LETTER

https://doi.org/10.1038/s41586-018-0868-6

# Computational periscopy with an ordinary digital camera

Charles Saunders[1,2], John Murray-Bruce[1,2] & Vivek K Goyal[1]*

**Computing the amounts of light arriving from different directions enables a diffusely reflecting surface to play the part of a mirror in a periscope—that is, perform non-line-of-sight imaging around an obstruction. Because computational periscopy has so far depended on light-travel distances being proportional to the times of flight, it has mostly been performed with expensive, specialized ultrafast optical systems[1–12]. Here we introduce a two-dimensional computational periscopy technique that requires only a single photograph captured with an ordinary digital camera. Our technique recovers the position of an opaque object and the scene behind (but not completely obscured by) the object, when both the object and scene are outside the line of sight of the camera, without requiring controlled or time-varying illumination. Such recovery is based on the visible penumbra of the opaque object having a linear dependence on the hidden scene that can be modelled through ray optics. Non-line-of-sight imaging using inexpensive, ubiquitous equipment may have considerable value in monitoring hazardous environments, navigation and detecting hidden adversaries.**

The ability to accurately image scenes or detect objects hidden from direct view has many potential applications. Active optical methods for such non-line-of-sight (NLOS) imaging have been developed recently, most of which depend on transient imaging. In a typical transient-imaging configuration, an imaging device consisting of a light source and a light detector lacks direct view of the scene but does have direct view of a diffusely reflecting surface that itself has direct view of the scene. Illumination of a small patch on the diffuse surface with a short light pulse creates transient illumination of the NLOS scene, which is observed indirectly through light that reaches the detector after reflection from the diffuse surface. Transient-imaging-based NLOS scene geometry recovery was first demonstrated through multilateration[1], and the model was extended to include variable reflectivity of scene elements and non-impulsive illumination[13]. Subsequent studies used transient imaging to infer shapes of objects with non-specular surfaces hidden from the direct view of the observer[2,3]. These early studies used femtosecond-laser illumination and picosecond-resolution streak cameras[1–3]. The cost of transient-imaging acquisition can be reduced dramatically with homodyne time-of-flight sensors[14–16], and the increasing availability of single-photon avalanche diode (SPAD) detectors and detector arrays with time-correlated single-photon counting (TCSPC) modules has enabled their use in transient-imaging-based NLOS imaging[4–12]. SPADs with TCSPC are common in many LIDAR (light detection and ranging) applications and were recently used for long-range three-dimensional (3D) imaging[17] and for capturing photometric and geometric information from as few as one detected photon per pixel[18–21]. In addition to lowering the cost of NLOS imaging systems, SPAD-based systems have facilitated the extension of previous round-trip distances of around 1 m to a few metres for NLOS hidden-object estimation[5] and to over 50 m for long-range human localization by coupling a telescope to a single-element SPAD[9]. Furthermore, room-geometry reconstruction by probing a single visible wall using a picosecond laser and an SPAD with TCSPC has been demonstrated[22]. Other established applications of transient imaging include NLOS estimation of object motion and size[23] and single-viewpoint estimation of angular reflectance properties[24].

To address the high cost and the impracticality of existing methods outside laboratory conditions, we developed a computational periscopy technique that uses only an ordinary digital camera. The imaging method is passive, with the radiosity of the NLOS scene caused by sources that are hidden from view and uncontrolled. The NLOS resolution is based on computational inversion of the influence of the scene of interest on the penumbra of an occluding object of known size and shape, which is in an a priori unknown position. Previous techniques exploiting penumbrae required precise knowledge of the occluder positions and used laser illumination and SPAD-based detection[25,26], required occluder motion[27], or had the more limited objective of producing a one-dimensional projection of the moving portion of a scene[28]. A very recent work used calibration measurements of a complex occluder in a light-field reconstruction[29]. NLOS tracking of a moving object using laser illumination—without image formation—has also been demonstrated[30]. Our method does not require calibration, controlled illumination, time-resolved light detection or scene motion, and obtains a full-colour two-dimensional (2D) image.

We demonstrate computational periscopy using an experimental setup consisting of a 4-megapixel digital camera, a 20-inch (1 inch = 2.54 cm) liquid-crystal display (LCD) colour monitor with 4:3 aspect ratio, and a black rectangular occluding object of size 7.7 cm × 7.5 cm supported by a black flat 7-mm-wide stand (Fig. 1). This occluder shape was chosen for computational convenience, but any known occluder shape and size could be incorporated similarly. Additional experiments using a three-dimensional, non-black occluder are presented in Supplementary Information. Light from the LCD monitor, originating from the unknown displayed scene and the monitor's background light, illuminates a visible white Lambertian surface placed in a direction fronto-parallel to the monitor, at a distance of 1.03 m. A monitor is used to allow convenient testing of multiple scenes; additional results using both 2D and 3D diffuse reflecting scenes are presented in Supplementary Information. The camera measurement, which includes shadows and penumbrae cast by the occluder, is a raw 14-bit, 2,016 × 2,016-pixel image with colour channels interleaved according to a Bayer filter RGBG pattern. After averaging of the two green channels and averaging each colour channel over 16 × 16 blocks, three 126 × 126 images (one per colour channel) are extracted and passed to a computer algorithm for occluder position and scene image recovery (Fig. 2).

With the occluder positioned at $\boldsymbol{p}_{\mathrm{o}}$ between the monitor and the visible wall, and with the monitor at distance $D$, the irradiance of a wall patch at $\boldsymbol{p}_{\mathrm{w}}$ is given by

$$I(\boldsymbol{p}_{\mathrm{w}}) = \int_{\boldsymbol{x}\in S} \left\{ \frac{\cos[\measuredangle(\boldsymbol{p}_{\mathrm{w}} - \boldsymbol{x}, \boldsymbol{n}_{x})]\cos[\measuredangle(\boldsymbol{x} - \boldsymbol{p}_{\mathrm{w}}, \boldsymbol{n}_{\mathrm{w}})]}{\left\|\boldsymbol{p}_{\mathrm{w}} - \boldsymbol{x}\right\|_2^2} \times V(\boldsymbol{x};\boldsymbol{p}_{\mathrm{w}};\boldsymbol{p}_{\mathrm{o}})\mu(\boldsymbol{x},\boldsymbol{p}_{\mathrm{w}})f(\boldsymbol{x}) \right\} \mathrm{d}\boldsymbol{x} + b(\boldsymbol{p}_{\mathrm{w}}) \qquad (1)$$

[1] Department of Electrical and Computer Engineering, Boston University, Boston, MA, USA. [2] These authors contributed equally: Charles Saunders, John Murray-Bruce. *e-mail: goyal@bu.edu

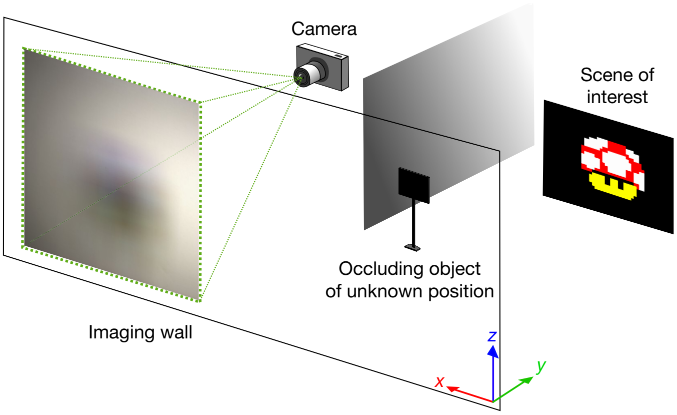

**Fig. 1 | Experimental setup for computational periscopy.** Controlled by a laptop PC, the standard digital camera obtains a snapshot of the irradiance distribution on a visible imaging wall, which is induced by the penumbra of an occluding object owing to light emanating from a scene of interest. The scene of interest is displayed on an LCD monitor for ease of performing experiments with many scenes. The snapshot is fed through a computer algorithm to recover an image of the scene of interest and an estimate of the position of the hidden occluder.

where $f(\boldsymbol{x})$ is the monitor scene radiosity, and integration over $\boldsymbol{x} \in S = \{(x, D, z): x, z \in \mathbb{R}\}$ represents the combination of contributions from the entire scene at $\boldsymbol{p}_{\mathrm{w}}$. The first weighting factor in the integrand models the radial falloff of the flux density (the denominator) and two foreshortening effects: from the wall patch relative to the direction of the incident light and from the monitor pixel relative to the viewing angle of that pixel, with $\boldsymbol{n}_{x}$ and $\boldsymbol{n}_{\mathrm{w}}$ the monitor and wall surface normals, respectively, $\measuredangle(\cdot,\cdot)$ denoting the angle between its vector arguments and $\|\boldsymbol{x}\|_2$ representing the Euclidean norm of a vector. The second weighting factor, $V(\boldsymbol{x}; \boldsymbol{p}_{\mathrm{w}}; \boldsymbol{p}_{\mathrm{o}})$, is a Boolean-valued visibility function that equals 1 when the path from $\boldsymbol{x}$ to $\boldsymbol{p}_{\mathrm{w}}$ is unoccluded and 0 otherwise. The factor $\mu(\boldsymbol{x}, \boldsymbol{p}_{\mathrm{w}})$ describes the radiometric model for the monitor's variation with viewing angle (see Supplementary Information). The final term, $b(\boldsymbol{p}_{\mathrm{w}})$, represents the contribution from sources outside the modelled scene area, $S$, at the visible wall. Equation (1) is the rendering equation for computer graphics adapted to our setting[31].

朗伯型

Assuming the reflection from the visible wall to be Lambertian, the reduced-resolution digital photograph of the visible wall is modelled by discretizing equation (1) with $\boldsymbol{p}_{\mathrm{w}}$ taking $(126)^2 = 15{,}876$ values in the camera's field of view (FOV). Namely, for each colour channel we obtain a simple affine model $\boldsymbol{y} = A(\boldsymbol{p}_{\mathrm{o}})\boldsymbol{f} + \boldsymbol{b}$, where the digital photograph is vectorized into a column vector, and the light transport matrix $A(\boldsymbol{p}_{\mathrm{o}})$ has 15,876 rows and a number of columns that depends on the attempted reconstruction resolution (see Supplementary Information). Forming an image of the hidden scene amounts to inverting the resulting linear system for each colour channel.

A：光传输矩阵

The visibility function—equivalently, the presence of the occluder—is central to the conditioning of the inversion. Without an occluder, the weighting factors in equation (1) depend too weakly on $\boldsymbol{x}$ for a well conditioned recovery of $f(\boldsymbol{x})$ (see Supplementary Information)[2,13,25]. The presence of an occluder introduces shadows and penumbrae that make some image formation possible, but everyday experience suggests that this is extremely limited. In discretized form, without an occluder, the rows of $A$ are too similar to enable well conditioned inversion. Variations in the visibility function $V(\boldsymbol{x}; \boldsymbol{p}_{\mathrm{w}}; \boldsymbol{p}_{\mathrm{o}})$ caused by the presence of an occluder improve the conditioning of $A(\boldsymbol{p}_{\mathrm{o}})$ for inversion because its columns become more different from each other. By treating a portion of the scene plane as resolvable if and only if this portion is visible in at least one camera measurement and invisible in at least one other, we define a computational FOV (Fig. 3; see Supplementary Information).

Recovering $\boldsymbol{p}_{\mathrm{o}}$ and $\boldsymbol{f}$ from the single-snapshot camera measurement $\boldsymbol{y}$ is a nonlinear problem. Because the number of measurements (rows of $A(\boldsymbol{p}_{\mathrm{o}})$) is large relative to the recoverable resolution of the hidden scene, the measurements $\boldsymbol{y}$ reside close to a low-dimensional affine subspace that is dependent on the occluder position $\boldsymbol{p}_{\mathrm{o}}$ and the background $\boldsymbol{b}$. The occluder position is estimated from $\boldsymbol{y}$ through

$$\hat{\boldsymbol{p}}_{\mathrm{o}} = \underset{\boldsymbol{p}_{\mathrm{o}}}{\operatorname{argmax}} \left\| A(\boldsymbol{p}_{\mathrm{o}})[A(\boldsymbol{p}_{\mathrm{o}})^{\mathrm{T}} A(\boldsymbol{p}_{\mathrm{o}})]^{-1} A(\boldsymbol{p}_{\mathrm{o}})^{\mathrm{T}} \boldsymbol{y} \right\|_2^2 \qquad (2)$$

where $A(\boldsymbol{p}_{\mathrm{o}})$ is the computed light-transport matrix for an occluder position $\boldsymbol{p}_{\mathrm{o}}$; the omission of the unknown $\boldsymbol{b}$ does not greatly degrade the estimate (see Supplementary Information). The three estimates obtained by solving this maximization for each colour channel are averaged to obtain a single $\hat{\boldsymbol{p}}_{\mathrm{o}}$.

Given the estimated occluder position $\hat{\boldsymbol{p}}_{\mathrm{o}}$, an estimate $\hat{A} = A(\hat{\boldsymbol{p}}_{\mathrm{o}})$ of the true light-transport matrix $A(\boldsymbol{p}_{\mathrm{o}})$ is computed. If the estimated occluder position were exactly correct and model mismatch and background contributions were inconsequential, pre-multiplying the vectorized measurements $\boldsymbol{y}$ (for each colour channel) by the pseudo-inverse $\hat{A}^{\dagger} = (\hat{A}^{\mathrm{T}}\hat{A})^{-1}\hat{A}^{\mathrm{T}}$ would yield the least-squares estimate of the hidden scene's RGB content. To improve robustness to noise and model mismatch, we exploit transverse spatial correlations that are prevalent in real-world scenes by promoting sparsity in the scene's gradient via total variation (TV) regularization[32]

$$\hat{\boldsymbol{f}} = \underset{\boldsymbol{f}}{\operatorname{argmin}} \ \|\hat{A}\boldsymbol{f} - \boldsymbol{y}\|_2^2 + \lambda \ \|\boldsymbol{f}\|_{\mathrm{TV}} \qquad (3)$$

where the operator $\|\cdot\|_{\mathrm{TV}}$ denotes the TV semi-norm and $\lambda$ is the TV-regularization parameter.

To further improve image quality, we take the differences of measurements $\boldsymbol{y}$ for (vertically) neighbouring blocks of 16 × 16 pixels and of the corresponding rows of $\hat{A}$ before solving an optimization problem analogous to equation (3) (see equation (7) in Methods). Because any light originating from outside the computational FOV has slow spatial variation, the background contribution $\boldsymbol{b}$ is approximately constant (that is, $b_{i+1} \approx b_i \approx b$) for neighbouring pixels and is thus approximately cancelled

$$y_{i+1} - y_i \approx (\boldsymbol{a}_{i+1}^{\mathrm{T}}\boldsymbol{f} + b) - (\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{f} + b) \approx (\boldsymbol{a}_{i+1} - \boldsymbol{a}_i)^{\mathrm{T}}\boldsymbol{f} \qquad (4)$$

where $\boldsymbol{a}_i^{\mathrm{T}}$ is the $i$th row of $A$. This also provides some robustness to ambient light, as verified by additional experiments described

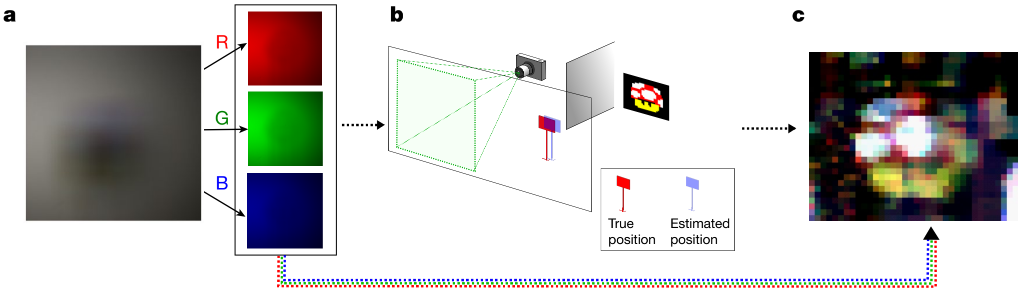

**Fig. 2 | Reconstruction procedure.** **a**, Camera measurements are de-interleaved to give RGB data from Bayer pattern measurements, with the green channels averaged. **b**, Occluder position is estimated. **c**, Hidden-scene reconstruction from camera measurements and from the estimated occluder position.

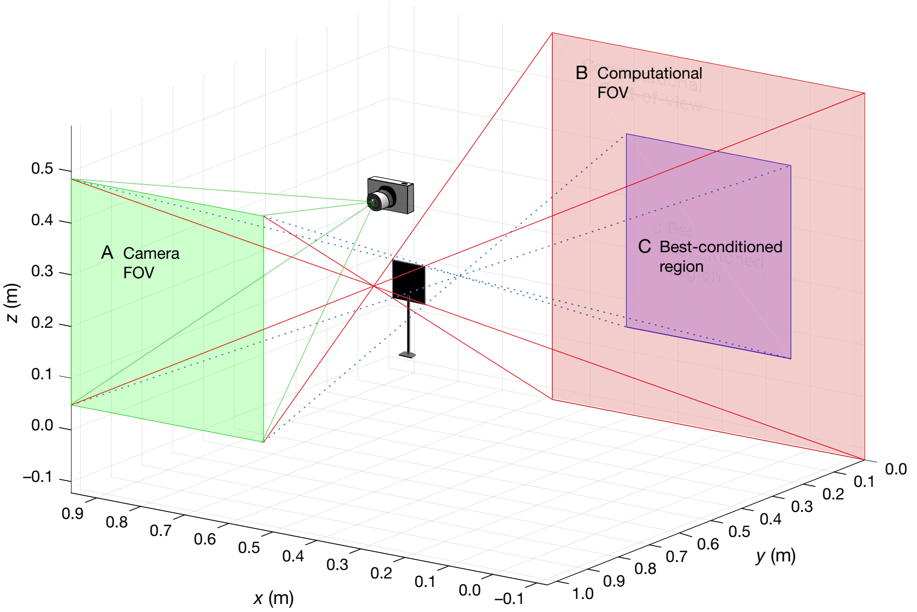

**Fig. 3 | Computational field of view.** Portions of the scene plane that are visible to a part of the camera FOV (region A) and occluded from another part of the camera FOV form the computational FOV (region B). Within the computational FOV, portions from which projections of multiple occluder edges fall within the camera FOV are better conditioned than those from which the projection of only one occluder edge falls within the camera FOV. The best-conditioned region is the portion of the scene plane from which projections of all four occluder edges fall within the camera FOV (region C).

in Supplementary Information. Ultimately, one is able to produce an image of the computational FOV because all appreciable variations in $y$ are due to this portion of the scene.

We propose also an alternative reconstruction method that exploits a property of the light-transport matrices as $\boldsymbol{p}_o$ is varied. Inaccurate occluder transverse position or depth leads to shifted or magnified (or minified) estimates of the displayed scene. This observation is exploited to produce a multiplicity of additional reconstructions, which are nonlinearly combined to produce a single noise-reduced image (see equation (8) in Methods).

Each hidden-scene patch is a block of 35 × 35 monitor pixels; because the monitor has 1,280 × 1,024 resolution, $\lfloor\lfloor 1{,}280/35\rfloor\rfloor \times \lfloor\lfloor 1{,}024/35\rfloor\rfloor = 36 \times 29$ rectangular scene pixels are obtained, each of size 1.12 cm × 1.05 cm ($\lfloor\lfloor\cdot\rfloor\rfloor$ denotes the floor function). The scene is aligned to the screen's top and left edges, occupying 30.5 cm vertically and 40.3 cm horizontally. The distance from the camera to the visible wall is approximately 1.5 m, such that its FOV is 43.7 cm × 43.7 cm, centred at (74.1 cm, 26.4 cm). With this configuration, several 36 × 29 scene-pixel test images are used to evaluate our computational periscope. One experimental scene is an anthropomorphic mushroom image with approximate dimensions of 26 cm × 19 cm (Fig. 4a, top). An exposure of 175 ms maximizes signal strength while avoiding saturation, and 20 such exposures are taken and averaged, yielding an effective exposure of 3.5 s. The snapshot is

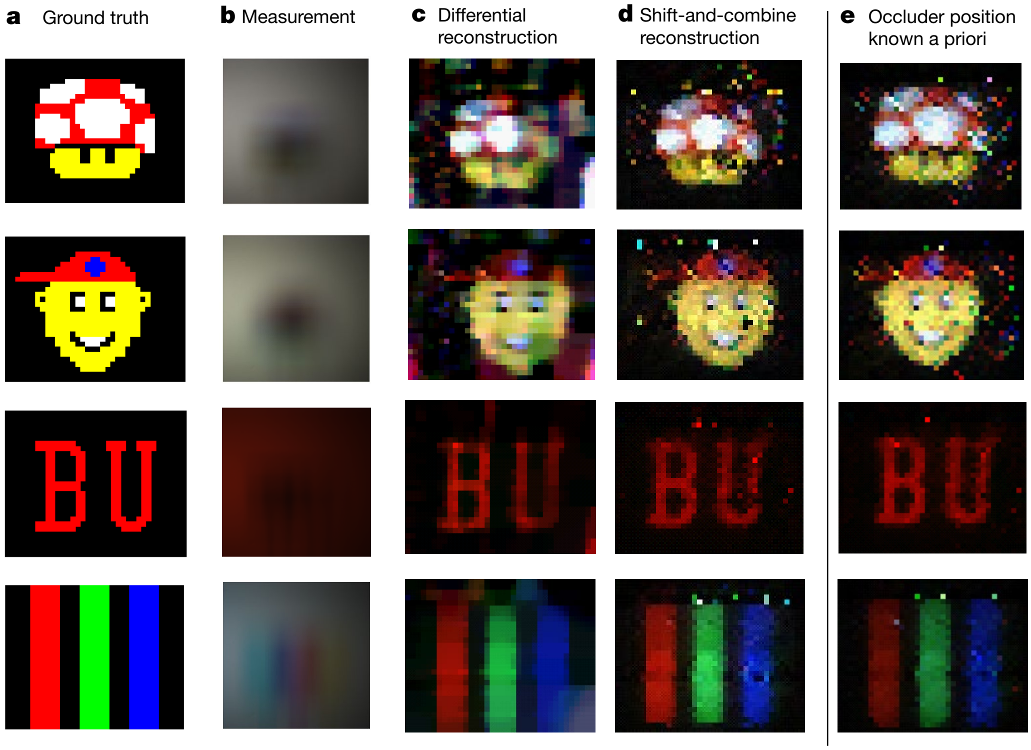
a Ground truth
b Measurement
c Differential reconstruction
d Shift-and-combine reconstruction
e Occluder position known a priori
BU

**Fig. 4 | Reconstructions of different hidden scenes. a**, Four ground-truth scenes displayed on the monitor. **b**, Camera measurement obtained for the corresponding scenes. **c**, A single reconstruction using the differential framework and the estimated occluder position. **d**, Reduced-noise final reconstructions, obtained by combining reconstructions acquired from each of 49 postulated occluder positions, around the estimated value. **e**, Reconstructions obtained assuming the true occluder position is a priori known and using the same algorithm as in **d**.

fed to a computer algorithm to produce a reconstruction (Fig. 4c, top) by estimation of the occluder position, vertical differencing of the data and the light-transport matrix, and image estimation. This snapshot (Fig. 4b, top) is also fed into another computer algorithm to produce a reconstruction (Fig. 4d, top) by estimation of the occluder position, formation of 49 image estimates for a 7 × 7 array of postulated occluder positions, and nonlinear combination of the 49 estimates. For comparison, a reconstruction (Fig. 4e, top) is formed from the snapshot along with the actual occluder position. Using unoptimized code on a desktop computer, the initial occluder estimation takes 18 min, and subsequent hidden-scene recovery (using the approximate background-cancellation method) takes an additional 48 s. Most of that computation time is for forming $A(\boldsymbol{p}_{\mathrm{o}})$ matrices (see Methods). Reducing the exposure time to under 1 s enables the capture of a one-frame-per-second movie (Supplementary Video).

Results are provided for three additional scenes (Fig. 4, bottom three rows). The exposure time required to maximize signal strength while avoiding saturation varies between 175 ms and 425 ms, and the averages of 20 such exposures give the inputs to the computational method (Fig. 4b). The estimated occluder positions are reported in Supplementary Table S1.

The results for the initial scene (Fig. 4, top) show that our computational imaging method clearly resolves moderately sized features, such as the white and red patches, along with larger features, such as the head and yellow face; smaller features, such as the eyes and unibrow, are visible but with worse accuracy. Similarly, for the second scene (Fig. 4, second row from the top), even the white teeth and blue plus on the hat are present in the reconstructions, along with larger features, such as the face and hat. These two scenes demonstrate that measurements that are difficult to distinguish visually (Fig. 4b) may yield distinct and clearly identifiable reconstructions.

The occluder position estimates have roughly centimetre accuracy (Supplementary Table S1). Reconstructions based on estimated occluder positions (Fig. 4c, d) have similar quality to those based on known occluder positions (Fig. 4e), demonstrating robustness to the lack of knowledge of the occluder position.

The results show that the penumbra cast by an object may contain enough information to both estimate the position of the object and computationally construct an image of the computational FOV created by the object. In such a setting, we demonstrate that 2D colour NLOS imaging is possible with an ordinary digital camera, without requiring time-varying illumination and high-speed sensing.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0868-6.

Received: 21 August 2018; Accepted: 12 December 2018;
Published online 23 January 2019.

1. Kirmani, A., Hutchison, T., Davis, J. & Raskar, R. Looking around the corner using transient imaging. In *Proc. 2009 IEEE 12th Int. Conf. Computer Vision* 159–166 (IEEE, 2009).
2. Velten, A. et al. Recovering three-dimensional shape around a corner using ultrafast time-of-flight imaging. *Nat. Commun*. **3**, 745 (2012).
3. Gupta, O., Willwacher, T., Velten, A., Veeraraghavan, A. & Raskar, R. Reconstruction of hidden 3D shapes using diffuse reflections. *Opt. Express* **20**, 19096–19108 (2012).
4. Xu, K. et al. Image contrast model of non-line-of-sight imaging based on laser range-gated imaging. *Opt. Eng*. **53**, 061610 (2013).
5. Laurenzis, M. & Velten, A. Nonline-of-sight laser gated viewing of scattered photons. *Opt. Eng*. **53**, 023102 (2014).
6. 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).
7. 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).
8. Klein, J., Laurenzis, M. & Hullin, M. Transient imaging for real-time tracking around a corner. In *Proc. SPIE Electro-Optical Remote Sensing X* 998802 (International Society for Optics and Photonics, 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**, 10109–10117 (2017).
10. Tsai, C.-y., Kutulakos, K. N., Narasimhan, S. G. & Sankaranarayanan, A. C. The geometry of first-returning photons for non-line-of-sight imaging. In *Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition* 7216–7224 (IEEE, 2017).
11. Heide, F. et al. Non-line-of-sight imaging with partial occluders and surface normals. Preprint at https://arxiv.org/abs/1711.07134 (2018).
12. O'Toole, M., Lindell, D. B. & Wetzstein, G. Confocal non-line-of-sight imaging based on the light-cone transform. *Nature* **555**, 338–341 (2018).
13. Kirmani, A., Jeelani, H., Montazerhodjat, V. & Goyal, V. K. Diffuse imaging: creating optical images with unfocused time-resolved illumination and sensing. *IEEE Signal Process. Lett*. **19**, 31–34 (2012).
14. Heide, F., Hullin, M. B., Gregson, J. & Heidrich, W. Low-budget transient imaging using photonic mixer devices. *ACM Trans. Graph*. **32**, 45 (2013).
15. Heide, F., Xiao, L., Heidrich, W. & Hullin, M. B. Diffuse mirrors: 3D reconstruction from diffuse indirect illumination using inexpensive time-of-flight sensors. In *Proc. 2014 IEEE Conf. Computer Vision and Pattern Recognition* 3222–3229 (IEEE, 2014).
16. Kadambi, A., Zhao, H., Shi, B. & Raskar, R. Occluded imaging with time-of-flight sensors. *ACM Trans. Graph*. **35**, 15 (2016).
17. Pawlikowska, A. M., Halimi, A., Lamb, R. A. & Buller, G. S. Single-photon three-dimensional imaging at up to 10 kilometers range. *Opt. Express* **25**, 11919–11931 (2017).
18. Kirmani, A. et al. First-photon imaging. *Science* **343**, 58–61 (2014).
19. Shin, D., Kirmani, A., Goyal, V. K. & Shapiro, J. H. Photon-efficient computational 3D and reflectivity imaging with single-photon detectors. *IEEE Trans. Comput. Imaging* **1**, 112–125 (2015).
20. Altmann, Y., Ren, X., McCarthy, A., Buller, G. S. & McLaughlin, S. Lidar waveform-based analysis of depth images constructed using sparse single-photon data. *IEEE Trans. Image Process*. **25**, 1935–1946 (2016).
21. Rapp, J. & Goyal, V. K. A few photons among many: unmixing signal and noise for photon-efficient active imaging. *IEEE Trans. Comput. Imaging* **3**, 445–459 (2017).
22. Pediredla, A. K., Buttafava, M., Tosi, A., Cossairt, O. & Veeraraghavan, A. Reconstructing rooms using photon echoes: a plane based model and reconstruction algorithm for looking around the corner. In *Proc. 2017 IEEE Int. Conf. Computational Photography* 1–12 (IEEE, 2017).
23. Pandharkar, R. et al. Estimating motion and size of moving non-line-of-sight objects in cluttered environments. In *Proc. 2011 IEEE Conf. Computer Vision and Pattern Recognition* 265–272 (IEEE, 2011).
24. Naik, N., Zhao, S., Velten, A., Raskar, R. & Bala, K. Single view reflectance capture using multiplexed scattering and time-of-flight imaging. *ACM Trans. Graphics* **30**, 171 (ACM, 2011).
25. Thrampoulidis, C. et al. Exploiting occlusion in non-line-of-sight active imaging. *IEEE Trans. Comput. Imaging* **4**, 419–431 (2018).
26. Xu, F. et al. Revealing hidden scenes by photon-efficient occlusion-based opportunistic active imaging. *Opt. Express* **26**, 9945–9962 (2018).
27. Torralba, A. & Freeman, W. T. Accidental pinhole and pinspeck cameras: revealing the scene outside the picture. *Int. J. Comput. Vis*. **110**, 92–112 (2014).
28. Bouman, K. L. et al. Turning corners into cameras: Principles and methods. In *Proc. 23rd IEEE Int. Conf. Computer Vision*, 2270–2278 (IEEE, 2017).
29. Baradad, M. et al. Inferring light fields from shadows. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 6267–6275 (2018).
30. Klein, J., Peters, C., Martín, J., Laurenzis, M. & Hullin, M. B. Tracking objects outside the line of sight using 2D intensity images. *Sci. Rep*. **6**, 32491 (2016).
31. Kajiya, J. T. The rendering equation. In *Proc. 13th Conf. Computer Graphics and Interactive Techniques* Vol. 20 143–150 (ACM, 1986).
32. Beck, A. & Teboulle, M. A fast iterative shrinkage-thresholding algorithm. *SIAM J. Imaging Sci*. **2**, 183–202 (2009).

**Acknowledgements** We thank F. Durand, W. T. Freeman, Y. Ma, J. Rapp, J. H. Shapiro, A. Torralba, F. N. C. Wong and G. W. Wornell for discussions. This work was supported by the Defense Advanced Research Projects Agency (DARPA) REVEAL Program contract number HR0011-16-C-0030.

**Reviewer information** *Nature* thanks M. Laurenzis, M. O'Toole and A. Velten for their contribution to the peer review of this work.

**Author contributions** V.K.G. conceptualized the project, obtained funding and supervised the research. C.S. and J.M.-B. developed the methodology, performed the experiments, wrote the software and validated the results. C.S. produced the visualizations. J.M.-B. wrote the original draft. C.S., J.M.-B. and V.K.G. reviewed and edited the paper.

**Competing interests** The authors declare no competing interests.

**Additional information**

**Supplementary information** is available for this paper at https://doi.org/10.1038/s41586-018-0868-6.

**Reprints and permissions information** is available at http://www.nature.com/reprints.

**Correspondence and requests for materials** should be addressed to V.K.G.

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

## METHODS

**Equipment.** The scenes were displayed on a Dell LCD monitor model 2001FP, which has 4:3 aspect ratio and 1,280 × 1,024 resolution. With no line of sight from the monitor to the camera, visibility was achieved via a white Elmer's foam board, which is visually diffuse; any specular component present was not modelled and thus not directly exploited. The camera was a FLIR Grasshopper3 model GS3-U3-41S4C-C, which has 2,016 × 2,016 resolution (4.1 megapixels) and was used with a Tamron M118FM16 lens with 16 mm focal length and f/1.4 aperture. The scene and the camera were controlled using a Lenovo ThinkPad P51s laptop computer.

**Data acquisition from the camera.** A Python script was used to control the data acquisition. It performed the following steps. First, a test scene was displayed on the scene monitor. Then, to form a snapshot with suppressed noise, multiple camera measurements were taken in succession and summed. A pre-calibrated shutter speed that utilized approximately the full dynamic range of the camera per exposure was used. The final measurement for each block was a raw, 14-bit, 2,016 × 2,016-pixel image with the three colour channels interleaved according to the RGBG Bayer filter pattern. In forming the three 1,008 × 1,008 colour-channel images, the two green channels were averaged. Each colour channel was further averaged over 16 × 16 blocks to produce a 126 × 126 data matrix that was reshaped to column vector $\boldsymbol{y}$ of length 15,876.

**Computing a light transport matrix $A(\boldsymbol{p}_\mathrm{o})$.** Recall that the element $[A(\boldsymbol{p}_\mathrm{o})]_{i,j}$ of the light transport matrix $A(\boldsymbol{p}_\mathrm{o})$ represents the weighting of the contribution of light from the hidden-scene pixel $i\in\{1, 2, \ldots, N\}$ to the camera FOV pixel $j\in\{1, 2, \ldots, M\}$, where $M = 15{,}876$ and $N$ is the resolution at which recovery of the hidden scene is attempted (for example, $N = 1{,}044$ for producing a 29 × 36 reconstruction). For any calculation of $A(\boldsymbol{p}_\mathrm{o})$, including the many computations for finding an estimate $\hat{\boldsymbol{p}}_\mathrm{o}$, we performed the calculation in equation (S6) in Supplementary Information with $L = 64$.

**Computing the estimate $\hat{\boldsymbol{p}}_\mathrm{o}$ of occluder position $\boldsymbol{p}_\mathrm{o}$.** Estimation of the occluder position $\boldsymbol{p}_\mathrm{o}$ was performed using the grid-search approach outlined in Algorithm 1 in Supplementary Information. The algorithm is based on the camera measurements $\boldsymbol{y}$ being made to reside near the range of $A(\hat{\boldsymbol{p}}_\mathrm{o})$, which is a low-dimensional subspace of the 15,876-dimensional space of downsampled measurements. The desired estimate of the position of the hidden occluder, $\hat{\boldsymbol{p}}_\mathrm{o}$, is the one that minimizes the Euclidean distance between $\boldsymbol{y}$ and the range space of $A(\hat{\boldsymbol{p}}_\mathrm{o})$ or, equivalently, maximizes the Euclidean norm of the orthogonal projection of $\boldsymbol{y}$ onto the range space[33] of $A(\hat{\boldsymbol{p}}_\mathrm{o})$. In practice, poor conditioning of $A(\boldsymbol{p}_\mathrm{o})$ for certain candidate occluder positions $\boldsymbol{p}_\mathrm{o}$ makes it more robust to orthogonally project to the smaller subspace spanned by the left singular vectors of $A(\hat{\boldsymbol{p}}_\mathrm{o})$ that are associated with the 'significant' singular values—that is, those that are within a factor $\kappa\in(0,1)$ of the largest singular value. (For instance, when $\boldsymbol{p}_\mathrm{o}$ is such that the occluder does not cast a shadow in the camera's FOV, $A(\boldsymbol{p}_\mathrm{o})$ is very poorly conditioned for inversion. Only a number $N_0 < N$ of the singular values will be substantially larger than zero. Hence, orthogonally projecting to the range of $A(\boldsymbol{p}_\mathrm{o})$ will retain $N - N_0$ dimensions of $\boldsymbol{y}$ that depend deterministically, but very erratically, on $\boldsymbol{p}_\mathrm{o}$; it is as if those directions are chosen uniformly at random, reducing the reliability of estimating the correct $\boldsymbol{p}_\mathrm{o}$.) Then, if $A(\boldsymbol{p}_\mathrm{o})$ is approximated by the truncated singular value decomposition $U\Sigma V^\mathrm{T}$ using only significant singular values, equation (2) can be written using

$$\begin{aligned}\left\|A(\boldsymbol{p}_\mathrm{o})[A(\boldsymbol{p}_\mathrm{o})^\mathrm{T}A(\boldsymbol{p}_\mathrm{o})]^{-1}A(\boldsymbol{p}_\mathrm{o})^\mathrm{T}\boldsymbol{y}\right\|_2^2 &= \|U\Sigma V^\mathrm{T}[(U\Sigma V^\mathrm{T})^\mathrm{T}(U\Sigma V^\mathrm{T})]^{-1}(U\Sigma V^\mathrm{T})^\mathrm{T}\boldsymbol{y}\|_2^2\\ &= \|U\Sigma V^\mathrm{T}(U\Sigma V^\mathrm{T}U\Sigma V^\mathrm{T})^{-1}U\Sigma V^\mathrm{T}\boldsymbol{y}\|_2^2\\ &= \|U\Sigma V^\mathrm{T}V\Sigma^{-2}V^\mathrm{T}V\Sigma U^\mathrm{T}\boldsymbol{y}\|_2^2\\ &= \|UU^\mathrm{T}\boldsymbol{y}\|_2^2 = (UU^\mathrm{T}\boldsymbol{y})^\mathrm{T}UU^\mathrm{T}\boldsymbol{y}\\ &= (U^\mathrm{T}\boldsymbol{y})^\mathrm{T}U^\mathrm{T}U(U^\mathrm{T}\boldsymbol{y}) = \|U^\mathrm{T}\boldsymbol{y}\|_2^2\end{aligned}$$

as in Algorithm 1 (see Supplementary Information).

For a given discretization in equation (S6), the cost of computing $A(\boldsymbol{p}_\mathrm{o})$ for a single occluder position is $\mathrm{O}(LMN)$ and the cost of its singular value decomposition[34] is $\mathrm{O}(N^2M)$ for $N < M$. Hence the cost of computing the occluder position estimate using Algorithm 1 is $\mathrm{O}[(LMN + N^2M)n]$, where $n$ is the total number of possible occluder positions considered.

Although this approach is effective (Supplementary Table S1), it can be computationally prohibitive for large $n$. With minimal loss in performance, considerable cost reduction can be achieved by first searching a coarse grid of, say, $m \ll n$ points to obtain an initial estimate $\hat{\boldsymbol{p}}_\mathrm{o}$ and then refining that estimate by searching within its neighbourhood. An initial coarse search is more accurate with projections to low-dimensional subspaces, and finer searches are more accurate with projections to higher-dimensional subspaces. We found three searches, with first $\kappa = 0.75$, then $\kappa = 0.5$ and finally $\kappa = 0.05$, to be effective. In addition, a coordinate ascent-based search provided further improvements in terms of computational complexity.

**Computing the scene estimate $\hat{\boldsymbol{f}}$ using single occluder position estimate $\hat{\boldsymbol{p}}_\mathrm{o}$.** Each time a scene estimate $\hat{\boldsymbol{f}}$ was to be computed from a single occluder position estimate, it was found by solving the TV-regularized optimization problem of equation (3) for each colour channel independently. Specifically, the isotropic total-variation semi-norm

$$\|\boldsymbol{f}\|_\mathrm{TV} = \sum_{i,j}\sqrt{(F_{i,j}-F_{i+1,j})^2 + (F_{i,j}-F_{i,j+1})^2} \qquad (5)$$

was used, where $F \in \mathbb{R}^{N_1\times N_2}$ is $\boldsymbol{f}$ reshaped to the dimensions of the $N_1 \times N_2$ image that we are reconstructing.

The optimization in equation (3) was performed using the fast iterative shrinkage–thresholding algorithm of Beck and Teboulle[32]. The algorithm requires an initial estimate $\boldsymbol{f}^{(0)}$, which was taken as the least-squares estimate $(\hat{A}^\mathrm{T}\hat{A})^{-1}\hat{A}^\mathrm{T}\boldsymbol{y}$. The regularization parameter $\lambda$ was chosen empirically for each test scene.

**Computing the final scene estimate with spatial differencing.** By noting that un-modelled, multi-bounce light and ambient background light will tend to be spatially slowly varying or close to constant in the camera measurements, we can augment the model with an approximately constant background term. Specifically, the model in equation (S3) in Supplementary Information

$$\boldsymbol{y} = A(\boldsymbol{p}_\mathrm{o})\boldsymbol{f} + \boldsymbol{b}$$

where $\boldsymbol{b} = [b_1, b_2, \ldots, b_i, b_{i+1}, \ldots, b_M]$, models the unknown background and $M = 15{,}876$ is the number of camera FOV pixels. Taking the difference between two neighbouring camera measurements, that is, $y_{i+1} - y_i$, gives

$$y_{i+1} - y_i \approx \boldsymbol{a}_{i+1}^\mathrm{T}\boldsymbol{f} + b_{i+1} - \boldsymbol{a}_i^\mathrm{T}\boldsymbol{f} - b_i \approx (\boldsymbol{a}_{i+1} - \boldsymbol{a}_i)^\mathrm{T}\boldsymbol{f} + (b_{i+1} - b_i)$$

Further imposing the slowly varying background assumption, $b_{i+1} \approx b_i$, implies that

$$y_{i+1} - y_i \approx (\boldsymbol{a}_{i+1} - \boldsymbol{a}_i)^\mathrm{T}\boldsymbol{f} \qquad (6)$$

Equation (6) can therefore be rewritten in matrix-vector form as

$$D\boldsymbol{y} = DA\boldsymbol{f}$$

where $D$ is the so-called difference matrix. Similarly to equation (3), we formulate and solve the optimization problem

$$\hat{\boldsymbol{f}} = \underset{\boldsymbol{f}}{\operatorname{argmin}}\ \|D\hat{A}\boldsymbol{f} - D\boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{f}\|_\mathrm{TV} \qquad (7)$$

obtained by combining the new linear forward model with the usual TV prior. This new approach empirically exhibits increased robustness to model mismatch. As such, only slight improvements can be made by combining multiple hidden-scene reconstructions using different postulated occluder locations. This slight improvement is considerably outweighed by the reduction in computational complexity by not having to compute a multiplicity of reconstructions.

**Computing the postulated occluder positions $\{\hat{\boldsymbol{p}}_{\mathrm{o},k}\}_{k=1}^{48}$ from the occluder position estimate $\hat{\boldsymbol{p}}_{\mathrm{o},0}$.** We generated a set of postulated occluder positions that give scene reconstructions with predetermined horizontal or vertical shifts. Shifts in the $x$ or $z$ components of $\hat{\boldsymbol{p}}_\mathrm{o}$ lead to shifts of the entire scene reconstruction by an amount proportional to $D/(\hat{\boldsymbol{p}}_{\mathrm{o},0})_y$, following from an application of the similar-triangles property.

Let $\boldsymbol{p}_{\mathrm{o},h,v} = ((\boldsymbol{p}_\mathrm{o})_x + hW(\boldsymbol{p}_\mathrm{o})_y/D, (\boldsymbol{p}_\mathrm{o})_y, (\boldsymbol{p}_\mathrm{o})_z + vH(\boldsymbol{p}_\mathrm{o})_y/D)$ denote an occluder position that results in an $h$-pixel horizontal shift and $v$-pixel vertical shift in the reconstructed scene, where $W$ is the width and $H$ the height of a scene pixel. Then lexicographic ordering of the set

$$\{\boldsymbol{p}_{\mathrm{o},h,v}: (h,v) \in \{-6,-4,\ldots,4,6\} \times \{-6,-4,\ldots,4,6\}, (h,v) \neq (0,0)\}$$

gives $\{\hat{\boldsymbol{p}}_{\mathrm{o},k}\}_{k=1}^{48}$ as required. We note that $\hat{\boldsymbol{p}}_{\mathrm{o},0}$ is precisely $\boldsymbol{p}_{\mathrm{o},0,0}$.

**Computing the final scene estimate from scene estimates $\{\hat{\boldsymbol{f}}_k\}_{k=0}^{48}$.** Once a set of scene estimates $\{\hat{\boldsymbol{f}}_k\}_{k=0}^{48}$ was computed, they were registered and combined. The scene estimates were generated using postulated occluder positions that resulted in intentional integer-pixel shifts in the reconstruction. Thus, to align each of the scene estimates, the reverse shifts were applied with zero padding to form a registered ensemble of estimates $\{\hat{\boldsymbol{f}}_k^\mathrm{reg}\}_{k=0}^{48}$. Examples are shown in Supplementary Figs. S1, S2.

To form the final estimate $\hat{\boldsymbol{f}}$, we combined the 49 registered estimates with a nonlinear procedure used independently for each pixel. Consider the set of registered estimates $\{\hat{f}_k^\mathrm{reg}\}_{k=0}^{48}$ for one pixel $\hat{f}_i$. To balance the outlier rejecting property of the median with the variance reduction property of the sample mean, we select a parameter $\theta$ (empirically set to 0.25) and use the sample mean of the samples that are within $\theta$ of the median. More explicitly, let $m_i$ denote the median of $\{\hat{f}_k^{\text{reg}}\}_{k=0}^{48}$. The estimate is

$$\hat{f}_i = \text{mean}(\{\hat{f}_{i,k}^{\text{reg}}, k \in \{0, 1, \ldots, 48\} : |\hat{f}_{i,k}^{\text{reg}} - m_i| < \theta\}) \qquad (8)$$

as illustrated in Supplementary Fig. S3. This method of combining the ensemble is inspired by the alpha-trimmed mean[35].

**Code availability.** The computer codes used to generate the results presented in this manuscript are available on GitHub at https://github.com/Computational-Periscopy/Ordinary-Camera. Documentation on how to use the codes to reproduce the results is also included therein.

## Data availability

Raw data captured with our digital camera during the experiments presented here are available on GitHub at https://github.com/Computational-Periscopy/Ordinary-Camera.

33. Vetterli, M., Kovačević, J. & Goyal, V. K. *Foundations of Signal Processing* (Cambridge Univ. Press, Cambridge, 2014).
34. Golub, G. H. & Van Loan, C. F. *Matrix Computations* 3rd edn (Johns Hopkins Univ. Press, Baltimore, 1989),
35. Bednar, J. B. & Watt, T. L. Alpha-trimmed means and their relationship to median filters. *IEEE Trans. Acoust. Speech Signal Process.* **32**, 145–153 (1984).