown geometry $\mathbf{X}$ from a set of spatio-temporal measurements $I_{\mathbf{s}}(\bar{\mathbf{p}}, t)$. Back-projection methods [3,9] aim to recover a discrete approximation $\bar{\mathbf{X}}$ of the unknown scene $\mathbf{X}$. Intuitively, only points $\mathbf{x} \in \mathbf{X}$ where there is an object will scatter light back towards $\mathbf{P}$. However, since we only know the spatial and temporal domain of the signal, given a measurement $I_{\mathbf{s}}(\bar{\mathbf{p}}, t)$ there is an infinite number of candidate points $\mathbf{x} \in \mathbf{X}$ that might have reflected light towards $\psi(\bar{\mathbf{p}}) = \mathbf{p}$ at instant t (see Fig. 1). This means that the scene cannot be recovered from a single measurement, so instead multiple measurements need to be taken.

In essence, the main idea is to build a probabilistic model where, assuming Lambertian reflectances, the probability of point $\mathbf{x}$ being part of the occluded geometry is modeled as

$$p(\mathbf{x}) = \iiint \frac{I_{\mathbf{s}}(\bar{\mathbf{p}}, t)\, \delta(t - \tau(\mathbf{s} \to \mathbf{x} \to \mathbf{p}))}{G(\mathbf{s} \to \mathbf{x} \to \mathbf{p})}\, dt d\mathbf{p} d\mathbf{s} \tag{3}$$

$$\approx \sum_{\mathbf{s} \in \mathbf{S}} \sum_{\bar{\mathbf{p}} \in \mathbf{P}} I_{\mathbf{s}}(\bar{\mathbf{p}}, \tau(\mathbf{s} \to \mathbf{x} \to \mathbf{p}))\, G(\mathbf{s} \to \mathbf{x} \to \mathbf{p})^{-1}, \tag{4}$$

where the signal $I_{\mathbf{s}}(\bar{\mathbf{p}}, t)$ is corrected by an estimate of $G(\mathbf{s} \to \mathbf{x} \to \mathbf{p})$, and δ is the delta function centered at zero. This probability map is later used to reconstruct the geometry, by using some operator over the voxelized representation, typically a three-dimensional Laplacian filter [3].

In practice, the most common straight forward form of computing the probability map consists of evaluating Eq. (4) for each unknown point $\mathbf{x} \in \bar{\mathbf{X}}$. This unfortunately is very expensive; the computational cost grows linearly with the number of pixels $\bar{P}$, lights S, and voxels $\bar{X}$, thus yielding $O(\bar{P} \times S \times \bar{X})$. In the following, we introduce our novel formulation, which significantly reduces the theoretical complexity of the computations, and allows for a very efficient implementation in commodity hardware.

3. Our method

Our method builds on the observation that the set of points that can potentially contribute to $I_{\mathbf{s}}(\bar{\mathbf{p}}, t)$ from a given laser shot hitting at $\mathbf{s}$ is defined by the ellipsoid $E(\mathbf{s}, \mathbf{p}, t)$ with focal points at $\mathbf{s}$ and $\mathbf{p}$, and focal distance $t \cdot c$ (see Fig. 1). This means that the more ellipsoids intersecting at point $\mathbf{x}$, the higher the probability $p(\mathbf{x})$ of having occluded geometry at $\mathbf{x}$, since more light arriving at pixel $\bar{\mathbf{p}}$ can have potentially been reflected at $\mathbf{x}$.

Following this observation, we pose Eq. (4) as an *intersection of ellipsoids* $E(\mathbf{s}, \mathbf{p}, t)$ with a voxelized representation of the scene, as

$$p(\mathbf{x}) = \sum_{\mathbf{s} \in \mathbf{S}} \sum_{\bar{\mathbf{p}} \in \mathbf{P}} I_{\mathbf{s}}(\bar{\mathbf{p}}, t)\, \mathrm{isect}(\mathbf{x}, E(\mathbf{s}, \mathbf{p}, t)), \tag{5}$$

where $t = \tau(\mathbf{s} \to \mathbf{x} \to \mathbf{p})$, and $\mathrm{isect}(\mathbf{x}, E(\mathbf{s}, \mathbf{p}, t))$ is a binary function returning 1 if the ellipsoid $E(\mathbf{s}, \bar{\mathbf{p}}, t)$ intersects voxel $\mathbf{x}$, and 0 otherwise. Note that here we are not correcting for the geometric attenuation as in Eq. (4); we opt for this approach to avoid some possible numerical singularities when $G(\mathbf{s} \to \mathbf{x} \to \mathbf{p}) \to 0$, and to keep the maximum values bounded (which is important in our implementation to avoid overflow, see Appendix A). Given that the probability map is latter processed by the Laplacian filter to reconstruct geometry, and the geometry term has almost no effect between neighboring voxels, we find that this does not affect the final result.

Our new formulation would be slower than Eq. (4) if computed naively, due to the need to calculate the intersection between the ellipsoid and point $\mathbf{x}$. However, the intersection operand

Optics EXPRESS

allows us to compute $p(\mathbf{x})$ by testing the ellipsoid-voxel intersection directly. Instead of evaluating each voxel $\mathbf{x}$ against the captured data, we now simply project the captured data into the voxelized scene representation.

Posing the problem this way has two main benefits: On the one hand, it significantly reduces the complexity of the required computations by only testing on locations with signal information, resulting in a theoretical complexity order of $O(\bar{P} \times S \times T)$, with T the temporal resolution. Note that this complexity is independent on the voxelization resolution, and that since the captured signal is in general sparse, in practice the complexity is even lower. On the other hand, our new formulation is equivalent to performing a voxelization of the full set of ellipsoids defined by the combination of tuples $\langle \mathbf{p}, \mathbf{s}, t \rangle$ (Fig. 1, right). Voxelization is a well-studied problem in computer graphics, which can be efficiently performed in commodity hardware [14]. In the following we describe the details of our implementation.

3.1. Fast back-projection

In order to perform the voxelization of the ellipsoids we rely on hardware-based voxelization [15], although other custom GPU-based voxelization methods could be used (e.g. [14, 16]).

To achieve this, we need to overcome two main problems: *i)* we need to create a large number of ellipsoids, which can be very expensive and memory consuming; and *ii)* hardware rasterization does not work with parametric surfaces beyond triangles, so we need to tessellate the ellipsoids before rasterization; this aggravates the cost/memory problem.

We address the first point by taking advantage of instanced rendering, which is standard in most modern GPUs, and allows to re-render the same primitive while applying a different linear transformation to each instance. For this, we create a base sphere, which is later transformed in the target ellipsoid i by scaling, translating and rotating it, by using a standard linear sphere-to-ellipsoid transformation matrix $\mathbf{M}_i$.

Before rendering, we apply recursive geodesic tessellation to the base sphere. Ideally, we would like all triangles' sizes to be smaller than the voxel size, so that high curvatures are accurately handled. However, since the transformations required for each ellipsoid might vary, the final size of each rendered triangle is not known in advance; this means that we cannot set a particular tessellation level for all ellipsoids. We instead precompute a set of spheres $\mathbf{O}$ with different tessellation levels o, and dynamically choose what level o will be used as

$$\text{argmin}_{o \in \mathbf{O}} \ (\alpha_o \max(\text{Eig}(\mathbf{M}_i)) < \epsilon) \ , \tag{6}$$

where α_o is the approximation error per triangle at tessellation level o, $\text{Eig}(\mathbf{M}_i)$ are the eigenvalues of transformation matrix $\mathbf{M}_i$, and ϵ is an error threshold, which we set to the voxel size. We offer additional implementation details in the Appendix.

3.2. Cost analysis

The computational cost of our method is linear with the number of ellipsoids $O(\bar{P} \times S \times T)$, which is independent on the resolution of the reconstruction space. Voxelizing each ellipsoid has a linear cost with the number of triangles per ellipsoid $\mathcal{T}$ times the cost of each triangle $\triangle$, which is approximately proportional to the number of voxels $\bar{X}_\triangle$ containing the triangle. Thus, the cost for each ellipsoid $O(E)$ is:

$$O(E) = \begin{cases} O(\mathcal{T} \times \bar{X}_\triangle) \approx O(\sqrt[3]{\bar{X}}) & \text{if } \bar{X}_\triangle \geq 1 \\ O(\mathcal{T}) & \text{elsewhere} \end{cases} . \tag{7}$$

As we will show later, setting ϵ [Eq. (6)] so that $\bar{X}_\triangle \approx 1$ gives the best trade-off between quality and cost. With this cost per ellipsoid, the total order of our technique is $O(\bar{P} \times S \times T \times \sqrt[3]{\bar{X}})$, which results in a speed-up with respect to traditional back-projection of $O(\frac{\sqrt[3]{\bar{X}^2}}{T})$. In the following section we demonstrate empirically the performance of our method.

Optics EXPRESS

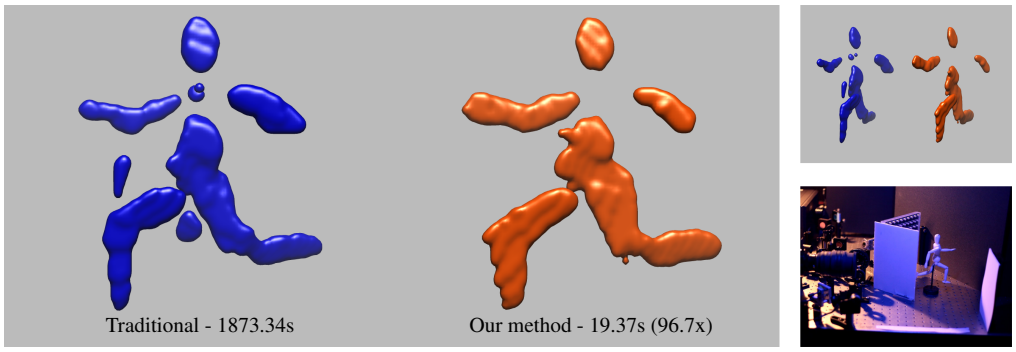

Fig. 2. Reconstruction of a mannequin (see bottom right) captured with a streak camera and a femtosecond laser [7], reconstructed using traditional back-projection (left, in blue) and our method (right, in orange), which is two orders of magnitude faster while yielding similar quality. The inset on the top-right shows the same reconstructed object under a different camera angle (inset from [3]).

4. Results

We evaluate our method by using datasets from three different sources (captured with femto-photography, a SPAD, and generated with transient rendering), each showing variable ranges of complexity, as well as different levels of signal quality. We compare our method against traditional back-projection [3], using Velten et al.'s optimized implementation. All our tests have been performed on an Intel i5-6500 @3.2GHz with 8 GB of RAM equipped with a GPU Nvidia GTX 1060.

Figure 2 shows the reconstruction of a hidden mannequin captured using a streak camera and a femtosecond laser [7]. A voxel grid of resolution $\bar{X} = 162^3$ is reconstructed from a set of $S = 59$ spatio-temporal measurements with different lighting position $\mathbf{s}$. Each measurement has spatial resolution of $\bar{P} = 336$ pixels, and temporal resolution of $T = 512$ frames. Although our method is mathematically equivalent to traditional back-projection (see Section 3), reconstructing this scene takes less than 20 seconds with our method, compared to more than half an hour with traditional back-projection. This represents a speed-up of 96.7x.

Figure 3 shows a similar reconstruction using recent data obtained using a single-photon avalanche diode (SPAD) [9]. The SPAD is a more affordable transient imaging device, at the cost of lower quality measurements (i.e. less spatio-temporal resolution and higher levels of noise). A voxel grid of resolution $\bar{X} = 150^3$ is reconstructed from a set of $S = 185$ temporal measurements with different lighting position $\mathbf{s}$. Each measurement has a single pixel $\bar{P} = 1$, and a temporal resolution of $T = 14000$ frames. Note that this is a difficult scenario for our method, given the high levels of noise in the captured signal, as well as the large temporal resolution; however, our method takes only 1.8s to reconstruct the scene, in comparison to 14.8s for traditional back-projection.

4.1. Analysis

We compare the performance of both our method and traditional back-projection [3] varying the three parameters involved in the reconstruction: The resolution of the output voxelized space $\bar{X}$, the input spatial and temporal resolution $\bar{P}$, and T. We use a synthetic scene (Fig. 4) using the time-resolved rendering framework by Jarabo et al. [17]. The scene is similar to scenes captured in previous works [8, 9]. We opt for a synthetic scene to eliminate errors that depend on the camera and the capture setup, and obtain a clean ground truth solution.

Figure 5 (top) shows a comparison of the cost of traditional back-projection and our method,

Optics EXPRESS

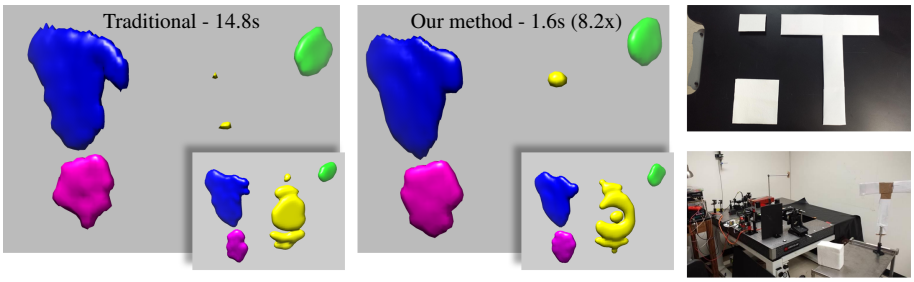

Fig. 3. Reconstruction of the scene captured using a SPAD [9] using traditional back-projection (left), and our method (right). The rightmost images show the capture setup (insets from [9]). We have used a different color to differentiate each object (blue: T; pink: big patch; green: small patch; yellow: noise from the camera). The quality of the reconstruction is almost identical. Note that the original dataset had a higher amount of camera noise, which we have filtered before reconstruction (the insets show the reconstruction with each method for the original unfiltered data). Our method takes only 1.8s to reconstruct the scene, in comparison to 14.8s for traditional back-projection.

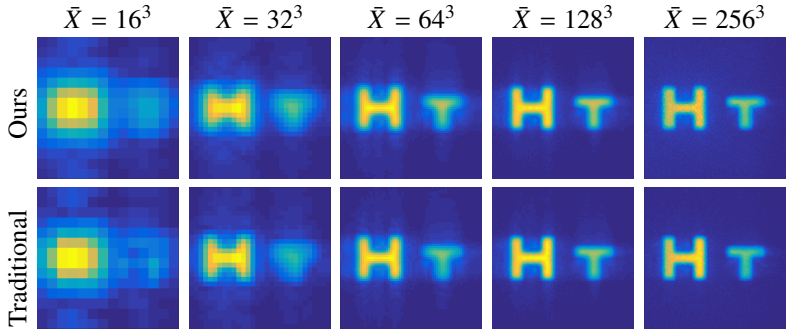

Fig. 4. Comparison between the reconstruction quality computed with our method (top) and traditional back-projection (bottom), for an increasing number of voxels $\bar{X}$. A comparison of the cost of both algorithms, as well as the progression of their numerical error can be found in Fig. 5, left.

varying one parameter and fixing the other two. In all cases our method is significantly more efficient for all practical resolutions. In particular, our method shows a much better scalability than the previous work with respect to the number of reconstructed voxels $\bar{X}$, showing speed-ups of up to 10000x for large resolutions, while having similar convergence with respect to the number of input pixels $\bar{P}$. On the other hand, our method scales linearly with the number of frames T, while the cost of traditional back-projection remains constant with respect to this parameter; however, note that even for large temporal resolutions (e.g. 10^4 frames) our method is still two orders of magnitude faster. In terms of reconstruction error, our algorithm scales similarly to traditional back-projection. Figure 4 shows the reconstruction result for a varying reconstruction resolution $\bar{X}$, while Fig. 5 (bottom) shows the error with respect to the ground truth for varying input parameters. Thus, despite the cost is significantly lower with our algorithm, the error introduced is always comparable with the previous work.

An important additional parameter of our method is the quality of the ellipsoids' tessellation [Eq. (6)]. This determines the number of triangles used to represent the ellipsoid during the voxelization, and has an impact in both the cost and the quality of the reconstruction. Figure 6 shows the result with varying number of triangles, compared with the traditional method for a reconstructed space of $\bar{X} = 256^3$ voxels. For tessellation levels leading to triangles of approxi-

Optics EXPRESS

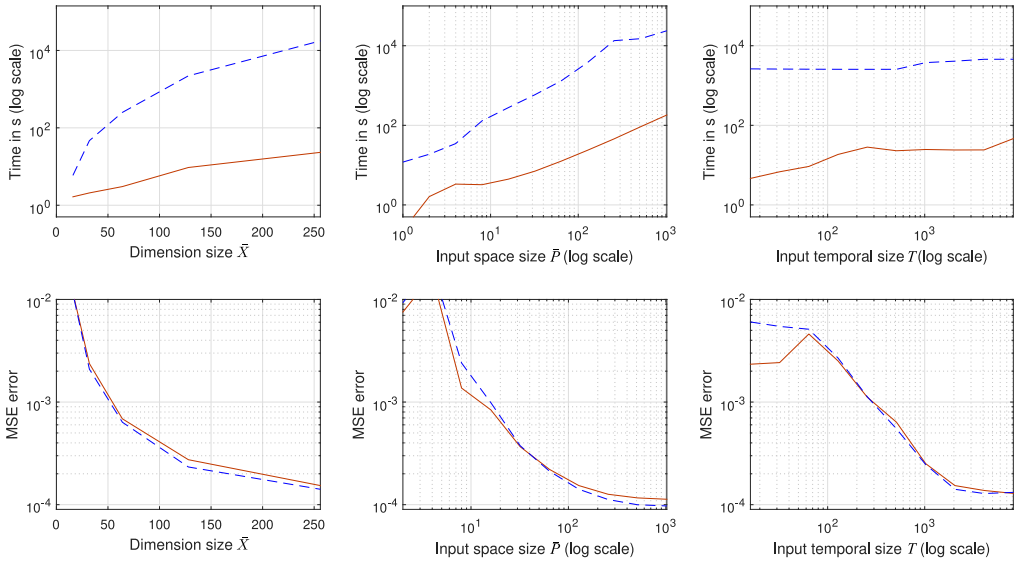

Fig. 5. Cost (top) and error with respect to the ground truth (bottom) comparisons between our method (solid orange) and traditional back-projection (dashed blue) for the synthetic scene shown in Fig. 4. Each graph varies one reconstruction parameters while fixing the other two, namely the number of voxels $\bar{X}$ (right), input spatial resolution $\bar{P}$ (middle) and input temporal resolution T (right). We fix the parameters to a reconstruction space of $\bar{X} = 256^3$ voxels, an input spatial resolution $\bar{P} = 128$ pixels, and temporal resolution $T = 1024$ frames. For all reconstruction we use 128 measurements with different virtual light positions $\mathbf{s}$. As shown in the bottom, the error introduced by our algorithm with respect to traditional back-projection is negligible. Note that at low spatial $\bar{P}$ and temporal T resolutions the error in both cases is large and scene dependent, which is the reason for the counter-intuitive behavior at very low $\bar{P}$ (bottom center) and T (bottom right).

mately the voxel size, our reconstruction achieves a reconstruction quality similar to traditional back-projection, more than a thousand times faster. This is further illustrated in Fig. 7: We can see that our reconstruction quality is bounded by the reconstruction resolution, and at certain point increasing the number of triangles does not improve the result, while the cost increases linearly. In our tests we found that setting the triangle size to roughly the voxel size leads to a sweet-spot in terms of reconstruction quality and cost.

4.2. *Discussion and limitations*

Our method takes advantage of the relatively sparse signal of time-resolved data to scale below the theoretic computational order (Section 3.2), by ignoring ellipsoids $E(\mathbf{s}, \mathbf{p}, t)$ with $I_{\mathbf{s}}(\bar{\mathbf{p}}, t) \approx 0$. Thus, our worst-case scenario occurs for non-sparse signals. While this is not problematic in surface-based NLOS reconstruction, it becomes dominant in the presence of participating media. In these cases, our technique still performs significantly faster than traditional back-projection: For an input resolution of $\bar{P} = 128^2$ and $T = 1024$ and a voxel resolution of $\bar{X} = 256^3$, traditional back-projection takes 475.78s, as opposed to 12.51s for for our method in the worst-case scenario.

Our algorithm shares some of the limitations from previous back-projection-based NLOS methods. In particular, our probabilistic model [Eq. (4)] assumes Lambertian surfaces, and ignores the effect of albedo. Moreover, Eq. (4) assumes that all light arriving the sensor is due to third-bounce reflection. While given the relative intensity between bounces is a sensible assumption, this might incur into errors in areas where higher-order scattering is dominant such as concavities.

Optics EXPRESS

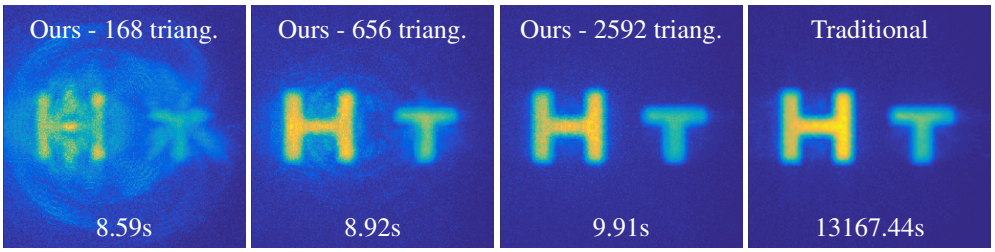

Fig. 6. Reconstruction results of our method with increasing ellipsoid tessellation quality (168, 656 and 2592 triangles per ellipsoid, respectively). The rightmost image shows the reconstruction result using using traditional back-projection. Our final reconstruction is comparable to traditional back-projection, computed with a speed-up of 1300x.

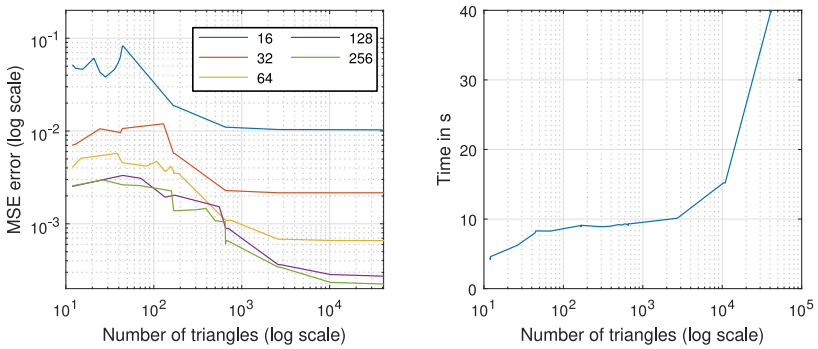

Fig. 7. Left: MSE of our method with respect to a ground truth reconstruction for different voxel resolutions $\bar{X}$, and for increasing number of triangles per ellipsoid. The minimum error is bounded by the reconstruction resolution, while increasing the number of triangles past a certain point does not improve the result. Right: Reconstruction time as a function of the number of triangles per ellipsoid used (reconstruction resolution of $\bar{X} = 256^3$). In practice we found that a near-optimal result can be obtained with a tessellation level of roughly the size of the voxel.

In addition, while our technique scales very well with respect the reconstruction resolution, the maximum reconstruction resolution is limited by the number of ellipsoids that can be generated from the input. If the input resolution is too low for the reconstruction voxelization, then several gaps due to insufficient ellipsoid coverage might appear. This can be solved by simply upsampling the input signal so that more ellipsoids can be generated, therefore increasing the cost in our algorithm. In this case the result would be similar to upsampling a lower-resolution reconstructed voxelized space. This imposes a maximum boundary on the resolution that can be achieved from a given input capture resolution.

5. Conclusions

Current reconstruction methods for NLOS reconstruction are still too slow for practical perfomance, becoming a key bottleneck of the process. In this work, we have presented a new method for NLOS reconstruction based on back-projection. Our work builds on the observation that the light incoming at a given pixel at instant t can have been reflected only from the points defined by an ellipsoid with poles at the observed point and the light source, allowing to define the probability of the NLOS geometry as the intersection of several of these ellipsoids, and exploits it by posing NLOS reconstruction as a voxelization problem. This allows us to reduce the computational cost significantly, and achieve speed-ups up to three orders of magnitude. Our

Optics EXPRESS

technique can be efficiently implemented on the GPU. We hope our work will help current and future NLOS reconstruction techniques, enabling their use in practical real-world setups: With that aim, we have made our code publicly available.

A. Implementation details

We have implemented our voxelization in OpenGL 4.4. We set the frame buffer to a 3D single channel `UINT32` texture, with depth testing disabled since we want *all* ellipsoids to be rendered. Most operations during rendering are performed in the geometry and fragment shaders: the former is used to select the best possible projection for tessellation (normal swizzling) [14], which is important to avoid possible holes. In the latter we perform the drawing by means of an atomic add, which updates the voxel's probability $p(\mathbf{x})$.

Due to limitations on current GPU architectures, we must set $p(\mathbf{x})$ as a `UINT32` value. To take this into account, we normalize the intensity values $I_s(\bar{\mathbf{p}}, t)$ to 255. We set this maximum value to prevent data overflow, which would be reached when intersecting more than 2^{24} ellipsoids in a worst-case scenario. Note that this assumes that the signal's intensity can be represented within this range. If we would like to compute different bounces, they should be computed separated, to accommodate for the large differences in intensity between them.

Although normal swizzling provides a good projection quality with almost no holes, there may still be some gaps in the probability map, due to fragment rejection when the triangle does not include the center of the projected pixel. While this could be fixed by conservative voxelization [18], multiple rasterizations of the same triangle may result in an overestimation of the probability map, so we choose not to apply it.

Finally, writing data in GPU using arbitrary access to graphics memory does not guarantee coherency, so we manage internal coherency manually by explicitly waiting for termination of write operations before filtering or rendering additional batches.

Funding

Defense Advanced Research Projects Agency - DARPA (HR0011-16-C-0025); European Research Council (ERC Consolidator Grant 682080); Spanish Ministerio de Economía y Competitividad (TIN2016-78753-P, TIN2014-61696-EXP).

Acknowledgements

We want to thank Andreas Velten for providing the captured data, as well as the reconstruction code from [3], and Julio Marco for his help setting up the figures.