

Deep learning optimized single-pixel LiDAR

Cite as: Appl. Phys. Lett. **115**, 231101 (2019); doi: [10.1063/1.5128621](https://doi.org/10.1063/1.5128621)

Submitted: 20 September 2019 · Accepted: 29 October 2019 ·

Published Online: 2 December 2019







View Online



Export Citation



CrossMark

Neal Radwell,¹ Steven D. Johnson,^{1,a)}  Matthew P. Edgar,¹ Catherine F. Higham,²  Roderick Murray-Smith,² 
and Miles J. Padgett^{1,b)} 

AFFILIATIONS

¹SUPA, School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, United Kingdom

²School of Computing Science, University of Glasgow, Glasgow G12 8QQ, United Kingdom

^{a)}Electronic mail: steven.johnson@glasgow.ac.uk

^{b)}Electronic mail: miles.padgett@glasgow.ac.uk

ABSTRACT

Interest in autonomous transport has led to a demand for 3D imaging technologies capable of resolving fine details at long range. Light detection and ranging (LiDAR) systems have become a key technology in this area, with depth information typically gained through time-of-flight photon-counting measurements of a scanned laser spot. Single-pixel imaging methods offer an alternative approach to spot-scanning, which allows a choice of sampling basis. In this work, we present a prototype LiDAR system, which compressively samples the scene using a deep learning optimized sampling basis and reconstruction algorithms. We demonstrate that this approach improves scene reconstruction quality compared to an orthogonal sampling method, with reflectivity and depth accuracy improvements of 57% and 16%, respectively, for one frame per second acquisition rates. This method may pave the way for improved scan-free LiDAR systems for driverless cars and for fully optimized sampling to decision-making pipelines.

© 2019 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/1.5128621>

Imaging our surroundings in 3D has become a key challenge across a range of sectors, including gaming, robotics, health-care, security, manufacturing, and automotive industries. Mature technologies such as radar and ultrasound sensing are effective at long and short ranges, respectively, but are unable to provide the spatially detailed maps necessary for many applications within the 10–100 m range. An attractive solution for these intermediate ranges is light detection and ranging (LiDAR), capable of providing millimetric depth precision at ranges of around 100 m with a good spatial resolution.¹

LiDAR is capable of recovering depth information using direct time-of-flight measurement,² frequency modulation continuous wave (FMCW),³ or the amplitude modulation continuous wave (AMCW)⁴ approaches. The FMCW method ramps the modulation frequency and measures the difference between the outgoing and return frequencies, and AMCW correlates the outgoing and return intensities to determine the distance to a target. Both these methods require a significant return signal, and therefore, high laser powers are required and the techniques cannot be used at long range. By contrast, time-of-flight methods measure the return delay of a short pulse and can be sensitive at the single photon level, with a temporal resolution in the tens of picoseconds. Time-of-flight LiDAR methods have therefore become the state-of-the-art for long-range, high-precision depth mapping.

To obtain millimetric precision for LiDAR, subnanosecond temporal resolutions are required, ruling out the use of conventional cameras. In recent years, there has been significant development of single photon avalanche detector (SPAD) array cameras, which combine multiple SPADs and their timing electronics on-chip.⁵ SPAD array technology allows scan-free LiDAR with longer integration times and consequently improved signal-to-noise (SNR) ratio. However, at present, the SPAD array technology suffers from high dark counts and readout noise, while also demanding stringent requirements on readout clock rates and data transfer rates, both of which can limit their performance and application readiness. Typical LiDAR systems are currently therefore more usually based on spot-scanning techniques, requiring mechanical scanning systems with a reconstruction resolution and acquisition rates limited by scan speed.⁶

Single-pixel imaging (SPI) techniques are an alternative imaging modality for recovering spatial information.^{7,8} SPI uses a single bucket-detector combined with a spatial modulator, most commonly a digital micromirror device (DMD), to measure the spatial overlap between the scene and the spatial patterns of an imaging basis. Simple orthogonal imaging bases such as the Hadamard basis⁹ are able to reconstruct an N pixel image from N sequential measurements, with the acquisition time traditionally limited by modulator speed.

However, we predict that in the case of single-photon counting (Geiger-mode) LiDAR measurements, the maximum count rate in photons per second (due to technical constraints of current single-photon detectors) becomes the limiting factor.

To explore whether the photon measurement rate is the limiting factor in the signal-to-noise (SNR), we make an order-of-magnitude estimate of the number of photons required to produce a recognizable image. If we consider a scene with an equal fraction of light and dark, and as each mask in our sampling patterns have 50% of the pixels “on,” the average overlap between the reflected signal and our pattern will be a quarter of the maximum possible signal for a fully reflecting scene. There will be an average measured signal for all patterns with fluctuations around this average level due to the varying back-scattered signals as the patterns of the basis are cycled. The information in SPI is contained within these fluctuations. If we say that the average signal level is $N/4$, then the degree of overlap between the pattern and the object gives fluctuations of order $\sqrt{N/4}$. The ratio of these fluctuations to the mean signal must be greater than SNR for the measurement. If each of these measurements is based upon P photons, the shot noise is \sqrt{P} , and therefore, we require that the shot noise is not greater than the information carrying fluctuations. We can state that the minimum number of photons required per pattern measurement is $\sqrt{P} > \sqrt{N/4}$. This implies that for an arbitrary scene, we will need $P \approx N$. For a fully sampled reconstruction of the image, we require N patterns and therefore a total of $\approx N^2$ photons per reconstruction. In a conventional scanning LiDAR or in a focal plane array, N photons are needed.

With our system, a 64×64 image will require $N = 4096$ photons per pattern, which will take $400 \mu\text{s}$ to acquire at 10^7 photons/s (a typical single-photon measurement rate of our commercially available detector), corresponding to a pattern rate of only 2.5 kHz and hence an image acquisition time of 1.6 s. Therefore, in this low photon number regime, the pattern modulation rate is no longer the limiting factor (which is typically tens of kilohertz); rather, maximizing the number of photons per pattern becomes essential. By reducing the number of patterns needed, the speed of acquisition can be increased while still producing an acceptable SNR.

Unlike scanning systems, the freedom to choose the sampling basis in SPI provides the opportunity to use compressed sensing techniques, where a high quality N pixel image can be reconstructed with fewer than N pattern projections and measurements.¹⁰ The most common compressed sensing approaches use sampling bases, which are maximally incoherent to the object basis, and use convex optimization techniques with a strong prior, such as total curvature minimization to recover the missing information. While effective, these techniques are computationally expensive and require highly sparse scenes, making them problematic for real-time (less than one frame per second) and therefore low-resolution applications. Several other approaches have been proposed for compressed sensing, which retain fast reconstruction speeds, including evolutionary compressed sensing,¹¹ Russian-dolls ordering,¹² terahertz imaging,¹³ and cake-cutting ordering¹⁴ and typically achieving compression ratios around 25%. More recently, a method of compressed sensing has been demonstrated using an optimized imaging basis and reconstruction algorithm derived from a trained convolutional neural network.¹⁵ This deep learning (DL) approach achieves a 4% compression ratio.

LiDAR systems have previously incorporated SPI techniques^{16,17} and photon counting.^{18,19} Here, we present a LiDAR system that uses SPI techniques with DL compressed sensing together with single-photon sensitive detection. Our single-pixel 3D imaging system is shown in Fig. 1. The illumination laser has a pulse length of 120 fs, a repetition rate of 80 MHz, and a center wavelength of 780 nm (frequency doubled Toptica FemtoFerb 1560). The laser illumination is expanded uniformly with a light integration tube, the output from which is reimaged onto a DMD. The high-speed DMD (Vialux model V-7001) provides time-varying structured illumination and consists of an array of 1064×768 micromirrors each approximately $10 \mu\text{m}$ in size, capable of pattern switching rates of over 20 kHz. Each micromirror is actuated to direct light either to the scene or to an internal beam stop, allowing binary amplitude modulation of the projected patterns. The DMD mirrors are grouped into larger “superpixels” to convert the 1064×768 micromirrors into the more typical 32×32 , 64×64 , or 128×128 resolutions for SPI. The DMD plane is imaged onto the scene using a standard high-quality camera lens, chosen to provide the appropriate field of view for the chosen range. A fast response Geiger-mode photomultiplier tube (Horiba PPD-900) detects back-scattered photons from the scene, which are spectrally filtered (Semrock LL01-780-12.5 and Thorlabs FB780-10 in series) to greatly reduce spurious photon events from background thermal light. The collection optics comprise a second camera lens to match the field-of-view of the detector to the structured illumination with a variable aperture. A time correlated single-photon counting (TCSPC) system is an efficient triggering device, which can resolve single-photon detection times with a jitter on the few tens of picoseconds range. Our TCSPC electronics (a customized Horiba DeltaHub) record the delay time for each detected photon “event” relative to the synchronization signal provided by the illuminating laser.

For each illumination pattern, the time-of-flight of many photon events is collected into a single histogram before being transferred to a laptop computer for image reconstruction. The TCSPC electronics enable continuous streaming of up to 20 000 histograms per second, each with 512 time bins with 25 ps widths, enabling real-time image

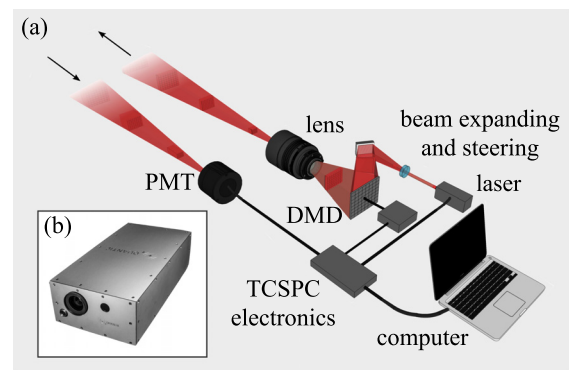


FIG. 1. LiDAR prototype. (a) A pulsed laser uniformly illuminates a digital-micromirror-device, used to provide structured illumination onto a scene, and the back-scattered light is collected onto a photomultiplier tube. The TCSPC electronics record photon arrival events and collect the delay times into histograms. The DMD triggers the histograms to be sent to the computer, where they are reconstructed into 3D point clouds. (b) Photograph of the enclosed LiDAR demonstrator unit.

construction. In practice, however, using all 512 time bins can result in reconstruction lagging behind acquisition, as each time bin has to be reconstructed individually. Therefore, in order to maintain real-time reconstruction, we group 8 time bins together, such that we have 64 distinct time bins, allowing 3D reconstruction within an acquisition time of 0.5 s. This pipeline bottleneck could be alleviated by computing the frames simultaneously with a graphics processing unit.

SPI sequentially illuminates the scene with patterned light, with each pattern forming part of a sampling basis, with many possible SPI sampling bases.^{9,20} In recent work,¹⁵ deep learning, specifically, a multilayer neural network called an autoencoder, was used to produce both an optimized highly compressed measurement basis and an algorithm capable of recovering real-time high-resolution images. In order for the measurement basis to be suitable for a general scene, a large image recognition dataset, the STL-10 dataset (100 000 images comprising 10 classes: airplane, bird, car, cat, deer, dog, horse, monkey, ship, and truck with natural backgrounds),²¹ was used for training. The multilayered autoencoder used a series of filters known as layers to encode the training images and decode the image from encoded measurements. This deep learning method was used to derive the optimal binary basis for this training dataset from the encoding layers of the neural network. A multilayered mapping algorithm to reconstruct an image from a series of spatially encoded measurements was developed by the decoding layers of the neural network. In this 3D reconstruction work, we use this general case measurement basis but adapt the reconstruction algorithm to recover 3D scene information rather than 2D images.

The basis set consists of 666 patterns with 128×128 pixels (i.e., 4% compression ratio), which were calculated using deep learning, specifically a deep convolutional neural network. We generalize the method from 2D scene reconstruction to 3D depth mapping by performing a reconstruction of every depth plane (time bin), and though this is somewhat outside of the scope of the 2D image DL training, we will show that the system is robust enough to provide impressive 3D results.

We test our DL basis against conventional fully sampled orthogonal imaging bases, namely, the Hadamard basis. To avoid patterns producing abnormally high, or low, return signals, we randomly permute the columns of H before reshaping. This “democratization” makes optimum use of the dynamic range of the detection. Both the Hadamard basis and the DL basis consist of -1 and 1 values, whereas our modulator produces 0 (no light) and 1 (light). By sequentially counting photons from a pattern and its contrast-reversed inverse pattern and taking the difference, the signal (photon number) corresponding to the -1 and 1 patterns can be deduced. For every new pattern displayed on the DMD, a trigger signal is sent to the DeltaHub histogrammer, transferring the current histogram to the computer and beginning acquisition of a new histogram.

Once a “set” of differential histograms (666 for DL sampling or N for Hadamard sampling) have been collected, the reconstruction is performed for each time bin of the histogram using our DL reconstruction or fast Hadamard transform. For Hadamard sampling, since the patterns are democratized, this returns an image in which the pixels are scrambled, and we unscramble this by inverting the democratization process. Reconstruction of an image at every time bin gives a 3D point-cloud of intensity. For typical scenes, each transverse position has a single depth (i.e., we are imaging nontransparent surfaces),

and so we choose to reduce noise by applying a $3 \times 3 \times 3$ Gaussian 3D smoothing kernel.²² This sparsity also lends itself to a much more compact representation by conversion of the 3D intensity map to a depth map where each transverse position (pixel) has a reflectivity and a depth determined as the time bin with the maximum; an example of this conversion can be seen in Fig. 2.

Quantitative analysis of the performance of the deep-learned and Hadamard imaging methods is gained by comparison to a ground truth reflectivity and depth map. This ground truth is estimated by fully sampling the same static scene using 16 384, 128×128 pixel Hadamard patterns and acquiring and averaging data for over 2 h (these are the data in Fig. 2). This provides low noise, high resolution reflectivity, and depth maps for comparing the short acquisition time data against, which will give a relative measure of performance though may of course still have absolute errors with respect to the physical reality. To estimate the difference between our test methods and this “ground truth,” we use a normalized root mean square error, which is defined as

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{N} \sum_{n=1}^N (R_n - GT_n)^2}}{\sqrt{\frac{1}{N} \sum_{n=1}^N (GT_n)^2}}, \quad (1)$$

where R_n is the n th pixel of the test reconstruction, GT_n is the n th pixel of the ground truth data, and N is the number of pixels (one depth tagged pixel for each lateral position in the image). We then define our depth and reflectivity “accuracy” as $1 - \text{NRMSE}$. For images with a lower resolution, we first upscale the test image to 128×128 by pixel repetition. We produce two separate accuracy figures for the reflectivity and depth maps. To improve the relevance of our depth accuracy estimates, we do not include depth accuracy for pixels with apparent reflectivity less than 5% of the maximum, as these pixels are likely noise and will contain no meaningful depth information. For reflectivity comparison, we first normalize the data, as the Hadamard and DL reconstructions return images with large differences in absolute reflectivity. We normalize the reflectivity of both our test image and ground truth by subtracting the mean value and dividing by the range. The

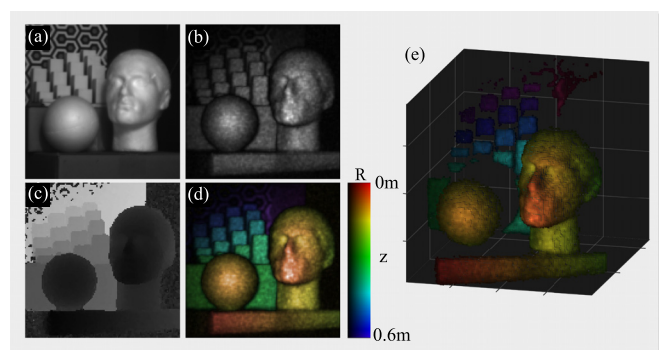


FIG. 2. Composite image breakdown. (a) shows a photograph of the target. (b) shows the intensity measured at the maximum signal for each pixel. (c) shows the depth measurement for each pixel. (d) shows the combined data in a single image. (e) shows the point cloud of the combined data showing the depth of the object.

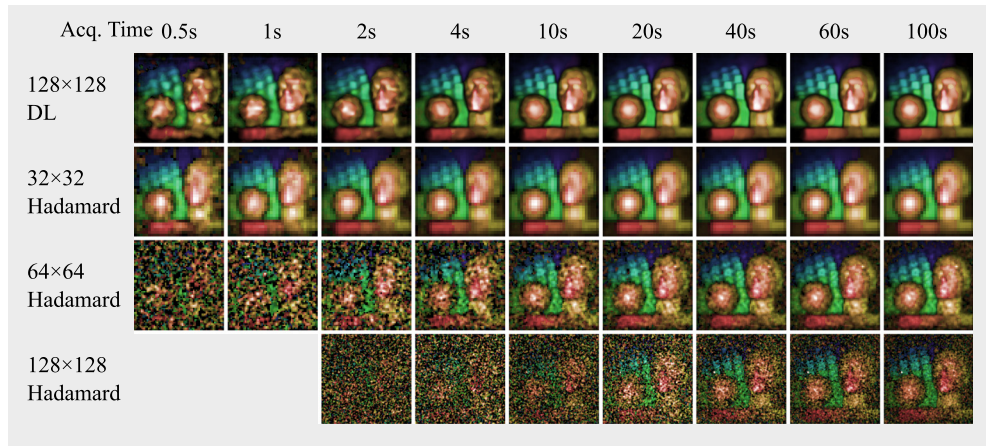


FIG. 3. Composite image reconstructions. Columns are reconstructions for varying acquisition times. Rows show different reconstruction methods. 128×128 Hadamards lack 0.5 s and 1 s acquisitions as the modulator speed is not sufficient to fully sample within those time limits. The color scale is the same as in Fig. 2.

depth accuracy is calculated with no normalization as the depth should be absolute.

We compare the DL reconstruction method against three reconstructions using a Hadamard imaging basis,⁹ with reconstruction resolutions of 32×32 , 64×64 , and 128×128 . The results while imaging a scene at a distance of 2 m are shown in Fig. 3 for acquisition times between 0.5 and 100 s. The Hadamard basis methods follow the predicted behavior that for the same acquisition time (photon number), higher resolutions suffer greatly increased noise. Our DL method, however, performs well even at short acquisition times, offering similar noise levels to the 32×32 Hadamard reconstructions but with a detail level much closer to the 64×64 reconstructions.

We quantify the quality of reconstructions using the accuracy metric defined previously, and the results for the data in Fig. 3 are shown in Fig. 4. The data were acquired as 200 individual 0.5 s acquisitions, and the longer acquisitions are formed by summing. The accuracy is given as the mean of the values from the whole 200 acquisition dataset, and the errors are given as the standard deviation. The reflectivity accuracy results show the expected trend of increasing accuracy with the acquisition time and reveal the DL reconstruction to

outperform even the 32×32 Hadamard basis for acquisition times above 0.5 s. The smaller error can be attributed to the finer details present in the (in principle) 128×128 DL reconstructions, which are smoothed over in the lower resolution 32×32 Hadamard reconstruction. The detail level of the DL reconstruction is qualitatively similar to the 64×64 reconstructions, and comparing the DL results with 64×64 Hadamards shows a 50% improvement for short acquisition times and 33% on average across all acquisition times. The depth accuracy data shown in Fig. 4(b) again show an advantage for the DL method. Absolute depth accuracies can be acquired directly from the error calculations, which for an acquisition time of 1 s are 40 mm, 49 mm, and 167 mm for DL, 32×32 Hadamards and 64×64 Hadamards methods, respectively, with an average DL advantage of 11 mm (75 mm) over 32×32 (64×64) across all acquisition times.

In addition to the lab image acquisitions presented, we have tested our DL LiDAR system at longer range in a corridor with ambient lighting. Measurements at 5 m and 28 m are shown in Fig. 5; for these measurements, the data have been taken with 32×32 Hadamard basis and the DL basis. Both long and short range data show the advantage of the DL imaging method, which again shows similar SNR levels to the 32×32 Hadamard reconstruction, but with increased image sharpness.

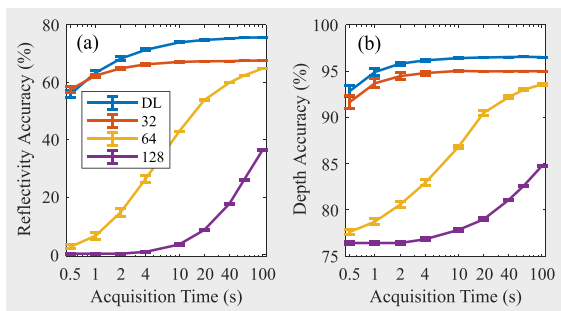


FIG. 4. Reconstruction results. (a) Reflectivity accuracy and (b) depth accuracy of the 4 reconstruction methods for varying acquisition times. DL: deep learned reconstruction method. 32, 64, and 128 are Hadamard reconstructions of 32×32 , 64×64 , and 128×128 resolution.

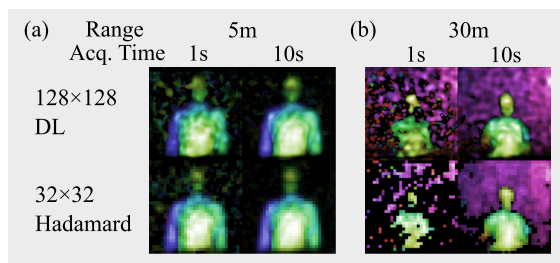


FIG. 5. Short and long range measurements. (a) Wide field of view measurement of a mannequin at 5 m. (b) Narrow field of view measurement of the mannequin against a flat background at a range of 30 m.

24 June 2024 12:31:23

We have demonstrated a single-pixel LiDAR system using a deep learning optimized imaging basis and reconstruction algorithm, which was derived from a trained convolutional neural network. This application of deep learning takes the decoding scheme developed for 2D images and demonstrates this methodology to a 3D LiDAR system. We show that this DL LiDAR system outperforms traditional Hadamard basis reconstruction methods in both reflectivity and depth accuracy. Comparing the reflectivity accuracy of our DL method with a similarly detailed 64×64 Hadamard reconstruction shows a 57% improvement. Depth accuracy is also improved for the DL method, providing an advantage for all acquisition times with a 16% improvement over 64×64 at an acquisition time of 1 s. We discussed the limitations of photon counting LiDAR with the photon acquisition rate, with a reduced number of patterns, and using a DL reconstruction, more precise images are produced when compared to conventional pattern sets. We hope that this can improve the performance of SPI LiDAR techniques and provide an alternative to 3D imaging methods based on scanning, thereby miniaturizing systems for applications such as autonomous vehicles. This can also pave the way for the development of compressed sensing technologies, which are tuned to meet the needs of down-stream decision making systems. Full 3D information may not be necessary to make effective decisions, and compressed sensing can greatly reduce the amounts of acquired and analyzed data in complex systems.

We wish to acknowledge the financial support from the Engineering and Physical Sciences Research Council (EPSRC) QuantIC (No. EP/M01326X/1) and the H2020 European Research Council (ERC) (TWISTS, No. 340507) (PhotUntangle, No. 804626).

REFERENCES

- ¹A. McCarthy, X. Ren, A. D. Frera, N. R. Gemmell, N. J. Krichel, C. Scarcella, A. Ruggeri, A. Tosi, and G. S. Buller, "Kilometer-range depth imaging at 1550 nm wavelength using an InGaAs/InP single-photon avalanche diode detector," *Opt. Express* **21**, 22098–22113 (2013).
- ²J. S. Massa, G. S. Buller, A. C. Walker, S. Cova, M. Umasuthan, and A. M. Wallace, "Time-of-flight optical ranging system based on time-correlated single-photon counting," *Appl. Opt.* **37**, 7298–7304 (1998).
- ³D. Pierrottet, F. Amzajerdian, L. Petway, B. Barnes, G. Lockard, and M. Rubio, "Linear FMCW laser radar for precision range and vector velocity measurements," *MRS Proc.* **1076**, 1076-K04-06 (2008).
- ⁴B. Behroozpour, P. A. M. Sandborn, M. C. Wu, and B. E. Boser, "Lidar system architectures and circuits," *IEEE Commun. Mag.* **55**, 135–142 (2017).
- ⁵X. Ren, P. W. R. Connolly, A. Halimi, Y. Altmann, S. McLaughlin, I. Gyongy, R. K. Henderson, and G. S. Buller, "High-resolution depth profiling using a range-gated CMOS SPAD quanta image sensor," *Opt. Express* **26**, 5541–5557 (2018).
- ⁶S. Xiang, S. Chen, X. Wu, D. Xiao, and X. Zheng, "Study on fast linear scanning for a new laser scanner," *Opt. Laser Technol.* **42**, 42–46 (2010).
- ⁷M. P. Edgar, G. M. Gibson, and M. J. Padgett, "Principles and prospects for single-pixel imaging," *Nat. Photonics* **13**, 13–20 (2019).
- ⁸M. J. Padgett and R. W. Boyd, "An introduction to ghost imaging: Quantum and classical," *Philos. Trans. R. Soc. A* **375**, 20160233 (2017).
- ⁹W. K. Pratt, J. Kane, and H. C. Andrews, "Hadamard transform image coding," *Proc. IEEE* **57**, 58–68 (1969).
- ¹⁰M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.* **25**, 83–91 (2008).
- ¹¹N. Radwell, K. J. Mitchell, G. M. Gibson, M. P. Edgar, R. Bowman, and M. J. Padgett, "Single-pixel infrared and visible microscope," *Optica* **1**, 285–289 (2014).
- ¹²M.-J. Sun, L.-T. Meng, M. P. Edgar, M. J. Padgett, and N. Radwell, "A Russian Dolls ordering of the Hadamard basis for compressive single-pixel imaging," *Sci. Rep.* **7**, 3464 (2017).
- ¹³W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, "A single-pixel terahertz imaging system based on compressed sensing," *Appl. Phys. Lett.* **93**, 121105 (2008).
- ¹⁴W.-K. Yu, "Super sub-Nyquist single-pixel imaging by means of cake-cutting Hadamard basis sort," *Sensors* **19**, 4122 (2019).
- ¹⁵C. F. Higham, R. Murray-Smith, M. J. Padgett, and M. P. Edgar, "Deep learning for real-time single-pixel video," *Sci. Rep.* **8**, 2369 (2018).
- ¹⁶M.-J. Sun, M. P. Edgar, G. M. Gibson, B. Sun, N. Radwell, R. Lamb, and M. J. Padgett, "Single-pixel three-dimensional imaging with time-based depth resolution," *Nat. Commun.* **7**, 12010 (2016).
- ¹⁷M. P. Edgar, S. D. Johnson, D. B. Phillips, and M. J. Padgett, "Real-time computational photon-counting LiDAR," *Opt. Eng.* **57**, 031304 (2017).
- ¹⁸G. A. Howland, P. B. Dixon, and J. C. Howell, "Photon-counting compressive sensing laser radar for 3D imaging," *Appl. Opt.* **50**, 5917–5920 (2011).
- ¹⁹G. A. Howland, D. J. Lum, M. R. Ware, and J. C. Howell, "Photon counting compressive depth mapping," *Opt. Express* **21**, 23822–23837 (2013).
- ²⁰Z. Zhang, X. Wang, G. Zheng, and J. Zhong, "Hadamard single-pixel imaging versus Fourier single-pixel imaging," *Opt. Express* **25**, 19619–19639 (2017).
- ²¹A. Coates, H. Lee, and A. Y. Ng, "An analysis of single layer networks in unsupervised feature learning," in Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (2011), Vol. 15.
- ²²S. D. Johnson, D. B. Phillips, Z. Ma, S. Ramachandran, and M. J. Padgett, "A light-in-flight single-pixel camera for use in the visible and short-wave infrared," *Opt. Express* **27**, 9829–9837 (2019).