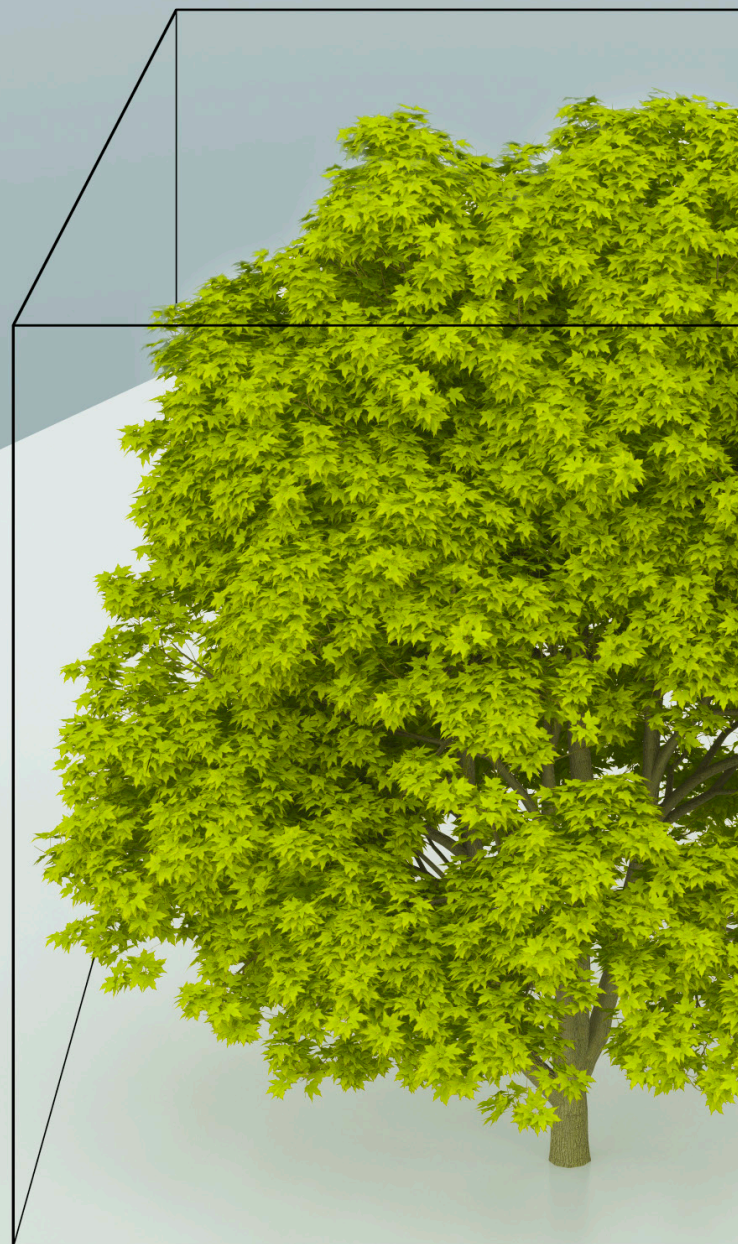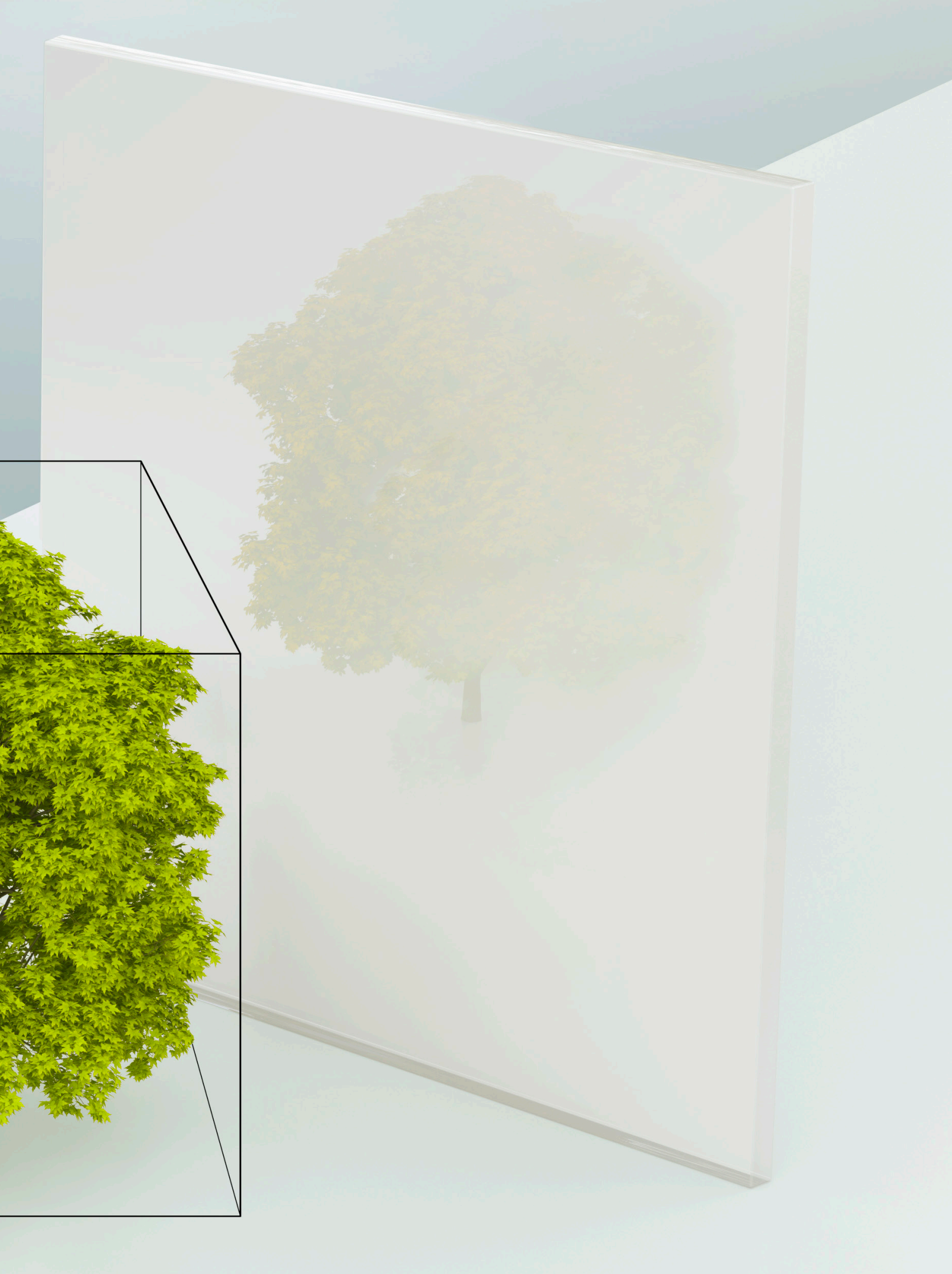# Toward Non-Line-of-Sight Videography

Xiaohua Feng and Liang Gao

Seeing objects hidden from sight, in real time, could significantly expand humanity's ability to explore the unknown.

Bending light with refractive lenses has revolutionized the way people picture the world. Until the late 19th century, pinhole cameras, which rely on straight-line propagation of light, were the mainstream technique for photography—but that technique was painfully slow. The invention of quality lenses to refract and focus light quickly eclipsed those cameras, allowing sharp images to be generated much faster.

Although the deterministic bending of light laid the foundation for modern imaging devices, it confined both human and machine vision to the direct line of sight. Even now, it remains challenging to image objects hidden behind an obscuring barrier. But all is not lost; rough surfaces around the hidden scene can relay the light originating from invisible objects to an observer. The problem is that the relay process is essentially random, bending light in a multitude of directions. This light-scattering process is the main culprit hampering vision around corners or into opaque media such as biological tissue.

Non-line-of-sight (NLOS) imaging attempts to deal with this issue by transforming a rough, light-scattering surface into a diffusive "mirror" that can fold hidden objects back into sight. A number of groups have demonstrated the ability to track sparse sets of obscured objects, or to estimate a 2D image of simple hidden objects with ordinary cameras under certain circumstances. Yet field-deployable NLOS imaging of large-scale, complex 3D scenes remains a great challenge. In this feature, we look at some recent efforts, including work in our own lab, to leverage new ultrafast cameras and computational techniques to move closer to real-time, video-rate NLOS imaging.
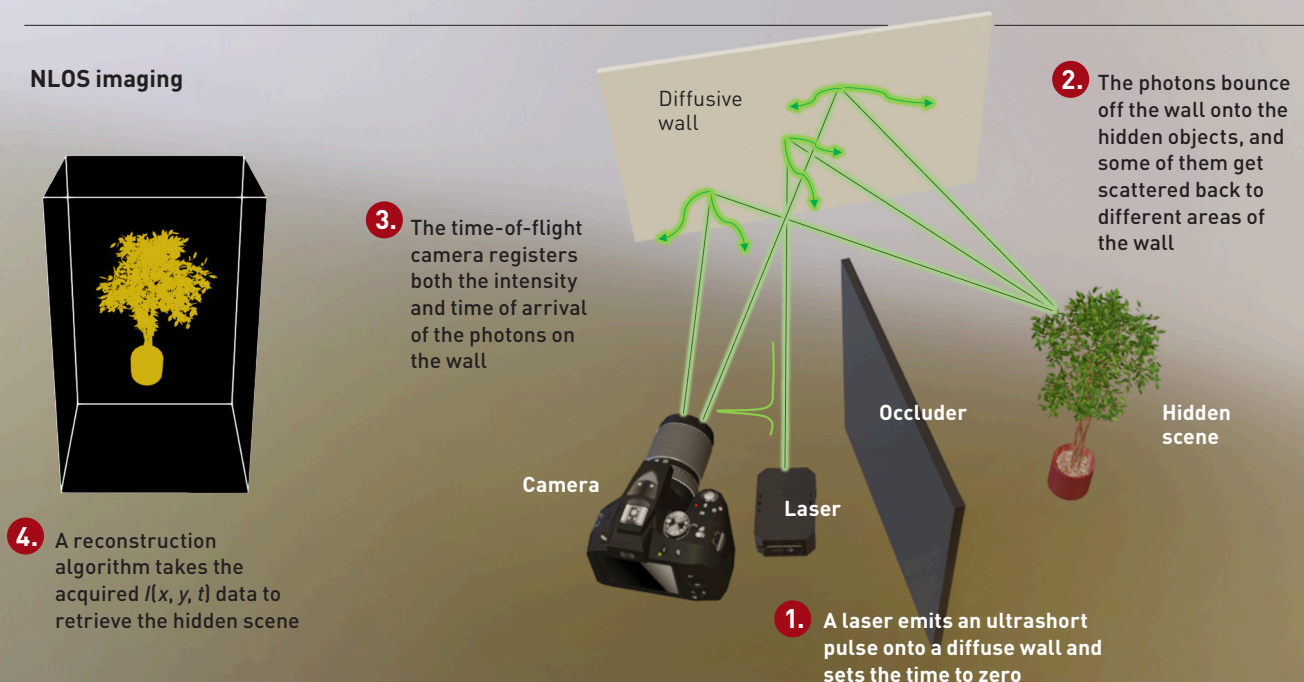
Echolocation and NLOS imaging

Echolocation uses the time delay between a transmitted sound wave and returned echoes to localize targets by triangulation. NLOS imaging, a similar concept, relies on the transmission and reception of photons, as mediated by a diffusive wall.

Echolocation

- Receiver
- Transmitter
- - - - transmission
- - - - echo

Triangulation
Receiver 1
Receiver 2
Target

NLOS imaging

Diffusive wall

2. The photons bounce off the wall onto the hidden objects, and some of them get scattered back to different areas of the wall

3. The time-of-flight camera registers both the intensity and time of arrival of the photons on the wall

Occluder

Hidden scene

Camera

Laser

4. A reconstruction algorithm takes the acquired $I(x, y, t)$ data to retrieve the hidden scene

1. A laser emits an ultrashort pulse onto a diffuse wall and sets the time to zero

The key to converting the light randomly bent by the surface into high-quality NLOS images was to record not only the intensity but also the exact time-of-arrival of detected photons.

From echolocation to NLOS imaging

Three-dimensional sensing of a scene without direct vision is common in nature. Bats, for example, can navigate in the dark via echolocation. The time delay between the emission of a sound wave and the returning echo signal allows the bat to calculate the distance from its current location to the target. Obtaining two such measurements at different but closely spaced spots—that is, the animal's two ears—enables the target to be localized accurately in 3D space via a simple triangulation.

In a seminal work published in 2012, Andreas Velten and colleagues demonstrated that a similar time-of-flight method, using light instead of sound, can provide high-quality 3D imaging of hidden scenes. The emission and reception of light waves were mediated by a diffusive surface, which scattered an incident laser pulse randomly onto hidden objects and relayed subsequently reflected photons back to the camera.

The key to converting the light randomly bent by the surface into high-quality NLOS images was to record not only the intensity but also the exact time-of-arrival of detected photons. Although the light's directions were scrambled, its time-of-arrival at different points of the surface remained distinct. Given a dense measurement of photons' time-of-arrival over a large area on the surface, the directions of reflected light could be computationally sorted to recover the 3D hidden scene in detail.

A big impediment, though, was that, in the absence of a suitable ultrafast camera, it took a very long time to acquire a dense time-of-flight dataset. Computationally reconstructing a hidden scene of any complexity also was time-prohibitive, even with powerful processors. Indeed, early NLOS imagers are even slower than the pinhole cameras of the late 19th century. A "lens revolution" for NLOS imaging was clearly needed to make these systems fast and practical.

Fortunately, the fascination for seeing beyond the direct line of sight, and the great potential of time-of-flight methods in tackling light scattering, has inspired researchers from various disciplines—signal processing, optical imaging and sensor design—to move NLOS imaging toward real-world applications. In the three years since this publication's last feature article on the topic (see "Non-Line-of-Sight Imaging," OPN, January 2019), breakthroughs have occurred in the design of ultrafast cameras and efficient reconstruction algorithms. The synergistic integration of those developments has, we believe, dramatically improved prospects of attaining real-time NLOS imaging in the field.
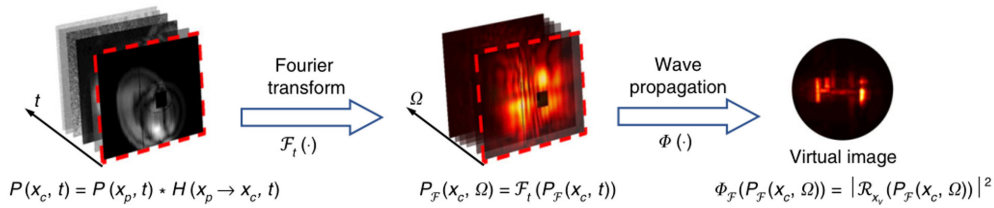
Efficient NLOS reconstruction

Early development of time-of-flight NLOS imaging strove to record a 5D dataset: the incident laser was scanned in 2D across the surface, and 3D time-of-flight data (x,y,t) were recorded at each position. To reconstruct the hidden scene with $N \times N \times N$ voxels, a back-projection algorithm computes the round-trip time light takes to propagate from the incident point to every voxel in the volume and then to return to a position (x,y) on the surface. This leads to a computational complexity of $o(N^6)$. Even with massively parallel computing, which can accelerate the reconstruction from hours to minutes, the imaging speed is still orders of magnitude slower than in ordinary cameras.

The first breakthrough toward practical NLOS imaging was the confocal method proposed in 2018 by Gordon Weinstein's group at Stanford University, USA, which reduces the time-of-flight dataset from 5D to 3D by co-locating the laser illumination and photon detection. More important, confocal NLOS facilitates a light-cone-transform imaging model. Such a model relates the time-of-flight data $I(x,y,t)$ to the 3D hidden scene via a convolution with a shift-invariant kernel, allowing the 3D hidden scene to be accurately retrieved by a Wiener filter–based deconvolution.

The key to vastly accelerating the reconstruction is to implement the deconvolution in the Fourier domain, via fast Fourier transforms (FFT). This slashes the complexity from $o(N^5)$ operations for a 3D time-of-flight dataset to only $o(N^3 \log N)$—an improvement of more than four orders of magnitude for NLOS imaging at a resolution of 100×100×100 voxels. When executed with parallel computing, the reconstruction can be completed within several milliseconds.

Fourier-domain reconstruction

$$P(x_c, t) = P(x_p, t) * H(x_p \rightarrow x_c, t) \qquad \mathcal{F}_t(\cdot) \qquad P_\mathcal{F}(x_c, \Omega) = \mathcal{F}_t(P_\mathcal{F}(x_c, t)) \qquad \Phi(\cdot) \qquad \Phi_\mathcal{F}(P_\mathcal{F}(x_c, \Omega)) = |\mathcal{R}_{x_v}(P_\mathcal{F}(x_c, \Omega))|^2$$

Imaging results

Speeding up with FFT

Implementing wave propagation in the Fourier domain, via efficient fast Fourier transform (FFT) algorithms, dramatically speeds up NLOS imaging reconstruction. The raw time-of-flight data are Fourier transformed along the time dimension and each temporal frequency component is then propagated in the spatial frequency domain to the hidden scene. This allows the phasor-field method to work well with both confocal and non-confocal datasets.

Adapted from Liu et al., Nat. Commun. **11**, 1645 (2020); CC-BY 4.0

The confocal method has demonstrated dynamic NLOS imaging of retroreflective objects at a few frames per second. Still, it remains restricted to a low image resolution, as it still relies on point-scanning acquisition, a mechanism that fundamentally impedes high-speed imaging. To attain real-time imaging, the reconstruction must efficiently accommodate the parallel acquisition of time-of-flight data—which is inevitably non-confocal.

The phasor-field framework takes a different approach to this problem. In this approach, the intensity measurement of time-of-flight data is transformed into a virtual, complex wave field, thereby allowing NLOS imaging to be modeled by diffractive wave propagation—a canonical and well-studied process in line-of-sight imaging systems. Decades of research and optimization have made highly efficient solvers for diffractive wave propagation readily available, allowing the phasor-field method to optimize imaging speeds with both confocal and non-confocal time-of-flight data.

Recent efforts by Xiaochun Liu and colleagues have further implemented the phasor-field method in the Fourier domain (again via FFT) to attain fast, high-quality NLOS imaging. The full potential of the phasor-field method, however, remains to be realized by parallel time-of-flight detectors—ideally a 2D ultrafast camera that can enable point-and-shoot NLOS imaging.

Snapshot acquisition of time-of-flight data

Long sought by various disciplines, an ultrafast camera capable of acquiring large-scale time-of-flight data without tedious scanning has remained elusive. For NLOS imaging in particular, a high-quality 3D reconstruction requires time-of-flight data with a resolution better than 100×100×1000 (2D space and time). The magnitude of the time-of-flight data cube and the need for a picosecond temporal resolution rules out most existing ultrafast cameras, especially the single-shot variety.

Compressive ultrafast photography (CUP), invented in 2014, has been the only approach capable of recording time-of-flight data at greater than 100×100×100 (2D space and time) resolution in a single shot, attaining a temporal resolution down to a few hundred femtoseconds. Yet scaling this technology further for NLOS imaging has been hindered by the technique's inherent spatial–temporal crosstalk and, above all, by a large compression factor in image acquisition.

Nonetheless, the idea of adapting CUP for NLOS imaging has prompted a rethinking of the need for densely sampling the time-of-flight data. In the phasor-field framework, after all, the relay surface is essentially the aperture of the NLOS imager, on which the images show drastically different characteristics than in conventional photography. A close examination of the instantaneous images on the surface originated from hidden scenes of different complexities clearly reveals

Essentially, LIFT reformulates photography as a computed-tomography problem that enables multiplexed image acquisition using measurement data only a fraction of the image size.

their compressibility, indicating the potential for a large reduction of the time-of-flight data without compromising NLOS imaging quality.

Yet simply sampling the surface on a low-resolution grid leads to an equally low-resolution reconstruction, as demonstrated by pioneering NLOS tracking experiments in 2016 from the lab of Optica Fellow Daniele Faccio, which used a 32×32 single-photon avalanche diode (SPAD) camera. The key is to find a compressive acquisition scheme that preserves sufficiently high-frequency information in instantaneous images.

A LIFT toward real-time NLOS

We believe that an approach currently being pursued in our lab, light-field tomography (LIFT), fills this void with a surprisingly simple and robust implementation. LIFT uses a cylindrical lens to acquire a parallel beam projection of the *en face* image onto a 1D sensor along its invariant axis (the one without optical power). As with the compound eye of insects, arranging an array of distinctly oriented cylindrical lenses with respect to the sensor allows both the light fields and the different projection data of the scene to be compactly sampled. An image can then be robustly recovered with an analytical or optimization-based algorithm.

Essentially, LIFT reformulates photography as a computed-tomography problem that enables multiplexed image acquisition using measurement data only a fraction of the image size. Moreover, it introduces light-field imaging to an ultrafast time scale, a feat previously thought unlikely due to the severely limited element counts of ultrafast detectors. Light-field capability is crucial for light-starved NLOS imaging, because, in practical applications, the camera must cope with arbitrarily curved surfaces showing large depth variations without compromising light throughput.

Of particular interest is LIFT's sampling behavior. Like X-ray computed

tomography, the projection data in LIFT are related to the image through the central-slice theorem—the Fourier transform of the projection data constitutes a slice of the image spectrum along the projection angle. With a sufficient number of projections, the image spectrum can be fully covered. But under compressive acquisition, the sampling pattern gets sparser as the spatial frequency increases—that is, as with natural photographs, the image energy decays with increasing frequency, with most energy being clustered at the low-frequency band. This enables LIFT to obtain high-quality NLOS imaging of complex scenes with only a small number of projections (simulation studies show that 14 projections will suffice), a drastic reduction in the measurement data.

It is worth noting that—given limited data bandwidth, and given the extra time dimension that quickly escalates the raw data volume of time-of-flight
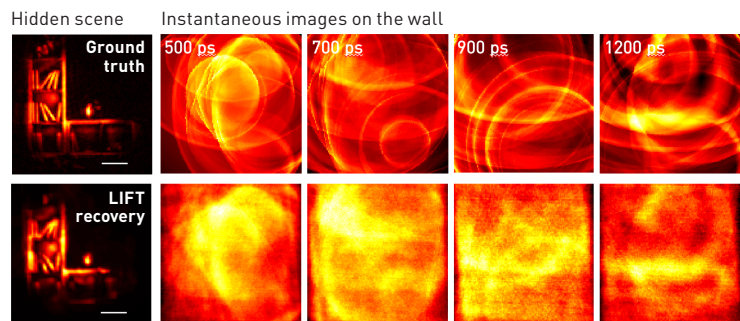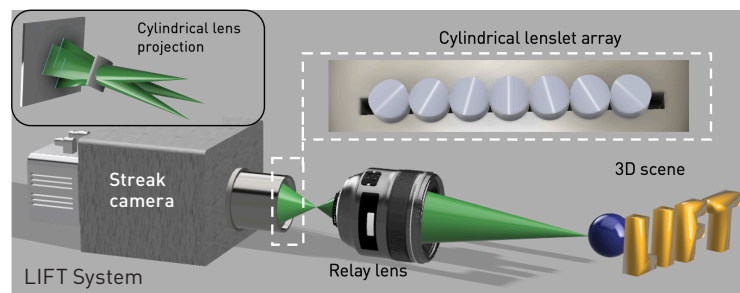
LIFT imaging approach

Light-field tomography (LIFT), analogous to computed-tomography systems, samples the light field in multiple different projections via a distinctly oriented cylindrical lens array (top). For NLOS imaging, the instantaneous images on the surface are highly compressible, which is exploited by LIFT for high-quality NLOS imaging with drastically reduced measurement data (bottom).

Adapted from X. Feng and L. Gao, Nat. Commun. **12**, 2179 (2021); CC-BY 4.0
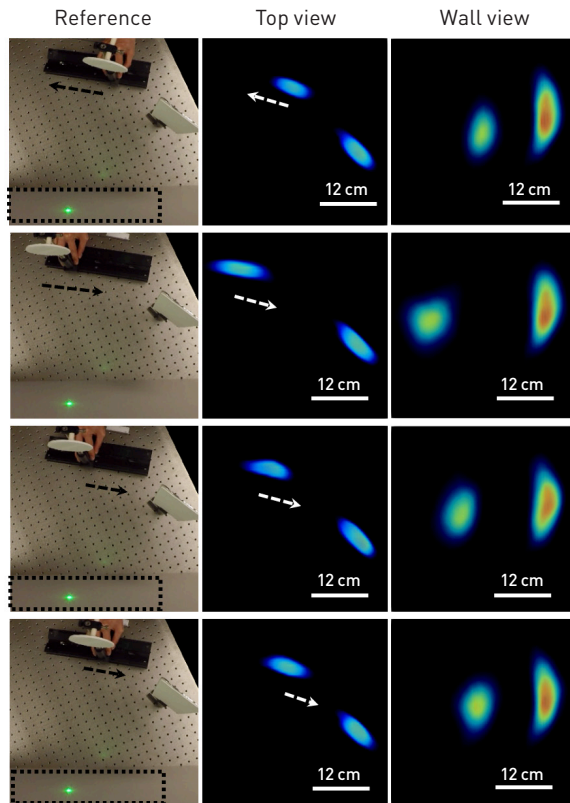
NLOS videography (30 Hz frame rate) depicting a circular plate being translated back and forth against a static strip at the background. Dashed arrows indicate movement direction.

measurement—compressive acquisition might be imperative for real-time imaging, even as high-resolution ultrafast cameras such as the megapixel SPAD sensor become widely available. Although several other schemes have recently been introduced to speed up NLOS imaging through sparse data acquisition, we believe that LIFT's efficient light-field imaging and sampling remain unique.

Machine learning and NLOS videography

Compressive data acquisition effectively shifts the imaging burden from the analog sensor to the digital-computation domain, the evolution of which has substantially outpaced that of sensor design. To make real-time NLOS imaging practical, image recovery from a compressive measurement must be made at least as efficient as current NLOS reconstruction. Yet iterative solutions for LIFT, as with inverse problems in general, are anything but fast. Efforts to speed up the solution have proceeded on various fronts, ranging from devising algorithms with faster convergence to parallelizing computations on graphical processing

units (GPUs) to, in the simplest case, relying on Moore's law, which promises approximately twice the computing power every two years.

A paradigm shift lies in a data-driven approach built upon machine-learning algorithms. These libraries take full advantage of high-performance GPUs and of free access to massive image datasets. In this new paradigm, measured imaging data are assumed to be related to a ground-truth image via an implicit function, represented by a deep neural network. The neural net is trained to learn the implicit function using thousands or even millions of measurement-and-image pairs, and the parameters of the network are dynamically adjusted until it yields a prediction of ground-truth images with sufficiently small errors.

Once so trained, the deep neural network operates as a proxy for the implicit function. It can thus directly map an input measurement to a corresponding image, thereby accomplishing the reconstruction at a speed orders of magnitude faster than iterative methods, and shortening the effective acquisition time (sampling plus reconstruction) of a technique such as LIFT.

One drawback of the machine-learning approach is its hunger for raw training data—datasets of close to a million are not unusual for solving vision-related problems. A large portion of such data can be obtained via simulations. For best results, however, the neural network typically needs further refinement from experimental data, which can be time-consuming or expensive to acquire. Fortunately, unlike most computer vision tasks such as face recognition, LIFT rests on a well-defined forward model relating the measurement data to the ground-truth image. As a result, blind end-to-end training of a deep neural network can be largely avoided, and the training dataset dramatically reduced.

With both efficient NLOS reconstruction algorithms and a compressive 2D ultrafast camera in place, we believe that the building blocks for real-time NLOS imaging are now complete. Using a low laser power of only 2 mW, LIFT has demonstrated 3D videography of a small-scale (roughly half a meter) dynamic scene that involves an object moving along an irregular path at a 30-Hz frame rate. The potential for LIFT in accommodating non-planar surfaces for NLOS imaging in the field has also been tested by digitally refocusing the image acquisition at different depths. The result showed LIFT's ability to obtain high-quality portraits of hidden scenes that are otherwise blurred when the relay surface exhibits a large depth variation.

We think point-and-shoot NLOS imaging—the ability to look "around the corner" in the field, and in real time—is on the cusp of becoming a reality.

The outlook for field-based NLOS videography

Moving proof-of-concept demonstrations of NLOS videography to the field will require overcoming several hurdles. Because of the r^4 decay of NLOS photons, large-scale NLOS imaging—for example, room-sized imaging—will need extremely sensitive detectors. SPAD sensors have emerged as the first choice owing to their exceptional sensitivity and compact size.

However, the tendency of SPAD sensors to experience pile-up distortions (that is, measurement distortions due to an excessive number of received photons at early times) makes detecting weak NLOS photons challenging under ambient illumination, such as daylight. Post-capture signal processing can mitigate this problem but is effective only under weak ambient light. Fast gating of the SPAD sensor can deal with stronger ambient illumination, when coupled with dynamically optimized light attenuation or special modulation of the detection process. However, these steps extend the data acquisition time. Thus large-scale NLOS videography in outdoor environments will require ingenious methods.

A lesser challenge is to adapt efficient reconstruction algorithms, which have hitherto been restricted to planar relay surfaces, to deal with arbitrarily curved surfaces. David Lindell and colleagues have proposed a transformation method to convert time-of-flight data from a curved surface to the corresponding data that would have come from a planar one, but the computation complexity is on par with the back-projection method and thus is not appealing. A hybrid approach synergizing the flexibility of time-domain wave propagation and the efficiency of Fourier-domain NLOS reconstruction might provide a solution.

One essential but often overlooked ingredient for field-deployable NLOS imaging is a real-time 3D imager for measuring the relay surface's geometry—currently obtained via nontrivial calibrations with a depth camera. Lidar would seem a natural solution, as it relies on similar hardware as an NLOS imager and, most important, is the only solution to handle the deployment of NLOS imaging at a large standoff distance. The downside of lidar is its currently low resolution and imaging speed.

However, impelled by the needs of the autonomous-vehicle community, large-format SPAD sensors and VCSEL laser diode arrays are becoming more accessible for time-of-flight measurements—benefiting both high-speed lidar and NLOS imaging.

With the rapid development of novel imaging techniques that cohesively synergize advances in signal-processing algorithms, efficient light generation and sensitive photon-counting detectors, we believe that these challenges can be met in the near future. As a result, we think point-and-shoot NLOS imaging—the ability to look "around the corner" in the field, and in real time—is on the cusp of becoming a reality. This capability for imaging beyond the direct line of sight could become an important complement to current human and machine vision systems—and could significantly expand our ability to explore the unknown. **OPN**

Xiaohua Feng (fengxiaohua@zhejianglab.com) is with the Research Center of Intelligent Sensing, Zhejiang Lab, Hangzhou, China. Liang Gao (gaol@ucla.edu) is with the Department of Bioengineering, University of California, Los Angeles, USA.

References and Resources

▸ A. Velten et al. "Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging," Nat. Commun. **3**, 745 (2012).

▸ L. Gao et al. "Single-shot compressed ultrafast photography at one hundred billion frames per second," Nature **516**, 74 (2014).

▸ G. Gariepy et al. "Detection and tracking of moving objects hidden from view," Nat. Photon. **10**, 23 (2016).

▸ M. O'Toole et al. "Confocal non-line-of-sight imaging based on the light-cone transform," Nature **555**, 338 (2018).

▸ D.B. Lindell et al. "Wave-based non-line-of-sight imaging using fast *f-k* migration," ACM Trans. Graph. **38**, 116:1 (2019).

▸ X. Liu et al. "Non-line-of-sight imaging using phasor-field virtual wave optics," Nature **572**, 620 (2019).

▸ C. Saunders et al. "Computational periscopy with an ordinary digital camera," Nature **565**, 472 (2019).

▸ X. Liu et al. "Phasor field diffraction based reconstruction for fast non-line-of-sight imaging systems," Nat. Commun. **11**, 1645 (2020).

▸ C.A. Metzler et al. "Deep-inverse correlography: Towards real-time high-resolution non-line-of-sight imaging," Optica **7**, 63 (2020).

▸ X. Feng and L. Gao. "Ultrafast light field tomography for snapshot transient and non-line-of-sight imaging," Nat. Commun. **12**, 2179 (2021).