



Synchronization-free top-down illumination photometric stereo imaging using light-emitting diodes and a mobile device

EMMA LE FRANCOIS,^{1,*}  JOHANNES HERRNSDORF,¹  JONATHAN J. D. MCKENDRY,¹  LAURENCE BROADBENT,² GLYNN WRIGHT,² MARTIN D. DAWSON,¹  AND MICHAEL J. STRAIN¹ 

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

²*Aralia Systems, Bristol Robotics Laboratory, Bristol BS16 1QY, UK*

**emma.le-francois@strath.ac.uk*

Abstract: Three dimensional reconstruction of objects using a top-down illumination photometric stereo imaging setup and a hand-held mobile phone device is demonstrated. By employing binary encoded modulation of white light-emitting diodes for scene illumination, this method is compatible with standard lighting infrastructure and can be operated without the need for temporal synchronization of the light sources and camera. The three dimensional reconstruction is robust to unmodulated background light. An error of 2.69 mm is reported for an object imaged at a distance of 42 cm and with the dimensions of 48 mm. We also demonstrate the three dimensional reconstruction of a moving object with an effective off-line reconstruction rate of 25 fps.

Published by The Optical Society under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/)

1. Introduction

Photometric stereo (PS) imaging [1] is one of the most common 3D imaging methods for indoor scenarios. It can achieve better resolution than structured illumination imaging [2–4] or state of the art laser scanners [5], offers fast image computation [6], and it can deal with objects in motion and untextured area [7,8]. Compared to stereovision [9,10], only one camera needs to be calibrated, which reduces the computational reconstruction speed, the footprint and the cost [10]. PS imaging relies on having one fixed camera perspective and different illumination directions to image an object in 3D. 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 object [1]. Surface normal components can then be integrated to recover the 3D shape. Most work on PS has been developed combining various methods in addition to PS imaging, such as multi-view PS imaging [11], non-calibrated PS imaging [12] or self-calibrating PS imaging [13]. All the techniques report a good reconstruction accuracy, within millimeter range [11,12,14] and include real-time reconstruction [15].

Even though current work on PS imaging tackles major challenges such as uncalibrated PS imaging [13] and non-Lambertian PS imaging [16–19], most PS methods fail to demonstrate an easily deployable imaging technique that could show imaging applications in already existing building infrastructures. If this were achieved, PS imaging can provide an attractive route to using 3D imaging in industrial settings for process control and robot navigation, in public spaces for security and surveillance applications, and for structural monitoring.

There are two major obstacles that inhibit the widespread use of PS imaging for these purposes, which are the compatibility of the PS specific illumination with indoors or outdoors lighting installations, and the cabling required to synchronize several luminaires with each other and with the camera, which may potentially be mobile. Usually the camera and the luminaires are placed in the same plane, and a particularly common configuration employs four luminaires surrounding the

camera in a top/bottom/left/right or X-shaped configuration [5,8,14]. While such a setup is known to provide high-fidelity imaging results, it is incompatible with an application scenario where the luminaires are installed at the ceiling to provide room lighting, and a wall-mounted or mobile camera views the scene from the side (see Fig. 1(a)). Current PS systems use cables between the camera and the luminaires to enable synchronization, which is an undesired complication when retrofitting to existing lighting fixtures. Use of a WiFi or optically encoded clock signal can be a solution to remove the cabling, though additional infrastructure would be needed to implement this and the transmitters, camera and clock signal must be synchronized. Achieving synchronization using a "self-clocking" Manchester-encoded modulation scheme makes the approach described here easy to use, does not require additional infrastructure and can work in environment where WiFi is not available. Finally, traditional PS imaging often has a strong visual flicker and illumination low duty cycle, which is detrimental for indoors or outdoors lighting.

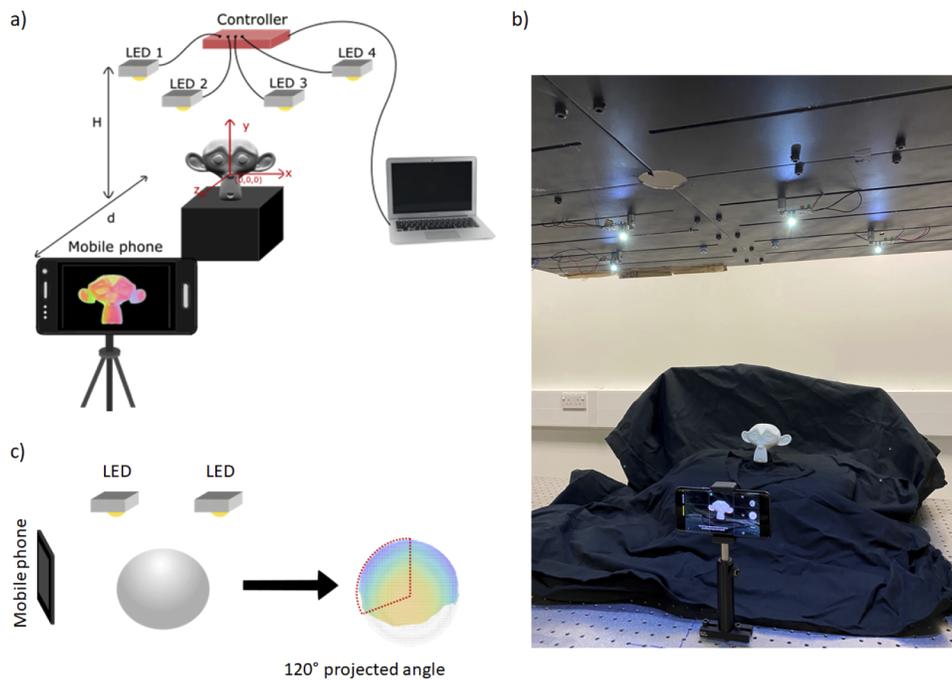


Fig. 1. Top-down illumination setup. a) Schematic of the photometric imaging setup, b) Picture of the photometric imaging setup and c) Schematic of the projected angle.

In this work, we present efforts to make PS imaging synchronization-free, reduce flicker, and demonstrate compatibility with ceiling lighting and both wall-mounted and mobile cameras. In this scenario, PS imaging would coexist with light fidelity (LiFi) networks [20] or visible light positioning (VLP) [21], potentially using the same light-emitting diode (LED) luminaires for all of these functions [22] as well as general lighting.

We demonstrate PS imaging using a hand-held mobile phone camera running at 960 fps and ceiling-mounted LEDs as illustrated in Fig. 1(a) and (b), operating in the presence of additional unmodulated lighting. Four LEDs operated by a controller board were modulated with a bespoke binary multiple access (MA) format, referred to as Manchester-encoded binary frequency division multiple access (MEB-FDMA), that removes the need for synchronization and reduces flicker through Manchester encoding while maintaining a 50% duty cycle. A mobile phone set within the scene acquired a stack of frames at 960 fps, which were then processed to obtain the surface normal components, and finally integrated to obtain the

topography of the object. As the object is static, we do not reconstruct the full 3D object but rather its topography, commonly known as 2.5D reconstruction. For a static 48 mm diameter sphere, we report an root mean square error (RMSE see Eq. (8)) of 2.69 mm at a distance of 40 cm with an angle of reconstruction from the top-down illumination of 120° , which represents 78 % reconstruction of the surface of the sphere. Finally, we also demonstrate a dynamic imaging scheme, using a high-speed camera, where an ellipsoid rotates at 7.5 rotations per minute (RPM) and report an effective off-line 2.5D reconstruction of 25 fps.

2. Orthogonal LED modulation

One key feature in our setup is the removal of the requirement to synchronize LEDs with the camera or among each other, both of which is needed in the time division multiple access (TDMA) that is used in conventional PS imaging. This is achieved through MA, and frequency division multiple access (FDMA) has been used by the authors before to achieve unsynchronized PS imaging [23]. However, FDMA has some drawbacks, in particular strong perceived flicker since some of the LEDs have to be modulated at a fraction of the camera frame rate, and the sinusoidal modulation requires analog control of the LED brightness. Here, we use a bespoke modulation scheme, called Manchester-encoded binary FDMA (MEB-FDMA), which works with direct digital modulation of the LEDs, has significantly reduced flicker compared to FDMA, and keeps the advantage of not having to synchronize sources and camera. A comparison of MEB-FDMA with other MA schemes (FDMA, code division multiple access CDMA [24], space division multiple access SDMA [25], wavelength division multiple access WDMA [26] and TDMA) that could be used for PS imaging is provided in Table 1. MEB-FDMA and some other MA schemes have the additional feature that they enable visible light positioning of receivers within the imaged scene through a relative signal strength approach (Sec. 3.3.4 in [21]).

Table 1. Comparison of MA schemes for N emitters

MA scheme	FDMA	CDMA	MEB-FDMA	SDMA	TDMA	WDMA
Modulation	real-valued	binary OOK	binary OOK	binary OOK	binary OOK	none
Duty cycle	50%	50%	50%	50%	$1/N$	100%
Frame length	$2N$	$2N$	2^{N+1}	$2 \log_2 N$ or $2\sqrt{N}$	N	1 (N fixed to 3)
Computation (flop/frame)	$\sim 10N \log_2 2N$	$4N^2 - N$	$(2^{3N+1} - 1) \sum_{i=1}^N 2^i$	0	0	0
Synchronization required LED \leftrightarrow LED	No	Yes	No	No	Yes	No
Synchronization required LED \leftrightarrow receiver	No	Yes	No	Yes	Yes	No
VLP enabled	3D	3D	3D	2D	3D	3D
PS enabled	Yes	Yes	Yes	No	Yes	Yes
Visual perception	Poor (flicker)	Good	Good	Good	Poor (flicker)	Poor (monochrome luminaires)
Compatible with constant background light	Yes	Yes	Yes	Yes	No	No

2.1. Phase invariant orthogonal modulation

The important property that allows us to use FDMA without having to synchronize LEDs and camera is that if one frequency carrier experiences an arbitrary phase shift due to the lack of synchronization, it still remains orthogonal to the other frequency carriers. We call this property “phase-invariant orthogonality” and introduce it here formally before describing an alternative modulation scheme that shares this property.

Consider N emitters illuminating the scene over n discrete time steps. Then the $N \times n$ signal matrix $s_{i,j} \in \{-1, 1\}$ describes the time-sequence of on/off states of the individual LEDs. Here, $s_{i,j} = +/ -1$ indicates that at time j the i^{th} LED element transmits a binary value of ‘1’/‘0’ in on-off keying (OOK), and after n binary values one 3D image frame is completed. Phase-invariant orthogonality requires that the rows of the matrix s remain orthogonal to each other even if they are time-shifted with respect to each other by an arbitrary phase Δj . Since camera pixels operate as integrating receivers, particularly at high camera frame rates, this requirement can be formalized to Eq. (1).

$$\sum_{j=1}^n s_{i,j} ((1 - \alpha)s_{i',1+(j-1+k)\%n} + \alpha s_{i',1+(j+k)\%n}) = 0 \quad (1)$$

$$\forall i \neq i', k = 1, \dots, n, \alpha \in [0, 1]$$

Here the phase shift between rows i and i' is $\Delta j = k + \alpha$, and $\%$ is the modulo operator. Equation (1) represents the requirement from the experimental layout, however, mathematically it is equivalent to the simpler Eq. (2).

$$\sum_{j=1}^n s_{i,j} s_{i',1+(j-1+k)\%n} = 0 \quad (2)$$

$$\forall i \neq i', k = 1, \dots, n$$

FDMA with square wave carriers is phase invariant orthogonal. If s is phase-invariant orthogonal, then its Manchester-encoded version $s^{(1)}$ given by Eq. (3) - where \otimes is the Kronecker product - is also phase-invariant orthogonal, *i.e.* all the benefits of Manchester encoding are readily available. Matrices of the form $s \otimes \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}$ are also phase-invariant orthogonal.

$$s^{(1)} = s \otimes \begin{bmatrix} -1 & 1 \end{bmatrix} \quad (3)$$

Decoding of phase-invariant orthogonal encoded signals is less trivial than for CDMA schemes with synchronization. Equation (2) effectively means, that the operation of phase-shifting scatters the source signals into orthogonal sub-spaces of \mathbb{R}^n . Therefore, to enable successful decoding, the rows of the matrix $s_{i,j}$ need to be complemented by appropriately chosen orthonormal vectors $e_{k,j}^{(i)}$ that together with the rows of $s_{i,j}$ span all of these sub-spaces.

2.2. Manchester-encoded binary FDMA

We construct the MEB-FDMA carriers by starting with binary-valued square-wave FDMA. In order to be phase-invariant orthogonal over a sampling period T , the frequencies ν_i of the square waves must be in a fixed relationship given by Eq. (4).

$$\nu_i = \frac{p_i}{T} \quad p_i \in \mathbb{N}^+ \quad (4)$$

A convenient choice of the integer values p_i is given by Eq. (5), in which $i = 1, \dots, N$ identifies each LED:

$$p_i = 2^{i-1} \quad (5)$$

This means that the frame length $n^{(0)}$ without Manchester encoding is $n^{(0)} = 2^N$. We then construct a binary FDMA emitter signal $s_{ij}^{(0)}$:

$$s_{ij}^{(0)} = (-1)^{\lceil j/p_i \rceil}, \quad i = 1, \dots, N \quad j = 1, \dots, n^{(0)} \quad (6)$$

When using $s^{(0)}$, individual emitters may have long on and off times, leading to unacceptable visual flicker. Therefore, we use Manchester encoding:

$$s_{i,j} = s_{i,j}^{(0)} \otimes \begin{bmatrix} -1 & 1 \end{bmatrix} = \begin{cases} (-1)^{\lceil j/2^i \rceil}, & j \text{ even} \\ (-1)^{1+\lceil (j+1)/2^i \rceil}, & j \text{ uneven} \end{cases} \quad (7)$$

If the emitters are modulated with $s_{i,j}$ according to Eq. (7) using OOK, then they provide MEB-FDMA. A decoding algorithm for MEB-FDMA and underlying mathematical proofs are given in the appendix.

2.3. Properties of MEB-FDMA

For MEB-FDMA, similar to FDMA, any DC offset can be added to the received signal without affecting the decoding result. This is a prerequisite for applying the scheme to LED illumination since the intensity-modulated LED emission has only positive values. Furthermore, it allows installation of additional lighting fixtures that either do not carry a modulation signal or carry one at a much higher frequency, *e.g.* for LiFi.

Another remarkable property is that the transmitter and receiver can use the same sampling rate. This is surprising because the scheme uses Manchester encoding and the Nyquist theorem requires oversampling by a factor 2 to reliably identify each Manchester encoded bit. However, by requiring the modulation to fulfil the stringent criterion Eq. (1), the scheme was implicitly designed such that not every single Manchester bit needs to be identified individually. This property of the scheme means that the frequency of the LED modulation is the same as the camera frame rate and thus flicker is significantly reduced.

The number of OOK-bits needed for a single frame in MEB-FDMA scales exponentially with the number of emitters. Therefore, this modulation scheme is suitable for modest numbers of modulated emitters illuminating the camera field of view, typically 4-6 emitters in our suggested application.

3. 3D reconstruction process

The process flow of PS imaging in our setup is illustrated in Fig. 2, which comprises MEB-FDMA modulation and demodulation, PS processing and surface normal integration.

3.1. Frame acquisition and surface normal map

The transmitted signal is encoded as a clock signal, hence no trigger signal is needed to start the acquisition. Therefore, the acquisition simply starts when the recording button of the mobile phone is pressed. In practice, the LEDs have a 50% duty-cycle and thanks to the known optical fingerprint of each LED a decoding matrix can be created see [Supplement 1](#). The received stack of images are therefore demodulated using the decoding matrix on each pixel of each image, see Fig. 2. At the end of the demodulation, four images corresponding to the four different illumination directions are retrieved.

The retrieved four images are then processed using established methods [1,14,27] to obtain the surface normal components N_x , N_y , N_z and the albedo A under the assumption of a Lambertian

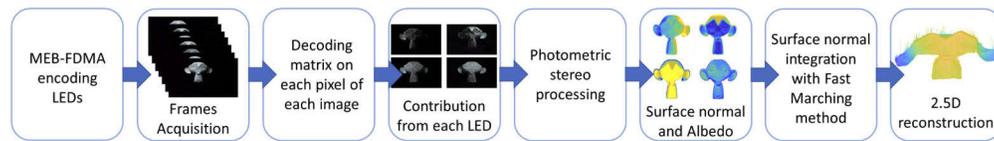


Fig. 2. Acquisition and Reconstruction program pipeline. LEDs are encoded with an MEB-FDMA scheme; the mobile device acquires a stack of images that are demodulated with a decoding matrix; four output images are then retrieved: one for each illumination direction; the photometric stereo processing determines the surface normal components and the albedo; afterwards the surface normals are integrated with a Fast Marching method to then obtain the 2.5D reconstruction of the object.

surface (see Fig. 2). As we are focusing our work on a new modulation scheme, we decided to use a conventional calibrated PS method to determine the surface normal map, hence the coordinates of each LED relative to the position of the object are needed to determine the lighting vectors.

3.2. Fast marching method

The next step of the reconstruction program is to integrate the surface normal vectors to obtain the topography of the object. Surface integration is a well known challenge and there are multiple methods in the literature for addressing it [28–30]. In this work, the surface normal vectors are integrated with the Fast Marching method [31–35] to take advantage of its reconstruction speed. The algorithm was implemented in Matlab and assessed on a data set from Yvain Queau [36], see details in the [Supplement 1](#). The reconstruction process takes a few minutes to run on a desktop PC.

4. Optical acquisition system

As illustrated in Fig. 1, our system consists of a mobile phone device (Samsung Galaxy 9), four white LEDs (Osram OSTAR Stage LE RTUDW S2W) placed on a gantry above the object at a height of $H = 46$ cm, a controller board (Arduino Uno) for the LED modulation and a computer to communicate with the controller board and run the reconstruction program [37]. A series of geometric solids are 3D printed, namely a sphere with a 48 mm diameter, a cube which is 75 mm wide, and a complex shape of a monkey head that is 130×94.5 mm² wide and 79 mm deep. 3D printing ensures that the ground truth shape of the objects is known. On the setup, the geometric center of the object is the reference (0,0,0) and the location of the LEDs is determined from this reference point. The phone and the LEDs are located in two different planes. The phone is in front of the object on the z axis at a distance of $d = 42$ cm from (0,0,0) with a field of view (FOV) of 43 degrees. The LEDs are located at (x,y,z) : LED1 (−27, 42, 10), LED2 (−14, 42, 35), LED3 (14, 42, 35) and LED4 (27, 42, 10), all in cm. The relative position of the LEDs to the camera and object positions have an impact on the extraction of the surface normal vectors. These coordinates are the best fit regarding the FOV of the scene and the object's size. A trade-off was made between the resolution of the reconstruction and the FOV. By placing the mobile phone at 42 cm from the object we can keep a mm-range depth resolution while assuming an orthogonal projection for the determination of the surface normal components. Moreover, a strict alignment of the mobile phone is not necessary in this work, as long as the FOV contains the front of the object then the orientation of the phone will not affect the accuracy of the surface normal components.

Each LED was modulated with an individual MEB-FDMA carrier signal at a on-off keying rate of 960 b/s. The phone captured frames with a resolution of 1280×720 at a rate of 960 fps for 0.2 s, with a black background to simplify image processing. The capture time is limited by the on-device data storage limit.

5. Results and discussion

5.1. Decoded frames

The decoded images of the three objects are displayed in Fig. 3. For the three cases, the light level clearly shows the different illumination directions. The brightness is slightly different depending on the position of the LEDs. This can be explained by the possibly imperfect match between the camera integration time and the MEB-FDMA scheme, i.e. the integration time may be shorter than the frame duration.

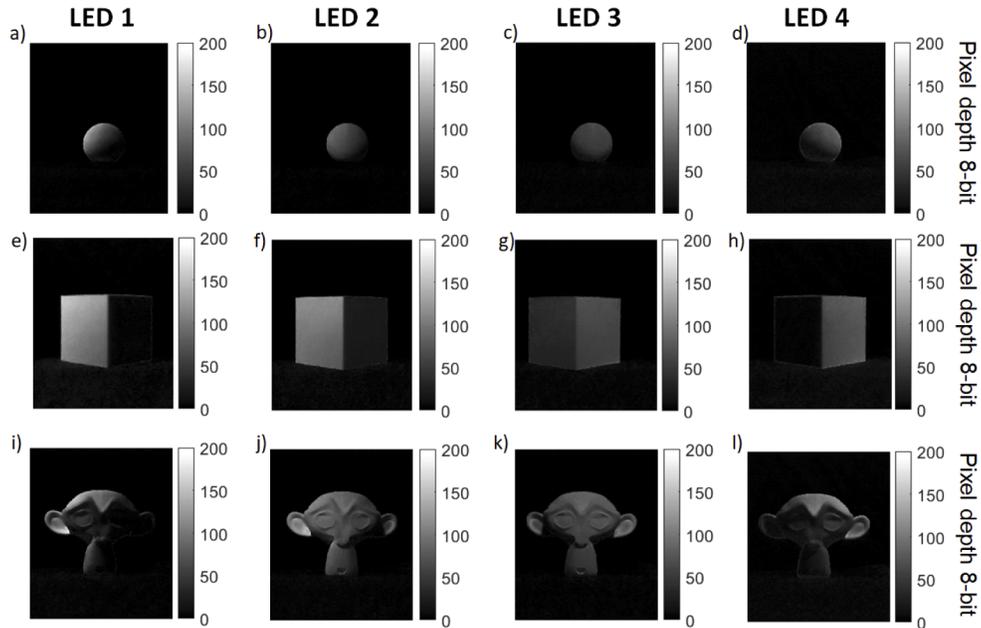


Fig. 3. Decoded images. Obtained after demodulation of the recorded frames for LED1, LED2, LED3 and LED4: a), b), c), d) for the sphere, e), f), g), h) for the cube corner and i), j), k), l) for the monkey head.

5.2. Surface normal vectors

From this set of images, surface normal components (N_x, N_y, N_z) and the albedo were calculated and are displayed in Fig. 4. For N_x , left and right facing surfaces are correctly distinguished as vector components with magnitude are ranging from -1 to 1 . Similarly, N_y indicates up and down facing surfaces correctly, albeit with lower fidelity, and its value range is limited to -0.2 to 1 instead of -1 to 1 . The poorer fidelity on N_y is due to the top-down illumination design as the bottom of each object is not suitably illuminated, which is also visible in the albedo plot. Moreover, as the camera is facing the object, N_z is positive and ranges from 0 to 1 with some variations due to the depth of the object. The albedo is normalized and is useful in understanding imperfections in the reconstruction. We notice that the albedo is more directional for the sphere and the cube corner than for the monkey, which is caused by the slight brightness variations seen in Fig. 3. Importantly, the surface normal components, which are the basis of the topography reconstruction, are observed to be less susceptible to these brightness variations than the albedo.

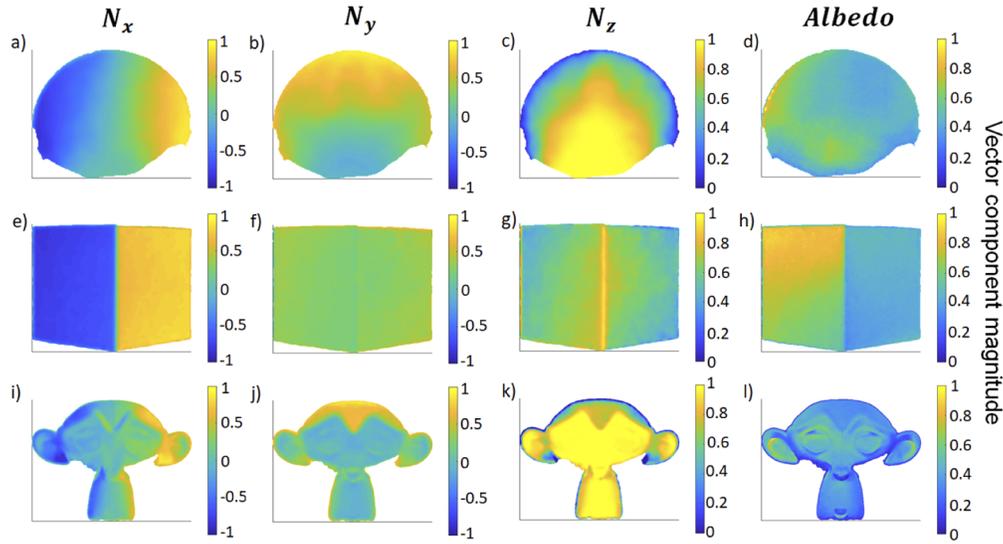


Fig. 4. Surface normal components and albedo. Obtained after running the photometric stereo algorithm, respectively N_x , N_y , N_z and Albedo: a), b), c), d) for the sphere, e), f), g), h) for the cuber corner, i), j), k), l) for the monkey head.

5.3. 3D reconstruction

Figure 5 and Fig. 6, respectively, plot the 2.5D reconstruction of the sphere, the cube corner and the monkey head in a perspective view, as well as a 3D rendered view (Blender). To render the reconstruction on Blender, a camera is set at a distance of 10 cm from the imported 2.5D reconstruction. For the sphere, Figs. 5(a)-(c) show a satisfactory global reconstruction for the top half of the object. The bottom half is poorly reconstructed and is "flat", which is also clearly shown on the rendered view. Because of the lack of information on the negative (downward facing) y axis, 78.4 % of the visible surface is reconstructed. The standard deviation is determined by the root mean square error (RMSE) and the normalised RMSE (NRMSE) which are defined as [14]:

$$RMSE = \sqrt{\left(\frac{1}{n} \sum_{i=1}^n z_i^2\right)}, \quad (8)$$

$$NRMSE = \frac{RMSE}{z_{max} - z_{min}}, \quad (9)$$

where n is the number of data pairs, z_i is the difference between measured depth values (along the z -axis) and reference values, and $(z_{max} - z_{min})$ is the range of measured values. According to the RMSE error map in Fig. 5(d), the most significant error is found at the bottom and on the edge of the sphere, while smaller errors in the top area are related to inaccuracies in N_z . Nonetheless, most of the error stays below 5 mm. Figure 1(c) shows the projected angle that can be retrieved using the top-down illumination setup where an angle of 120° is retrieved. An RMSE of 2.69 mm and an NRMSE of 5.61 % are obtained within the 78.4 % of the surface reconstructed. Despite the lack of information on downward facing facets, both standard deviation errors for the sphere are within the same range as in [14].

For the cube corner, despite the unequal partition of light on the cube and the important gradient variation, the reconstruction can retrieve the shape of the cube corner. Nonetheless, the top view shows that the reconstruction on the edge is deteriorating at the bottom of the object. This is explained with the top-down illumination configuration. Overall, by comparing

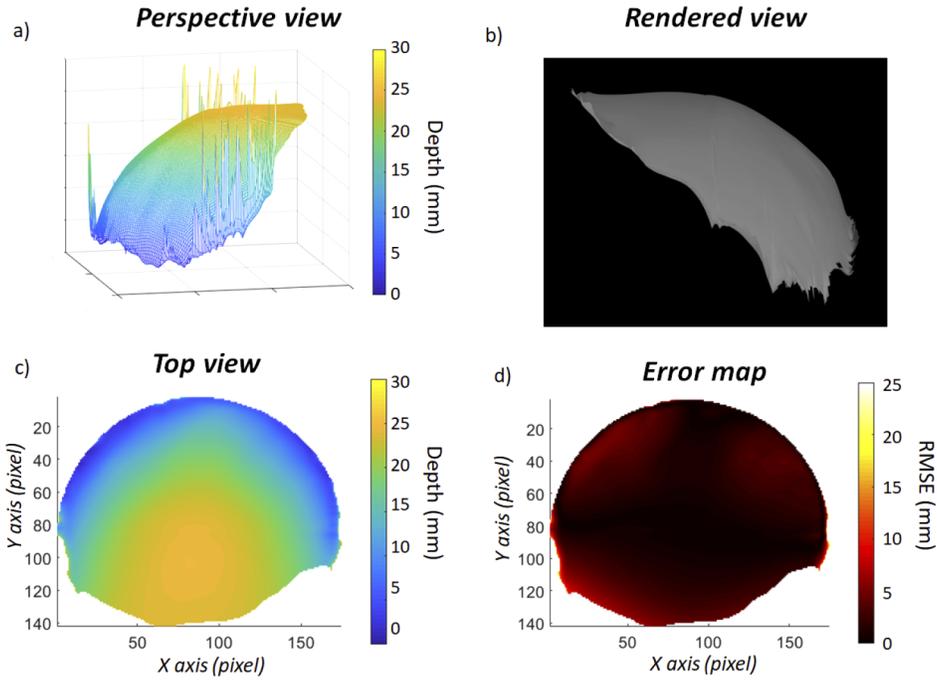


Fig. 5. 2.5D reconstruction of the sphere. a) Perspective view, b) Rendered view, c) Top view, d) RMSE error map.

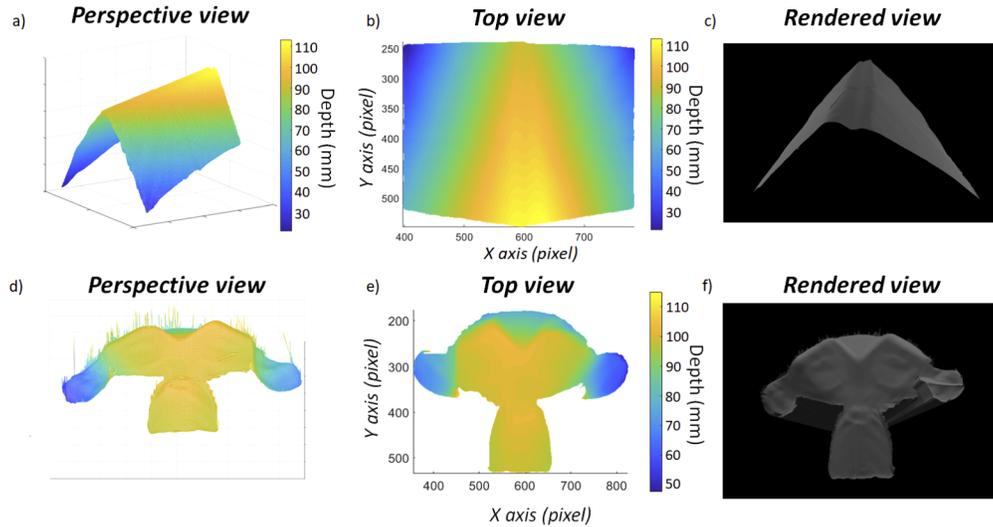


Fig. 6. 2.5D reconstruction of the cube corner and the monkey head. Cube corner: a) Perspective view, b) Top view, c) Rendered view. Monkey head: d) Perspective view, e) Top view, f) Rendered view.

the reconstruction of Figs. 6(a)-(c) with the images of Fig. 3 a good match is obtained. On the rendered view, an indent is observed and can be explained by the position of the starting point of the Fast Marching reconstruction process, from which the gradient values are integrated in a propagating fashion. At points with a strong gradient variation, the propagation will happen with very different gradient values in different directions and can create an indent on the row or the column of the starting point.

Finally, the monkey head has been chosen for its complex features, such as the eyes, the nose and the top of the head. The discontinuity between the face and the ears is a challenge for the Fast Marching algorithm to deal with. Indeed, the 2.5D reconstruction in Figs. 6(d)-(f) shows the shape of the nose, the eyes and also the upper head. However, the shape of the ears is harder to determine and the depth is substantially decreasing, which demonstrates the difficulty of the algorithm for dealing with those discontinuities. To quantify the error, a few features on the monkey face are measured: the horizontal size of an eye is 22 mm, the size of the nose is 1.1 mm and the distance between the eye orbits is 27 mm. After calibration, the same features are measured on the 2.5D reconstruction and we obtained the following measurements: 22.5 mm, 9.3 mm and 29.8 mm respectively. A few millimeters difference can be observed, which is close to the RMSE values obtained for the sphere. The top view reconstruction gives an idea of the percentage of the surface that is correctly reconstructed. The relative position between discontinuous regions is not handled well. However, features within each region are reproduced with good fidelity, such as the ears, on the rendered view. Small depth details, within mm, such as the earlobes and the eyes, are detectable and well reconstructed. Moreover, the distance between the face and the ears is about 50 mm and on the top view of the reconstruction the distance between the two features is also about 50 mm.

The 3D reconstruction relies on the surface normals, therefore most of the error on the reconstructed object topography will be dominated by the error on the surface normal vectors. Whenever N_z values are close to zero, the gradient integration during the Fast Marching process is numerically ill-conditioned. This phenomenon is clearly shown by the artefacts on the different reconstructions in Fig. 6.

5.4. Signal to noise measurement

Our modulation scheme can operate in the presence of additional unmodulated lighting. In order to assess its robustness, the reconstruction of the sphere is tested with different levels of background light in the room. For this experiment, the ceiling light of the room is illuminated and a voltage divider has been added on the LEDs in order to control the brightness and hence modify the signal power. After measuring the optical power of the signal and background light at the object, the SNR, in dB, is determined following Eq. (10):

$$SNR = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right) \quad (10)$$

with P_{signal} the optical power in Watts of the LEDs and P_{noise} the optical power in Watts of the ceiling light.

Figure 7(a) shows a graph of the RMSE and the percentage of surface reconstructed versus the SNR for the sphere in the out-of-plane configuration. The SNR ranges from 0 dB to 5 dB. Across the SNR, the RMSE does not show a specific trend and the error ranges from 4 mm to 6 mm which is acceptable in our range of application. The percentage of surface reconstruction that is achieved over the measured range of SNR does show a dependance on the SNR. Figure 7(b) shows that as the SNR decreases, the reconstruction of the bottom part of the sphere is more and more challenging, which is a consequence of the top-down illumination configuration. This is explained by our reconstruction process. To avoid high error values being incorporated in the final image, a threshold is set 10 mm above the expected reconstruction value. Any values in the

final reconstruction above the threshold are discarded. This means that as the SNR decreases, we can reconstruct less area of the object, but the portion that is reconstructed is not affected by the SNR. Nonetheless, 65 % of the object view can be reconstructed even with an SNR just below 0 dB.

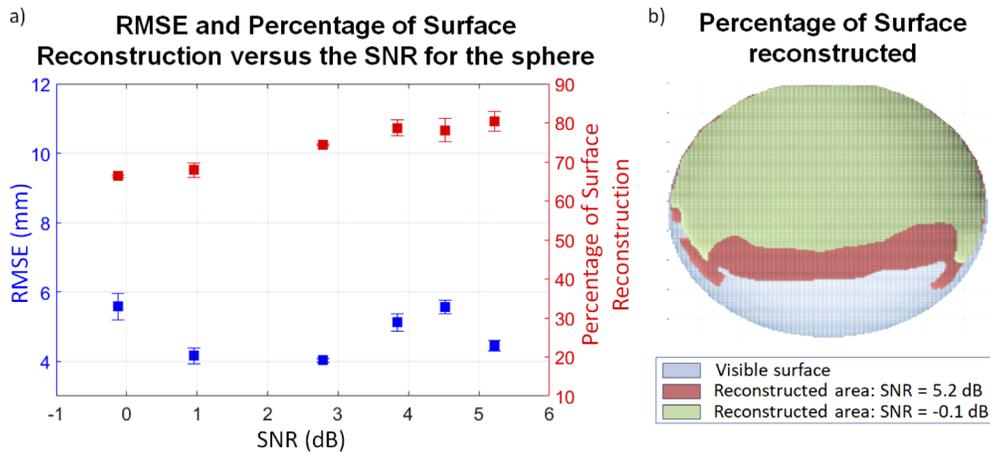


Fig. 7. Signal to noise result. a) Graph plotting the RMSE error and the percentage of the surface reconstructed regarding the signal to noise ratio for the sphere. b) Superposition of the different reconstruction of the sphere depending on the SNR.

5.5. Dynamic imaging

Finally, we reconstructed a moving object using our top-down illumination setup. A stepper motor (RS PRO Hybrid, Permanent Magnet Stepper Motor, 1.8° step) was added in order to rotate a 3D printed ellipsoid which was 100 mm long and 60 mm wide. A high-speed camera (Photron MiniUx100) replaced the mobile phone for its larger video capture memory thus enabling a longer acquisition time. The camera frame rate was set at 1000 fps with a shutter speed of 1 ms and was matched to the LED modulation rate. The stepper motor rotated at a speed of 7.5 RPM. Therefore, the acquisition ran for 8 s and the real-time video of the ellipsoid in motion can be found under the folder 'Dynamic imaging' in [38] (see [Visualization 1](#)). The 3D reconstruction is done off-line with the same reconstruction program pipeline as Fig. 2. The reconstruction requires at least 32 frames for a full reconstruction, and here we chose to record 40 camera frames for each 3D frame to match the motor step duration and to achieve effective full 3D reconstruction at a standard video rate of 25 fps.

Detailed analysis has been carried out on 21 representative 3D video frames, all separated by an angle of 18°. Therefore, our full 3D reconstruction relies on 21 topographies of the ellipsoid out of 200 possible reconstructions. To be able to clearly see the reconstruction and to match it with the display speed of the real-time video of the ellipsoid in motion, we decided to display the surface normal components and the reconstruction at 3 fps. This video display rate is 10 times slower than the effective rate we can achieve but it still represents a real-time 3D reconstruction video matching the object in motion. Two videos, one for the surface normal components and one for 3D reconstruction, can be found under the folder 'Dynamic imaging' in [38] (see [Visualization 2](#) and [Visualization 3](#) respectively).

The surface normals behave very similar to the static situation in that a high fidelity is observed in N_x , while N_y and N_z have lower but still useful fidelity. It is important to notice that the ellipsoid is constantly moving and despite its motion, its boundaries are sharp and well-defined which shows that the imaging rate is adequate.

In the 3D reconstruction video, the 21 reconstructed frames are repeated three times, showing both a color-coded plot of the reconstruction as well as a rendered view. Some errors can be observed at the edge of the ellipsoid which are artefacts caused by poor numerical condition due to $N_z \approx 0$. A flat reconstruction of the bottom of the object is also visible which is expected with the top-down illumination. Similarly to the sphere result with the static configuration, the ellipsoid is not entirely reconstructed. Some of the bottom part is missing which is not only due to the top-down illumination but also to the support piece that has been used to hold the ellipsoid in a tilted position. Nonetheless, the global 2.5D reconstruction of each view is satisfactory and of comparable quality to the static scenes.

6. Conclusion

In this work, we have been able to demonstrate an accurate 2.5D reconstruction of objects with different shape complexity with a RMSE error of 2.69 mm using a new photometric stereo imaging configuration that can readily be employed in conventional room lighting scenarios. We have also shown the 3D reconstruction of a moving object with an off-line effective 3D frame rate of 25 fps. Most importantly, MEB-FDMA encoding enables simple installation through removing the need for synchronization, and as importantly, modulates LEDs above the visual flicker recognition threshold, thus significantly simplifying the deployment, which was not possible before with successively flashed LEDs. Furthermore, we demonstrated that this method can be implemented using commercially available and hand-held mobile devices. Our work on synchronization-free top-down illumination photometric stereo imaging is currently at a proof-of-concept stage. However, future work will be focus on applying the method to digital lighting applications in public areas or industrial applications for surveillance, and also process and structural monitoring.

Funding

Engineering and Physical Sciences Research Council (EP/M01326X/1, EP/S001751/1); QuantIC (EP/T00097X/1); Fraunhofer UK (studentship of Emma Le Francois).

Acknowledgments

The development of the Fast Marching algorithm has been possible thank to Dr Juan Cardelino who shared his work on MathWorks File Exchange. We thank John Leck from Fraunhofer UK for his help on 3D printing the three different objects: sphere, cube and monkey head. We thank Dr Adam Polak for his help on the project. We thank Mark Stonehouse for his help on the experimental setup.

Disclosures

The authors declare no conflicts of interest.

See [Supplement 1](#) for supporting content.

References

1. R. J. Woodham, "Photometric Method For Determining Surface Orientation From Multiple Images," *Opt. Eng.* **19**(1), 139–144 (1980).
2. D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1 (2003).
3. M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan, "Structured light 3D scanning in the presence of global illumination," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* pp. 713–720 (2011).

4. S. M. Haque, A. Chatterjee, and V. M. Govindu, "High quality photometric reconstruction using a depth camera," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* pp. 2283–2290 (2014).
5. Z. Lu, Y. W. Tai, F. Deng, M. Ben-Ezra, and M. S. Brown, "A 3D imaging framework based on high-resolution photometric-stereo and low-resolution depth," *Int. J. Comput. Vis.* **102**(1-3), 18–32 (2013).
6. J. Ackermann and M. Goesele, "A survey of photometric stereo techniques," *Foundations Trends Comput. Graph. Vis.* **9**(3-4), 149–254 (2015).
7. S. Herbot and C. Wöhler, "An introduction to image-based 3D surface reconstruction and a survey of photometric stereo methods," *3D Res.* **2**(3), 4–17 (2011).
8. J. Herrnsdorf, L. Broadbent, G. C. Wright, M. D. Dawson, and M. J. Strain, "Video-Rate Photometric Stereo-Imaging with General Lighting Luminaires," *IEEE Photonics Conference* pp. 483–484 (2017).
9. A. Lipnickas and A. Knys, "A stereovision system for 3-D perception," *Elektronika ir Elektrotechnika* pp. 99–102 (2009).
10. S. E. Ghobadi, "Real time object recognition and tracking using 2D/3D images," Ph.D. thesis, University of Siegen (2010).
11. D. Vlasic, P. Peers, I. Baran, P. Debevec, J. Popović, S. Rusinkiewicz, and W. Matusik, "Dynamic shape capture using multi-view photometric stereo," *ACM Trans. Graph.* **28**(5), 1–11 (2009).
12. E. Salvador-Balaguer, P. Latorre-Carmona, C. Chabert, F. Pla, J. Lancis, and E. Tajahuerce, "Low-cost single-pixel 3D imaging by using an LED array," *Opt. Express* **26**(12), 15623 (2018).
13. G. Chen, K. Han, B. Shi, and Y. M. K.-y. K. Wong, "Self-calibrating Deep Photometric Stereo Networks," *IEEE CVF CVPR* pp. 8731–8739 (2019).
14. Y. Zhang, G. M. Gibson, R. Hay, R. W. Bowman, M. J. Padgett, and M. P. Edgar, "A fast 3D reconstruction system with a low-cost camera accessory," *Sci. Rep.* **5**(1), 10909 (2015).
15. G. Schindler, "Photometric Stereo via Computer Screen Lighting for Real-time Surface Reconstruction, International Symposium on 3D Data Processing," *Visualization and Transmission* pp. 1–6 (2008).
16. G. Chen, K. Han, B. Shi, Y. Matsushita, and K.-Y. K. Wong, "Deep Photometric Stereo for Non-Lambertian Surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8828**, (2020).
17. M. Li, C. yu Diao, D. qing Xu, W. Xing, and D. ming Lu, "A non-Lambertian photometric stereo under perspective projection," *Front. Inf. Technol. Electron. Eng.* **21**(8), 1191–1205 (2020).
18. K. H. Cheng and A. Kumar, "Revisiting outlier rejection approach for non-Lambertian photometric stereo," *IEEE Trans. on Image Process.* **28**(3), 1544–1555 (2019).
19. Q. Zheng, A. Kumar, B. Shi, and G. Pan, "Numerical reflectance compensation for non-Lambertian photometric stereo," *IEEE Trans. on Image Process.* **28**(7), 3177–3191 (2019).
20. H. Haas, L. Yin, Y. Wang, and C. Chen, "What is LiFi?" *J. Lightwave Technol.* **34**(6), 1533–1544 (2016).
21. T.-H. Do and M. Yoo, "An in-depth survey of visible light communication based positioning systems," *Sensors* **16**(5), 678 (2016).
22. J. Herrnsdorf, J. McKendry, M. Stonehouse, L. Broadbent, G. C. Wright, M. D. Dawson, and M. J. Strain, "Lighting as a service that provides simultaneous 3D imaging and optical wireless connectivity," in *2018 IEEE Photonics Conference (IPC)*, (2018), pp. 1–2.
23. J. Herrnsdorf, J. McKendry, M. Stonehouse, L. Broadbent, G. C. Wright, M. D. Dawson, and M. J. Strain, "LED-based photometric stereo-imaging employing frequency-division multiple access," in *2018 IEEE Photonics Conference (IPC)*, (2018), pp. 1–2.
24. J. K. Park, T. G. Woo, M. Kim, and J. T. Kim, "Hadamard Matrix Design for a Low-Cost Indoor Positioning System in Visible Light Communication," *IEEE Photonics J.* **9**(2), 1–10 (2017).
25. J. Herrnsdorf, M. J. Strain, E. Gu, R. K. Henderson, and M. D. Dawson, "Positioning and space-division multiple access enabled by structured illumination with light-emitting diodes," *J. Lightwave Technol.* **35**(12), 2339–2345 (2017).
26. C. Hernández, G. Vogiatzis, G. J. Brostow, B. Stenger, and R. Cipolla, "Non-rigid photometric stereo with colored light," *Proceedings of the IEEE International Conference on Computer Vision* pp. 1–8 (2007).
27. M. Chantler and J. Wu, "Rotation Invariant Classification of 3D Surface Textures using Photometric Stereo and Surface Magnitude Spectra," *BMVC2000* pp. 49.1–49.10 (2013).
28. Y. Quéau, J. D. Durou, and J. F. Aujol, "Normal Integration: A Survey," *J. Math. Imaging Vis.* **60**(4), 576–593 (2018).
29. R. T. Frankot and R. Chellappa, "A Method for Enforcing Integrability in Shape from Shading Algorithms," *IEEE Trans. Pattern Anal. Machine Intell.* **10**(4), 439–451 (1988).
30. K. M. Lee and C.-C. Jay Kuo, "Surface reconstruction from photometric stereo images," *J. Opt. Soc. Am. A* **10**(5), 855 (1993).
31. J. Ho, J. Lim, M.-H. Yang, and D. Kiriegm, "Integrating surface normal vectors using fast marching method," *Computer Vision-ECCV* pp. 239–250 (2006).
32. J. A. Sethian, *Level Set Methods and Fast Marching Methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and material science*, Cambridge University (Cambridge monographs on applied and computational mathematics, 1999).

33. J. A. Sethian, "A fast marching level set method for monotonically advancing fronts," *Proc. Natl. Acad. Sci.* **93**(4), 1591–1595 (1996).
34. R. Sedgewick, *ALGORITHMS* (Addison-Wesley Publishing Company, 1983).
35. E. Ziegel, W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes: The Art of Scientific Computing* (Cambridge University, 1987).
36. Y. Quéau, J. D. Durou, and J. F. Aujol, "Variational Methods for Normal Integration," *J. Math. Imaging Vis.* **60**(4), 609–632 (2018).
37. E. L. Francois, J. Herrnsdorf, L. Broadbent, M. D. Dawson, and M. J. Strain, "Top-down illumination photometric stereo imaging using light-emitting diodes and a mobile device," in *Frontiers in Optics + Laser Science APS/DLS*, (Optical Society of America, 2019), p. JTU3A.106.
38. E. L. Francois, "<https://doi.org/10.15129/44b337ab-9fd9-4983-9079-a0dcfaae1d84>," DOI (2020).