Ranging and light field imaging with transparent photodetectors

Miao-Bin Lien [ID], Che-Hung Liu, Il Yong Chun [ID], Saiprasad Ravishankar [ID], Hung Nien [ID], Minmin Zhou, Jeffrey A. Fessler [ID]*, Zhaohui Zhong [ID]* and Theodore B. Norris [ID]*

The core of any optical imaging system is a photodetector. Whether it is film or a semiconductor chip in a camera, or indeed the retina in an eye, conventional photodetectors are designed to absorb most of the incident light and record a projected two-dimensional (2D) distribution of light from a scene. The intensity distribution of light from 3D objects, however, can be described by a 4D light field, so optical imaging systems that can acquire higher dimensions of optical information are highly desirable[1–3]. Here, we report a proof-of-concept light field imaging scheme using transparent graphene photodetector stacks. On a transparent substrate we fabricate a photodetector using graphene as the light-sensing layer, the conducting channel layer, the gate layer and interconnects, enabling sensitive light detection and high transparency at the same time. This technology opens up the possibility of developing sensor arrays that can be stacked along the light path, enabling entirely new configurations of optical imaging devices. We experimentally demonstrate depth ranging using a double stack of transparent detectors and develop a method for computational reconstruction of a 4D light field from a single exposure that can be applied following the successful fabrication of dense 2D transparent sensor arrays.

Optical sensors in the vast majority of today's imaging devices are flat devices that record the intensity of the impinging light at each pixel on the sensor, typically for three particular colours (red, green and blue). Because light is detected in only a single plane, all information about the direction of the light rays is lost. As a result, the recorded images are 2D projections of the actual 3D object in real space, with a single plane of focus and hence a finite depth of field (that is, only a limited region of the object space is in precise focus). The ultimate imaging system would produce a complete representation of the 3D scene, with an infinite depth of field. For any given wavelength and point in time, the light rays emanating from 3D objects in a scene contain 5D information, namely the intensity at each location in space and the angular direction (θ, φ) of propagation. Such an imaging system will collect these light rays and propagate them to an optical sensor array. At any given plane in the system, the light propagation may be described by a 4D function corresponding to the intensity of the light at each transverse position (x, y) and the direction of propagation described by angles (u, v). This 4D representation of the propagation through the imaging system is known as the light field[2], and knowledge of the complete light field enables computational reconstruction of objects in the image space, for example digital refocusing to different focal planes, novel view rendering, depth estimation and synthetic aperture photography (see refs. [1,3,4] and references therein). Indeed, the co-development of novel optical systems and computational photography is

opening up exciting new frontiers in imaging science, well beyond the traditional camera and its biological inspiration, the eye.

Various schemes for light field imaging have been proposed and demonstrated. For example, one may employ an array of microlenses at the focal plane of the imaging lens, in conjunction with a 2D detector array, to obtain the angular information necessary to reconstruct the light field. The first prototype was implemented in 2005, and imaging devices of this type are called plenoptic cameras[3,5]. However, this approach inherently trades off spatial resolution for angular resolution. Schemes incorporating programmable apertures[6,7], multi-camera arrays[8] and other mask-based designs[9–11] attempt to solve the issue of low resolution, but they suffer from signal-to-noise limitations or require multiple images to be acquired and are thus unsuitable for recording dynamic scenes. Implementing a system with full sensor resolution, high signal-to-noise ratio (SNR) and real-time light field imaging remains a challenging problem.

This Letter proposes and demonstrates a novel photodetector concept that enables light field imaging in a single exposure, based on recent breakthroughs in optoelectronic materials[12–17]. Unlike previous approaches, we have fundamentally changed the detector design itself rather than relying on complex optical systems to achieve light field imaging. The key technology at the centre of the proposed scheme is a transparent photodetector that enables multiple sensor arrays to be stacked along the path of the light rays. Such an optical configuration retains directional information, so the light field can be computationally reconstructed from a single exposure. Figure 1a presents a conceptual design of a focal stack light field imaging system. A bundle of light rays emitted from an object point is collected by a lens and focused on a conventional 2D sensor array. At several intermediate positions, additional 2D sensor arrays are positioned. The sensors must of course absorb some of the light to obtain the intensity profile in each (x, y) plane, but pass sufficient light that several (perhaps up to 10) sensor planes may be positioned in front of the final sensor (a conventional opaque sensor with high pixel density).

The necessary first step towards developing a focal stack light field camera is to demonstrate a high-responsivity, transparent, single-pixel photodetector. The detector should ideally only absorb 5–10% of the incident light, yet achieve a responsivity of $>1\,A\,W^{-1}$. We have accomplished this goal using a modification of the graphene-based photodetector described previously in ref. [12].

To achieve high transparency, we fabricated all-graphene heterojunction photodetectors by using graphene for all the functional components of the detector, including the transistor channel, light-sensing top layer, gates and interconnects (Fig. 1c). The device was fabricated on a transparent glass substrate by chemical vapour
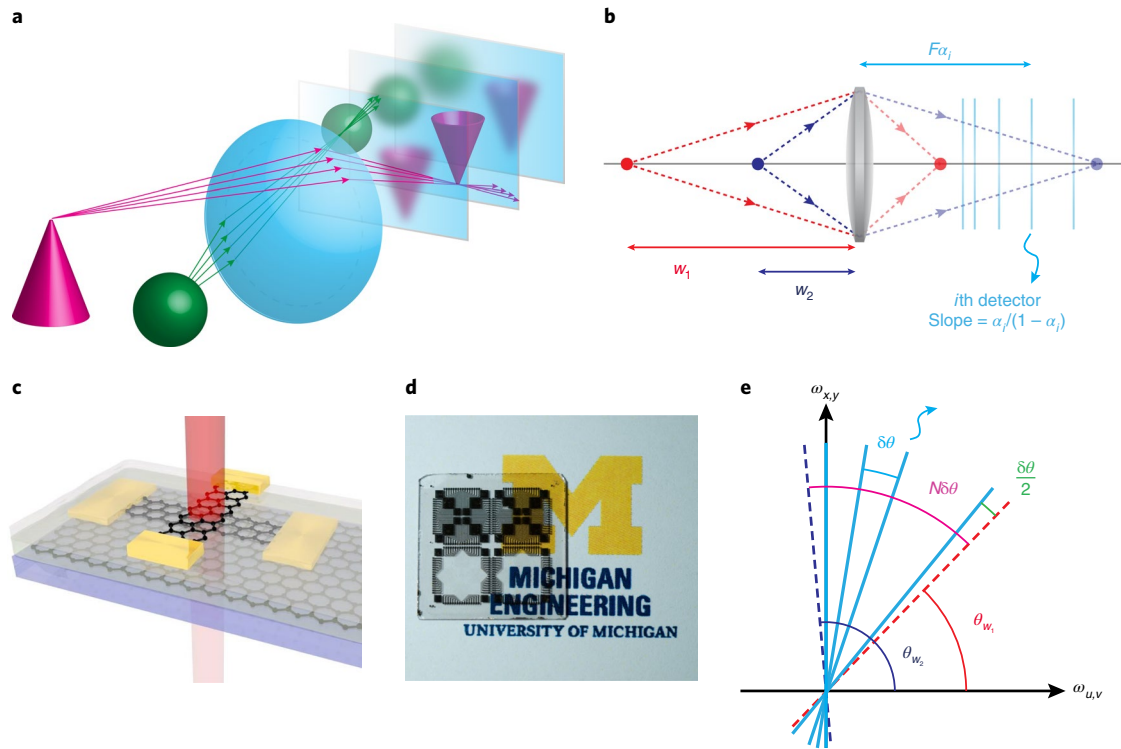
Department of Electrical Engineering, University of Michigan, Ann Arbor, MI, USA. *e-mail: fessler@umich.edu; zzhong@umich.edu; tnorris@umich.edu

Fig. 1 | Light field imaging system enabled by focal stacks of highly transparent photodetectors. a, Conceptual design of a focal stack light field imaging system. **b**, Schematic of the system design with N detectors and a working range of $[w_1, w_2]$ ($w_1 > w_2$). F is the distance between the camera lens and farthest detector. The ith sensor is placed at distance $\alpha_i F$ from the imaging lens and $\alpha_N = 1$. **c**, Schematic of the all-graphene heterojunction photodetector with graphene functioning as the light-absorbing layer, the conducting channel layer and the gate layer, fabricated on a transparent substrate. **d**, An image of the actual fabricated photodetectors, resting on a sheet of paper with the University of Michigan 'M' logo to illustrate the high degree of transparency. Four devices are shown here, with the top two devices having a transparent window size of $200 \times 200\ \mu m^2$ and the bottom two $4.2 \times 4.2\ mm^2$. **e**, Radial slice sampling in the Fourier domain, where the detectors are placed such that their corresponding radial sampling lines have equal angular spacing.

deposition (CVD) of graphene and standard graphene transfer/photolithography processes[12]. Between the bottom gate and middle channel layers we deposited 40 nm Al_2O_3 as the backgate dielectric using atomic layer deposition (ALD), and between the middle channel and top absorption layers we sputtered either 6 nm Ta_2O_5 or 6 nm intrinsic Si as the tunnel barrier. A transparent window is necessary for light rays to pass through the stacked detector arrays in our light field photodetector design. To accomplish this, we eliminated the use of metal within the transparent window by replacing metal interconnects with graphene. We fabricated devices with two different transparent window sizes, $200 \times 200\ \mu m^2$ and $4.2 \times 4.2\ mm^2$; the metal pads for wire bonding were all located outside the transparent windows. Figure 1d shows the actual fabricated all-graphene transparent photodetectors with both designs of transparency windows, resting on a printed University of Michigan 'M' logo to illustrate the high transparency of the devices. Although each graphene layer absorbs 2.3% of light in the visible and near-infrared region, the major part of the area within the transparent windows was either not covered by graphene or only covered by one layer of graphene. We have shown previously that a similar three-layer all-graphene device on a transparent substrate can achieve an overall transparency of ~95%[13].

We characterized the graphene transparent photodetector performance in the near-infrared using femtosecond pulse lasers with wavelengths of 1.2 μm and 2.4 μm. We first considered the detector's photoresponse at 1.2 μm excitation. Figure 2a shows the gate-dependent transistor current under different laser intensities. The Dirac point gate voltages shift to the left under increasing incident

light intensity, consistent with our previous observation on non-transparent graphene detectors[12]. The device photocurrent response can be extracted by subtracting the dark current curve, as shown in Fig. 2b. We then calculated the photoresponsivity of the transparent detector; the prototype device achieved a photoresponsivity greater than 20 A W⁻¹ under 1 V bias voltage, comparable to previous non-transparent graphene photodetectors. We next considered the device photoresponse at the longer excitation wavelength of 2.4 μm. The photocurrent dependence on the gate voltage and light intensity show similar behaviour, as shown in Fig. 2c,d. The detector photoresponsivity at 2.4 μm drops to 0.45 A W⁻¹ at 1 V bias in our prototype device. Nevertheless, the photoresponsivity of the transparent detector is still comparable to the previous non-transparent design[12]. Overall, the all-graphene transparent photodetectors demonstrate a photoresponsivity that is comparable to non-transparent graphene detectors, thus enabling the use of the transparent detector for light field photodetection and imaging.

Proceeding from demonstrating a single pixel to a dense 2D sensor array will be a major technological step, so a full implementation of light field imaging using transparent focal stacks is not yet possible. Nevertheless, we can demonstrate a key operating principle of focal stack light field imaging, particularly the ability to section (image a particular focal plane) or perform optical ranging. To accomplish this we used two single-pixel transparent graphene photodetectors. The set-up and experimental scheme are presented in Fig. 3a,c.

For this demonstration, the set-up comprises a 100 mm focal length front imaging lens and two transparent graphene detectors
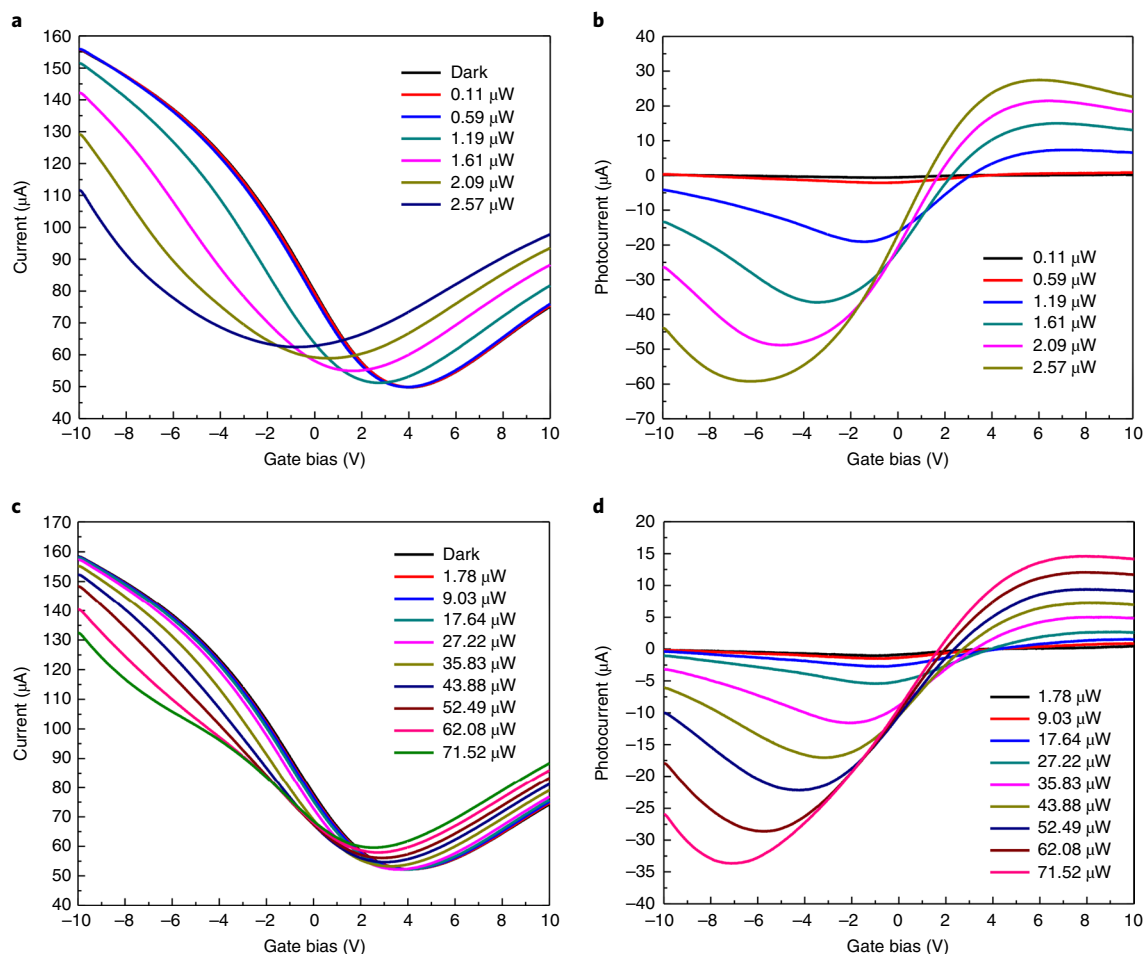
Fig. 2 | Photoresponse characterization of the all-graphene heterojunction photodetectors. a,b, $I–V_g$ transfer curves (**a**) and photocurrents (**b**) under different laser power illumination at 1.2 μm wavelength. **c,d,** $I–V_g$ transfer curves (**c**) and photocurrents (**d**) under different laser power illumination at 2.4 μm wavelength.

separated by 2 mm. Behind the focal stack, a microscope provides separate confirmation that the test object is perfectly imaged at the centre of each graphene detector pixel. The test object is a point source formed by illuminating a 30 μm pinhole with a focused 632 nm HeNe laser. The point source, the centre of the imaging lens, the two single-pixel detectors and the optical axis of the microscope system are well aligned on the same optical axis (referred to as the z axis). When the point source is very far from the imaging lens, the real image is completely out of focus on both graphene detector sheets. We then move the point source towards the imaging lens with a linear stage. At some point, the real image of the point source is perfectly focused on the front detector sheet (1st focal sheet in Fig. 3b) while staying out of focus on the back detector sheet; the signal current of the front detector reaches its maximum, as it is proportional to the optical intensity illuminated, corresponding to the leftmost data points in Fig. 3b. As we continue to move the point source toward the imaging lens, the intensity of the real image decreases on the front detector sheet and increases on the back detector. At some point, the point source is sharply imaged on the back detector sheet (2nd focal sheet in Fig. 3b) while staying out of focus on the front detector sheet, corresponding to the rightmost data points in Fig. 3b. The curves in Fig. 3b demonstrate optical ranging or sectioning—with knowledge of the lens focal length and sensor positions, one can determine the longitudinal position of the object. In the figure, we normalize the object position to the Rayleigh range of the imaging system. In principle, high axial

resolution can be obtained by imaging with a lens with short focal length (for example, a microscope objective). The axial resolution is determined simply by the Rayleigh range/depth of field of the optic used in the imaging system, as in a traditional confocal microscope.

Knowing the intensity profile along the z axis, we can extract the depth information of the scene from the data. The design can be readily extended to a 3D focal stack when high-quality transparent detector arrays become available. We note that the development of such transparent detector arrays will require transparent interconnects as well as high-performance transparent detector elements. Adopting graphene as the interconnects, in addition to the active sensing elements, could have added benefits in minimizing unwanted optical effects, for example by minimizing scattering and speckle from coherent light sources.

The experimental demonstration of high-performance transparent graphene detectors and the successful demonstration of 1D ranging with these single-pixel detectors lays the ground work for implementing these individual detectors into imaging arrays and systems. To this end, it is crucial to investigate, using computational tools, any conceptual barriers for implementing the focal stack imaging system with transparent arrays for 3D imaging and reconstruction of light field images. To motivate the development of stacked transparent detector arrays and show how such an optical system can enable light field imaging, we consider here a model of the imaging system and examine a suitable reconstruction algorithm using a synthetic scene. The general strategy of the algorithm
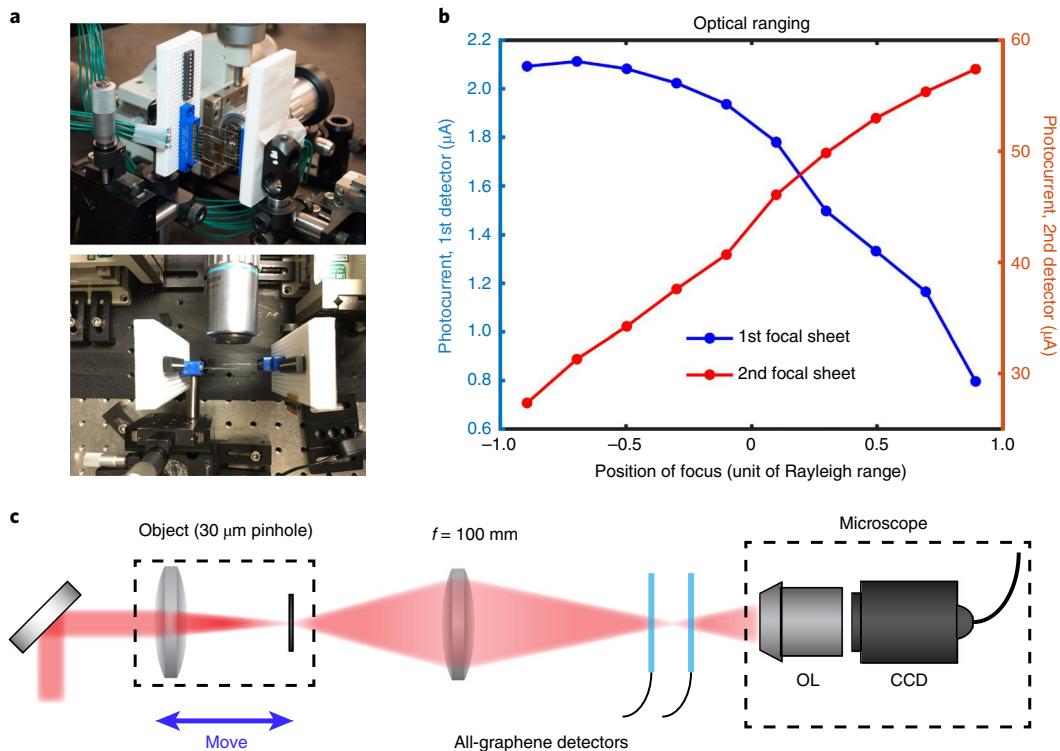
Fig. 3 | Experimental demonstration of depth ranging using a double stack of transparent graphene detectors. a, Cross-sectional (top) and top-down (bottom) views of two all-graphene transparent photodetectors stacked along the light propagation direction. **b**, Demonstration of optical ranging. Photocurrent is measured for the two all-graphene photodetectors (separated by 2 mm) at various positions of the point object. **c**, Schematic of the optical imaging set-up for the demonstration of 1D ranging using the light field detection system. OL, objective lens; CCD, charge-coupled device.

is as follows. We first construct a forward model of the optical system that describes the intensity profile measured in each plane of the focal stack in terms of the light field propagating through the system[18,19]. The reconstruction process then corresponds to solving an inverse problem to determine the unknown object of an initial light field from the captured focal stack data.

Under the paraxial approximation, light field transport in an optical system can be traced with optical ray-transfer matrices. With the in-camera parameterization (detailed in the Supplementary Information), the scene light field ℓ_{scene} and the in-camera light field ℓ_{cam} are related through a linear transformation H. The sensor image formed is the integration of ℓ_{cam} over the aperture plane, and, after discretization, the integration is replaced by summation. Discrete image formation on the sensor can be understood as a 4D convolution followed by a 4D-to-2D slicing, where the kernel shape and the direction of slicing both depend on the linear transformation H (refs. [4,19]). Hence, the complete process of forming a 3D focal stack from a 4D light field can be modelled by a linear operation of the form $f = A\ell + n$, where A is the forward model, n is the detection noise and f is the resulting measured focal stack. The problem of reconstructing the scene light field from the focal stack data can then be posed as the least-square minimization problem of $\hat{\ell} = \min_{\ell} \| f - A\ell \|_2^2$, which can be solved efficiently with linear iterative methods such as gradient descent.

An important consideration in designing a focal stack of transparent sensor arrays is the specific arrangement of the sensor planes. The analysis leading to an optimal configuration is performed in the Fourier domain, where a photograph is simply a 2D slice of the 4D light field. This is formally stated as the Fourier slice theorem under the Lambertian scene and full aperture assumptions[20]. Now consider a focal stack camera with an objective lens followed by N transparent detector arrays, where the distance

between the lens and the ith detector array is F_i. We first specify a depth range of the scene from the main lens, $[w_1, w_2]$ ($w_1 < w_2$), over which the reconstruction of the scene information is desired (Fig. 1b). A direct consequence of the Fourier slice theorem is that the ith focal stack sheet will radially sample ω_x, ω_u and the ω_y, ω_v Fourier domains along a line with slope $\alpha_i/(1 - \alpha_i)$, where $\alpha_i = F_i/F_N$ (see Supplementary Information). Without prior information about the scene, we can arrange the N focal stack sheet locations such that they correspond to the radial sampling lines that have equal angular spacing within the sector $[\theta_{w_1}, \theta_{w_2}]$ defined by the designed working range, illustrated in Fig. 1e. Using the thin lens formula and trigonometric relations, θ_{w_1} and θ_{w_2} can be represented as functions of w_1, w_2, f and F_N. With w_1, w_2, f and N being design parameters, F_N can be numerically solved from the constraint of $\delta\theta/2 = \theta_{w_2} - \pi/2$, where $\delta\theta = (\theta_{w_2} - \theta_{w_1})/N$. Once F_N becomes known, θ_{w_1}, θ_{w_2} and $\delta\theta$ are directly derived and the detector array positions $F_i = \alpha_i F_N$ can be calculated with the condition $\alpha_i/(1 - \alpha_i) = \tan[(\theta_{w_1} + \delta\theta/2) + (i - 1)\delta\theta]$.

The effectiveness of the sampling strategy is demonstrated with a set-up consisting of two patterned disks perpendicular to the optical axis. The two disks are placed at different depths (467 and 687.1 mm) from the camera lens, transversely displaced such that one disk occludes part of the other. In this numerical example, the working range of the camera is set between 30 cm and 3 m, and the focal length of the imaging lens is 50 mm. The five transparent detectors are placed at 51.65, 53.31, 55.07, 56.95 and 58.95 mm based on the above design rules. Figure 4a displays the second, third and fourth captured focal stack images (from the nearest to the farthest), with scaling corrected for visual comparison. Note that the back disk appears to be identical on the second and third focal sheet, the front disk appears to be identical on the third and fourth focal sheet, and they appear to be equally blurry on the third focal sheet.
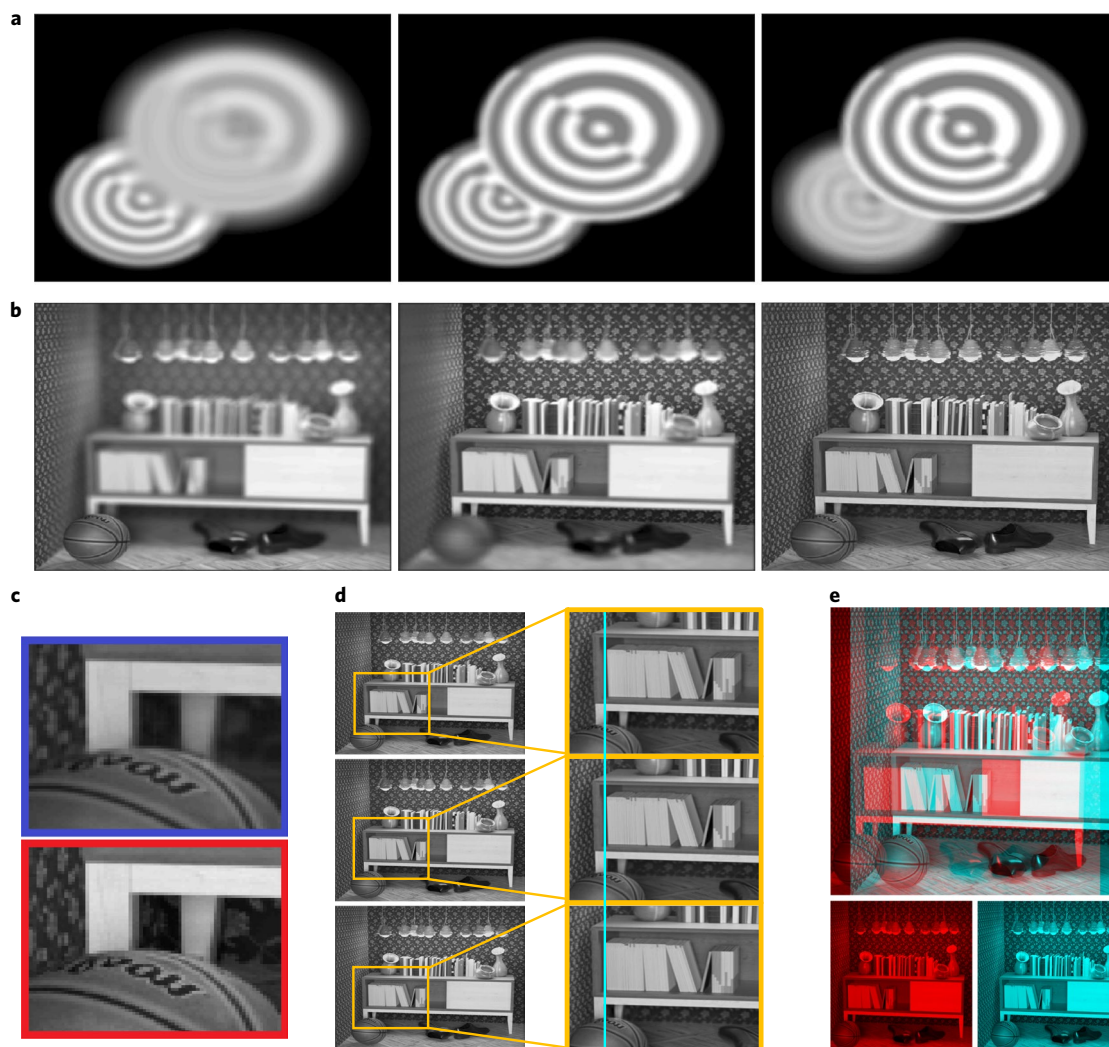
Fig. 4 | Computational reconstruction of the 4D light field using focal stack data. a, The second, third and fourth focal stack images (left to right) from a scene of two patterned disks (transversely displaced) located at different depths. The five transparent detectors were deployed based on Fourier slice sampling. **b**, Light field reconstruction of the photorealistic scene, showing the farthest sensor film (left) and nearest sensor film (centre) of the focal stack images, as well as the all-in-focus image from the reconstructed light field data (right). **c**, Enlarged views at the occlusion region of the all-in-focus image from the reconstructed data (blue) and the ground truth data (red). **d**, The perspective shifts, where the enlarged views show the disparities within the yellow boxes. Top: left viewpoint. Middle: central viewpoint. Bottom: right viewpoint. **e**, Top: a stereo image, constructed from sub-aperture images at different viewpoints with a baseline of 10.67 cm. Bottom left: left viewpoint. Bottom right: right viewpoint.

This is a manifestation of the dispersion of the spectral energy of the back disk along the spectral line that has equal angular spacing between the two spectral sampling lines determined by the second and third focal sheets.

Now we consider the light field reconstruction of a photorealistic scene, where the scene light field is adapted from ref. [21] and the scene depth ranges from 1.97 to 2.56 m from the camera lens. The five sensor arrays are arranged according to the Fourier sampling design rule, and the light field reconstruction from these focal stack data was performed by the iteration $\ell^{i+1} = \ell^i - 2\alpha A^T(A\ell^i - \mathit{f})$, where α is a tunable parameter representing the step size and superscript i indicates the iteration number. Figure 4b displays the first (left) and fifth (centre) focal stack images as well as the reconstructed all-in-focus image $\ell[m, n, p_c, q_c]$, extracted from the subset of the reconstructed light field at the central view (p_c, q_c) (right). Apart from the blurring of some spatial details, the reconstruction result matches well the ground truth image, and no obvious artefact is found at the occlusion region (Fig. 4c). On the other hand, the reconstruction

of the light field corresponds to the full reconstruction of the sub-aperture images, that is, estimates of pinhole images from all different viewpoints. Novel views aside from the central viewpoint are readily accessible, and Fig. 4d illustrates perspective shifts from the reconstructed left, central and right viewpoints. One direct application is multi-view stereo, where the depth can be perceived visually with disparity, as Fig. 4e illustrates.

In conclusion, this Letter proposes a novel light field imaging scheme that consists of a camera lens followed by multiple transparent sensor arrays stacked along the path of the light rays, enabling the simultaneous capture of 2D images under different focusing conditions. The key enabling technology is the highly transparent graphene photodetector, where graphene is used both as the photoconductive gain material and the circuit interconnects. A proof-of-concept single-pixel focal stack light field camera was built and its key operating principle to perform optical ranging was demonstrated experimentally. To demonstrate how light field imaging can be performed using focal stacks once transparent detector arrays

are available, we generated focal stack data for a synthetic photo-realistic scene, with the transparent detector arrays deployed based on Fourier slice sampling. A simulation of 4D light field reconstruction was then performed, where all-in-focus image extraction and multi-view stereo were demonstrated. Physically and computationally, this work opens up a new approach to the realization of novel single-exposure optical systems with detectors at different focal planes. This work focuses on the essential elements for monochromatic light field imaging; specific applications may require different algorithms or different optical designs appropriate to the desired field of view, projection angle, depth of focus, correction on image distortion or colour information. Furthermore, the performance of both the individual sensor and the optical system, including detection linearity, spectral range, dynamic range and responsivity, can be greatly enhanced by the rapid advances in 2D material quality, dielectric and interface defect control, and new sensor designs utilizing other types of 2D semiconductor. This work only represents the beginning of a new frontier in 2D material-based optical systems and optoelectronic applications.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41566-019-0567-3.

References

1. Levoy, M. Light fields and computational imaging. *IEEE Computer* **39**, 46–55 (2006).
2. Levoy, M. & Hanrahan, P. Light field rendering. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive* Techniques (SIGGRAPH) 31–42 (ACM, 1996).
3. Ng, R. et al. *Light Field Photography with a Hand-held Plenoptic Camera* Technical Report CTSR 2005-02 (Stanford University, 2005).
4. Blocker, C. J., Chun, I. Y. & Fessler, J. A. Low-rank plus sparse tensor models for light-field reconstruction from focal stack data. In *Proceedings of 2018 IEEE Image, Video, and Multimedia Signal Processing Workshop (IVMSP)* 1–5 (IEEE, 2018).
5. Georgiev, T., Yu, Z., Lumsdaine, A. & Goma, S. Lytro camera technology: theory, algorithms, performance analysis. *Proc. SPIE* **8667**, 1J (2013).
6. Liang, C.-K., Lin, T.-H., Wong, B.-Y., Liu, C. & Chen, H. H. Programmable aperture photography: multiplexed light field acquisition. *ACM Trans. Graphics* **27**, 55 (2008).
7. Nagahara, H., Zhou, C., Watanabe, T., Ishiguro, H. & Nayar, S. K. Programmable aperture camera using LCoS. In *Proceedings of 2010 IEEE European Conference on Computer Vision (ECCV)* 337–350 (IEEE, 2010).
8. Venkataraman, K. et al. PiCam: an ultra-thin high performance monolithic camera array. *ACM Trans. Graphics* **32**, 166 (2013).
9. Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A. & Tumblin, J. Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Trans. Graphics* **26**, 69 (2007).
10. Xu, Z., Ke, J. & Lam, E. High-resolution lightfield photography using two masks. *Opt. Express* **20**, 10971–10983 (2012).
11. Lin, X., Suo, J., Wetzstein, G., Dai, Q. & Raskar, R. Coded focal stack photography. In *Proceedings of 2013 IEEE International Conference on Computational Photography (ICCP)* 1–9 (IEEE, 2013).
12. Liu, C.-H., Chang, Y.-C., Norris, T. B. & Zhong, Z. Graphene photodetectors with ultra-broadband and high responsivity at room temperature. *Nat. Nanotechnol.* **9**, 273–278 (2014).
13. Lee, S., Lee, K., Liu, C.-H., Kulkarni, G. S. & Zhong, Z. Flexible and transparent all-graphene circuits for quaternary digital modulations. *Nat. Commun.* **3**, 1018 (2012).
14. Sun, Z. et al. Generalized self-assembly of scalable two-dimensional transition metal oxide nanosheets. *Nat. Commun.* **5**, 3813 (2014).
15. Hao, X. et al. High detectivity and transparent few-layer MoS_2/glassy-graphene heterostructure photodetectors. *Adv. Mater.* **30**, 1706561 (2018).
16. Konstantatos, G. et al. Hybrid graphene–quantum dot phototransistors with ultrahigh gain. *Nat. Nanotechnol.* **7**, 363–368 (2012).
17. Xia, F. et al. Photocurrent imaging and efficient photon detection in a graphene transistor. *Nano Lett.* **9**, 1039–1044 (2009).
18. Levin, A. & Durand, F. Linear view synthesis using a dimensionality gap light field prior. In *Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* 1831–1838 (IEEE, 2010).
19. Nien, H. *Model-based X-ray CT Image and Light Field Reconstruction using Variable Splitting Methods*. Thesis, Univ. Michigan (2014).
20. Ng, R. Fourier slice photography. *ACM Trans. Graphics* **24**, 735–744 (2005).
21. Honauer, K., Johannsen, O., Kondermann, D. & Goldluecke, B. A dataset and evaluation methodology for depth estimation on 4D light fields. In *Proceedings of the Asian Conference on Computer Vision* 19–34 (Springer, 2016).

Author contributions

M.-B.L. and C.-H.L. conducted the optical experiments with help from M.Z. Simulation work was performed by M.-B.L., with help from S.R., H.N. and I.Y.C. I.Y.C. analysed the proposed light field imaging system and provided its sampling property. Device fabrication and measurements were performed by C.-H.L., with help from M.Z. The research was conceived and directed by T.B.N., Z.Z. and J.A.F. All authors contributed to preparation of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41566-019-0567-3.

Correspondence and requests for materials should be addressed to J.A.F., Z.Z. or T.B.N.

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

Ranging and light field imaging with transparent photodetectors

Miao-Bin Lien , Che-Hung Liu, Il Yong Chun , Saiprasad Ravishankar , Hung Nien ,
Minmin Zhou, Jeffrey A. Fessler *, Zhaohui Zhong * and Theodore B. Norris *

Department of Electrical Engineering, University of Michigan, Ann Arbor, MI, USA. *e-mail: fessler@umich.edu; zzhong@umich.edu; tnorris@umich.edu

SUPPLEMENTARY INFORMATION

Materials and Methods

The graphene films used in this work were grown by chemical vapor deposition (CVD) on copper foil or purchased through commercial suppliers. The all-graphene transparent photodetectors were fabricated using the same layer transfer techniques described in ref [12], using a transparent glass substrate, and repeatedly transfer/pattern graphene layers three times using conventional photolithography processes (see Fig. S1 below). The three layers of graphene serve as bottom gate, middle channel, and top absorption layer, respectively. Between the bottom gate and middle channel, we deposited 40 nm Al_2O_3 using atomic layer deposition (ALD); between the middle channel and top absorption layer, we sputtered 6 nm Ta_2O_5 as the tunnel barrier. The detectors were situated within a transparent window on the device to allow unabsorbed light to pass through for light field sensing. Within the transparent windows, no metal was used and all electrical interconnects were fabricated with patterned graphene. Outside the transparent window, metal was used for wire-bonding leads and connections to the leads. We fabricated devices with two different transparent window sizes, 200×200 μm^2 and 4.2×4.2 mm^2.
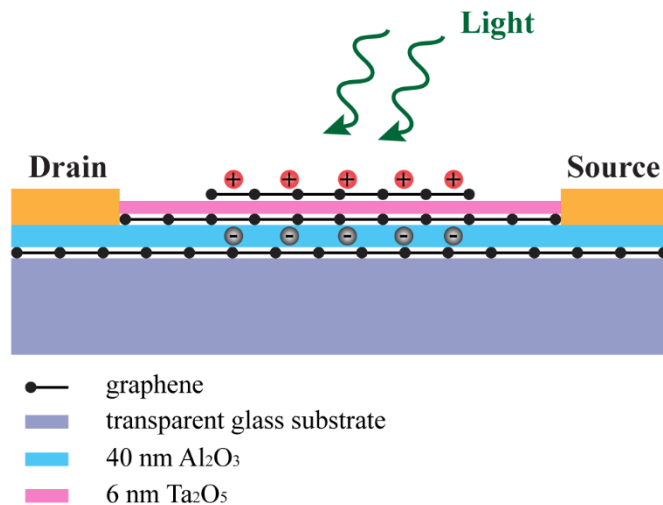

Fig. S1. Schematic of the all-graphene heterojunction photodetector.

Camera Forward Model

Here we present the formulation of the imaging process on a detector array in 2D (position x and angle u) and then extend the expressions to 4D (position x,y and angles u,v).
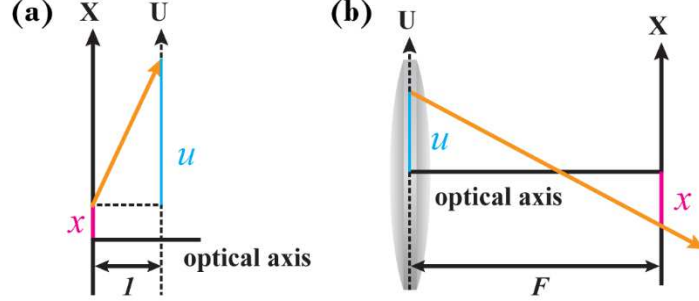

Fig. S2. Light field parameterizations. (a) plane-plane parameterization (b) in-camera parameterization.

For free space propagation over a distance d, the initial (2D) light field (ℓ) and the final light field (ℓ') are related through (see Fig. S2)

$$\ell'\left(\begin{bmatrix}x\\u\end{bmatrix}\right) = \ell\left(T_d\begin{bmatrix}x\\u\end{bmatrix}\right), \text{ where } T_d = \begin{bmatrix}1 & -d\\0 & 1\end{bmatrix}.$$

Under the paraxial approximation, the effect of the lens refraction with focal length f on the light field can be expressed by

$$\ell'\left(\begin{bmatrix}x\\u\end{bmatrix}\right) = \ell\left(R_f\begin{bmatrix}x\\u\end{bmatrix}\right), \text{ where } R_f = \begin{bmatrix}1 & 0\\\frac{1}{f} & 1\end{bmatrix}.$$

Now consider the process of scene light field transport onto the sensor plane in a camera. Assume a linear scene corresponding to a 1D object, with a scene intensity using plane-plane parameterization. The light field first travels a distance z to a lens with focal length f and is then imaged onto a 1D sensor that is distance F behind the lens. In the absence of any aperture, the light field on the sensor film can be traced by the serial application of optical transfer matrices:

$$\ell_{\text{sensor}}\left(\begin{bmatrix}x\\u\end{bmatrix}\right) = \ell_{\text{scene}}\left(C_{zfF}\begin{bmatrix}x\\u\end{bmatrix}\right), \text{ where } C_{zfF} = T_z R_f T_F = \begin{bmatrix}1-\frac{z}{f} & zF\left(\frac{1}{f}-\frac{1}{z}-\frac{1}{F}\right)\\\frac{1}{f} & 1-\frac{F}{f}\end{bmatrix}.$$

When a finite size aperture is present, the in-camera parameterization is introduced to rectify the coordinates (see Fig. S2), making dimension x on the sensor plane (spatial) and dimension u on the aperture plane (directional) so that the image formation $i(x)$ on the sensor plane is the integration of the camera light field over the aperture plane:

$$\mathcal{I}(x) = \int_{\text{aperture}} \ell_{\text{cam}}(x, u) du$$

With change of variables, the relationship between the in-camera and the plane-plane parameterization can be derived as ref [22]:

$$\ell_{\text{cam}}\left(\begin{bmatrix} x \\ u \end{bmatrix}\right) = \frac{1}{F} \ell_{\text{sensor}}\left(\boldsymbol{P}_F \begin{bmatrix} x \\ u \end{bmatrix}\right), \text{ where } \boldsymbol{P}_F = \begin{bmatrix} 1 & 0 \\ \frac{1}{F} & -\frac{1}{F} \end{bmatrix}.$$

The prefactor $1/F$ originates from radiance conservation, which is unimportant in this analysis and will be omitted hereafter for simplicity. With the re-parameterization, the resulting light field in the camera coordinates now becomes

$$\ell_{\text{cam}}\left(\begin{bmatrix} x \\ u \end{bmatrix}\right) = \ell_{\text{scene}}\left(\boldsymbol{H} \begin{bmatrix} x \\ u \end{bmatrix}\right), \text{ where } \boldsymbol{H} = \boldsymbol{C}_{zfF}\boldsymbol{P}_F = \begin{bmatrix} -\frac{z}{F} & 1-\frac{z}{f}+\frac{z}{F} \\ \frac{1}{F} & \frac{1}{f}-\frac{1}{F} \end{bmatrix}.$$

Now consider a focal stack camera with N transparent detector planes. Let ℓ_N denote the camera light field on the Nth (last) detector plane. We express ℓ_i, the light field of the ith film, in terms of ℓ_N:

$$\ell_i\left(\begin{bmatrix} x \\ u \end{bmatrix}\right) = \ell_{\text{scene}}\left(\boldsymbol{H}_i \begin{bmatrix} x \\ u \end{bmatrix}\right) = \ell_N\left(\boldsymbol{H}_N^{-1}\boldsymbol{H}_i \begin{bmatrix} x \\ u \end{bmatrix}\right) = \ell_N\left(\boldsymbol{H}_i^F \begin{bmatrix} x \\ u \end{bmatrix}\right),$$

where $\boldsymbol{H}_i^F = \boldsymbol{H}_N^{-1}\boldsymbol{H}_i = \begin{bmatrix} \frac{F_N}{F_i} & 1-\frac{F_N}{F_i} \\ 0 & 1 \end{bmatrix}$ is the refocusing transformation.

The ith focal stack image is $\mathcal{I}_i(\boldsymbol{x}) = \int \ell_N\left(\boldsymbol{H}_i^F \begin{bmatrix} x \\ u \end{bmatrix}\right) du$, where the integration is over the aperture. In linear algebra notation, this model is approximated by (see Sec. Model-based Reconstruction)

$$\mathcal{I}_i = \mathcal{A}_i(\ell_N)$$

for some linear operator $\mathcal{A}_i$. Explicitly, the focal stack now becomes

$$[\mathcal{I}_1; \dots \mathcal{I}_i, ; \dots \mathcal{I}_N] = [\mathcal{A}_1\ell_N; \dots \mathcal{A}_i\ell_N; \dots \ell_N]$$

(where the semicolon denotes vertical stacking as in, *e.g.*, Matlab) and can be cast into the following form:

$$f\!\!\!f = \mathcal{A}(\ell),$$

where $f\!\!\!f = [\mathcal{I}_1; \dots \mathcal{I}_i; \dots \mathcal{I}_N]$, $\mathcal{A} = [\mathcal{A}_1; \dots \mathcal{A}_i; \dots \mathcal{A}_N]$ and $\ell = \ell_N$ is the light field on the last detector plane.

The formulation can be directly generalized to the 3D space (4D light field):

$\ell_{\text{cam}}\left(\begin{bmatrix} x \\ u \end{bmatrix}, \begin{bmatrix} y \\ v \end{bmatrix}\right) = \ell_{\text{scene}}\left(H\begin{bmatrix} x \\ u \end{bmatrix}, H\begin{bmatrix} y \\ v \end{bmatrix}\right)$, where the coordinate transformation matrix H

is defined as above. This can be rewritten in the form of more concise notations:

$$\ell_{\text{cam}}(x, u) = \ell_{\text{scene}}(Hx, Hu)$$

where the bold characters $x = \begin{bmatrix} x \\ y \end{bmatrix}$ and $u = \begin{bmatrix} u \\ v \end{bmatrix}$ represent the 2D spatial and angular

vectors respectively. With the presence of a (circular) aperture, this is modified as

$$\ell_{\text{cam}}(x, u) = \ell_{\text{scene}}(Hx, Hu)B_u(x, u)$$

where $B_u(x, u) = \text{rect}(\, |u|/D\,)$ and D is the diameter of the aperture.

Fourier Slice Sampling

Since a photograph can be understood as a 2D slice of the 4D light field, we analyze the proposed light-field imaging system in the Fourier domain. This is stated formally in the following theorem that is a direct consequence of the Fourier slice photography theorem [20] based on Lambertian scene and full aperture assumptions:

Theorem: Fourier slice sampling with focal stack

The measurement at the ith detector is given by

$$\mathcal{I}_i(x, y) = \frac{\gamma^{i-1}}{F^2} \mathcal{F}_{2D}^{-1}\left\{ S_i\left\{ \mathcal{F}_{4D}\{\ell(x, y, u, v)\} \right\} \right\}, \quad i = 1, \dots, N,$$

where $\mathcal{F}_{4D}\{\ell(x, y, u, v)\} = L(\omega_x, \omega_y, \omega_u, \omega_v)$ and $\mathcal{F}_{2D}\{\ell(x, y)\} = L(\omega_x, \omega_y)$ are the 4D and 2D Fourier transforms respectively. The slicing operator $S_i\{\cdot\}$ is defined by

$$S_i\{W\}(\omega_x, \omega_y) := W(\alpha_i \omega_x, \alpha_i \omega_y, (1 - \alpha_i)\omega_x, (1 - \alpha_i)\omega_y)$$

where F is the distance between the camera lens and the Nth (furthest) detector; $\alpha_i F$ corresponds to the distance between the camera lens and the ith detector (i.e., $\alpha_N = 1$); $\gamma \in [0,1]$ is the transparency of the detectors; $\omega_x, \omega_y, \omega_u, \omega_v \in \mathbb{R}$ are spatial frequencies and N is the number of detectors.

The theorem suggests that the number of 4D Fourier samples of the light field increases as the number of detectors N increases, thereby predicting the improvement of reconstruction quality with N. A direct consequence of the theorem is that the ith focal stack sheet radially samples the $\omega_x - \omega_u$ and the $\omega_y - \omega_v$ Fourier domains along the line with slope $\alpha_i/(1 - \alpha_i)$. Now suppose we are to design a light field camera of size F (distance between the last focal stack sheet and the camera lens) and the working range is between w_1 and w_2 ($w_1 > w_2$) from the imaging lens. Point sources at $w_{1,2}$ will be sharply imaged at $\alpha_{w_{1,2}} F$, which satisfies $\frac{1}{f} = \frac{1}{w_{1,2}} + \frac{1}{\alpha_{w_{1,2}} F}$, see Fig 1(b). They would confine a sector between θ_{w_1} and θ_{w_2} in the $\omega_x - \omega_u$ and $\omega_y - \omega_v$ domain, where

$$\theta_{w_1} = \tan^{-1} \frac{\alpha_{w_1}}{1 - \alpha_{w_1}}$$

$$\theta_{w_2} = \pi + \tan^{-1} \frac{\alpha_{w_2}}{1 - \alpha_{w_2}}$$

Without prior information about the scene (and therefore the corresponding 4D Fourier spectrum), we can arrange the N focal stack sheet locations such that they correspond to the radial sampling lines that are spaced with equal angular separations of $\delta\theta = \frac{\theta_{w_2} - \theta_{w_1}}{N}$ within the sector confined by the designed working range, see Fig 1(e).

As a numerical example, if our light field camera is expected to work in the range of 30 - 300cm, with a focal length of the imaging lens of 50 mm and the system has 5 transparent detectors, then they should be placed at [51.65, 53.31, 55.07, 56.95, 58.95] mm based on the above design rules.

Model-based Reconstruction

To perform model-based reconstruction, the light field is first discretized as:

$$\ell(x, u) \approx \sum_{\mathbf{m}} \sum_{\mathbf{p}} \ell[\mathbf{m}, \mathbf{p}] \cdot \text{rect}_{\Delta_x}(x - \mathbf{m}\Delta_x)\text{rect}_{\Delta_u}(u - \mathbf{p}\Delta_u)$$

where $\mathbf{m}$ and $\mathbf{p}$ are index vectors of dimension 2 that correspond to x and u respectively, $\text{rect}_w(\cdot)$ denotes the rectangular basis function with the width w, Δ_x and Δ_u denote widths of the rectangular basis function in x- and u- dimensions respectively. We set the simulation parameters as follows: for Fig 4 (a), $D = 6(mm)$, $\Delta_x = 0.01(mm)$ and $\Delta_u = D/10$; for Fig 4(b), $D = 160(mm)$, $\Delta_x = 0.04(mm)$ and $\Delta_u = D/8$.

The integration is then replaced by summation for the image formation of $\ell[\mathbf{m}, \mathbf{p}]$. The image formation $\mathcal{I}^H[\mathbf{m}]$ on the sensor with the transformed light field $\ell^H[\mathbf{m}, \mathbf{p}]$ can be derived as (see, for example, ref [4,19]):

$$\mathcal{I}^H[\mathbf{m}] = (\ell * g)[\mathbf{m}', \mathbf{p}']\big|_{\mathbf{m}'=a\mathbf{m}, \mathbf{p}'=\frac{c\Delta_x}{\Delta_u}\mathbf{m}}, \text{ where}$$

$g[\mathbf{m}, \mathbf{p}] = (s * t)(\mathbf{m}\Delta_x, \mathbf{p}\Delta_u)$, $*$ denotes a convolution operator,

$s(x, u) = \text{rect}_{\Delta_x}(x)\text{rect}_{\Delta_u}(u)$ and $t(x, u) = \frac{1}{|\det(H)|}\text{rect}_{|\det(H)|\Delta_x}(dx - bu)$,

and a, b, c, d are the matrix elements of $\boldsymbol{H} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

This is a linear transformation of 4D convolution followed by a 2D slicing and resampling (for $a \neq 1$). As we have derived the refocusing transformation $\boldsymbol{H}_i^F$, the forward model for a single focal sheet, $\mathcal{A}_i$, can be explicitly defined with the above equations, and the discrete focal stack images can finally be cast in the linear closed form:

$$\mathbf{f} = \mathcal{A}\ell,$$

where $\mathbf{f}[\mathbf{m}] = \left[\mathcal{I}^{H_1^F}[\mathbf{m}]; \dots \mathcal{I}^{H_i^F}[\mathbf{m}]; \dots \mathcal{I}^{H_N^F}[\mathbf{m}]\right]$ is the discrete focal stack, $\ell[\mathbf{m}, \mathbf{p}]$ is the discrete light field on the last film, and $\mathcal{A} = [\mathcal{A}_1; \dots \mathcal{A}_i; \dots \mathcal{A}_N]$ is the forward model where $\mathcal{A}_i$ is defined as above.

Note however, the $\mathcal{A}_i$ matrices other than $\mathcal{A}_N$ may require resampling since objects are magnified by different factors on different detectors. Practically focal stack images are first upscaled to the same field of view (FOV) as the reference (last) detector, and then we solve for an equivalent synthetic camera system that consists of a main lens with N synthetic focal lengths, which has the transformation of

$$\boldsymbol{H}_i^f = \begin{bmatrix} 1 & \dfrac{F_N}{f_i} - \dfrac{F_N}{f} \\ 0 & 1 \end{bmatrix}$$

for each synthetic focal length where $f_i = \dfrac{1}{\frac{1}{f} + \frac{1}{F_N} - \frac{1}{F_i}}$ to link the two systems. In this simpler form of refocusing transformation ($a = 1$ always), the forward and backward projections can be implemented without resampling and computed efficiently with 2D FFTs on the 2D slice; in addition, the convolution kernel can be precomputed and stored in the memory.

The problem of reconstructing the scene light field from the focal stack data can then be posed as the least-squares minimization problem:

$$\hat{\ell} = \min_{\ell} \| f - \mathcal{A}\ell \|_2^2$$

which can be solved with linear iterative methods such as gradient descent:

$$\ell^{i+1} = \ell^i - 2\gamma \mathcal{A}^{\mathrm{T}}(\mathcal{A}\ell^i - f)$$

where γ is a tunable parameter representing the step size. Further improvements can be expected from a development of proper light field regularizations such as 4D total variation, tensor-based model [4], or learning-based methods.

The time complexity is discussed as follow. Let N_{xy} denote the spatial resolution on each sensor and N_{uv} denotes the angular resolution on the aperture. Recall the forward/backward projection of the camera system is essentially a 4D convolution followed by a 2D slicing, which can be implemented by the 2D FFTs on the 2D slice. The $\mathcal{O}$ complexity is therefore $\mathcal{O} \sim N_{uv}N_{xy} \log N_{xy}$ for the computation of a single sensor. Let F be the number of sensors and n be the number of computation iterations. The overall $\mathcal{O}$ complexity is therefore $\mathcal{O} \sim nFN_{uv}N_{xy} \log N_{xy}$.

Refocusing Error and Number of Films

Given the ground truth light field (ℓ_{true}) and the reconstructed light field (ℓ_{recon}), we can refocus to some position p by the application of an operator F_p, i.e.,

$$i_{true}(p) = F_p \ell_{true}$$

$$i_{recon}(p) = F_p \ell_{recon}$$

where $i_{true}(p)$ and $i_{recon}(p)$ are the 2D refocused images. Define the refocusing error (RE) at p to be:

$$RE(p) = \|i_{recon}(p) - i_{true}(p)\| / \|i_{true}(p)\|,$$

where $\|\cdot\|$ is the Euclidean norm. For a given number of arrays in the stack, N, we are able to plot refocusing error versus p. In addition, we can plot multiple curves of this kind corresponding to different numbers of N. An example is illustrated in Fig. S3 below:
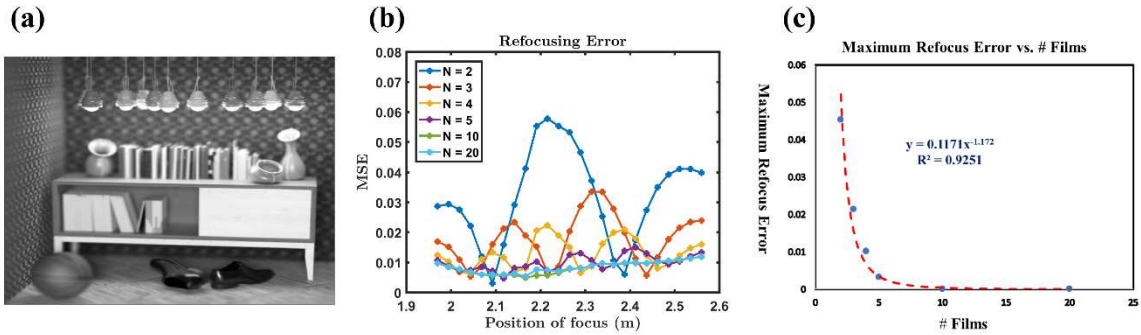

Fig. S3. Refocusing error and number of films. (a) sample scene for the illustration, (b) refocusing error vs. position of focus (c) maximum refocusing error vs. number of films.

While RE versus p is ultimately scene dependent with the iterative reconstruction, we are still able to observe how increasing the number of films efficiently reduces the maximum refocusing error in a power trend (take the value at $N = 20$ as offset):

$$RE \sim a \cdot N^{-b}, \text{ where } 1 < b < 2$$

References

22. Liang, C. K., Shih, Y. C. & Chen, H. H. Light field analysis for modeling image formation. *IEEE Trans. Image Process.* **20**, 446–460 (2011).