SCIENTIFIC REP♦RTS

OPEN

Neural network identification of people hidden from view with a single-pixel, single-photon detector

Piergiorgio Caramazza[1,2], Alessandro Boccolini[1], Daniel Buschek[3], Matthias Hullin[4], Catherine F. Higham [5], Robert Henderson[6], Roderick Murray-Smith[5] & Daniele Faccio [2]

Light scattered from multiple surfaces can be used to retrieve information of hidden environments. However, full three-dimensional retrieval of an object hidden from view by a wall has only been achieved with scanning systems and requires intensive computational processing of the retrieved data. Here we use a non-scanning, single-photon single-pixel detector in combination with a deep convolutional artificial neural network: this allows us to locate the position and to also simultaneously provide the actual identity of a hidden person, chosen from a database of people ($N = 3$). Artificial neural networks applied to specific computational imaging problems can therefore enable novel imaging capabilities with hugely simplified hardware and processing times.

Recent years have seen a surge of interest and corresponding advances in the ability to image objects that are not visible within the direct line of sight. In particular we refer to the situation in which the object is hidden behind a wall, a corner or inside a room to which we do not have access[1–13]. The majority of the techniques that attempt to image a scene or object that is hidden behind an obstacle have relied on active imaging, i.e. the scene is actively illuminated using a light source that is controlled by the observer. Although recent work used continuous illumination[6], the most common approach is to use a pulsed light source, for example a laser. The basic functioning principle is then very similar to listening to sound echoes reflected from multiple surfaces: the laser beam is scattered off a surface that lies within the direct line of sight, but also such that the scatter may enter the hidden environment. By then synchronising the detection system to the emitted pulses and measuring the return time for each echo, it is possible to determine the distance of the object that created/reflected the signal. If one wants to build a full image of the hidden environment or object, simply measuring return times from a single point is not sufficient: multiple pixel information is required and is built up by either directly imaging and/or scanning the imaging optics across the surface where the reflected echoes are detected, or by scanning the illumination spot on the first scattering surface. Both approaches, followed by computational processing of the collected data can provide full 3D reconstruction of the hidden environment.

However, the overall constraints on the problem make it extremely hard to achieve 3D imaging of hidden objects with high resolution (few mm or less), at significant distances (1 m or more) and within reasonable (e.g. less than several seconds) time-frames. The main limitations are: the very low return signal which will typically decay as $\sim 1/d^6$ (d is the distance of the hidden object from the imaging system)[9]; the very high temporal resolution (10–100 ps or less) required for the detector to obtain sub-cm precision; the requirement to scan either the laser or the detector lengthens the acquisition times to minutes or even hours[3,5]. Currently, there is no obvious and single solution to all of these problems: 3D imaging can be obtained at the expense of acquisition and processing time or tracking can be obtained at higher frame rates, albeit with limited resolution and hence the impossibility to reconstruct actual 3D shapes of the hidden objects. Moreover, very little work has been performed to date on large objects and actual people, with results limited to tracking of location and movement detection[9].

[1]Institute of Photonics and Quantum Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, UK. [2]School of Physics and Astronomy, Kelvin Building, University of Glasgow, Glasgow, G12 8QQ, UK. [3]Media Informatics Group, University of Munich (LMU), Munich, Germany. [4]Institute for Computer Science II, University of Bonn, Friedrich-Ebert-Allee 144 53113, Bonn, Germany. [5]School of Computing Science, University of Glasgow, Glasgow, G12 8QQ, UK. [6]School of Engineering, Institute for Integrated Micro and Nano Systems, University of Edinburgh, Edinburgh, EH9 3JL, UK. Correspondence and requests for materials should be addressed to D.F. (email: Daniele.Faccio@glasgow.ac.uk)
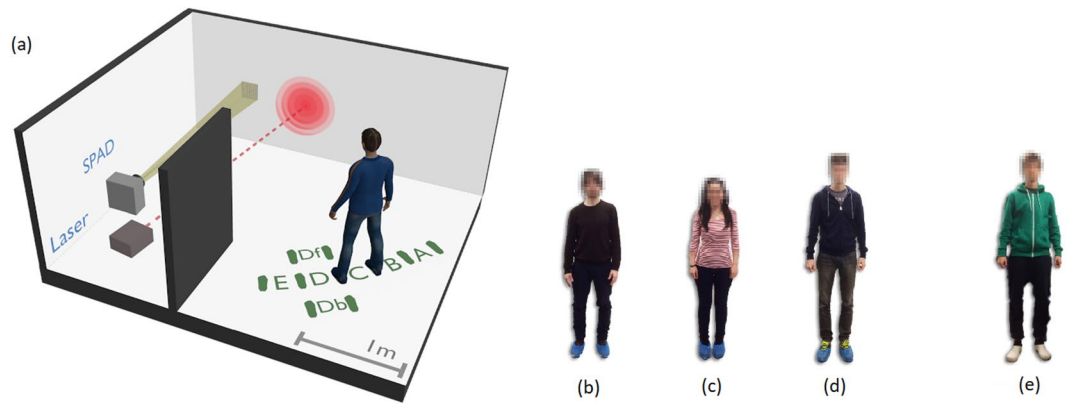
Figure 1. Experimental details: layout is shown in (**a**). A pulsed laser light source illuminates a wall which scatters light behind the obscuring wall, thus illuminating the hidden person. The return echoes are collected by the SPAD array, aimed at the first scattering wall. The three people used for the experiments, referred to in the text and figures as "n.1", "n.2" and "n.3" are shown in (**b**), (**c**), (**d**), respectively. Their relative heights are: 1.68 m, 1.57 m and 1.87 m. Two different cases were verified: different clothing, (**b**–**d**), and same clothing (**e**) (only individual "n.3" is shown for simplicity). Measurements where repeated 5 times across all 7 different positions (A, B, C, Db, Df, E and F).

If we then ask the question, "is it possible to both locate and even identify a person who is hidden behind a wall?", we are faced with even larger hurdles and the answer is that such a capability is well out of the reach of all currently adopted approaches.

The key point of this work is to show that by moving beyond the current paradigms of non-line-of-sight (NLOS) imaging, it is actually possible to perform the feat questioned above on at least a limited set of chosen targets. In order to do this we use data from a single-photon, single-pixel detector that is analysed by a previously trained artificial neural network (ANN). Use of just one single pixel detector is highly advantageous due to commercial availability and optimised specifications in terms of photon sensitivity (can be close to 100%, whilst quantum efficiency remains similar) and temporal resolution (can be lower than 20 ps). The ANN is able to recognise the position of a hidden person (from a training set of 7 positions) and even provide the identity of the person (from a training set of 3 people). Remarkably, person-identification is successful even when all three individuals have the same clothing, thus hinting that the ANN can recognise the more subtle changes in the physiognomy from one person to another. This is all the more remarkable when we note that the temporal resolution of the detector (120 ps, corresponding to 1.8 cm depth resolution) would not be sufficient to precisely reconstruct the full 3D shape of a face, even after raster scanning. The ANN therefore extends the capability of the imaging system and by removing the requirement of any forward modelling or computational post-processing, allows fast location and identification of hidden people that would otherwise not be possible. We underline that currently there are no other techniques that can identify people from behind a corner, even within the strong limitations demonstrated here.

Experiments

Data is collected with the set-up shown in Fig. 1. A femtosecond pulsed laser beam, $\lambda = 808$ nm and Rep. Rate $\lambda = 80$ MHz, is pointed towards a wall where a first scattering occurs. Thus, the light illuminates the hidden individual so that the backscattered light can be captured by the detection system. Three different people have been used for this experiment, and both same clothing and different clothing data were acquired (see Fig. 1). Moreover, we tested seven different positions of which, those from "A" to "E" share similar photon time-of-flight (i.e. the individual has a similar distance from the first scattering point on the wall). This situation was chosen so as to ensure that the ANN did not train solely on arrival time of the return photon echoes but rather focused on actual features within the temporal shape of the echo. For each person in each position, five separate measurements were taken, alternating people and position for each measurement.

In order to provide the large amount of data required for the neural network training, we make use of a single photon avalanche diode (SPAD) segmented array[14]. This consists of a 32×32 array of single-pixel SPAD detectors, each characterized by an instrumental temporal response function (IRF) of ~120 ps (this is the total IRF, i.e. includes both laser pulse duration and all electronic jitter effects). Thus, the pixels are treated as independent observers that are looking at roughly the same position on the wall (within the 3×3 cm^2 imaged area on the wall). We note that the field of view covered by the array was such that there was some variability in the data but that this variability did not significantly distort the histograms from one pixel to the next. Indeed, we noticed in initial tests that with a field of view three times larger that the classification results were significantly worse by a factor 2–3. We also acquire a background signal under identical conditions but with the individual removed: this is then subtracted from the measurements. We note however, that if for example the individual is moving, then it is possible to acquire a background from a time-averaged signal, as demonstrated in[4] therefore providing a practical approach to situations where it is not possible to acquire a pre-recorded background signal. Furthermore, we eliminate "hot" pixels and thus obtain 800 temporal histograms from each measurement. A subset example of
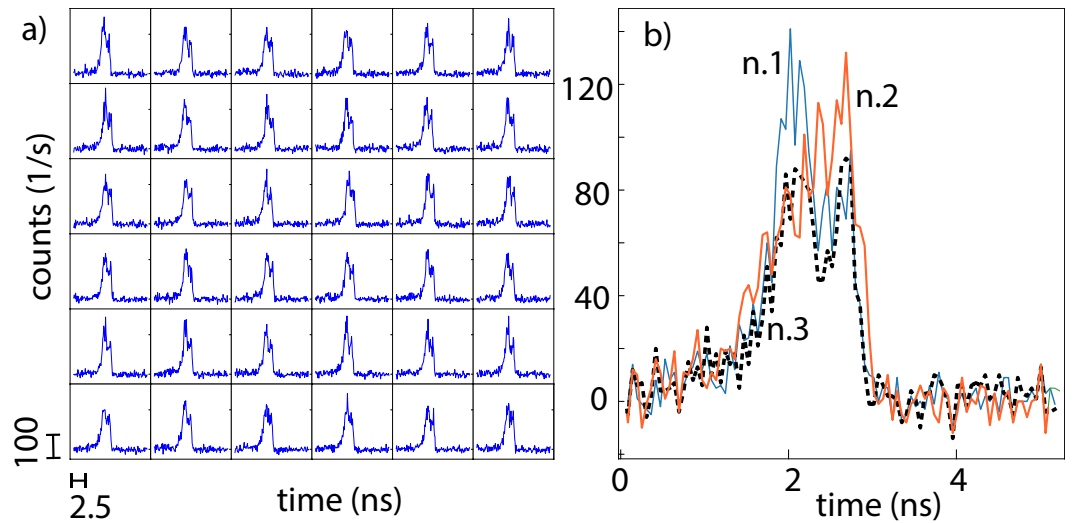
Figure 2. (**a**) Example of the input data for the ANN, showing a subset 6×6 taken from the full 32×32 array. This data is a single measurement of individual "n.1" in position "C". (**b**) For comparison, we show the temporal histograms for individuals n.1, 2 and 3 for the same pixel. The width of the temporal binning is 55 ps for both (**a**) and (**b**).

data from a single, background-subtracted measurement (for individual "n.1" in position "C") is shown in Fig. 2a). The SPAD array is triggered directly from the laser external trigger and is thus synchronised to the emission of each individual laser pulse: a single acquisition takes two seconds, equivalent to integration over 160×10^6 laser pulses.

Analysis of Experimental Data

The architecture we propose is inspired by the physics involved in the experiment. Here, the underlying assumption is that the information about position and shape of the hidden individuals are both encoded in the photon time-of-flight and final temporal shape of the return echo. In Fig. 2b) we show a typical example of the histograms for the three individuals tested in this work (labeled as n.1, n.2 and n.3). As can be seen, there are clear differences between the three temporal histograms yet there is no unique feature that stands out as distinguishing one from the other. They have similar heights, total photon counts (physically connected to the overall target reflectivity) and widths (physically connected to the overall height of the target individual). This means that any data-driven classification approach has to learn to identify the overall ensemble of more subtle differences and classify the data accordingly. As previously introduced, we underline that SPAD single-pixel detectors in our setup are treated independently. Therefore no spatial information is actually used. Indeed, the aim of our algorithm is to identify and locate the individuals from just one time-binned pixel.

Neural Network Classifier

Artificial neural networks (ANN) are mathematical models loosely inspired by the human brain, which have provided an important contribution to both scientific and technological research for their capacity to learn distinctive features from large amounts of data. Deep convolutional neural networks are computational models which are concerned with learning representations of data with multiple levels of abstraction. They have been broadly employed for tasks such as regression, classification, unsupervised learning, and are proving very successful at discovering features in high-dimensional data arising in many areas of science, with breakthroughs in image processing and time-series analysis[15–17]. The performance increases are due to increases in processing power, algorithm improvements better quality flexibility software, and the availability of large collections of training data.

Very recent studies have started to look at the use of ANNs in the area of computational imaging with applications in phase-object identification[18], pose-identification of human-shaped objects[19] and number/letter identification[20] from behind a diffusive screen. In this work, we rely on supervised machine learning algorithms in order to classify people hidden from our line of sight and located in varying positions.

We therefore build a nonlinear classifier aiming to correctly identify the label of the histograms resulting from the acquisition of pulsed laser light backscattered from three different people in seven different positions. We use a supervised approach where we pair the temporal histogram as input to the ANN and create an output vector encoding the class of the person and the target location. Both class and location are treated as categorical classification tasks, and encoded using a 'one-hot' encoding such that we use N_c binary outputs for N_c classes, and N_l binary outputs for location positions. In this work $N_c = 3$, $N_l = 7$. The cost function minimised during learning is the categorical cross-entropy[21] see SM).

As it is shown in Fig. 3, our ANN architecture processes input data in parallel through: a fully-connected layer on one side, in order to retrieve more information about the distance, and in parallel, convolutional layers which due to their translation invariant nature, will focus more on the temporal histogram shape and features (see SM for more details).
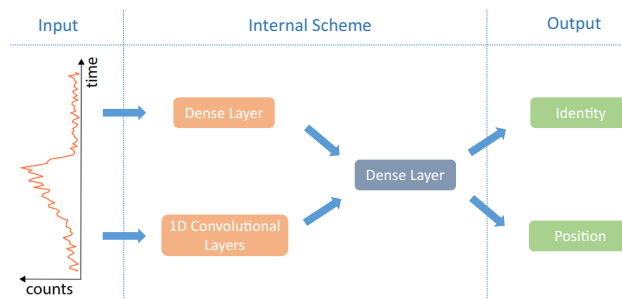
Figure 3. Network architecture tested and discussed in the text (1Dconv + Dense). Other architectures are discussed in the SM.
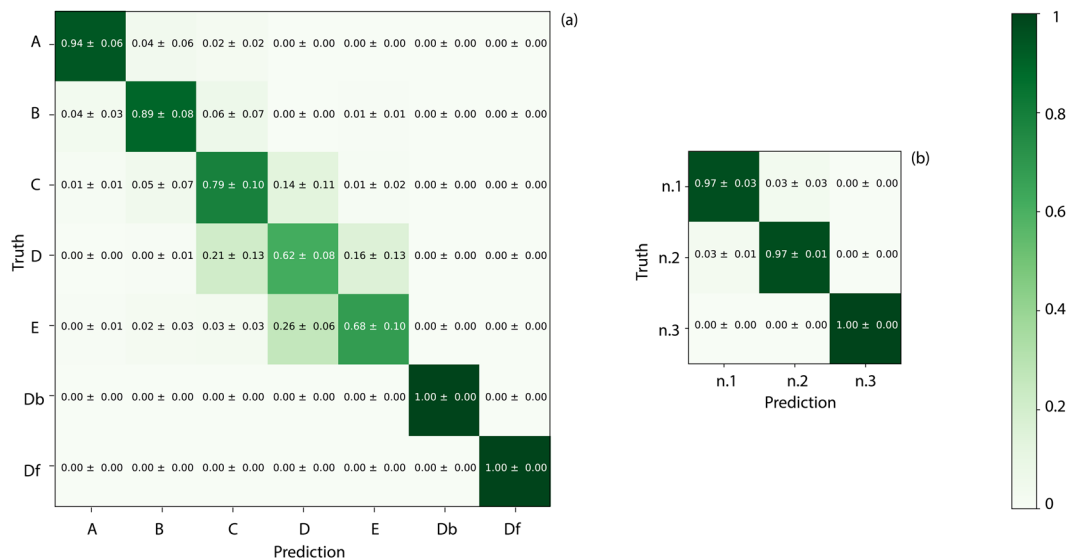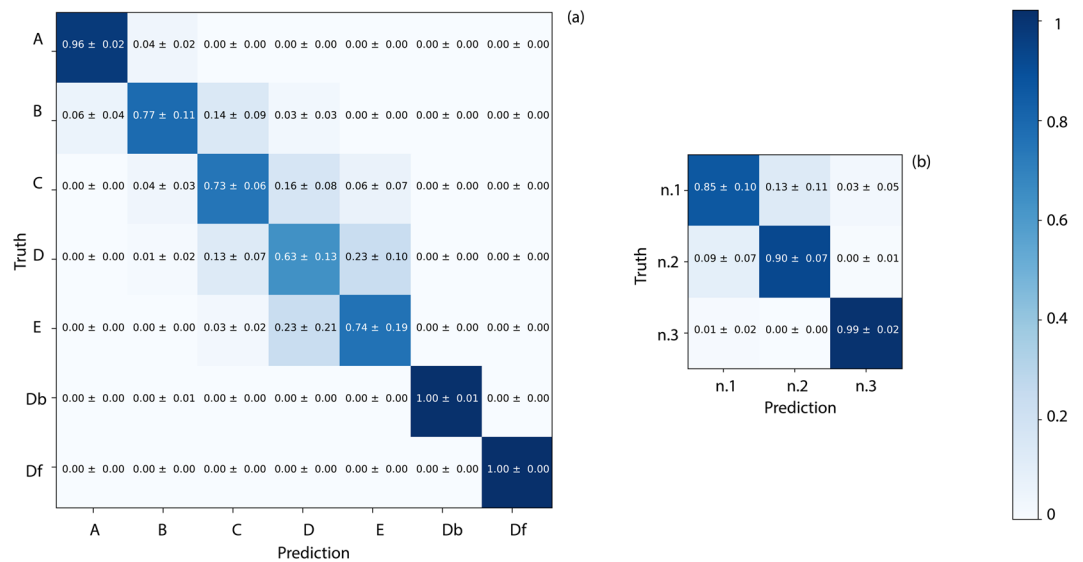

Figure 4. Results for the case in which the three individuals (n.1, n.2, n.3) have *different* clothing. Confusion matrices are shown for: (**a**) retrieval of position (averaged over all individuals) and (**b**) retrieval of the individual identities. The ANN was trained with 4 measurements and tested with one, repeating over every permutation and averaging.

Results

After optimisation was completed, we tested the performance of the classifier on new data, taken under the same conditions as the training data but is not used during the training process. To test the robustness of the optimisation process, we use a leave-one-out cross-validation process, where we extract all data from one measurement to use as test data, then train on data associated with all other measurements. We then average the classification results in a per-pixel basis. For this training set with 5 measurements of ca. 800 pixels each, the classification average calculated over 5 runs obtained by training the ANN on data from 4 measurements and testing on the remaining one (repeating the procedure by all permutations of the 4 training and single test data sets).

Firstly, we show the results for the case in which all three individuals have "different clothing", Fig. 4. Data are reported in a confusion matrix that compares the actual classes (vertical axis, "truth") with the predicted classes (horizontal axis). Since our ANN predicts simultaneously position and identity relative to a testing histogram, the positions' confusion matrix will contain information from all the three characters. Analogously, the people's confusion matrix will contain prediction on characters disposed on all the seven positions. The matrix values represent the normalised average calculated over all 5 cross-validation runs. A "0" indicates no overlap between training and test data and a "1" indicates 100% statistical certainty in the correct classification.

As we can observe in Fig. 4(a), positions "Db" and "Df" are identified with 100% certainty, whereas among the positions with similar photon time-of-flight, the classification fidelity is slightly worse, thus indicating a certain role played by the overall arrival time of the photons. However, the classification still shows a very good agreement for all positions with the ground truth. In Fig. 4(b) we show the confusion matrix for person identification indicating that the classifier is able to correctly identify the three people. Since it is possible in this case that both shape and reflectivity of the bodies (that are clothed differently) play a key role, we try to isolate one of these two degrees of freedom by repeating the experiments with all individuals in the same clothing, see Fig. 5. The classification of the individual's position occurs with similar fidelity as in the "different clothing" case, Fig. 5(a). However,

Figure 5. Results for the case in which the three individuals (n.1, n.2, n.3) have the *same* clothing. Confusion matrices are shown for: (**a**) retrieval of position averaged over results for all three individuals and (**b**) retrieval of the individual's identity.

person identification is indeed now more challenging with a higher confusion between individuals. The notable result from these tests is that even when individuals have the same clothing, a single-pixel is sufficient to identify them. All of the information is encoded in the temporal shape of the return photon echo. Yet, as the temporal resolution of 120 ps corresponds to a spatial depth resolution of 1.8 cm which is insufficient to create a 3D map of a human face, so we suspect that the classifier is not focusing solely on facial features, but is relying on overall physiognomy, including body height, width and skin reflectivity.

Finally, we compared the results from different ANN architectures. The results (see SM) tend to not show any particular sensitivity to the specific ANN architecture employed among those designs compared in this final stage, suggesting the robustness of the model structures chosen for this problem. We anticipate that further improvement in performance would need to come from larger and more controlled training sets. However, one interesting outcome is that the performance summary does seem to suggest that classifying location and identities *jointly* is consistently better than trying to deal with these individually. This is probably because the internal representations learned to predict class can then be useful to help predict location more accurately, and vice versa.

Instead of taking average classification performance at a per-pixel level, we can base classification on a majority verdict, all ca 800 per-pixel classifications for a single measurement. The results of this approach (see SM) indicate that misclassification occurs within certain measurements rather than across measurements suggesting that increasing the variation in the training data should improve the classifier.

Conclusion

One-dimensional temporal histograms obtained by capturing laser echoes backscattered from a hidden body contain information that reliably allows the identification of different people in different positions. We underline that once the ANN is trained, the classification process is just the result of multiple matrix multiplications and vector function evaluations and can thus proceed extremely quickly, with millisecond processing times. Thus, a data-driven classifier, such as the ANN used here, can achieve identification with a precision that is not possible with any other currently available approach, and it does so with speeds that are orders of magnitude faster than even the best NLOS reconstruction shown to date. The method proposed here is based on supervised training. We therefore require knowledge on the person from which we are receiving data for the training phase. Once the training is complete however, we no longer require access to the hidden environment. We underline that with the classification scheme presented here, one is obviously limited to identify individuals that are part of the training database. New individuals that have never been seen before cannot be classified or identified with this method. The impact for example of changing clothing (that has not been seen during the training phase) also needs to be assessed in future work.

Interesting questions arise from this work, such as exactly how many individuals may be used for training and then successfully identified with a single temporal histogram that has a given IRF, and the impact of target movement on the classification performance. Furthermore, we have not considered the impact of pose, i.e. the impact of the person standing at an angle, facing away from the wall or indeed, different "arrangements" of a single individual (e.g. hair style, colour etc.). This is an interesting additional degree of freedom that should be considered in future work. The information content analysis of temporal photon echoes remains an open question for future work as well. Recent work has also shown how device-independent training is possible within the context of computational imaging problems, promising the ability to either train and test using completely different detectors or even to train using forward modelling as in[19]. The latter would be particularly significant in the context of

classification of hidden environments where the forward model is possibly easier to fully characterise and model with respect to people. These results pave the way to exciting novel scenarios for machine learning applications, such as identification of groups of individuals (for example distinguishing adults from children) but also of entire environments by means of a single pixel temporal measurements.

Data. All experimental data is available at http://dx.doi.org/10.17861/c9ec32bf-0ff2-4b9f-a48a-5ee2aed59c61.

References
1. Kirmani, A., Hutchison, T., Davis, J. & Raskar, R. Looking around the corner using ultrafast transient imaging. *Int. journal of computer vision* **95**, 13–28 (2011).
2. Gupta, O., Willwacher, T., Velten, A., Veeraraghavan, A. & Raskar, R. Reconstruction of hidden 3D shapes using diffuse reflections. *Opt. express* **20**, 19096–19108 (2012).
3. Velten, A. *et al.* Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nat. Commun.* **3**, 745 (2012).
4. Gariepy, G., Tonolini, F., Henderson, R., Leach, J. & Faccio, D. Detection and tracking of moving objects hidden from view. *Nat Photon* **10**, 23–26 (2016).
5. Buttafava, M., Zeman, J., Tosi, A., Eliceiri, K. & Velten, A. Non-line-of-sight imaging using a time-gated single photon avalanche diode. *Opt. express* **23**, 20997–21011 (2015).
6. Klein, J., Peters, C., Martn, J., Laurenzis, M. & Hullin, M. B. Tracking objects outside the line of sight using 2D intensity images. *Sci. reports* **6** (2016).
7. Katz, O., Small, E. & Silberberg, Y. Looking around corners and through thin turbid layers in real time with scattered incoherent light. *Nat. photonics* **6**, 549–553 (2012).
8. Kadambi, A., Zhao, H., Shi, B. & Raskar, R. Occluded imaging with time-of-flight sensors. *ACM Transactions on Graphics (TOG)* **35**, 15 (2016).
9. Chan, S., Warburton, R. E., Gariepy, G., Leach, J. & Faccio, D. Non-line-of-sight tracking of people at long range. *Opt. express* **25**, 10109 (2017).
10. O'Toole, M., Lindell, D. B. & Wetzstein, G. Confocal non-line-of-sight imaging based on the light-cone transform. *Nat.* (2018).
11. Bouman, K. L. *et al.* Turning corners into cameras: Principles and methods. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2270–2278 (2017).
12. Torralba, A. & Freeman, W. T. Accidental pinhole and pinspeck cameras: Revealing the scene outside the picture. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 374–381 (IEEE, 2012).
13. Cohen, A. L. Anti-pinhole imaging. *Opt. Acta: Int. J. Opt.* **29**, 63–67 (1982).
14. Richardson, J. *et al.* A 32x32 50ps resolution, 10 bit time to digital converter array in 130nm CMOS for time correlated imaging. *Cust. Integr. Circuits Conf. 2009, IEEE CICC 09* 77–80 (2009).
15. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Networks* **61**, 85–117 (2015).
16. LeCun, Y., Bengio, Y. & Hinton, G. E. Deep learning. *Nat.* **521**, 7553 (2015).
17. Goodfellow, I., Bengio, Y. & Courville, A. *Deep learning* (MIT press, 2016).
18. Sinha, A., Lee, J., Li, S. & Barbstathis, G. Lensless computational imaging through deep learning. *arXiv:1702.08516* (2017).
19. Satat, G., Tancik, M., Gupta, O., Heshmat, B. & Raskar, R. Object classification through scattering media with deep learning on time resolved measurement. *Opt. Express* **25**, 17466–17479 (2017).
20. Lyu, M., Wang, H., Li, G. & Situ, G. Exploit imaging through opaque wall via deep learning. *arXiv:1708.07881* (2017).
21. Chollet, F. K. https://github.com/fchollet/keras (2015).

Acknowledgements

Author Contributions
D.F. and M.H. conceived and P.C., A.B. performed the experiments. P.C., D.B., C.F.H., R.M.-S. analysed the data. R.H. developed the SPAD camera. All authors contributed to the manuscript.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-30390-0.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.