Creating pictures by measuring time

A radical new method of imaging which harnesses artificial intelligence to turn time into visions of 3D space could help cars, mobile devices and health monitors develop 360-degree awareness

Photos and videos are usually produced by capturing photons – the building blocks of light – with digital sensors. For instance, digital cameras consist of millions of pixels that form images by detecting the intensity and colour of the light at every point of space. 3D images can then be generated either by positioning two or more cameras around the subject to photograph it from multiple angles, or by using streams of photons to scan the scene and reconstruct it in three dimensions. Either way, an image is only built if one gathers spatial information of the scene.

In a new paper published in July in the journal Optica, researchers based in the UK, Italy and the Netherlands describe how they have found an entirely new way to make animated 3D images – by capturing temporal information about photons instead of their spatial coordinates.

Stopwatch for photons

Their process begins with a simple, inexpensive single-point detector tuned to act as a kind of stopwatch for photons. Unlike cameras, measuring the spatial distribution of colour and intensity, the detector only records how long it takes the photons produced by split-second flash of a pulse of laser light to bounce off each object in any given scene and reach the sensor. The further away an object is, the longer it will take each reflected photon to reach the sensor.

The information about the timings of each photon reflected in the scene – what the researchers call the temporal data - is collected in a very simple graph.

Those graphs are then turned into a 3D image with the help of a sophisticated neural network algorithm. The researchers 'trained' the algorithm by showing it thousands of different conventional photos of the team moving and carrying objects around the lab, alongside temporal data captured by the single-point detector at the same time.

Eventually, the network had learned enough about how the temporal data corresponded with the photos that it was capable of creating highly accurate images from the temporal data alone. In the proof-of-principle experiments, the team managed to construct moving images at about 10 frames per second from the temporal data, although the hardware and algorithm used has the potential to produce thousands of images per second.
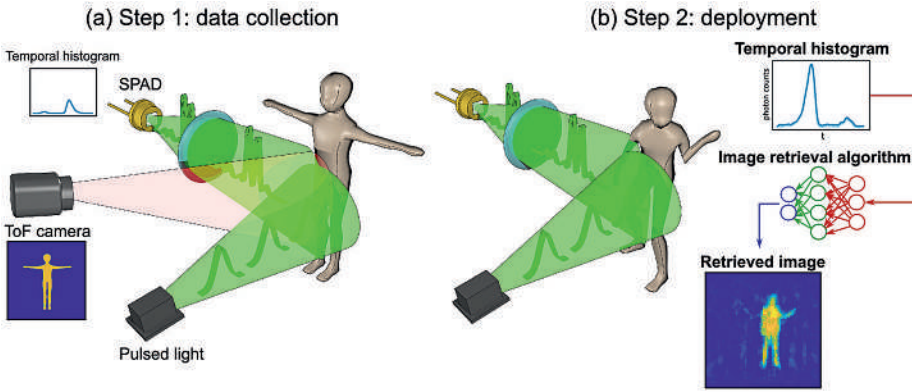
Fig. 1. 3D imaging with single-point time-resolving sensors. The approach is divided into two steps: In step 1, a pulsed laser beam flash-illuminates the scene, and the reflected light is collected with a single-point sensor - in this case, a Single-Photon Avalanche Diode (SPAD) - that provides a temporal 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 Artificial Neural Network (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.
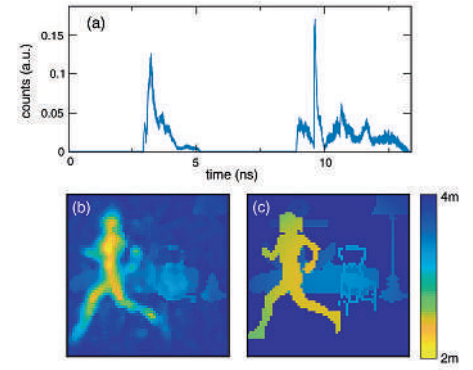

Fig. 2. Numerical results showing 3D imaging from a single temporal histogram recorded with a single-point time-resolving detector. (a) Temporal trace obtained from the scene [shown in (c) as a colour-encoded depth image]. (b) 3D image obtained from image retrieval algorithm when fed with the histogram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

Dr. Alex Turpin, Lord Kelvin Adam Smith Fellow in Data Science at the University of Glasgow's School of Computing Science, led the University's research team together with Prof. Daniele Faccio, with support from colleagues at the Polytechnic University of Milan and Delft University of Technology.

Dr. Turpin said: "Cameras in our cell-phones form an image by using millions of pixels. Creating images with a single pixel alone is impossible if we only consider spatial information, as a single-point detector has none. However, such a detector can still provide valuable information about time. What we've managed to do is find a new way to turn one-dimensional data – a simple measurement of time – into a moving image which represents the three dimensions of space in any given scene.

"The most important way that differs from conventional image-making is that our approach is capable of decoupling light altogether from the process. Although much of the paper discusses how we've used pulsed laser light to collect the temporal data from our scenes, it also demonstrates how we've managed to use radar waves for the same purpose.

Adaptable method

Dr Turpin continues: "We're confident that the method can be adapted to any system which is capable of probing a scene with short pulses and precisely measuring the return 'echo'. This is really just the start of a whole new way of visualising the world using time instead of light."

Currently, the neural net's ability to create images is limited to what it has been trained to pick out from the temporal data of scenes created by the researchers. However, with further training and even by using more advanced algorithms, it could learn to visualise a much varied range of scenes, widening its potential applications in real-

Dr. Alex Turpin

world situations.

Dr. Turpin added: "The single-point detectors which collect the temporal data are small, light and inexpensive, which means they could be easily added to existing systems like the cameras in autonomous vehicles to increase the accuracy and speed of their pathfinding.

"Alternatively, they could augment existing sensors in mobile devices like the Google Pixel 4, which already has a simple gesture-recognition system based on radar technology. Future generations of our technology might even be used to monitor the rise and fall of a patient's chest in hospital to alert staff to changes in their breathing, or to keep track of their movements to ensure their safety in a data-compliant way.

Next step

"We're very excited about the potential of the system we've developed, and we're looking forward to continuing to explore its potential. Our next step is to work on a self-contained, portable system-in-a-box and
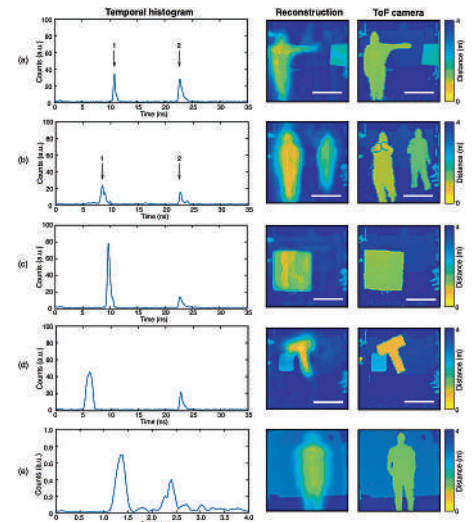

Fig. 3. Experimental results showing the performance of the system recovering the 3D image from temporal histograms in different scenarios. The first column shows temporal histograms recorded with the SPAD sensor and TCSPC electronics [rows (a)–(d)] or with the RADAR transceiver [row (e)], while the last column represents 3D images measured directly with the ToF camera for comparison to the reconstructed images (second column). The colour bars describe the colour-encoded depth map. The white scale bar corresponds to 80 cm at 2m distance.

we're keen to start examining our options for furthering our research with input from commercial partners."

Note: Although the described experiments used a ToF camera, the paper notes 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.