

High-accuracy, high-speed 3D structured light imaging techniques and potential applications to intelligent robotics

**Beiwen Li, Yatong An, David Cappelleri,
Jing Xu & Song Zhang**

**International Journal of Intelligent
Robotics and Applications**

ISSN 2366-5971

Volume 1

Number 1

Int J Intell Robot Appl (2017) 1:86-103

DOI 10.1007/s41315-016-0001-7

Volume 1 • Number 1 • February 2017

International Journal of Intelligent Robotics and Applications

Editor-in-Chief: Kok-Meng Lee
Co-Editors-in-Chief: Honghai Liu, Zhouping Yin

Focused Section: Intelligent Robotics for Rehabilitation
and Human Assistance

Lead Guest Editor: Jingang Yi



Your article is protected by copyright and all rights are held exclusively by Springer Singapore. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

High-accuracy, high-speed 3D structured light imaging techniques and potential applications to intelligent robotics

Beiwen Li¹ · Yatong An¹ · David Cappelleri¹ · Jing Xu² · Song Zhang¹

Received: 15 August 2016 / Accepted: 2 November 2016 / Published online: 9 January 2017
© Springer Singapore 2017

Abstract This paper presents some of the high-accuracy and high-speed structured light 3D imaging methods developed in the optical metrology community. These advanced 3D optical imaging technologies could substantially benefit the intelligent robotics community as another sensing tool. This paper mainly focuses on one special 3D imaging technique: the digital fringe projection (DFP) method because of its numerous advantageous features compared to other 3D optical imaging methods in terms of accuracy, resolution, speed, and flexibility. We will discuss technologies that enabled 3D data acquisition, reconstruction, and display at 30 Hz or higher with over 300,000 measurement points per frame. This paper intends to introduce the DFP technologies to the intelligent robotics community, and casts our perspectives on potential applications for which such sensing methods could be of value.

Keywords 3D optical sensing · 3D optical imaging · Micro robotics · Human robotic interaction · Perception and vision

1 Introduction

Advances in two-dimensional (2D) optical imaging and machine/computer vision have provided integrated smart sensing systems for numerous applications. By adding one more dimension, advanced three-dimensional (3D) optical imaging and vision technologies have much greater impact on scientific research and industrial practices including intelligent robotics.

3D optical imaging techniques do not require surface contact and are thus suitable for remote sensing, where no external force is applied to objects being tested. 3D optical imaging can be broadly divided into two major categories: passive methods and active methods. Passive methods rely solely on naturally depicted information (e.g., texture, color) to determine cues for 3D information recovery, while active methods require active emission of signals to find cues for 3D reconstruction. The most extensively adopted passive methods in the robotics community are probably the stereo vision methods (e.g., Bumblebee). A stereo vision method (Dhond and Aggarwal 1989) recovers 3D information through triangulations by identifying the corresponding pairs between camera images from different perspectives through analysis of texture information. To more accurately and more efficiently determine corresponding pairs, stereo vision methods often calibrate system geometric constraints through epipolar analysis (Scharstein and Szeliski 2002; Hartley and Zisserman 2000). A stereo vision system has the following obvious advantages with its extensive adoption in many different fields: (1) hardware design is compact and cost is low since only two cameras are required for 3D reconstruction; and (2) hardware integration and calibration are fairly straightforward since camera calibration has been well studied. However, the measurement accuracy is

This research was supported by National Science Foundation research Grant Numbers CMMI-1523048 and CMMI-1531048.

✉ Song Zhang
szhang15@purdue.edu

¹ School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA

² Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China

limited if the surface texture is not rich (e.g., variation is slow from one point to another). For applications where accuracy is not a primary concern (e.g., rough navigation under controlled or complex environments), stereo vision is an excellent option. However, for detecting the correspondence pairs between images, the stereo vision method fails if the image is close to being uniform or there are naturally present repetitive textures.

Instead of relying on natural information, active methods require at least an optical emitter to send known signals to the object surface to conquer the fundamental correspondence problem of the stereo vision method. Basically, the correspondence is established by finding the emission information from camera images. Historically, researchers have tried different types of active information including a single dot, a line, even fixed, or a predefined structured pattern(s) (Salvi et al. 2010). Laser scanning technology (Keferstein and Marxer 1998; Sirikasemlert and Tao 2000; Manneberg et al. 2001) typically sweeps a laser dot or a line (e.g., Intel RealSense) across the object surface, and this technique recovers 3D information by finding corresponding points from the laser dots or lines captured by the camera. The accuracy of laser triangulation methods can be high, albeit the speed is slow due to point or line scanning. Therefore, laser triangulation methods are mainly used in application areas where sensing speed is not the primary concern. The first generation of Kinect developed by Microsoft (Zhang 2012) uses a pre-defined unique dot distribution for establishment of stereo correspondence. This technique achieves a reasonably high speed by using dedicated hardware and a limited number of sensing points. As one may recognize, Kinect has a fairly low spatial resolution and depth accuracy, making it good for coarse gesture motion capture for applications like human–computer interactions.

The availability and affordability of digital video projectors has enabled the structured light method, which is nowadays among the most popular 3D imaging methods. The structured light method is very similar to a stereo vision method since both use two devices, but they differ in that the structured light method replaces one of the cameras of a stereo vision system with a digital video projector (Salvi et al. 2010). Due to their flexibility and programmability, structured light systems allow versatile emission patterns to simplify establishment of correspondence. Over the years, researchers have attempted pseudorandom patterns (Morano et al. 1998), speckle patterns (Huang et al. 2013), binary coded structured patterns (Salvi et al. 2010), multi-gray-level patterns (Pan et al. 2004), triangular patterns (Jia et al. 2007), trapezoidal patterns (Huang et al. 2005), and sinusoidal patterns (Zhang 2016), among others. Though all these structured patterns have proven successful for 3D

reconstruction, using phase-shifted sinusoidal structured patterns is overwhelmingly advantageous because such patterns are continuous in both the horizontal and vertical directions, and the correspondence can be established in phase instead of in the intensity domain (Zhang 2010). In the optics community, sinusoidal patterns are often called *fringe patterns* and the structured light method using digital fringe patterns for 3D imaging is often called *digital fringe projection* (DFP).

This paper will explain thoroughly the differences between two of the most extensively adopted binary structured patterns and sinusoidal patterns to help readers recognize the superiority of the DFP method, albeit it is not easy to understand at the very beginning. The main purposes of this paper are to (1) provide a “tutorial” for those who have not had substantial experience in developing 3D imaging system, (2) lay sufficient mathematical foundations for 3D imaging system developments, and (3) cast our perspectives on how high-speed and high-accuracy 3D imaging technologies could be another sensing tool to further advance intelligent robotics.

Section 2 explains the principles of the DFP technique and its differences from stereo vision and binary coding, as well as system calibration and 3D reconstruction. Section 3 shows some representative 3D imaging results captured by DFP systems. Section 4 presents our thoughts on potential applications, and Sect. 5 summarizes this paper.

2 Principle

This section explains the principles of structured light 3D imaging techniques and how to achieve accurate high-spatial and -temporal resolution with the same hardware settings.

2.1 Basics of structured light

Structured light techniques evolved from a well-studied stereo vision method (Dhond and Aggarwal 1989) which imitates a human vision system for 3D information

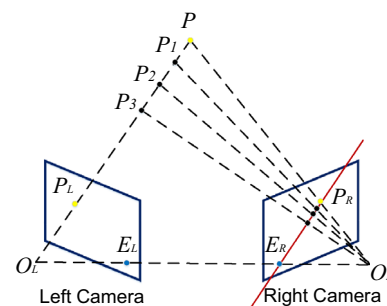


Fig. 1 Schematic of a stereo vision system

recovery. Figure 1 schematically shows a stereo vision system that captures 2D images of a real world scene from two different perspectives. The geometric relationships between a real-world 3D point, P , and its projections on 2D camera image planes, P_L and P_R form a triangle, and thus triangulation is used for the 3D reconstruction. In order to use triangulation, one needs to know (1) the geometric properties of two camera imaging systems, (2) the relationship between these two cameras, and (3) the precise corresponding point on one camera to a point on the other camera. The first two can be established through calibration, which will be discussed in Sect. 2.2. The correspondence establishment is usually not easy solely from two camera images.

To simplify the correspondence establishment problem, the geometric relationship, the so-called *epipolar geometry*, is often used (Scharstein and Szeliski 2002; Hartley and Zisserman 2000). The epipolar geometry essentially constructs the single geometric system with two known focal points O_L and O_R of the lenses to which all image points should converge. For a given point P in a 3D world, its image point P_L together with points O_L and O_R should form a plane, called the *epipolar plane*. The intersection line $\overline{P_R E_R}$ of this epipolar plane with the imaging plane of the righthand camera is called the *epipolar line* (red line). For point P_L , all possible corresponding points on the righthand camera should lie on the epipolar line $\overline{P_R E_R}$. By establishing the epipolar constraint, the correspondence point searching problem becomes 1D instead of 2D, and thus more efficient and potentially more accurate. Yet, even with this epipolar constraint, it is often difficult to find a correspondence point if the object surface texture does not vary drastically locally and appears random globally. For example, if two cameras capture a polished metal plate, images captured from two different cameras do not provide enough cues to establish correspondence from one camera image to the other.

Structured light techniques resolved the correspondence finding problems of stereo vision techniques by replacing one of the cameras with a projector (Salvi et al. 2010). Instead of relying on the natural texture of the object surface, the structured light method uses a projector to project pre-designed structured patterns onto the scanned object, and the correspondence is established by using the actively projected pattern information. Figure 2a illustrates a typical structured light system using a phase-based method, in which the projection unit (D), the image acquisition unit (E), and the object (B) form a triangulation base. The projector illuminates one-dimensional varying encoded stripes onto the object. The object surface distorts the straight stripe lines into curved lines. A camera captures distorted fringe images from another perspective.

Following the same epipolar geometry as shown in Fig. 2b, for a given point P in a 3D world, its projector image point P_L lies on a unique straight stripe line on the projector sensor; on the camera image plane, the corresponding point P_R is found at the intersecting point of the captured curved stripe line with the epipolar line.

2.2 Structured light system calibration

The structured light illumination can be used to ease the difficulty in detecting correspondence points. To recover 3D coordinates from the captured images through triangulation, it is required to calibrate the structured light system. The goal of system calibration is to model the projection from a 3D world point (x^w, y^w, z^w) to its corresponding image point (u, v) on the 2D sensor (e.g., camera's CCD or projector's DMD).

Mathematically, such a projection usually uses a well-known pinhole model if a non-telecentric lens is used (Zhang 2000). Figure 3 schematically shows an imaging system. Practically, the projection can be described as

$$s[u, v, 1]^T = A[R|t][x^w, y^w, z^w, 1]^T. \quad (1)$$

Here, s stands for the scaling factor. $[u, v, 1]^T$ denotes the homogeneous image coordinate on the camera image plane, T is the matrix transpose, and $[R|t]$ represents the extrinsic parameters

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \quad (2)$$

$$t = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}. \quad (3)$$

The extrinsic parameters transform the 3D world coordinate $X^w = (x^w, y^w, z^w)$ to the camera lens coordinate through a 3×3 rotation matrix R and a 3×1 translation vector t . The lens coordinate is then projected to the 2D imaging plane through the intrinsic matrix A ,

$$A = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

f_u and f_v are the effective focal lengths of the camera along the u and v axes; γ models the skew factor of these two axes; and (u_0, v_0) is the principal point (the intersection of the optical axis with the image plane).

The camera calibration procedure is used to estimate the intrinsic and extrinsic parameters to establish geometric relationships from some known geometric points. Nowadays, the most popular camera calibration method is to use

Fig. 2 Principles of the structured light technique. **a** Schematic of a structured light system; **b** correspondence detection through finding the intersecting point between the distorted phase line and the epipolar line

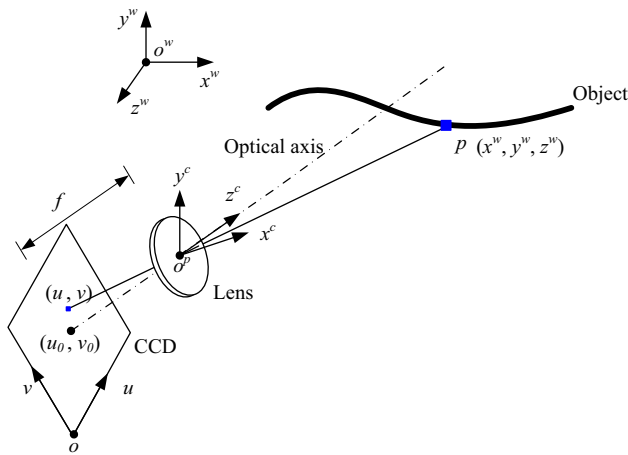
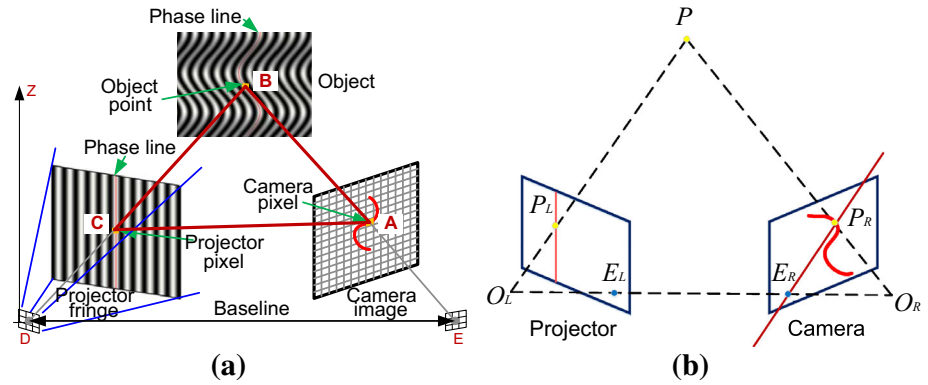


Fig. 3 The pinhole camera model

a 2D calibration target, as developed by Zhang (2000), with known feature points on the plane. After capturing a set of images with different poses, the optimization software algorithm can be applied to estimate the numerical parameters. Zhang's method is popular because of its flexibility and the ease of the calibration target fabrication.

The structured light calibration involves both camera calibration and projector calibration. As discussed in Sect. 2.1, a structured light system is constructed by replacing one of the stereo cameras with a projector. In fact, the projector and the camera share the same pinhole model except that the optics are mutually inverted (the projector projects images instead of capturing images) (Zhang and Huang 2006). Therefore, we can model a structured light system using the following two sets of equations:

$$s^c[u^c, v^c, 1]^T = A^c[R^c|t^c][x^w, y^w, z^w, 1]^T, \quad (5)$$

$$s^p[u^p, v^p, 1]^T = A^p[R^p|t^p][x^w, y^w, z^w, 1]^T. \quad (6)$$

Here, superscript c and p respectively represent the camera and the projector. The calibration of a projector can be quite difficult given that a projector cannot capture images by itself. The enabling technology, developed by Zhang and

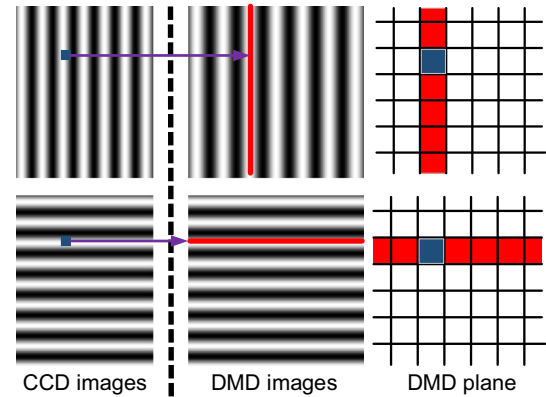


Fig. 4 One-to-one mapping establishment between a camera pixel and a projector pixel by using two sets of structured patterns

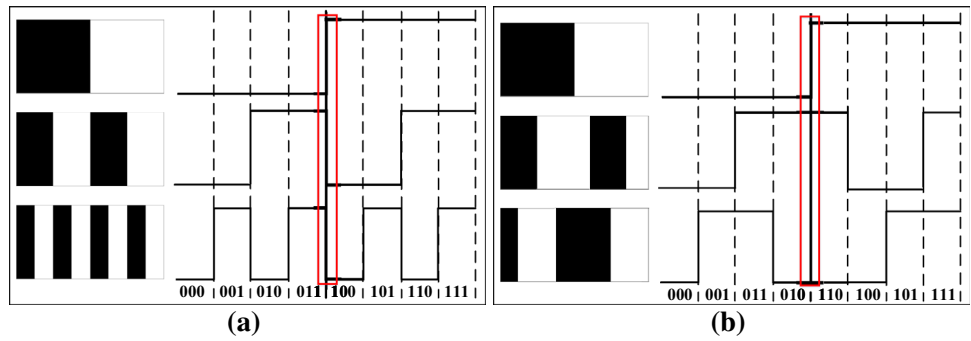
Huang (2006), establishes one-to-one pixel correspondence between the camera and projector sensor by using two sets of structured patterns with different orientations (e.g., one set of horizontal patterns and one of vertical patterns).

Figure 4 illustrates how to establish the one-to-one mapping. The top row shows the scenario in which the camera captures the vertical pattern illuminated by the projector. Suppose we pick a pixel (blue) (u^c, v^c) on the camera image; this changes if the pixel shifts horizontally but does not change if it shifts vertically. The horizontal shifts can be uniquely defined by using a set of continuously coded structured patterns, to be discussed in Sect. 2.4. Since the encoded information is the same from the projector space, by using this set of coded patterns, one can correlate one camera pixel (u^c, v^c) uniquely to one vertical line on the projector sensor u^p ,

$$u^p = f_h(u^c, v^c), \quad (7)$$

where f_h denotes a one-to-many correspondence function of (u^c, v^c) for given coded patterns. Similarly, if the camera captures the horizontal pattern illuminated by the projector as shown in the bottom row of Fig. 4, for the same camera pixel, we can correlate that pixel with a horizontal line v^p on the projector's DMD sensor,

Fig. 5 Two different types of binary coding methods. **a** Three bits simple coding, and **b** corresponding gray coding



$$v^p = f_v(u^c, v^c). \quad (8)$$

Combining these two one-to-many mappings, we can find the unique corresponding point (u^p, v^p) for any camera pixel (u^c, v^c) . By this means, the camera can assist the projector in capturing images, and since the projector can capture images, the structured light system calibration becomes a well-established stereo camera system calibration. The camera and the stereo vision system calibration can be readily carried out by capturing a set of images of the calibration target, and it uses an open-source camera calibration software toolbox (e.g., Matlab or OpenCV).

Once the system is calibrated, we have obtained the intrinsic matrices A^c and A^p and the extrinsic matrices $[R^c|t^c]$ and $[R^p|t^p]$ for the camera and the projector. A simplified model can be obtained by combining the intrinsic and extrinsic matrices:

$$M^c = A^c[R^c|t^c], \quad (9)$$

$$M^p = A^p[R^p|t^p]. \quad (10)$$

The 3D reconstruction process is used to obtain the 3D coordinates (x^w, y^w, z^w) of a real object from each camera pixel (u^c, v^c) . Equations (5) and (6) provide six equations yet with seven unknowns: the 3D coordinates (x^w, y^w, z^w) ; the mapping projector pixel (u^p, v^p) for each camera pixel; and the scaling factors s^c and s^p . To obtain an additional equation to solve for the 3D coordinates, we only need to project one-directional (e.g., horizontal) patterns to establish one-dimensional mapping and use Eq. (7) to provide the last equation to uniquely solve (x^w, y^w, z^w) coordinates for each camera pixel (u^c, v^c) as,

$$\begin{bmatrix} x^w \\ y^w \\ z^w \end{bmatrix} = \begin{bmatrix} m_{11}^c - u^c m_{31}^c m_{12}^c - u^c m_{32}^c m_{13}^c - u^c m_{33}^c \\ m_{21}^c - u^c m_{31}^c m_{22}^c - u^c m_{32}^c m_{23}^c - u^c m_{33}^c \\ m_{11}^p - u^p m_{31}^p m_{12}^p - u^p m_{32}^p m_{13}^p - u^p m_{33}^p \end{bmatrix}^{-1} \times \begin{bmatrix} u^c m_{34}^c - m_{14}^c \\ v^c m_{34}^c - m_{24}^c \\ u^p m_{34}^p - m_{14}^p \end{bmatrix}, \quad (11)$$

where m_{ij}^c and m_{ij}^p denote the matrix parameters in M^c and M^p in the i -th row and j -th column.

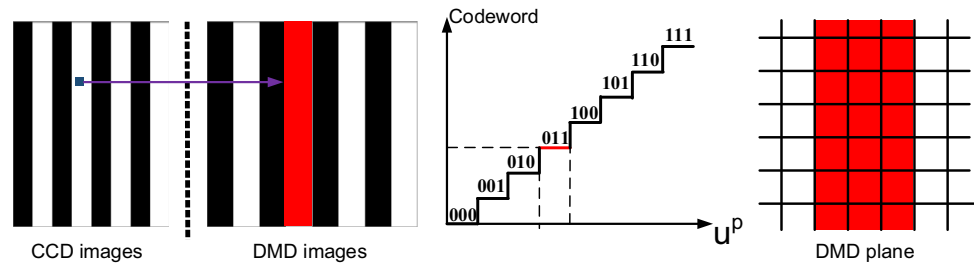
2.3 3D imaging with binary coding methods

As discussed in Sect. 2.1, in order to perform 3D reconstruction through triangulation, at least one-dimensional mapping (or correspondence) is required. Namely, we need to map a point on the camera to a line (or a predefined curve) on the projector. One straightforward method is to assign a unique value to each unit (e.g., a stripe or a line) that varies in one direction. The unique value here is often regarded as the *codeword*. The codeword can be represented by a sequence of black (intensity 0) or white (intensity 1) structured patterns through a certain coding strategy (Salvi et al. 2010). There are two commonly used binary coding methods: simple coding and gray coding.

Figure 5a illustrates a simple coding example. The combination of a sequence of three patterns, as shown on the left of Fig. 5a, produces a unique codeword for each stripe made up of 1s and 0s, (e.g., 000, 001, ...), as shown on the right of Fig. 5a. The projector sequentially projects this set of patterns, and the camera captures the corresponding patterns distorted by the object. If these three captured patterns can be properly binarized (i.e., converting camera grayscale images to 0s and 1s), for each pixel, the sequence of 0s and 1s from these three images forms the codeword which is defined from the projector space. Therefore, by using these images, the one-to-many mapping can be established and thus 3D reconstruction can be carried out.

Gray-coding is another way of encoding information. Figure 5b illustrates an example of using three images to represent the same amount of information as simple coding. The major difference between gray coding and simple coding is that, at a given location, gray coding only allows one bit of codeword status change (e.g., flip from 1 to 0 or 0 to 1 on one pattern), yet the simple coding method does not have such a requirement. For the example illustrated in the red bounding boxes of Fig. 5a and b, simple binary coding has three bit changes while gray coding only has

Fig. 6 1D correspondence detection through binary coding: one camera pixel maps to multiple lines of projector pixels sharing the same binary codeword



one. Fewer changes at a point means less chance of errors, and thus gray coding tends to be more robust for codeword recovery.

The binary coding methods are simple and rather robust since only two binary states are used for any given point, but the achievable spatial resolution is limited to be larger than single camera and projector pixels. This is because the narrowest stripes must be larger than one pixel from the projector space to avoid sampling problems, and each captured stripe width also needs to be larger than one camera pixel to be able to properly find the binary state from the captured image. Figure 6 illustrates that the decoded codewords are not continuous, but discrete with a stair width larger than one pixel. The smallest achievable resolution is the stair width, since no finer correspondence can be precisely established. The difficulty of achieving pixel-level spatial resolution limits the use of binary coding methods for high-resolution and high-accuracy measurement needs.

2.4 3D imaging using digital fringe projection (DFP)

Digital fringe projection (DFP) methods resolve the limitation of the binary coding method and achieve camera pixel spatial resolution by using continuously varying structured patterns instead of binary patterns. Specifically, sinusoidally varying structured patterns are used in the DFP system, and these sinusoidal patterns are often regarded as fringe patterns. Therefore, the DFP technique is a special kind of structured light techniques by using sinusoidal or fringe patterns. The major difference of the DFP technique lies in the fact that it does not use intensity for coding but rather uses phase. And one of the most popular methods to recover phase is the phase-shifting-based fringe analysis technique (Malacara 2007). For example, a three-step phase-shifting algorithm with equal phase shifts can be mathematically formulated as

$$I_1(x, y) = I'(x, y) + I''(x, y) \cos(\phi - 2\pi/3), \quad (12)$$

$$I_2(x, y) = I'(x, y) + I''(x, y) \cos(\phi), \quad (13)$$

$$I_3(x, y) = I'(x, y) + I''(x, y) \cos(\phi + 2\pi/3). \quad (14)$$

Here $I'(x, y)$ denotes the average intensity, $I''(x, y)$ stands for the intensity modulation and ϕ is the phase to be extracted. The phase can be computed by simultaneously solving Eq. (12)–(14):

$$\phi(x, y) = \tan^{-1} \left[\sqrt{3}(I_1 - I_3)/(2I_2 - I_1 - I_3) \right]. \quad (15)$$

The extracted phase ϕ ranges from $-\pi$ to $+\pi$ with 2π discontinuities due to the nature of the arctangent function. To remove the 2π discontinuities, a spatial (Ghiglia and Pritt 1998) or temporal phase unwrapping algorithm is necessary which detects 2π discontinuities and removes them by adding or subtracting the integer $k(x, y)$ of 2π , e.g.,

$$\Phi(x, y) = \phi(x, y) + k(x, y) \times 2\pi. \quad (16)$$

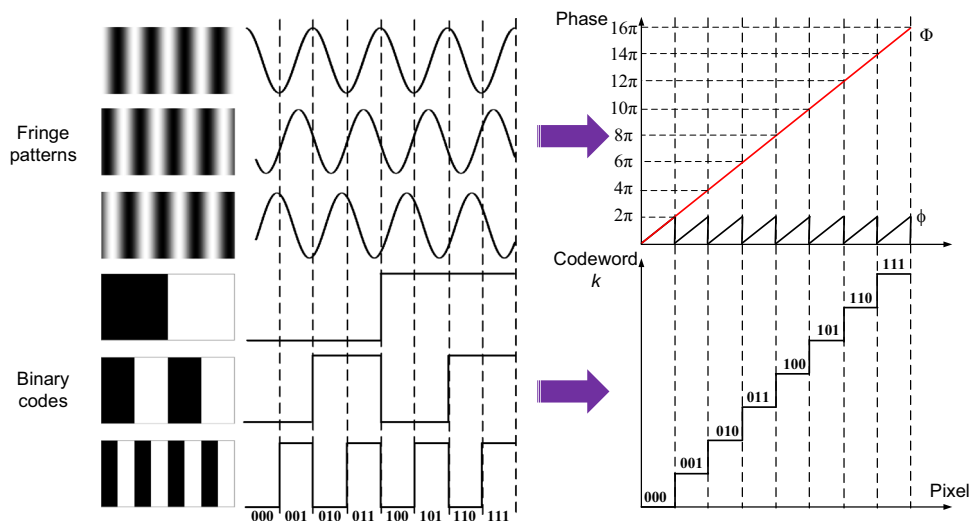
Here, the integer $k(x, y)$ is often called the *fringe order*, and Φ is the unwrapped phase.

A spatial phase unwrapping algorithm determines fringe order $k(x, y)$ relative to the starting point within a connected component, and thus only generates a continuous phase map *relative* to that pixel, called the *relative phase*. The relative phase cannot be used for correspondence establishment since it cannot be used to uniquely determine the phase on the projector space. Therefore, additional information is required to rectify the relative phase to be absolute such that it can then be uniquely defined.

A temporal phase unwrapping algorithm retrieves the absolute fringe order $k(x, y)$ per pixel by acquiring additional information, and therefore generates the *absolute phase*. One of the commonly adopted temporal phase unwrapping methods is by means of encoding the fringe order $k(x, y)$ with binary coded patterns, discussed in Sect. 2.3. Such a temporal phase unwrapping method recovers the absolute phase by capturing additional binary patterns in addition to the sinusoidal patterns.

Figure 7 illustrates the procedures for absolute phase recovery. First, the wrapped phase map ϕ with 2π discontinuities is obtained by applying the phase-shifting algorithm; then the fringe order $k(x, y)$ is recovered by analyzing the binary coded fringe patterns, and finally, the absolute phase Φ is recovered by applying Eq. (16). The red line in Fig. 7 shows the unwrapped phase without 2π discontinuities.

Fig. 7 Absolute phase recovery combining phase shifting with binary coding; the phase-shifting method extracts the wrapped phase ϕ with 2π discontinuities; the binary coded patterns encodes the fringe order k ; the phase is finally unwrapped by adding integer k multiples of 2π to the wrapped phase ϕ to remove 2π discontinuities



There are other temporal phase unwrapping methods developed in optical metrology fields, such as multi-frequency (or multi-wavelength) phase-shifting methods (Cheng and Wyant 1984, 1985; Towers et al. 2003; Wang and Zhang 2011). The multi-frequency phase-shifting methods essentially capture fringe patterns with different frequencies and uses phases from all frequencies to determine absolute fringe order $k(x, y)$.

The absolute phase map can be used as the codeword to establish one-to-many mapping in the same way as binary coding methods. However, since the phase map obtained here is continuous NOT discrete, the mapping (or correspondence) can be established at camera pixel level. Figure 8 illustrates the concept of pixel level correspondence using phase-shifting methods. For a selected camera pixel (u^c, v^c), its absolute phase value $\Phi(u^c, v^c)$ uniquely maps to the projector's line u^p with exactly the same phase value on the projector space. If horizontal sinusoidal patterns are used, Eq. (7) becomes

$$u^p = \Phi(u^c, v^c) \times P / (2\pi), \quad (17)$$

where P is the number of pixels for a single period of sinusoids which corresponds to 2π in phase. The scaling factor $P/(2\pi)$ simply converts the phase to the projector line in pixels. The continuous and differentiable nature of

an absolute phase map makes it possible to achieve pixel-level correspondence between the camera and the projector. Once the correspondence is known, (x^w, y^w, z^w) can be computed using Eq. (11).

Besides recovering 3D geometry, Eqs. (12)–(14) can also generate texture information I_t ,

$$I_t(x, y) = \frac{I_1 + I_2 + I_3}{3} + \frac{\sqrt{(I_1 - I_3)^2 + (2I_2 - I_1 - I_3)^2}}{3}. \quad (18)$$

The texture information, which appears like an actual photograph of the imaged scene, can be used for object recognition and feature detection purposes.

Figure 9 shows an example of measuring a complex statue with the gray coding method. Figure 9a shows a photograph of the object to be measured. A sequence of nine gray-coded binary patterns are captured to recover the codeword map. Figure 9b–f shows five of these images from the wider to denser structured patterns. From this sequence of patterns, the codeword map is then recovered, shown in Fig. 9g. This binary coded map can then be used to recover the 3D shape of the object, and Fig. 9h shows the result.

The same object is then measured again by the DFP method; Fig. 10a–c shows three phase-shifted fringe

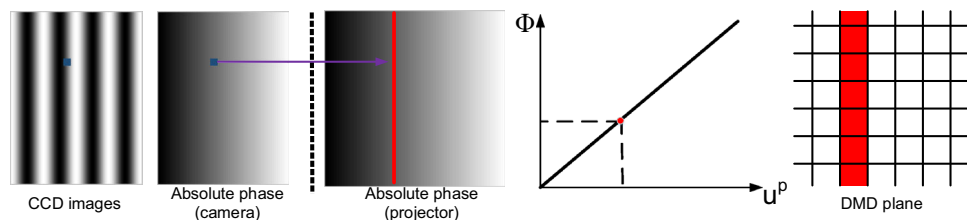


Fig. 8 1D correspondence detection through DFP: one camera pixel map has a unique absolute phase value, which maps to a unique phase line on the projector absolute phase, and a pixel line on the projector DMD plane

Fig. 9 Example of measuring a complex statue with the gray coding method. **a** Photograph of the measured object; **b–f** gray coded binary patterns from wider to denser stripes; **g** recovered codeword map; **h** 3D recovered geometry

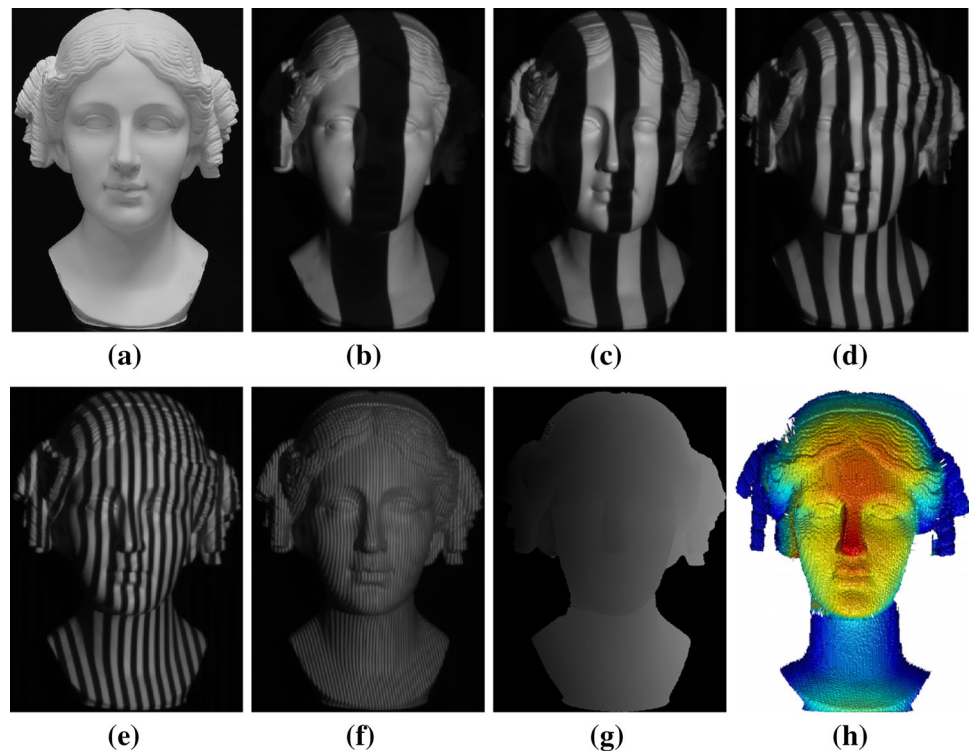
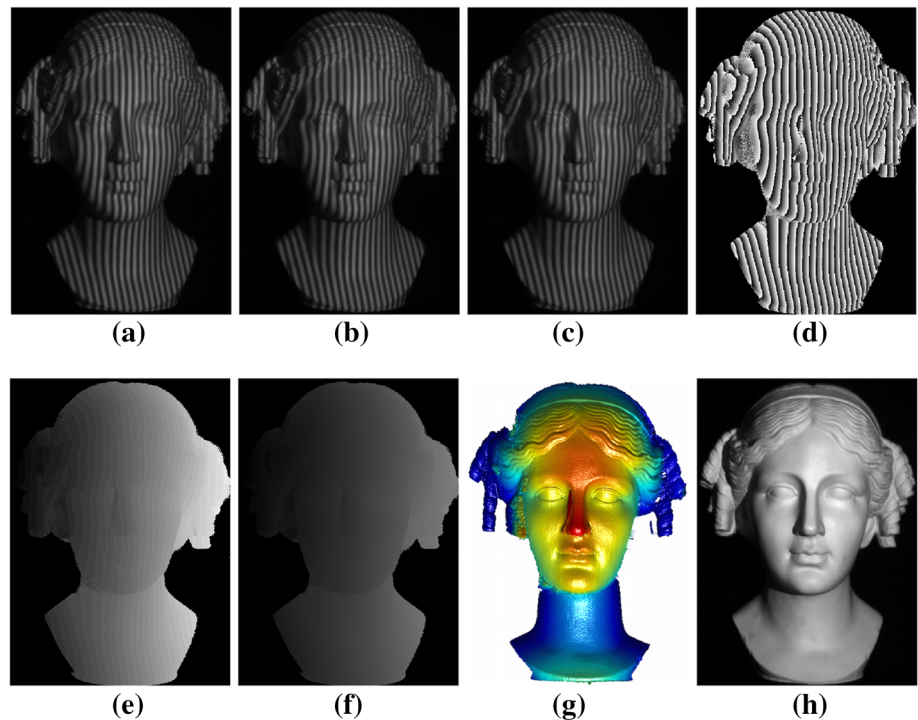


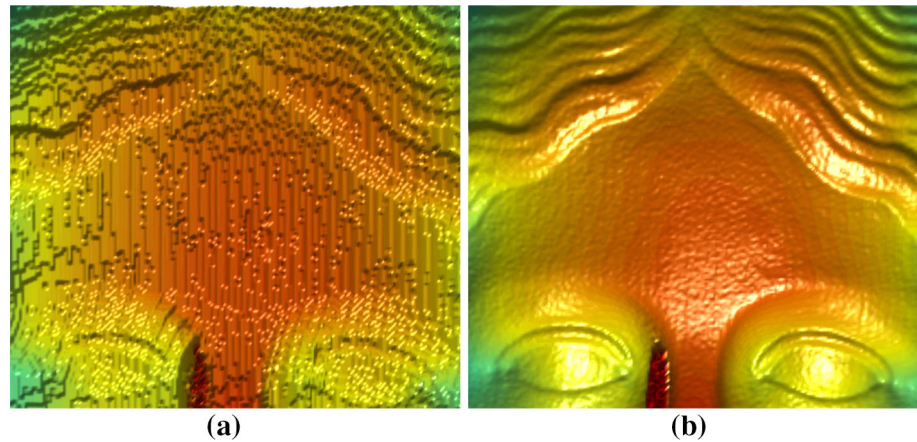
Fig. 10 The same object as in Fig. 9 measured by the DFP method. **a–c** Three phase-shifted fringe images; **d** wrapped phase map; **e** fringe order map; **f** unwrapped phase map; **g** 3D recovered geometry; **h** texture image



patterns. Applying Eq. 15 to these phase-shifted fringe patterns will generate the wrapped phase map, shown in Fig. 10d. In the meantime, we use six gray coded binary patterns captured for the binary coding method to generate a fringe order map $k(x, y)$, as shown in Fig. 10e. The unwrapped phase can then be obtained pixel by pixel by

applying Eq. 16. Figure 10f shows the unwrapped phase. Once the unwrapped phase is known, the 3D shape can be recovered pixel by pixel. Figure 10g shows the resultant 3D geometry. In addition to 3D geometry, the same three phase-shifted fringe patterns shown in Fig. 10a–c can be used to generate the texture image using Eq. 18. Figure 10h

Fig. 11 Comparing results from binary coding and DFP methods. **a** Zoom-in view of 3D geometry, shown in Fig. 9h, by the binary coding method; **b** zoom-in view of 3D geometry, shown in Fig. 10g, by the DFP method



shows the texture image which is perfectly aligned with 3D geometry point by point.

To better visualize the difference between these binary coding methods and the phase-shifting method, Fig. 11 shows the zoom-in view of the same region of the recovered 3D geometry. Clearly, 3D geometry recovered from the DFP method has a lot more detail than that using the binary coding method, yet it requires fewer acquired images, demonstrating the merits of using the DFP method for 3D imaging.

In summary, compared with binary coding methods, the DFP technique based on phase-shifting algorithms has the following advantages:

- *High spatial resolution* From Eqs. (12)–(15), one can see that the phase value of each camera pixel can be independently computed, and thus 3D measurement can be performed at camera pixel spatial resolution.
- *Less sensitive to ambient light* The phase computation numerator and denominator take the differences of the captured images, and the ambient light embedded in $I'(x, y)$ is automatically cancelled out. In theory, ambient light does not affect phase at all, albeit it will affect the signal-to-noise ratio (SNR) of the camera image and thus may reduce the measurement quality.
- *Less sensitive to surface reflectivity variations* Since surface reflectivity affects all three fringe patterns at the same scale for each pixel, the pixel-by-pixel phase computation (Eq. (15)) also cancels out the influence of reflectivity.
- *Perfectly aligned geometry and texture* Since pixel-wise 3D geometry and texture are obtained from exactly the same set of fringe patterns, they are perfectly aligned without any disparities.

2.5 High-speed 3D imaging

Real-time 3D imaging includes three major components: 3D image acquisition, reconstruction and visualization,

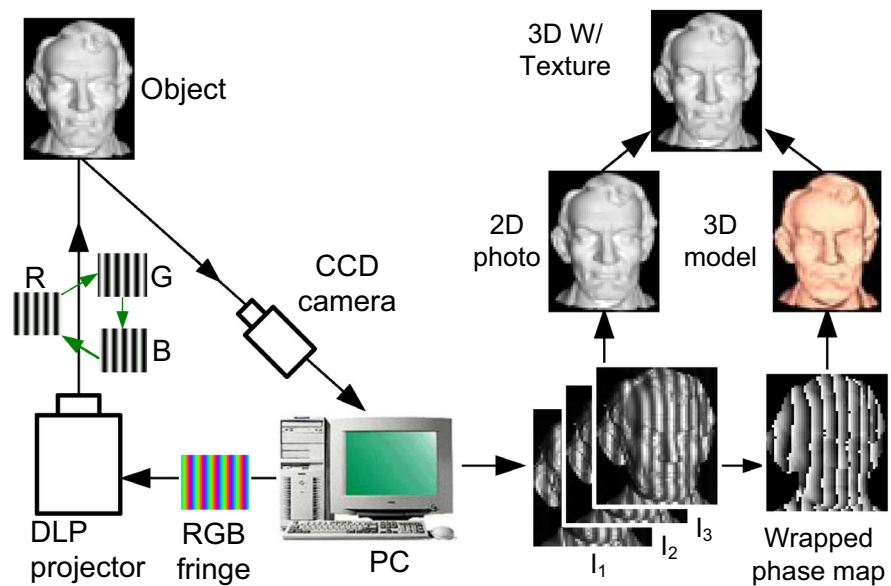
which are all done simultaneously in real time. Real-time 3D imaging can be applied to numerous areas, including manufacturing, entertainment, and security. For intelligent robotics, real-time 3D imaging technology is also of great value as a non-contact optical sensing tool. DFP techniques have been one of the best available methods due their advantageous properties, discussed in Sect. 2.4.

The advancement of real-time 3D imaging using DFP methods has evolved with hardware improvements. Earlier technologies (Zhang and Huang 2006) mainly used the single-chip DLP technology and encoded three phase-shifted fringe patterns into the three primary color channels of the DLP projector. Due to its unique projection mechanism, the single-chip DLP projection system allows the camera to capture three primary color channels separately and sequentially. Figure 12 shows the layout of such a real-time 3D imaging system. The single-chip DLP projector projects three phase-shifted patterns rapidly and sequentially in grayscale (when color filters are removed), and the camera, when precisely synchronized with the projector, captures each individual channel for 3D reconstruction. Since a DLP projector typically projects fringe patterns at 120 Hz, such a technology allows 3D shape measurement at a speed of up to 120 Hz.

Being limited to encoding three primary color channels, only three phase-shifted fringe patterns can be used for such a real-time 3D imaging technology, and thus absolute phase recovery has to be realized by encoding a marker on these fringe patterns (Zhang and Yau 2006), while only single *smooth* geometry can be measured. This method also requires substantial projector modifications, and sometimes these modifications are impossible without the projector manufacturer's involvement (Bell and Zhang 2014). Although DLP Discovery platforms have been available for a long time, they have been too expensive for wide adoption in academia or industry.

Fortunately, with more than a decade of effort on high-speed 3D imaging from our research community, projector

Fig. 12 Layout of a real-time 3D imaging system using a single-chip DLP projector. Three phase-shifted fringe patterns are encoded into *R*, *G* and *B* color channels of the projector are sequentially and repeatedly projected onto the object surface. The high-speed camera used to capture the images which is precisely synchronized with each individual 8-bit pattern projection. A three-step phase-shifting algorithm is applied to the combined channel images to compute the phase for 3D reconstruction. The projector refreshes typically at 120 Hz in color, 360 Hz for individual channels



manufacturers are finally recognizing the opportunities in industry for producing affordable specialized projectors for this field: for example, LogicPD has produced the LightCommander, and WinTech Digital the LightCrafter series. With such specialized DLP projectors, it is much easier to employ more robust algorithms for real-time 3D imaging, such as those absolute phase recovery methods discussed in Sect. 2.4.

Once data acquisition becomes fast enough and ready for use, the second challenge is to speed up data processing. High-speed 3D data processing starts with fast phase-shifting algorithms including phase wrapping (Huang and Zhang 2006) and unwrapping (Zhang et al. 2007) coupled with advanced graphics processing unit (GPU) technology developments. GPU is a dedicated graphics rendering device for a personal computer or games console. Current GPUs are not only very efficient at manipulating and displaying computer graphics but their highly parallel structure also makes them more effective than typical CPUs for a range of complex algorithms. Although CPUs have been increasing their performance over time, they have encountered severe bottlenecks for progressing since increasing the clock frequency has fundamental physics limitations. GPUs boost the performance of the CPU by employing a massive number of lower-frequency simple processors in parallel. Due to the simpler architecture, the fabrication cost is much lower, making them available now for almost all graphics cards. Naturally, researchers have endeavored to bring GPU technologies to the optical imaging field, such as Zhang et al. (2006), Liu et al. (2010) and Karpinsky et al. (2014). Faster than real-time (e.g., 30

Hz) 3D data processing speeds have been successfully achieved even with an integrated graphics card on a laptop.

With advanced GPU technologies, real-time 3D image visualization becomes straightforward since all data are readily available on the graphics card, and so can be displayed immediately on the screen. It is important to note that the amount of data to be visualized is very large since DFP techniques recover 3D coordinates and texture for each camera pixel. Therefore, the deciding factors of real-time 3D imaging efficiency are the number of pixels on the camera sensor and the processing power of the computer. For all obtained data points, it is very challenging to send them directly to a graphics card through the data bus between the video card and the computer, and thus, in order to achieve real-time visualization, 3D reconstruction typically has to be done on the graphics card with GPU.

It is always desirable to achieve higher-speed 3D image acquisition to reduce motion artifacts and to more rapidly capture changing scenes. Lei and Zhang (2009) developed the 1-bit binary defocusing method to break the speed bottleneck of high-speed 3D imaging methods. Using 1-bit binary patterns reduces the data transfer rate and thus making it possible to achieve a 3D imaging rate faster than 120 Hz with the same DLP technology. For example, the DLP Discovery platform introduced by Texas Instruments can switch binary images at a rate up to over 30,000 Hz, and thus kHz 3D imaging is feasible (Li et al. 2014). This method is based on the nature of defocusing: evenly squared binary patterns appear to be sinusoidal if the projector lens is properly defocused. Therefore, instead of directly projecting 8-bit sinusoidal patterns, we can

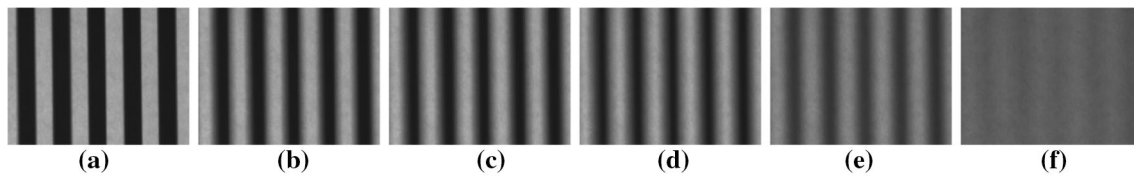


Fig. 13 Sinusoidal fringe generation by defocusing binary structured patterns. **a** When the projector is in focus; **b–f** gradual resultant fringe patterns when the projector's amount of defocusing increases

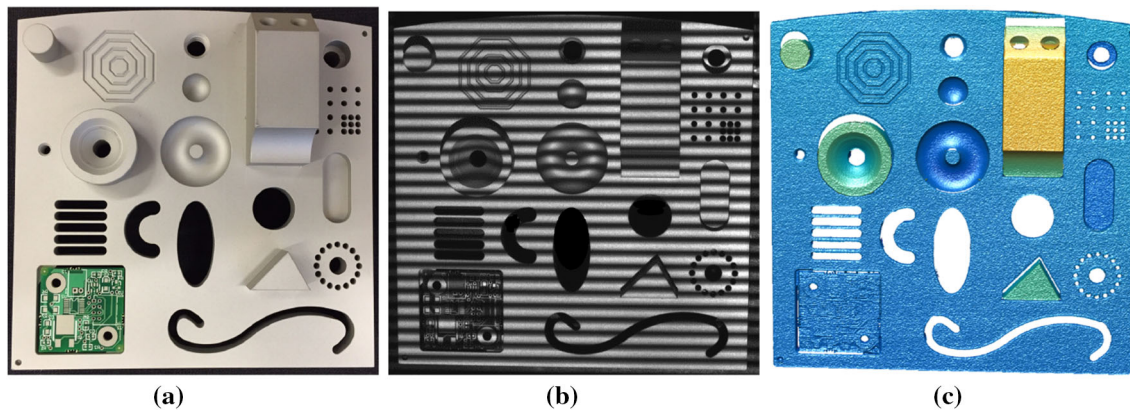


Fig. 14 3D imaging of a complex mechanical part. **a** The captured part; **b** one of the fringe patterns; **c** 3D reconstruction

approximate sinusoidal profiles through projecting 1-bit binary patterns and properly defocusing the projector.

Figure 13 shows some captured fringe images with the projector at different defocusing levels. As one can see, when the projector is in focus, as shown in Fig. 13a, it preserves apparent squared binary structures, but when the projector is properly defocused (see Fig. 13c), the squared binary structure will appear to have an approximately sinusoidal profile. Without a doubt, the sinusoidal structures will gradually diminish if the projector is overly defocused, which results in low fringe quality. Once sinusoidal patterns are generated, a phase-shifting algorithm can be applied to compute the phase and thus 3D geometry after system calibration.

3 Measurement examples

This section shows some representative 3D measurement results using DFP techniques, ranging from static to high speed, and from micro- to macro- and to large-scale scene captures.

3.1 Complex 3D mechanical part measurement

Figure 14 shows the result of measuring the static mechanical part. Figure 14a shows that the part has a variety of different shapes, blocks, and color on its surface. Figure 14b and c respectively show one of the captured

fringe patterns and the reconstructed 3D geometry. The 3D result clearly shows that the fine details are well recovered under our 3D shape measurement system. The system includes a digital CCD camera (Imaging Source DMK 23UX174) with a resolution of 1280×1024 pixels, and a DLP projector (Dell M115HD) with a resolution of 1280×800 pixels. The camera has a lens with 25-mm focal length (Fujinon HF25SA-1).

3.2 Real-time 3D shape measurement

Figure 15a shows the result captured by a system developed more than 10 years ago (Zhang et al. 2006). The right part of the image shows a subject and the left side shows the simultaneously recovered 3D geometry on the computer screen. Recently, the laptop computer has proven powerful enough to perform real-time 3D imaging (Karpinsky et al. 2014). Figure 15b shows a recently developed system that used a laptop computer (IBM Lenovo laptop with a Intel i5 3320M 2.6 GHz CPU and NVIDIA Quadro NVS5400M GPU) to achieve 800×600 image resolution (or 480,000 points per frame) at 60 Hz. The entire system cost is also fairly low due to the reduced hardware component cost.

Facial expressions carry a lot of information including emotions, and thus the capability of capturing facial expression details is of great interest to different communities potentially including robotics. Figure 16 shows a few example frames captured by our real-time 3D shape

Fig. 15 Real-time 3D imaging. **a** Real-time 3D imaging system on a desktop computer developed over a decade ago; **b** real-time 3D imaging on a laptop computer developed recently

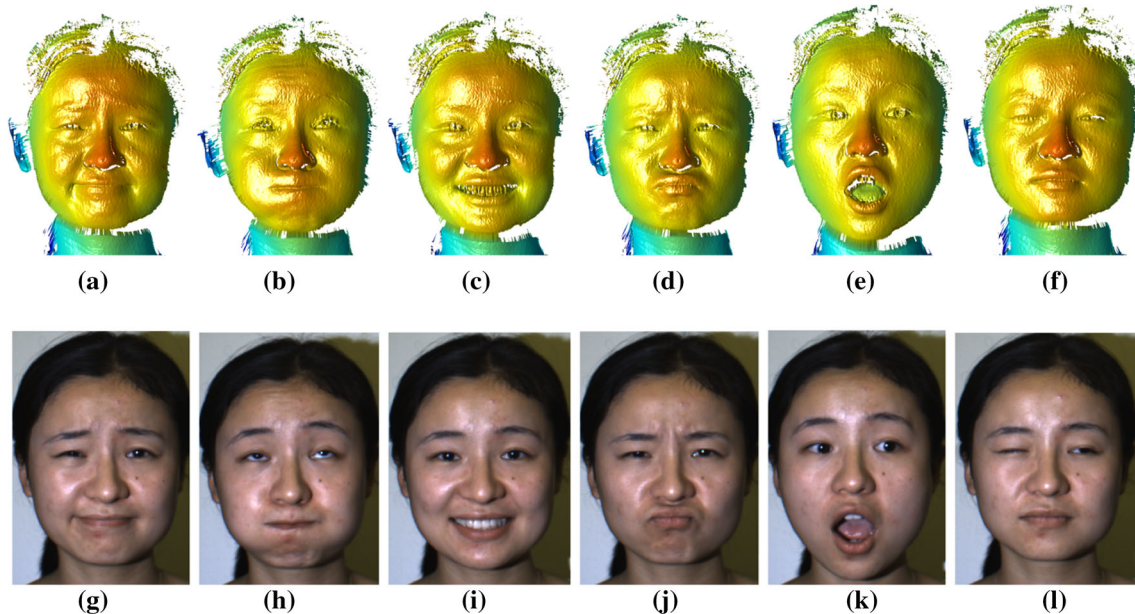
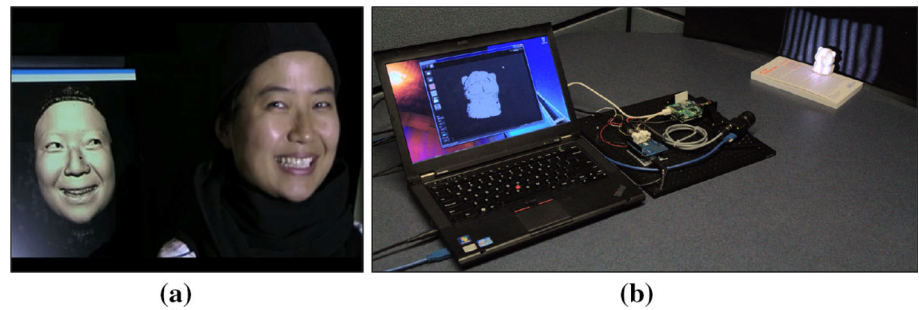


Fig. 16 Capturing different facial expressions. The *first row* shows the 3D geometry and the *second row* shows the corresponding texture that is perfectly aligned with the 3D geometry

measurement system. This system includes a USB3.0 camera (Point Grey Grasshopper3) and a LightCrafter 4500 projector. The acquisition speed was chosen to be 50 Hz, and the image resolution was set at 640×480 . As discussed earlier, the same set of fringe patterns can also be used to recover texture (or a photograph of the object). The second row of Fig. 16 shows the corresponding color texture that is perfectly aligned with the 3D geometry shown above.

Hands are very important parts of the human body for interactions, manipulations, and communications. We have used the same system shown in Fig. 16 for facial data acquisition to capture hands. Figure 17 shows the results of different hand gesture in 3D; and, similarly, color texture is also available for immediate use. The color texture was not included in this paper because it is straightforward to understand.

Human body gesture motion dynamics also provide rich information for communication. Figure 18 shows that the DFP system can also be used to measure the human body.

This system includes a USB3.0 camera (Point Grey Grasshopper3) and a LightCrafter 4500 projector. The acquisition speed was 33 Hz, the image resolution was 1280×960 , and the pixel size is approximately 1 mm at the object space. As discussed earlier, the same set of fringe patterns can also be used to recover texture (or a photograph of the object).

3.3 Superfast 3D imaging

Figure 19 shows an example of using a kHz 3D imaging system (Li and Zhang 2016) to capture object deformation, in which three sample frames of a fluid flow surface topological deformation process are shown. As one can see, the geometric deformation of the imaged droplet is well recovered with the kHz binary defocusing technique, which could potentially bring additional information for fluid mechanics analysis. This superfast 3D imaging technique is also applicable to other applications, such as vibration analysis (Zhang et al. 2010) in mechanical engineering or

Fig. 17 Capturing different hand gestures. *First row* shows single hand examples, and the *bottom row* shows two hands examples

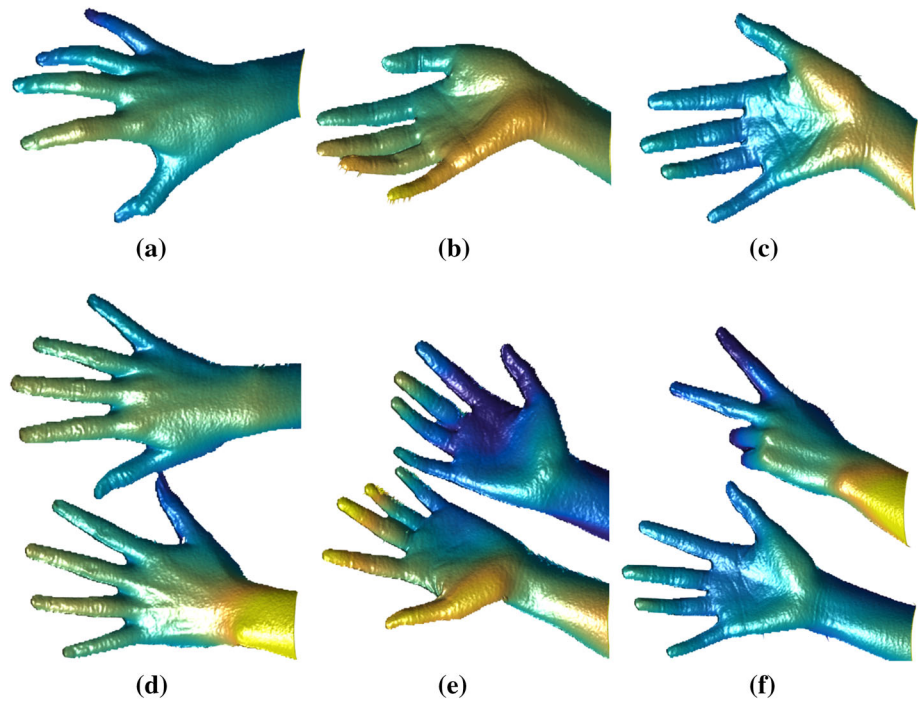
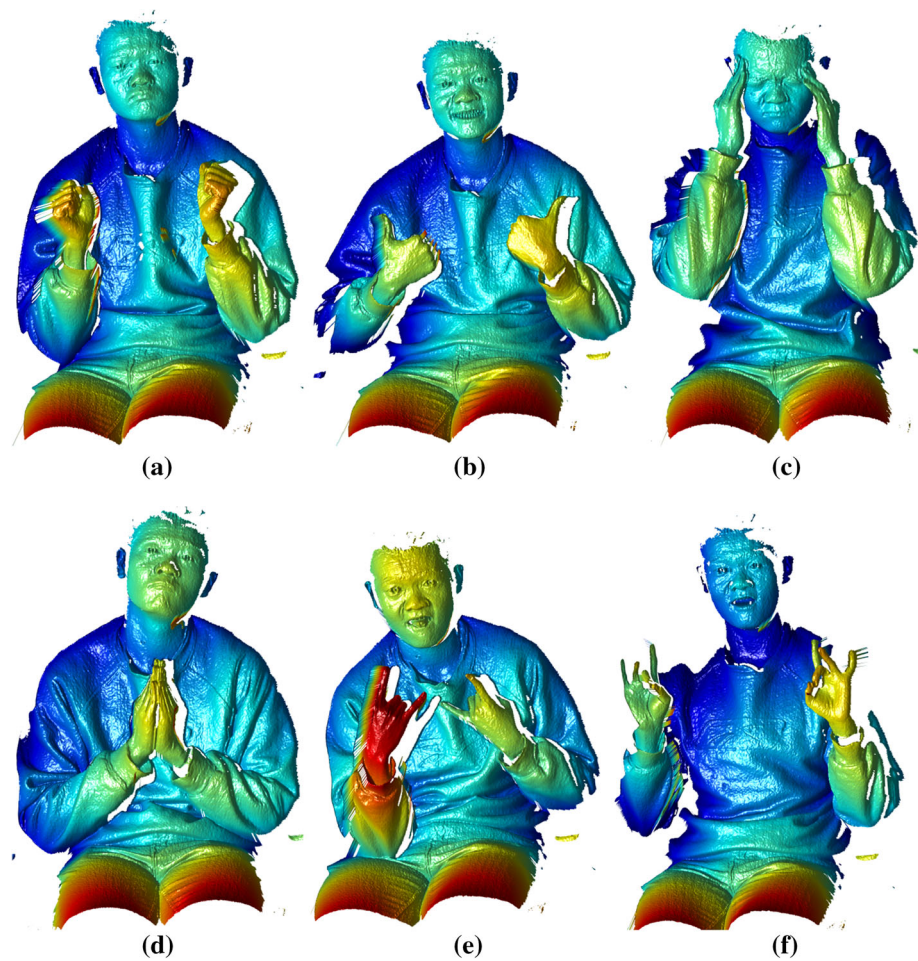


Fig. 18 Capturing different poses of the human body



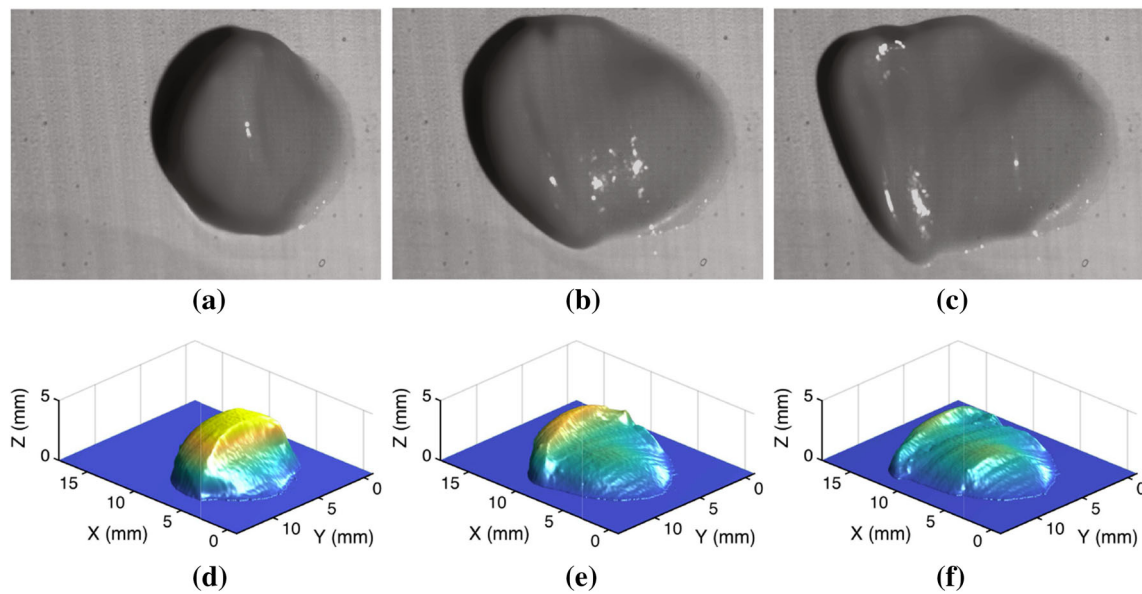
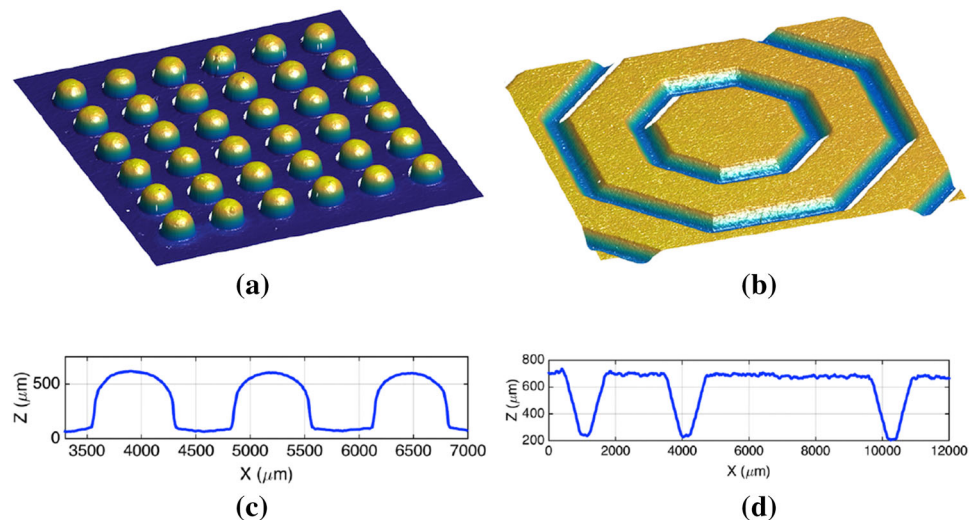


Fig. 19 Superfast 3D imaging of fluid surface dynamics. **a–c** Three sample frames of texture; **d–f** three frames of the 3D geometry

Fig. 20 3D imaging of micro-structures. **a** 3D result of a PCB board; **b** 3D result of a mechanical part; **c** cross-section of (a); **d** cross-section of (b)



cardiac mechanics analysis (Wang et al. 2013) in biomedical science.

3.4 3D Microstructure measurement

The real-time to superfast 3D imaging techniques can also be used to measure micro-structures by different optics. For example, we have developed micro-scale 3D imaging with dual telecentric lenses, and achieved μm measurement accuracy (Li and Zhang 2016). This system used the Wintech PRO4500 for pattern projection and the Imaging Source DMK 23U274 camera for data acquisition. The camera resolution was 1600×1200 , and the pixel size is approximately $16 \mu\text{m}$. Figure 20 shows two measurement

examples. For this system setup, the measurement accuracy was found to be approximately $\pm 5 \mu\text{m}$ for a volume of $10(H) \text{ mm} \times 8(W) \text{ mm} \times 5(D) \text{ mm}$.

4 Potential applications

With the development of computer data analysis, high-resolution and high-accuracy 3D data captured by these optical 3D imaging techniques could be an integrated part of future intelligent robots. In this section, we cast our view over potential applications of the high-accuracy and high-speed 3D optical imaging techniques in the field of intelligent robotics.

4.1 Robot manipulation and navigation

The intelligent robot has the capability of performing some tasks autonomously without any teaching or programming. The prerequisite of autonomous action is to know the environment around the robot by sensors, followed by decision making and dynamic planning. The visual sensors are commonly employed to reconstruct the 3D environment for robot manipulation and navigation. To reconstruct the 3D environment, the corresponding matching always relies on the optical properties of the environment, such as texture, reflectance, and illumination, resulting in decreased reliability. To address this problem, structured light-based visual sensors have increasingly been used in the field of robotics because they can provide a high-accuracy, high-speed, and high-reliability 3D environment. For example, with integrated advanced 3D imaging sensors, intelligent robots could figure out a complex and unknown assembly task by a single one or a team after a certain level of training. Of course, to be able to do that, advanced machine learning techniques have to be developed, and the miniaturized 3D imaging techniques have to be embedded onto the robot itself.

For robot manipulation, a geometry-based Xpoint primitive can be designed to achieve high-accuracy location regardless of invariant surface reflectivity (Xu et al. 2012). Commercial sensor such as Kinect, RealSense, and Tango have started being used, albeit their accuracy is still limited for precision manipulations. Structured light system has proven to be able to achieve μm measurement accuracy (Li and Zhang 2016), making it possible to precisely tell where a particular object is as well as the geometry of those features. Figure 14 shows an example of part measurement at tens of μm accuracy. Once these accurately measured 3D data are available, one can measure the distance between two holes for inspection, and also precisely tell the layout of the features on the surface.

With such high-accuracy measurement and further data analytics tools for feature detection and path planning, we believe that future robots will be able to use such high-accuracy 3D imaging techniques to learn unknown parts and then precisely manipulate those parts. For mobile robot navigation, a wide field of view (FOV) is preferred to avoid obstacle, especially in a narrow space. Thus, extending real-time structure light techniques to be omnidirectional (e.g., Zhang et al. 2012) would add more value.

4.2 Human robot interaction

Extremely high-resolution and high-accuracy 3D imaging could also potentially help robots to understand humans, allowing humans and robots to interact with each other naturally and thus collaborate more seamlessly. In modern,

especially smart, factories, where humans and robots typically coexist, the safety of persons working in such industrial settings is a major concern. If smart sensors can precisely measure where people are close by, they can send signals to the robots to avoid accidents.

Human body language can tell a lot about the current emotional and physical status of a person. Thus, understanding human facial expressions could infer whether a person is happy or sad; understanding the gesture motion dynamics (e.g., walking) could provide information about the physical strength of an aged person; and understanding the hand and arm dynamics of a worker could give cues about their reliability. By sensing such cues of human partners, robots could make decisions about whether they should provide assistance to them (e.g., support an aged person before s/he falls).

4.3 Mobile microrobotics

High-accuracy, and high-speed 3D imaging also has a great potential to conquer some fundamental challenges in the microrobotics field. Robots with the size of several microns have numerous applications in medicine, biology, and manufacturing (Diller et al. 2013; Abbott et al. 2007). Simultaneous independent locomotion of multiple robots and their end-effectors at this scale is difficult since the robot itself is too small to carry power, communication, and control on board. However, high-accuracy, high-speed 3D structured light imaging may be the key to unlocking the potential of these systems. Mobile microrobots have an overall size (footprint) of less than a millimeter, and their motions are no longer dominated by inertial (gravitational) forces (Chowdhury et al. 2015a). Thus, microrobots have to overcome the size restrictions that do not allow for on-board actuation, power, and control, and due to the unique interaction forces, the conventional actuation principles utilizing the gravitational forces typically do not work.

Researchers typically rely on off-board or external global fields for power and actuation of mobile microrobots (Kummer et al. 2010; Steager et al. 2013; Floyd et al. 2008; Jing et al. 2011, 2012, 2013a, b). Using an external magnetic field is a popular actuation method due to its high actuation force, compact system size, and low cost. Researchers have long been trying to control multiple microrobots independently using these global magnetic fields. However, it has primarily resulted in coupled movements of the robots in the workspace (Pawashe et al. 2009a, b; Diller et al. 2012; Frutiger et al. 2010; DeVon and Bretl 2009; Cheang et al. 2014). Recently, researchers have developed a specialized substrate with an array of planar microcoils to generate local magnetic fields for independent actuation of multiple microrobots (Cappelleri et al. 2014; Chowdhury et al. 2015c, b). While some new

microrobot designs are emerging with soft end-effectors (Jing and Cappelleri 2014a, b, c), they are passive, and current man-made microrobots cannot actively deform. Another control modality is needed if one is to control an active end-effector, i.e., a micro-gripper, at the end of the magnetic microrobot. This is an opportunity for 3D structured light techniques.

By taking advantage of the wireless, scalable and spatiotemporally selective capabilities that light allows, Palagi et al. (2016) show that soft microrobots consisting of photoactive liquid-crystal elastomers can be driven by structured light to perform sophisticated biomimetic motions. Selectively addressable artificial microswimmers that generate travelling-wave motions to self-propel, as well as microrobots capable of versatile locomotion behaviors on demand, have been realized. The structured light fields allow for the low-level control over the local actuation dynamics within the body of microrobots made of soft active materials. This same technique can be applied to actuate end-effectors made from similar materials attached to a magnetic microrobot body, like the ones in Jing and Cappelleri (2014a, b, c). Magnetic fields can be used for position and orientation control while the structured light can be used for end-effector actuation control.

Structured light exposure can also be used for shape-shifting soft microrobots (Fusco et al. 2015) into different configurations. Huang et al. (2016) demonstrated that individual microrobots can be selectively addressed by NIR light and activated for shape transformation, yielding the microrobot's "shape" as an extra degree of freedom for control. Thus, the principle of using structured light has great potential in the mobile microrobotics community, and it can be extended to other microrobotic applications that require microscale actuation with spatiotemporal coordination.

5 Summary

This paper has presented the high-speed and high accuracy 3D imaging techniques using the digital fringe projection method, a special yet advantageous structured light method. We have elucidated the details of these techniques to help beginners to understand how to implement such techniques for their applications. We have also presented some representative measurement results to demonstrate the capabilities of the DFP techniques for different scale and resolution measurements. Finally, we cast our perspective over potential applications of the DFP techniques in the robotics field. We hope that this paper is a good introduction of DFP techniques mainly developed in the optical metrology community to the intelligent robotics community.

Acknowledgements We would like to thank many current and former students working in our laboratory for their tremendous work in advancing 3D imaging technologies to their current state. In particular, we would like to thank Dr. Yajun Wang for his development of binary dithering and optimal pulse width modulation methods; Dr. Nik Karpinsky for implementation of graphics processing unit (GPU) programming; Jae-Sang Hyun and Zexuan Zhu for their hardware system design and developments; and Maggie Hao, Chufan Jiang and Ziping Liu for serving as models to test our systems.

References

- Abbott, J., Nagy, Z., Beyeler, F., Nelson, B.: Robotics in the small. *IEEE Robot. Autom. Mag.* **14**, 92–103 (2007)
- Bell, T., Zhang, S.: Towards superfast three-dimensional optical metrology with digital micromirror device (dmd) platforms. *Opt. Eng.* **53**(11), 112206 (2014)
- Cappelleri, D., Efthymiou, D., Goswami, A., Vitoroulis, N., Zavlanos, M.: Towards mobile microrobot swarms for additive micromanufacturing. *Int. J. Adv. Robot. Syst.* **11**(9), 150 (2014).
- Cheang, U.K., Lee, K., Julius, A.A., Kim, M.J.: Multiple-robot drug delivery strategy through coordinated teams of microswimmers. *Appl. Phys. Lett.* **105**(8), 083705 (2014)
- Cheng, Y.Y., Wyant, J.C.: Two-wavelength phase shifting interferometry. *Appl. Opt.* **23**, 4539–4543 (1984)
- Cheng, Y.Y., Wyant, J.C.: Multiple-wavelength phase shifting interferometry. *Appl. Opt.* **24**, 804–807 (1985)
- Chowdhury, S., Jing, W., Cappelleri, D.J.: Controlling multiple microrobots: recent progress and future challenges. *J. Micro Bio Robot.* **10**(1), 1–11 (2015a)
- Chowdhury, S., Jing, W., Cappelleri, D.J.: Towards independent control of multiple magnetic mobile microrobots. *Micromachines* **7**(1), 3 (2015b)
- Chowdhury, S., Jing, W., Jaron, P., Cappelleri, D.: Path planning and control for autonomous navigation of single and multiple magnetic mobile microrobots. In: *Proceedings of the ASME 2015 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*. Boston, Massachusetts (2015c)
- DeVon, D., Bretl, T.: Control of many robots moving in the same direction with different speeds: a decoupling approach. In: *American Control Conference. ACC'09.*, pp. 1794–1799. IEEE (2009)
- Dhond, U.R., Aggarwal, J.K.: Structure from stereo—a review. *IEEE Trans. Syst. Man. Cybern.* **19**(6), 1489–1510 (1989)
- Diller, E., Floyd, S., Pawashe, C., Sitti, M.: Control of multiple heterogeneous magnetic microrobots in two dimensions on nonspecialized surfaces. *IEEE Trans. Robot.* **28**(1), 172–182 (2012)
- Diller, E., Sitti, M., et al.: Micro-scale mobile robotics. *Found. Trends Robot.* **2**(3), 143–259 (2013)
- Floyd, S., Pawashe, C., Sitti, M.: An untethered magnetically actuated micro-robot capable of motion on arbitrary surfaces. In: *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, pp. 419–424. IEEE (2008)
- Frutiger, D.R., Vollmers, K., Kratochvil, B.E., Nelson, B.J.: Small, fast, and under control: wireless resonant magnetic micro-agents. *Int. J. Robot. Res.* **29**(5), 613–636 (2010)
- Fusco, S., Huang, H.W., Peyer, K.E., Peters, C., Hberli, M., Ulbers, A., Spyrogiani, A., Pellicer, E., Sort, J., Pratsinis, S.E., Nelson, B.J., Sakar, M.S., Pan, S.: Shape-switching microrobots for medical applications: The influence of shape in drug delivery and locomotion. *ACS Appl. Mater. Interfaces* **7**(12), 6803–6811 (2015)

- Ghiglia, D.C., Pritt, M.D. (eds.): Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software. Wiley, New York (1998)
- Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2000). ISBN:0521623049
- Huang, H.W., Sakar, M.S., Riederer, K., Shamsudhin, N., Petruska, A., Pan, S., Nelson, B.J.: Magnetic microrobots with addressable shape control. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 1719–1724 (2016)
- Huang, P.S., Zhang, S.: Fast three-step phase-shifting algorithm. *Appl. Opt.* **45**(21), 5086–5091 (2006)
- Huang, P.S., Zhang, S., Chiang, F.P.: Trapezoidal phase-shifting method for three-dimensional shape measurement. *Opt. Eng.* **44**(12), 123601 (2005)
- Huang, Y., Shang, Y., Liu, Y., Bao, H.: Handbook of 3D Machine Vision: Optical Metrology and Imaging, 1st edn., chap. 3D shapes from Speckle, pp. 33–56. CRC, USA (2013)
- Jia, P., Kofman, J., English, C.: Two-step triangular-pattern phase-shifting method for three-dimensional object-shape measurement. *Opt. Eng.* **46**(8), 083201 (2007)
- Jing, W., Cappelleri, D.: Incorporating in-situ force sensing capabilities in a magnetic microrobot. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2014), pp. 4704–4709 (2014a)
- Jing, W., Cappelleri, D.: A magnetic microrobot with in situ force sensing capabilities. *Robotics* **3**(2), 106–119 (2014b)
- Jing, W., Cappelleri, D.J.: Towards functional mobile magnetic microrobots. In: Small-Scale Robotics. From Nano-to-Millimeter-Sized Robotic Systems and Applications, pp. 81–100. Springer, New York (2014c)
- Jing, W., Chen, X., Lytle, S., Fu, Z., Shi, Y., Cappelleri, D.: A magnetic thin film microrobot with two operating modes. In: 2011 IEEE International Conference on Robotics and Automation (ICRA), pp. 96–101 (2011)
- Jing, W., Pagano, N., Cappelleri, D.J.: A micro-scale magnetic tumbling microrobot. In: ASME 2012 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pp. 187–196 (2012)
- Jing, W., Pagano, N., Cappelleri, D.J.: A novel micro-scale magnetic tumbling microrobot. *J. Micro Bio Robot.* **8**(1), 1–12 (2013a)
- Jing, W., Pagano, N., Cappelleri, D.J.: A tumbling magnetic microrobot with flexible operating modes. In: 2013 IEEE International Conference on Robotics and Automation (ICRA), pp. 5514–5519 (2013b)
- Karpinsky, N., Hoke, M., Chen, V., Zhang, S.: High-resolution, real-time three-dimensional shape measurement on graphics processing unit. *Opt. Eng.* **53**(2), 024105 (2014)
- Keferstein, C.P., Marxer, M.: Testing bench for laser triangulation sensors. *Sensor Rev.* **18**(3), 183–187 (1998)
- Kummer, M.P., Abbott, J.J., Kratochvil, B.E., Borer, R., Sengul, A., Nelson, B.J.: Octomag: An electromagnetic system for 5-dof wireless micromanipulation. *Robot. IEEE Trans.* **26**(6), 1006–1017 (2010)
- Lei, S., Zhang, S.: Flexible 3-d shape measurement using projector defocusing. *Opt. Lett.* **34**(20), 3080–3082 (2009)
- Li, B., Wang, Y., Dai, J., Lohry, W., Zhang, S.: Some recent advances on superfast 3d shape measurement with digital binary defocusing techniques. *Opt. Laser Eng.* **54**, 236–246 (2014)
- Li, B., Zhang, S.: Microscopic structured light 3d profilometry: Binary defocusing technique vs. sinusoidal fringe projection. *Opt. Laser Eng.* (2016) (in press)
- Liu, K., Wang, Y., Lau, D.L., Hao, Q., Hassebrook, L.G.: Dual-frequency pattern scheme for high-speed 3-d shape measurement. *Opt. Express* **18**, 5229–5244 (2010)
- Malacara, D. (ed.): Optical Shop Testing, 3rd edn. Wiley, New York (2007)
- Manneberg, G., Hertegard, S., Liljencrantz, J.: Measurement of human vocal fold vibrations with laser triangulation. *Opt. Eng.* **40**(9), 2041–2044 (2001)
- Morano, R.A., Ozturk, C., Conn, R., Dubin, S., Zietz, S., Nissanov, J.: Structured light using pseudorandom codes. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 322–327 (1998)
- Palagi, S., Mark, A.G., Reigh, S.Y., Melde, K., Qiu, T., Zeng, H., Parmeggiani, C., Martella, D., Sanchez-Castillo, A., Kapernaum, N., Giesselmann, F., Wiersma, D.S., Lauga, E., Fischer, P.: Structured light enables biomimetic swimming and versatile locomotion of photoresponsive soft microrobots. *Nat. Mater.* **15**(6), 647–653 (2016)
- Pan, J., Huang, P., Zhang, S., Chiang, F.P.: Color n-ary gray code for 3-d shape measurement. In: 12th International Conference on Experimental Mechanics. Politecnico di Bari, Italy (2004)
- Pawashe, C., Floyd, S., Sitti, M.: Modeling and experimental characterization of an untethered magnetic micro-robot. *Int. J. Robot. Res.* **28**(8), 1077–1094 (2009a)
- Pawashe, C., Floyd, S., Sitti, M.: Multiple magnetic microrobot control using electrostatic anchoring. *Appl. Phys. Lett.* **94**(16), 164108 (2009b)
- Salvi, J., Fernandez, S., Pribanic, T., Llado, X.: A state of the art in structured light patterns for surface profilometry. *Patt. Recogn.* **43**(8), 2666–2680 (2010)
- Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comp. Vis.* **47**(1–3), 7–42 (2002)
- Sirikasemlert, A., Tao, X.: Objective evaluation of textural changes in knitted fabrics by laser triangulation. *Text. Res. J.* **70**(12), 1076–1087 (2000)
- Steager, E.B., Sakar, M.S., Magee, C., Kennedy, M., Cowley, A., Kumar, V.: Automated biomanipulation of single cells using magnetic microrobots. *Int. J. Robot. Res.* **32**(3), 346–359 (2013)
- Towers, D.P., Jones, J.D.C., Towers, C.E.: Optimum frequency selection in multi-frequency interferometry. *Opt. Lett.* **28**, 1–3 (2003)
- Wang, Y., Laughner, J.I., Efimov, I.R., Zhang, S.: 3d absolute shape measurement of live rabbit hearts with a superfast two-frequency phase-shifting technique. *Opt. Express* **21**(5), 5822–5832 (2013)
- Wang, Y., Zhang, S.: Superfast multifrequency phase-shifting technique with optimal pulse width modulation. *Opt. Express* **19**(6), 5143–5148 (2011)
- Xu, J., Zhao, J., Luo, R.: A 3d imaging sensor for mobile manipulation. IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems (CYBER), pp. 315–320. Bangkok, Thailand (2012)
- Zhang, C., Xu, J., Xi, N., Jia, Y., Li, W.: Development of an omnidirectional 3d camera for robot navigation. In: IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), pp. 262–267. Kaohsiung, Taiwan (2012)
- Zhang, S.: Recent progresses on real-time 3-d shape measurement using digital fringe projection techniques. *Opt. Laser Eng.* **48**(2), 149–158 (2010)
- Zhang, S.: High-speed 3D imaging with digital fringe projection technique, 1st edn. CRC, New York (2016)
- Zhang, S., Huang, P.S.: High-resolution real-time three-dimensional shape measurement. *Opt. Eng.* **45**(12), 123601 (2006)
- Zhang, S., Huang, P.S.: Novel method for structured light system calibration. *Opt. Eng.* **45**(8), 083601 (2006)
- Zhang, S., Li, X., Yau, S.T.: Multilevel quality-guided phase unwrapping algorithm for real-time three-dimensional shape reconstruction. *Appl. Opt.* **46**(1), 50–57 (2007)
- Zhang, S., Royer, D., Yau, S.T.: Gpu-assisted high-resolution, real-time 3-d shape measurement. *Opt. Express* **14**(20), 9120–9129 (2006)

- Zhang, S., van der Weide, D., Oliver, J.: Superfast phase-shifting method for 3-d shape measurement. *Opt. Express* **18**(9), 9684–9689 (2010)
- Zhang, S., Yau, S.T.: High-resolution, real-time 3-d absolute coordinate measurement based on a phase-shifting method. *Opt. Express* **14**(7), 2644–2649 (2006)
- Zhang, Z.: A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000)
- Zhang, Z.: Microsoft kinect sensor and its effect. *IEEE Multimed.* **19**(2), 4–10 (2012)



Beiwen Li is a Ph.D. student and research assistant of School of Mechanical Engineering at Purdue University. He received his M.S. degree in Mechanical Engineering from Iowa State University in 2014 and B.S. degree in Information Engineering from Beihang University in 2012. His major research of interest is high-speed macro-scale and microscale 3D optical metrology.



machine learning.

Yatong An is currently a Ph.D. student in the School of Mechanical Engineering at Purdue University. He received his M.Phil. degree in computer science from the University of Hong Kong in August 2015. In 2013, he received his B.S. degree from the Department of Control Science and Engineering in Zhejiang University, China. His research interests include optical measurement, 3D reconstruction, computer vision, control, robotics and



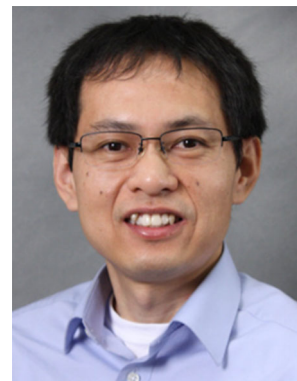
Engineering at Stevens Institute of Technology, Hoboken, NJ. Prof.

Cappelleri is a recipient of the Villanova University Engineering Alumni Society John J. Gallen Memorial Award (2013), National Science Foundation CAREER Award (2012), Harvey N. Davis Distinguished Assistant Professor Teaching Award (2010) and the Association for Lab Automation (ALA) Young Scientist Award (2009). His current research focuses on robotic and automation systems at various length scales: macro-scale (cm to m), meso-scale (~100's of um to a few mm's), micro-scale (10's of um to 100's of nm), and nano-scale (nm).



His research interests include vision-guided manufacturing, image processing, and intelligent robotics.

Jing Xu is currently an associate professor in the Department of Mechanical Engineering, Tsinghua University, Beijing, China. He received his B.E. degree in mechanical engineering from Harbin Institute of Technology, Harbin, China, in 2003, and Ph.D. degree in mechanical engineering from Tsinghua University, Beijing, China, in 2008. He was a post-doctoral researcher in the Department of Electrical and Computer Engineering, Michigan State University, East Lansing.



Song Zhang is an associate professor of mechanical engineering at Purdue University. He received his Ph.D. degree in mechanical engineering from Stony Brook University in 2005. He was a postdoctoral fellow at Harvard University for three years and joined Iowa State University as an assistant professor in 2008 and promoted to associate professor in 2014 before moved to Purdue in 2015. He authored on book entitled high-speed 3D imaging

with digital fringe projection techniques, and the editor of another book entitled handbook of 3-D machine vision: optical metrology and imaging. He won the AIAA Best Paper Award, Best of SIGGRAPH by the Walt Disney, NSF CAREER award, Stony Brook University's 40 under Forty Alumni Award, and Discovery in Mechanical Engineering Award. He is a fellow of SPIE-the International Society for Optics and Photonics, and a senior member of Optical Society of America (OSA). His current research focuses on developing super-fast, super-resolution 3-D imaging technologies and on exploring their applications.