Spatial images from temporal data

ALEX TURPIN,[1,*] GABRIELLA MUSARRA,[2] VALENTIN KAPITANY,[2] FRANCESCO TONOLINI,[1] ASHLEY LYONS,[2] ⒾILYA STARSHYNOV,[2] Ⓘ FEDERICA VILLA,[3] ENRICO CONCA,[3] Ⓘ FRANCESCO FIORANELLI,[4] RODERICK MURRAY-SMITH,[1] AND DANIELE FACCIO[2,5] Ⓘ

[1]School of Computing Science, University of Glasgow, Glasgow G12 8QQ, UK
[2]School of Physics & Astronomy, University of Glasgow, Glasgow G12 8QQ, UK
[3]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milano, Italy
[4]Department of Microelectronics, TU Delft, 2628CD, The Netherlands
[5]e-mail: daniele.faccio.@glasgow.ac.uk
*Corresponding author: alex.turpin@glasgow.ac.uk

Traditional paradigms for imaging rely on the use of a spatial structure, either in the detector (pixels arrays) or in the illumination (patterned light). Removal of the spatial structure in the detector or illumination, i.e., imaging with just a single-point sensor, would require solving a very strongly ill-posed inverse retrieval problem that to date has not been solved. Here, we demonstrate a data-driven approach in which full 3D information is obtained with just a single-point, single-photon avalanche diode that records the arrival time of photons reflected from a scene that is illuminated with short pulses of light. Imaging with single-point time-of-flight (temporal) data opens new routes in terms of speed, size, and functionality. As an example, we show how the training based on an optical time-of-flight camera enables a compact radio-frequency impulse radio detection and ranging transceiver to provide 3D images.

1. INTRODUCTION

The most common approach to image formation is obvious and intuitive: a light source illuminates the scene, and the back-reflected light is imaged with lenses onto a detector array (a camera). A second paradigm, single-pixel imaging, relies instead on the use of a single pixel for light detection, while the structure is placed in the illumination by spatially scanning the scene in some form [1–6]. Three-dimensional (3D) imaging can be obtained with these approaches, gathering depth information via stereovision, holographic, or time-of-flight (ToF) techniques [7–10]. In ToF approaches, the depth information is estimated by measuring the time needed by light to travel from the scene to the sensor [11]. Many recent imaging approaches, ranging from sparse-photon imaging [12–14] to non-line-of-sight (NLOS) imaging [15–20], also rely on computational techniques for enhancing imaging capabilities. Among the various possible computational imaging algorithms [21], machine learning (ML) [22] and, in particular, deep learning [23], provides a statistical or data-driven model for enhanced image retrieval [24]. State-of-the-art deep learning techniques have been applied in computational imaging problems such as 3D microscopy [25,26], super-resolution imaging [27,28], phase recovery [29,30], lensless imaging [31,32], and imaging through complex media [33–38].

All of these imaging approaches use either a detector array or scanning/structured illumination to retrieve the transverse spatial

information from the scene. This requirement is clear if we consider the inverse problem that needs to be solved if data is collected only from a single-point, non-scanning detector with no structure in the illumination: there are infinite possible scenes that could give the same measurement on the single-point sensor, thus rendering the inverse problem very strongly ill-posed.

Here, we introduce a new paradigm for spatial imaging based on single-point, time-resolving detectors. In our approach, the scene is flood-illuminated with a pulsed laser, and the return light is focused and collected with a single-point single-photon avalanche diode (SPAD) detector, which records only the arrival time of the return photons from the whole scene in the form of a temporal histogram. During the measurement, no spatial structure is imprinted at any stage, either on the detector or the illumination source. Then, an artificial neural network (ANN) reconstructs the 3D scene from a single temporal histogram. We demonstrate 3D imaging of different objects, including humans, with a resolution sufficient to capture scene details and up to a depth of 4 m. We prove that using the background of the scenes is a key element to detect, identify, and image moving objects, and we exploit it for our application. Our approach is a conceptual change with respect to the common mechanisms for image formation, as spatial images are obtained from a single temporal histogram. This result lends itself to cross-modality imaging, whereby training based on ground-truth from an optical system can be applied to data from

a completely different platform. As an example, we show that a single radio-frequency (RF) impulse radio detection and ranging (RADAR) transducer together with our ML algorithm is enough to retrieve 3D images.

2. SINGLE-POINT 3D IMAGING APPROACH

Previous work has shown that, in addition to the object–sensor distance, the 3D profile of objects manifests through a particular temporal footprint that makes them classifiable even in cluttered environments [39]. Here, we extend this concept to full imaging using only photon arrival time from the scene. It is simple to construct a forward model, where all points in the scene that are at some distance, $\mathbf{r}_i = (x_i, y_i, z_i)$, from the detector provide a related photon arrival time, $t_i = c^{-1}|\mathbf{r}_i| = c^{-1}\sqrt{x_i^2 + y_i^2 + z_i^2}$ (where c is the speed of light). By recording the number of photons arriving at different times t, we can build up a temporal histogram that contains information about the scene in 3D.

However, solving the inverse problem is a much harder task. Indeed, obtaining the 3D coordinates $\mathbf{r}_i$ of objects, the reflected photons of which contribute to a one-dimensional (1D) temporal histogram (i.e., containing no spatial information in any form), is an extremely ill-posed problem. This data problem becomes even harder to solve when one realizes that photons reflected from objects at coordinates placed within a spherical dome represented by the equation $(ct_i)^2 = x^2 + y^2 + z^2$ have the exact same arrival time t_i at the detector, and, as a consequence, they will contribute with equal probability to the same time bin on the histogram. Therefore, a single temporal histogram is not enough, in principle, to obtain a unique solution for the inverse problem and to retrieve meaningful shape or depth information that resembles the actual scene. This problem arises due to a lack of additional information, priors, or constraints on the scene (that is usually provided by using multiple light sources or detectors/pixels). This lack of information and priors can be accounted for and brought into an image retrieval process in different ways. For instance, by following the methodology of common computational imaging algorithms, it is possible to have a forward model that generates different scenes compatible with the experimentally recorded temporal histogram and then use an iterative algorithm that estimates the degree of compatibility of these scenes with the data. However, if no prior information of the types of scenes is provided (e.g., imaging one or more humans continuously moving in an empty room), the number of solutions compatible with this approach is infinite, and the algorithm would hardly converge towards the correct answer. In our work, we take a different approach, where this additional information is provided through priors based on data-sets containing the type of images that we aim to retrieve, and a supervised ML algorithm that is trained for that purpose.

In more detail, the 3D imaging approach is depicted in Fig. 1 and consists of three main elements: (i) a pulsed light source, (ii) a single-point time-resolving sensor, and (iii) an image retrieval algorithm. The scene is flood-illuminated with the pulsed source, and the resulting back-scattered photons are collected by the sensor. We use a single-point SPAD detector operated together with time-correlated single-photon counting (TCSPC) electronics to form a temporal histogram [Fig. 1(b)] from the photons arrival time to objects placed at different positions within the scene, and objects with different shapes provide different distributions of arrival times at the sensor [39].

The histogram, h, measured by the single-point sensor can be mathematically described as $h = \mathcal{F}(S)$, where $S = S(\mathbf{r})$ represents the distribution of objects within the scene. The problem to solve is the search for the function $\mathcal{F}^{-1}$ that maps the temporal histograms onto the scene. We adopt a supervised training approach (see Supplement 1 for details) by collecting a series of temporal histograms corresponding to different scenes together with the corresponding ground-truth 3D images collected with a commercial ToF camera. The ANN is then trained to find an approximate solution for $\mathcal{F}^{-1}$ and is finally used to reconstruct 3D images only from time-resolved measurements of new scenes that have not been seen during the training process. We recall that this is one of the key reasons for using a ML approach: once the algorithm has been trained (which happens only once), this can be used with unseen temporal histograms straight away, i.e., no further training is required. Moreover, the trained algorithm could be implemented on portable platforms for fast and lightweight applications, as it is extremely light computationally speaking.
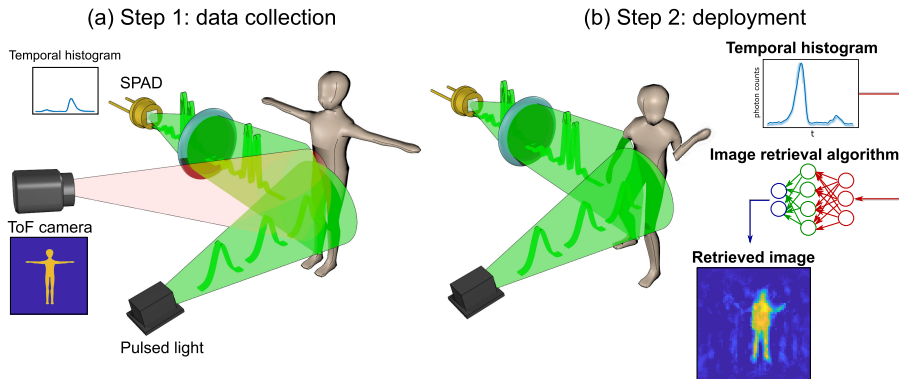

Fig. 1. 3D imaging with single-point time-resolving sensors. Our approach is divided into two steps: (a) a data collection step and (b) the deployment phase. During step 1, a pulsed laser beam flash-illuminates the scene, and the reflected light is collected with a single-point sensor (in our case, SPAD) that provides a temporal histogram via time-correlated single-photon counting (TCSPC). In parallel, a time-of-flight (ToF) camera records 3D images from the scene. The ToF camera operates independently from the SPAD and pulsed laser system. The SPAD temporal histograms and ToF 3D images are used to train the image retrieval ANN. Step 2 occurs only after the ANN is trained. During this deployment phase, only the pulsed laser source and SPAD are used: 3D images are retrieved from the temporal histograms alone.

3. NUMERICAL RESULTS

To evaluate the validity of our approach, we first analyzed its performance with numerical simulations. We consider human-like objects with different poses, moving within a scene of 20 m³, which is represented as a color-encoded depth image, as shown in Fig. 2(c). We assume flash illumination of the scene with a pulsed light source (with a duration that is much shorter than all other timescales in the problem) and then calculate the photons' arrival-time from every point of the scene. Simulating different scenes allows us to obtain multiple 3D images and temporal histograms pairs that are used to train the image retrieval algorithm (details about the structure of the ANN and training parameters can be found in Supplement 1).

Typical scenes consist of a static background with moving human figures in different poses, as shown in Supplement 1. After training the ANN, single temporal histograms are tested to reconstruct the related 3D scenes. To evaluate the potential performance in idealized conditions, for these simulations, we assumed that the time bin width $\Delta t = 2.3$ ps is also the actual temporal resolution (impulse response function, IRF) of the full system. The minimum resolvable transverse feature size or lateral object separation δ that can be distinguished with our technique depends on both the IRF, Δt, and the distance from the sensor, d:

$$\delta(d, \Delta t) = c\,\Delta t \sqrt{\frac{2d}{c\,\Delta t} + 1}, \qquad (1)$$

where c is the speed of light. In the depth direction, the spatial resolving power is determined only by the ToF resolution (as in standard LiDAR), i.e., $\delta_z = c\,\Delta t$. At a distance of 4 m from the detector, for $\Delta t = 2.3$ ps, we can expect a transverse image resolution of 7 cm, which will degrade to 77 cm for $\Delta t = 250$ ps. The impact of the latter realistic time response will be shown in the following experimental results. Figure 2(a) shows one example of a temporal histogram constructed from the scene in Fig. 2(c).
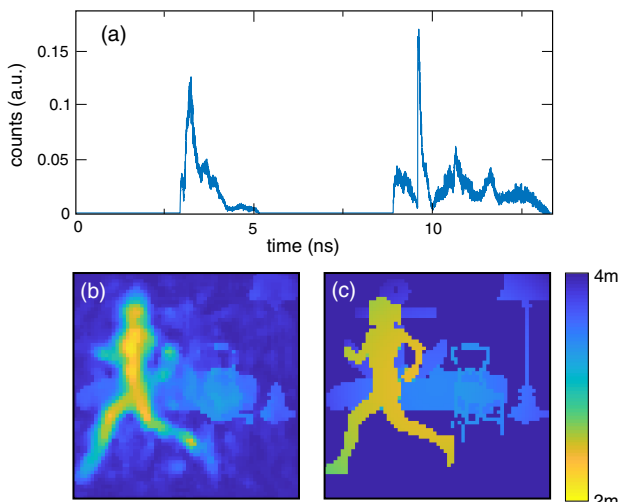

Fig. 2. Numerical results showing 3D imaging from a single temporal histogram recorded with a single-point time-resolving detector. (a) Temporal trace obtained from the scene [shown in (c) as a color-encoded depth image]. (b) 3D image obtained from our image retrieval algorithm when fed with the histogram from (a). The color bars describe the color-encoded depth map.

Figure 2(b) shows the scene reconstructed using the numerically trained ANN and highlights the relatively precise rendition of both depth and transverse details in the scene.

4. EXPERIMENTAL RESULTS

A. Optical Pulses

After numerically demonstrating the concept of 3D imaging with single-point time-resolving detectors, we test its applicability in an experimental environment. We flood-illuminate the scene with a pulsed laser source at (550 ± 25) nm, with pulse width of $\tau = 75$ ps. Our scenes are formed by a variety of fixed background objects, up to a depth of 4 m (the maximum distance allowed by our ToF camera), while different additional objects (people, large-scale objects) are moving dynamically around the scene. Although in our experiments we use a ToF camera, we note that any other 3D imaging system, such as LiDAR, stereoimaging, or holography devices, could be used for collecting the ground-truth data for the training process. The ToF camera is synchronized with a SPAD detector equipped with TCSPC electronics that provide temporal histograms from the back-reflected light that is collected from the whole scene with an angular aperture of ~30° and an IRF $\Delta t = 250$ ps (measured with a small 2 cm mirror in the scene).

In Fig. 3, the first column shows examples of recorded temporal histograms, the second column shows the images reconstructed from these temporal histograms, and the last column shows the ground-truth images obtained with the ToF camera for comparison. Full movies with continuous movement of the people and objects within the scenes, acquired at 10 fps, are shown in Visualization 1. As can be seen, even with the relatively long IRF of our system, it is possible to retrieve the full 3D distribution of the moving people and objects within the scene from the single-point temporal histograms. Compared to the numerical results, the larger IRF leads to the loss of some details in the shapes, such as arms or legs that are not fully recovered.

We can see these limitations, for example, in the reconstruction of the letter 'T,' row (d) of Fig. 3 (with dimensions 39 cm × 51 cm) and especially in Visualization 1, where the algorithm is able to detect the object, but struggles to obtain the correct shape. This lateral resolution power [Eq. (1)] would be improved for example to 25 cm with $\Delta t = 25$ ps and increases with a square-root law with distance, implying a relatively slow deterioration versus distance (for example, resolution would be 50 cm at 20 m distance).

Specific shape information is retrieved from all features in the scene, both dynamic (e.g., moving people) and static (e.g., objects in the background). This can be seen in Figs. 3(a) and 3(b), where both temporal histograms have peaks that are placed at similar positions, yet the reconstructed scenes are different. On the one hand, in Fig. 3(a), the ANN recognizes the box at the right of the image (corresponding to peak 2 in the histogram) as a static background object that was present in all of the training data, while the person (peak 1) is identified as a dynamic object with a certain shape given by the peak structure. In contrast, the ANN recognizes both temporal peaks 1 and 2 in Fig. 3(b) as people moving dynamically through the scene (as these were not constant/static in the training data).

This example highlights the role of the background, which is also key in removing ambiguities that would arise in the presence
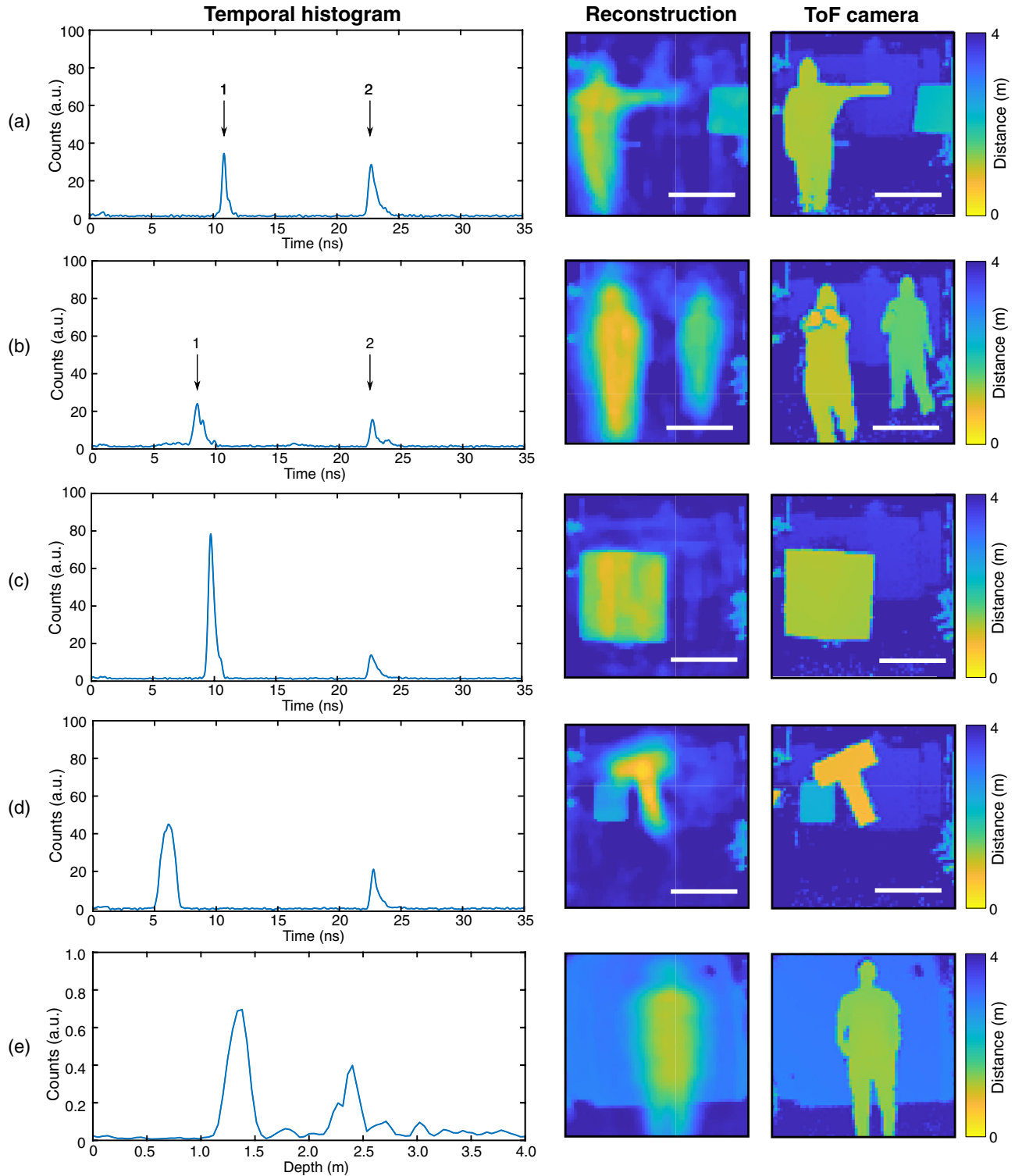
Fig. 3. Experimental results showing the performance of our system recovering the 3D image from temporal histograms in different scenarios. The first column shows temporal histograms recorded with the SPAD sensor and TCSPC electronics [rows (a)–(d)] or with the RADAR transceiver [row (e)], while the last column represents 3D images measured directly with the ToF camera for comparison to the reconstructed images (second column). The color bars describe the color-encoded depth map. The white scale bar corresponds to 80 cm at 2 m distance. Full videos are available in the supplementary information (Visualization 1 and Visualization 2).

of a single isolated object moving in front of a uniform background (for instance, a single isolated histogram peak could be interpreted as a person placed either to the left or to the right of the scene—see Supplement 1 for details).

B. Gigahertz Pulses

We further analyze the impact of the IRF on the image quality reconstruction by repeating the reconstruction with temporal histograms that are convolved with Gaussian point-spread functions

with different temporal widths. The results (Supplement 1) show that although longer IRFs degrade image reconstruction, the shape and 3D location of people in the scene are still easily recognizable even when using an IRF of 1 ns (corresponding nominally to a lateral spatial resolution of ≈ 1.4 m at 4 m distance). This opens the possibility for cross-modality imaging, understood here in the context of detecting signals in one modality or domain and extracting information from a completely different modality. In particular, our approach offers new perspectives for sensing with pulsed sources outside the optical domain, which typically have nanosecond (ns)-scale IRF, providing 3D images with resolutions typical of those obtained in the optical domain, e.g., with the ToF camera. To demonstrate this, we replaced the pulsed laser source, SPAD detector, and other optical elements with an impulse RADAR transceiver emitting at 7.29 GHz (see Supplement 1 and Visualization 2). After re-training the ANN using the RADAR transceiver for the time-series data and the optical ToF camera for the ground-truth images, we can retrieve a person's shape and location in 3D from a single RADAR histogram [see Fig. 3(e)], thus transforming a ubiquitous RADAR distance sensor into a full imaging device.

5. DISCUSSION

Although the maximum resolution of the reconstructed images is limited by the resolution of the 3D sensor used during training, overall, the quality of the final image is essentially determined by the temporal resolution of the single-point time-resolving detector. With state-of-the-art sensors currently heading towards 10 ps or better resolution, there is potential for 3D imaging with spatial resolution better than 10 cm at distances of 10 m or more. The precision in the image reconstruction is also determined by the reconstruction algorithm with improvement possible by using more advanced algorithms (including non-ML-based ones) and also fusing the ToF flight data with other sensor data, e.g., a standard CCD/CMOS camera. However, using single-point SPADs has promising potential for high-speed implementations. After the algorithm is trained (which is performed only once), the image reconstruction problem has two different time-frames: (i) the algorithm reconstruction time and (ii) the histogram data collection time. On the one hand, (i) is easy to be measured with the computer directly, providing times on the order of $10-30$ µs for the algorithm used here. On the other hand, to account for (ii), different factors need to be considered. First, typical time-to-digital converters (TDCs) run at 10 MHz. Our experiments indicate that about 1000 photons per temporal histogram are needed to retrieve a meaningful image, which leads to histogram recording frame rates of ≈ 10 kHz, which is reduced to 1 kHz or less if we are in the photon starved regime, and we account for data transfer to an electronics board. This frame rate could be increased further if instead of a single pixel, a SPAD array is used as a "super-pixel" by adding all the outputs into a single histogram, thus collecting hundreds of photons for each illumination pulse (e.g., with a 32×32 array). Such devices are commercially available and can run at 100 kHz, which would therefore define the rate at which we could collect single temporal histograms These estimates indicate a clear potential for imaging at 1–100 kHz with no scanning parts and a retrieval process that can match this rate even when running with standard software and hardware.

Although the above-discussed advantages of our approach for imaging in terms of data processing and hardware are important, the key message in this work is the potential of using temporal data gathered with just a single pixel for spatial imaging. This approach broadens the remit of what is traditionally considered to constitute image information. The same concept is therefore transferable to any device that is capable of probing a scene with short pulses and precisely measuring the return "echo," for example, RADAR and acoustic distance sensors, indicating a different route for achieving, for example, full 360° situational awareness in autonomous vehicles and smart devices or wearable devices.

6. CONCLUSIONS

Current state-of-art imaging techniques use either a detector array or scanning/structured illumination to retrieve the transverse spatial information from the scene, i.e., they relate spatial information directly to some type of spatial sensing of the scene. In this work, we have demonstrated an alternative approach to imaging that suppresses all forms of spatial sensing and relies only on a data-driven image retrieval algorithm processing a single time series of data collected from the imaged scene. The experiments were carried out in scenes where objects were moving in front of a static background. This makes our approach well suited for applications where the device needs to be placed at a fixed position during operation, i.e., with a fixed background. There are multiple situations where operating in a fixed environment is useful. Examples are surveillance, security in public spaces, etc. These are examples where the background (e.g., walls of the room, buildings) do not change at all, and they are also very widespread scenarios. Currently, cities have spaces that are constantly monitored with CCTV cameras that also potentially record information from which it is possible to extract information that breaches data protection policies. Our approach is therefore highly indicated for cases where one requires human activity in a fixed area and in a data-compliant way. The approach shown here would be also valid in a slowly changing environment, where training could, in principle, be continuously updated. Indeed, background objects will appear static if they change at a slower rate (and/or are at a larger distance) with respect to the dynamic elements of the scene or slower than the acquisition rate of the sensor. An interesting route for future research is of course to also investigate methods that account for dynamic backgrounds.

Finally, an interesting extension would be to NLOS imaging, especially given the latest developments exploiting computational techniques for image information retrieval from temporal data [16,40–43] and the availability of public data-sets [44,45].

Disclosures. The authors declare no conflicts of interest.

See Supplement 1 for supporting content.

REFERENCES

1. J. H. Shapiro, "Computational ghost imaging," Phys. Rev. A **78**, 061802 (2008).

2. M. 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**(2), 83–91 (2008).

3. M. P. Edgar, G. M. Gibson, and M. J. Padgett, "Principles and prospects for single-pixel imaging," Nat. Photonics **13**, 13–20 (2019).

4. D. B. Lindell, M. O'Toole, and G. Wetzstein, "Single-photon 3D imaging with deep sensor fusion," ACM Trans. Graph. **37**, 1–12 (2018).

5. N. Radwell, A. Selyem, L. Mertens, M. P. Edgar, and M. J. Padgett, "Hybrid 3D ranging and velocity tracking system combining multi-view cameras and simple LiDAR," Sci. Rep. **9**, 5241 (2019).

6. C. Callenberg, A. Lyons, D. den Brok, R. Henderson, M. B. Hullin, and D. Faccio, "EMCCD-SPAD camera data fusion for high spatial resolution time-of-flight imaging," in *Computational Optical Sensing and Imaging* (Optical Society of America, 2019), paper CTh2A–3.

7. S. T. Barnard and M. A. Fischler, "Computational stereo," ACM Comput. Surv. **14**, 553–572 (1982).

8. Y. Frauel, T. J. Naughton, O. Matoba, E. Tajahuerce, and B. Javidi, "Three-dimensional imaging and processing using computational holographic imaging," Proc. IEEE **94**, 636–653 (2006).

9. B. Sun, M. P. Edgar, R. Bowman, L. E. Vittert, S. Welsh, A. Bowman, and M. J. Padgett, "3D computational imaging with single-pixel detectors," Science **340**, 844–847 (2013).

10. M.-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).

11. P. Dong and Q. Chen, *LiDAR Remote Sensing and Applications* (CRC Press, 2017).

12. A. Kirmani, D. Venkatraman, D. Shin, A. Colaco, F. N. C. Wong, J. H. Shapiro, and V. K. Goyal, "First-photon imaging," Science **343**, 58–61 (2014).

13. P. A. Morris, R. S. Aspden, J. E. C. Bell, R. W. Boyd, and M. J. Padgett, "Imaging with a small number of photons," Nat. Commun. **6**, 5913 (2015).

14. J. Tachella, Y. Altmann, N. Mellado, A. McCarthy, R. Tobin, J.-Y. T. Gerald, S. Buller, and S. McLaughlin, "Real-time 3D reconstruction from single-photon lidar data using plug-and-play point cloud denoisers," Nat. Commun. **10**, 4984 (2019).

15. 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).

16. G. Gariepy, F. Tonolini, R. Henderson, J. Leach, and D. Faccio, "Detection and tracking of moving objects hidden from view," Nat. Photonics **10**, 23–26 (2016).

17. M. O'Toole, D. B. Lindell, and G. Wetzstein, "Confocal non-line-of-sight imaging based on the light-cone transform," Nature **555**, 338–341 (2018).

18. C. Jin, J. Xie, S. Zhang, Z. Zhang, and Y. Zhao, "Reconstruction of multiple non-line-of-sight objects using back projection based on ellipsoid mode decomposition," Opt. Express **26**, 20089–20101 (2018).

19. V. Arellano, D. Gutierrez, and A. Jarabo, "Fast back-projection for non-line of sight reconstruction," Opt. Express **25**, 11574–11583 (2017).

20. G. Musarra, A. Lyons, E. Conca, Y. Altmann, F. Villa, F. Zappa, M. Padgett, and D. Faccio, "Non-line-of-sight 3D imaging with a single-pixel camera," Phys. Rev. Appl. **12**, 011002 (2019).

21. Y. Altmann, S. McLaughlin, M. J. Padgett, V. K. Goyal, A. O. Hero, and D. Faccio, "Quantum-inspired computational imaging," Science **361**, eaat2298 (2018).

22. M. I. Jordan and T. M. Mitchell, "Machine learning: trends, perspectives, and prospects," Science **349**, 255–260 (2015).

23. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).

24. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," Optica **6**, 921–943 (2019).

25. L. Waller and L. Tian, "Computational imaging: machine learning for 3D microscopy," Nature **523**, 416–417 (2015).

26. Y. Rivenson, Z. Göröcs, H. Günaydin, Y. Zhang, H. Wang, and A. Ozcan, "Deep learning microscopy," Optica **4**, 1437–1443 (2017).

27. E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-storm: super-resolution single-molecule microscopy by deep learning," Optica **5**, 458–464 (2018).

28. H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydin, L. A. Bentolila, C. Kural, and A. Ozcan, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," Nat. Methods **16**, 103–110 (2019).

29. Y. Rivenson, Y. Zhang, H. Günaydin, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light. Sci. Appl. **7**, 17141 (2018).

30. A. Goy, K. Arthur, S. Li, and G. Barbastathis, "Low photon count phase retrieval using deep learning," Phys. Rev. Lett. **121**, 243902 (2018).

31. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," Optica **4**, 1117–1125 (2017).

32. Y. Rivenson, H. Ceylan Koydemir, H. Wang, Z. Wei, Z. Ren, H. Gunaydn, Y. Zhang, Z. Gorocs, K. Liang, D. Tseng, and A. Ozcan, "Deep learning enhanced mobile-phone microscopy," ACS Photon. **5**, 2354–2364 (2018).

33. S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, "Imaging through glass diffusers using densely connected convolutional networks," Optica **5**, 803–813 (2018).

34. N. Borhani, E. Kakkava, C. Moser, and D. Psaltis, "Learning to see through multimode fibers," Optica **5**, 960–966 (2018).

35. Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," Optica **5**, 1181–1190 (2018).

36. A. Turpin, I. Vishniakou, and J. D. Seelig, "Light scattering control in transmission and reflection with neural networks," Opt. Express **26**, 30911–30929 (2018).

37. B. Rahmani, D. Loterie, G. Konstantinou, D. Psaltis, and C. Moser, "Multimode optical fiber transmission with a deep learning network," Light. Sci. Appl. **7**, 69 (2018).

38. P. Caramazza, O. Moran, R. Murray-Smith, and D. Faccio, "Transmission of natural scene images through a multimode fibre," Nat. Commun. **10**, 2029 (2019).

39. P. Caramazza, A. Boccolini, D. Buschek, M. Hullin, C. F. Higham, R. Henderson, R. Murray-Smith, and D. Faccio, "Neural network identification of people hidden from view with a single-pixel, single-photon detector," Sci. Rep. **8**, 11945 (2018).

40. D. Faccio, A. Velten, and G. Wetzstein, "Non-line-of-sight imaging," Nat. Rev. Phys. **2**, 318–327 (2020).

41. M. O'Toole, D. B. Lindell, and G. Wetzstein, "Confocal non-line-of-sight imaging based on the light-cone transform," Nature **555**, 338–341 (2018).

42. X. Liu, I. Guillén, M. La Manna, J. H. Nam, S. A. Reza, T. H. Le, A. Jarabo, D. Gutierrez, and A. Velten, "Non-line-of-sight imaging using phasor-field virtual wave optics," Nature **572**, 620–623 (2019).

43. J. Iseringhausen and M. B. Hullin, "Non-line-of-sight reconstruction using efficient transient rendering," ACM Trans. Graph. **39**, 1–14 (2020).

44. M. Galindo, J. Marco, M. O'Toole, G. Wetzstein, D. Gutierrez, and A. Jarabo, "A dataset for benchmarking time-resolved non-line-of-sight imaging," in *ACM SIGGRAPH 2019 Posters* (2019), pp. 1–2.

45. J. Klein, M. Laurenzis, D. L. Michels, and M. B. Hullin, "*NLoS Benchmark*" (2019).