Optics Letters

Combining time of flight and photometric stereo imaging for 3D reconstruction of discontinuous scenes

EMMA LE FRANCOIS,^{1,*} ALEXANDER D. GRIFFITHS,¹ JONATHAN J. D. MCKENDRY,¹ HAOCHANG CHEN,² DAVID DAY-UEI LI,³ ROBERT K. HENDERSON,⁴ JOHANNES HERRNSDORF,¹ MARTIN D. DAWSON,^{1,2} AND MICHAEL J. STRAIN¹

¹Institute of Photonics, Department of Physics, University of Strathclyde, Glasgow, G1 1RD, UK

²Fraunhofer Centre for Applied Photonics, Fraunhofer UK Research Ltd, Glasgow G1 1RD, UK

³Faculty of Engineering, Department of Biomedical Engineering, University of Strathclyde, Glasgow G1 1XQ, UK

⁴Institute for Integrated Micro and Nano Systems, School of Engineering, University of Edinburgh, Edinburgh, EH9 3JL, UK

*Corresponding author: emma.le-francois@strath.ac.uk

Received 10 March 2021; revised 21 May 2021; accepted 22 June 2021; posted 23 June 2021 (Doc. ID 424000); published 23 July 2021

Time of flight and photometric stereo are two three-dimensional (3D) imaging techniques with complementary properties, where the former can achieve depth accuracy in discontinuous scenes, and the latter can reconstruct surfaces of objects with fine depth details and high spatial resolution. In this work, we demonstrate the surface reconstruction of complex 3D fields with discontinuity between objects by combining the two imaging methods. Using commercial LEDs, a single-photon avalanche diode camera, and a mobile phone device, high resolution of surface reconstruction is achieved with a RMS error of 6% for an object auto-selected from a scene imaged at a distance of 50 cm. © 2021 Optical Society of America

<https://doi.org/10.1364/OL.424000>

Three-dimensional (3D) imaging is of major interest, as the need to reconstruct and analyze the surrounding environment has become important in different domains including defense and security, robot navigation, autonomous vehicle systems, facial recognition, and surveillance. 3D imaging is becoming well established using a range of different techniques, namely, time-of-flight (ToF) [1–4], light detection and ranging (LiDAR) [5,6] and photometric stereo (PS) imaging [7–9]. A trade-off is usually needed between some 3D imaging requirements such as depth accuracy, spatial resolution, capture speed, or processing time. For high-fidelity results, several techniques may be combined, and we note that two methods in particular, ToF and PS, complement each other particularly well.

ToF relies on the time correlation of a reflected optical signal with the outgoing pulse to obtain a range map of the illuminated scene [1]. Laser-based ToF excels at long range target recognition, because the signal-to-noise ratio limited range can be as much as 45 km [6], and the depth resolution is determined by the timing accuracies of the output pulse and the detector, allowing centimeter-scale depth resolution with light-emitting

diode (LED) sources [3] and sub-centimeter accuracy at kilometer ranges with laser sources [2]. Recent advances in ToF were enabled by single-photon avalanche diode (SPAD) detectors fabricated in a complementary metal oxide semiconductor (CMOS) process with on-chip digital logic [4,10,11]. Each pixel of a SPAD array can extract a distance measurement from a light pulse emitted on a scene and reflected back on the sensor [3,4]. LED-based ToF imaging can achieve high depth accuracy of discontinuous scenes but is limited by the resolution of current single-photon camera systems, or acquisition time of scanning systems [3].

On the other hand, PS imaging is a 3D imaging method that relies on having one fixed camera perspective and different illumination directions [7]. This technique determines the surface normal vectors and surface albedo at each pixel of the captured images assuming a perfectly diffuse (Lambertian) surface of the imaged objects [7]. PS uses conventional imagers and therefore has high spatial resolution in all three dimensions [8]. Although this technique does not deal well with discontinuities or give absolute range, it can easily be applied to moving objects and to real-time imaging [12,13].

We therefore propose and investigate here a hybrid technique of ToF for absolute range and discontinuity in object sets and PS for local details and spatial resolution. Previous work on sensor fusion includes a fusing technique where high-intensity images can be used to improve a ToF camera's low-resolution depth [14] and 3D mapping procedures [15,16] or perform segmentation and tracking [17]. However, our hybrid method focuses on improving our recent work on PS imaging [13] using eye-safe, compact, power-efficient LED emitters and a commercial mobile camera along with a SPAD array for meter-range applications with all the benefits noted previously.

Here we report the 3D reconstruction of different objects placed at different depths within the image scene using a combination of ToF and PS imaging, where ToF enables detection of the boundaries and absolute depth of each object, while PS imaging yields a high resolution surface profile within

these boundaries. A time-correlated single photon-counting (TCSPC) SPAD camera [18] is used with pulsed blue LEDs to obtain a range map of the scene. For PS imaging, four modulated white LEDs illuminate the object while a mobile phone captures frames at 960 fps. The LEDs are modulated at 960 Hz with an orthogonal multiple access carrier scheme such that visible flicker is minimal, and no electronic synchronization is needed between the LEDs and the phone [13]. Our results of the surface reconstruction show a root mean square error (RMSE) of 6.2%.

Time-of-flight and PS imaging. LED illumination for ToF is provided by a blue commercial LED operated in a pulsed manner with respect to a trigger signal from the SPAD sensor. To perform a distance measurement of the scene, the reflected pulses are captured by the SPAD camera. The timing functionality is performed on the SPAD sensor, producing a single time-of-arrival value per frame for each pixel. The TCSPC histogram is constructed and analyzed off-line on the laptop. By taking a reference histogram at a known distance, sub-nanosecond-scale resolution of the time delay can be achieved, resulting in ~ 3.4 cm depth resolution. The histogram build-up program and cross-correlation work are fully explained in our previous work on ToF imaging [3].

The surface reconstruction pipeline of the PS imaging remains the same as in our previous work [13]. Four white LEDs illuminate the scene from four distinct illumination directions and a mobile phone records the frame in a super slow-motion mode. The LEDs are electronically interfaced with a distinct modulation fingerprint for each, so the camera can readily determine which LED generated which image, thus facilitating the 3D reconstruction. Our modulation scheme is self-clocking so the camera operation can be self-synchronized with the LEDs. The recorded stack of frames is demodulated off-line on a laptop and generates four distinct images, one for each LED. The obtained images are processed by a PS algorithm to determine the surface normal components and the albedo. The surface normal components are then integrated within the corresponding mask, previously determined using the range map, which speeds up the surface reconstruction [13,19].

Experimental setup. As illustrated in Fig. 1, our ToF system consists of a 192×128 pixels SPAD sensor, of which we used 167×128 pixels, placed at 80 cm from the sphere. Each pixel is $18.4 \times 9.2 \mu\text{m}^2$ in area and can be operated with TCSPC functionality (see [3,18]). The camera is fitted with a set of optics with a focal length of 8 mm, providing a field of view (FOV) of 16 deg. Upon a trigger signal from the SPAD camera, the blue LED array emits a train of 10 pulses with 11.3 ns pulse

width and a pulse spacing of 30 ns. Both the LED array and SPAD camera are controlled with field-programmable gate array modules.

The PS experimental setup consists of a mobile phone device (Samsung Galaxy 9) mounted on a tripod at a distance of 50 cm with a FOV of 32 deg. Four commercial white LEDs were placed 25 cm away from the object in an X shape and connected to a controller board for LED modulation. A series of geometric solids was 3D printed, namely, a sphere with a 48 mm diameter, a cube 75 mm wide, and a complex shape of a monkey head $130 \times 94.5 \text{ mm}^2$ wide and 79 mm deep. On the setup, the geometric center of the scene is the reference (0,0,0), and the location of the LEDs is determined from this reference point. The LEDs are located at (x, y, z) : LED1 (−14.5, 9, 25), LED2 (−14.5, −5, 25), LED3 (14.5, 9, 25), and LED4 (14.5, −5, 25), all in centimeters. Each LED was modulated with an individual Manchester encoded binary frequency division multiple access carrier signal [13] at an on–off keying rate of 960 b/s, which is above visual flicker recognition. The phone captured frames at a rate of 960 fps for 0.2 s, where the focal distance remains constant. The exposure time is limited by the image memory of the phone; however, a fast frame acquisition means a higher effective frame rate of the 3D image.

One-time calibration. The SPAD camera and the mobile phone have different resolutions and pixel sizes. In this experiment, we proceed via a one-time calibration of the image depth using a screen with a checkerboard pattern at each object location. In real life application, the ToF range map will be used to select the correct calibration from a pre-determined calibration table. The checkerboard pattern can be identified automatically from images taken with both the SPAD camera and the mobile phone to then determine the relative pixel scales and frame alignment. Parameters from this calibration are saved and applied to the mask obtained in the imaging process described below.

Object masks. The processed range map of the scene is shown in Fig. 2(a) where depth ranges from 0 m to 0.7 m. The available size array of the SPAD is 167×128 pixels. The absolute range of the sphere, monkey head, and cube are equal to 20 cm, 30 cm, and 40 cm, respectively. Figure 2(b) shows a cropped image of the scene obtained with the mobile phone. By selecting the distance of each object with the range map, a mask can be created on the mobile phone images to select within which PS imaging can be performed. The corresponding masks of the sphere, the monkey head, and the cube are shown in Figs. 2(c), 2(d), and 2(e), respectively. A centimeter-scale

Fig. 1. (a) Top-down schematic of the experimental setup; (b) picture of the experimental setup.

Fig. 2. (a) Range map obtained with the SPAD camera; (b) picture of the scene from the mobile phone; (c)–(e) respective masks of the sphere, monkey, and cube from the range map superimposed on the mobile phone frame.

resolution from ToF imaging is good enough to mask each object.

Surface normal vectors and 3D reconstruction. Figure 3 displays the surface normal components (N_x , N_y , N_z) and the albedo, obtained after PS processing, of the sphere, the monkey head, and the cube. N_x correctly distinguishes left and right facing surfaces, N_y indicates up and down facing surfaces, and N_z shows depth variations as the camera is facing the objects. The albedo is normalized and is useful in understanding imperfections in the reconstruction, especially on the border of the object.

Figures 4 and 5 plot the 2.5D reconstruction of the sphere, the monkey head, and the cube in a perspective view, a top view, and a rendered view obtained with Blender, a 3D creation open source software. The shapes of all three objects are well reconstructed, and millimeter-range eye details of the monkey head are visible on the rendered view. The standard deviation of the surface reconstruction is determined by the RMSE and normalized RMSE (NRMSE) [9]. According to the RMSE error map in Fig. 4(d), the error is minimal across the center of the sphere and ranges up to 10 mm in the bottom, which is due to the base that creates a shadow in this area. Nonetheless, most of the error remains under 4 mm, and the calculated RMSE over the sphere is equal to 3 mm, which corresponds to a NRMSE of 6.2%. This reconstruction error is comparable to our standalone PS work [13], which means that the auto-ToF masking does not introduce any additional error and handles the discontinuity issue well. However, we can notice some difficulty for the algorithm to properly reconstruct the boundary area that is affected by discontinuities and shadows.

The results demonstrate that ToF can be used as a masking tool to overcome discontinuous issues of the PS method, using illumination with LEDs and CMOS-based image detectors in both cameras. The experimental work focused on demonstrating a simple combination technique that works well without

Fig. 3. Surface normal components and albedo, obtained after running the photometric stereo algorithm, respectively, N_x , N_y , N_z , and albedo: (a)–(d) for the sphere, (e)–(h) for the monkey head, and (i)–(l) for the cube corner.

Fig. 4. 2.5D reconstruction of the sphere: (a) perspective view, (b) rendered view, (c) top view, and (d) RMSE error map.

having to change the PS imaging process. Similarly, PS can be used to improve the resolution of LED-based ToF where depth resolution is limited by the characteristic nanosecond pulse duration, as compared with laser-based ToF that can achieve millimeter-range resolution. With some development, this technique could be used to correct for significant discontinuity errors in PS imaging internal to complex, single objects. In other words, ToF ranging would set the coarse topology of objects, including areas of high gradient in the z direction, and PS could be used to give local detail, with a full image stitched together from these sub-regions. However, for real-life applications, shadows affecting different objects within the scene will affect the determination of the surface normal components, hence the

Fig. 5. 2.5D reconstruction of the monkey head and the cube corner. Monkey head: (a) perspective view, (b) top view, and (c) rendered view. Cube corner: (d) perspective view, (e) top view, and (f) rendered view.

surface reconstruction of an object. For example, Fig. 5 shows that the monkey head is slightly impacted by a shadow from the sphere and results in a lower-resolution reconstruction around the mouth area. Some methods currently exist where shadow areas are being detected and where the PS method is adjusted to the number of illumination directions made available in the affected area [20,21]. These methods should be readily applicable to our hybrid approach, requiring a modest modification of the image processing routine.

An approach to combining ToF and PS imaging for high-resolution surface reconstruction of discontinuous scenes has been demonstrated. A centimeter-scale range map obtained with a SPAD camera is used to mask the object area that is then reconstructed with finer depth details using a previously demonstrated PS imaging technique. The experimental implementation shows that a standard deviation of 3 mm can be achieved for an object at a distance of 50 cm. Complex shapes with millimeter-range details can be reconstructed with this technique. In future work, the method could be easily adapted to a larger scale by increasing the object size, angular spread of the illuminating LEDs, and the working distance used to image the scene.

Funding. QuantIC (EP/T00097X/1); Engineering and Physical Sciences Research Council (EP/M01326X/1, EP/S001751/1).

Acknowledgment. The authors thank the Engineering and Physical Sciences Research Council (EPSRC) and QuantIC, the UK Quantum Technology Hub in Quantum Imaging. J. Herrnsdorf acknowledges funding from EPSRC.

Disclosures. The authors declare no conflicts of interest.

Data Availability. Data underlying the results presented in this Letter are available in [22].

REFERENCES

1. I. Eichhardt, D. Chetverikov, and Z. Jankó, *Mach. Vis. Appl.* **28**, 267 (2017).
2. A. M. Pawlikowska, A. Halimi, R. A. Lamb, and G. S. Buller, *Opt. Express* **25**, 11919 (2017).
3. A. D. Griffiths, H. Chen, D. D.-U. Li, R. K. Henderson, J. Herrnsdorf, M. D. Dawson, and M. J. Strain, *Opt. Express* **27**, 35485 (2019).
4. A. Tontini, L. Gasparini, and M. Perenzoni, *Sensors* **20**, 5203 (2020).
5. M. Edgar, S. Johnson, D. Phillips, and M. Padgett, *Opt. Eng.* **57**, 031304 (2017).
6. Z. P. Li, X. Huang, Y. Cao, B. Wang, Y. H. Li, W. Jin, C. Yu, J. Zhang, Q. Zhang, C. Z. Peng, F. Xu, and J. W. Pan, *Photon. Res.* **8**, 1532 (2019).
7. R. J. Woodham, *Opt. Eng.* **19**, 139 (1980).
8. R. Cipolla, S. Battiato, and G. M. Farinella, *Computer Vision: Detection, Recognition and Reconstruction* (Springer, 2010), Vol. **285**.
9. Y. Zhang, G. M. Gibson, R. Hay, R. W. Bowman, M. J. Padgett, and M. P. Edgar, *Sci. Rep.* **5**, 10909 (2015).
10. C. Zhang, S. Lindner, I. M. Antolovic, M. Wolf, and E. Charbon, *Sensors* **18**, 4016 (2018).
11. K. Morimoto, A. Ardelean, M.-L. Wu, A. C. Ulku, I. M. Antolovic, C. Bruschini, and E. Charbon, *Optica* **7**, 346 (2020).
12. G. Schindler, in *International Symposium on 3D Data Processing, Visualization, and Transmission* (2008), pp. 1–6.
13. E. L. Francois, J. Herrnsdorf, J. J. D. McKendry, L. Broadbent, G. Wright, M. D. Dawson, and M. J. Strain, *Opt. Express* **29**, 1502 (2021).
14. D. B. Lindell, M. O'Toole, and G. Wetzstein, *ACM Trans. Graph.* **37**, 113 (2018).
15. P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox, *Int. J. Robot. Res.* **31**, 647 (2012).
16. B. Haefner, S. Peng, A. Verma, Y. Queau, and D. Cremers, *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 2453 (2019).
17. A. Bleiweiss and M. Werman, *Lect. Notes Comput. Sci.* **5742**, 58 (2009).
18. R. K. Henderson, N. Johnston, F. Mattioli, D. Rocca, H. Chen, D. D.-U. Li, G. Hungerford, R. Hirsch, D. Mcloskey, P. Yip, and D. J. S. Birch, *IEEE J. Solid-State Circuits* **54**, 1907 (2019).
19. J. Ho, J. Lim, M.-H. Yang, and D. Kiriakos, in *European Conference on Computer Vision* (2006), pp. 239–250.
20. Q. Zhang, M. Ye, R. Yang, Y. Matsushita, B. Wilburn, and H. Yu, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2012), pp. 2472–2479.
21. C. Hernández, G. Vogiatzis, and R. Cipolla, in *European Conference on Computer Vision* (2010), pp. 290–303.
22. E. Le Francois, M. Dawson, M. Strain, and J. Herrnsdorf, "Data for: 'Combining time of flight and photometric stereo imaging for surface reconstruction of discontinuous scenes'," University of Strathclyde Glasgow (2021), <https://doi.org/10.15129/efad4754-5aeb-45c0-99f2-bac7ce7858ed>.