

High-resolution, real-time 3D absolute coordinate measurement based on a phase-shifting method

Song Zhang and Shing-Tung Yau

Mathematics Department, Harvard University, One Oxford Street, Cambridge, MA 02143

szhang@fas.harvard.edu, yau@math.harvard.edu

Abstract: We describe a high-resolution, real-time 3D absolute coordinate measurement system based on a phase-shifting method. It acquires 3D shape at 30 frames per second (fps), with 266K points per frame. A tiny marker is encoded in the projected fringe pattern, and detected by software from the texture image and the gamma map. Absolute 3D coordinates are obtained from the detected marker position and the calibrated system parameters. To demonstrate the performance of the system, we measure a hand moving over a depth distance of approximately 700 mm, and human faces with expressions. Applications of such a system include manufacturing, inspection, entertainment, security, medical imaging.

© 2006 Optical Society of America

OCIS codes: (110.6880) Three-dimensional image acquisition; (120.5800) Scanners; (150.6910) Three-dimensional sensing; (120.6650) Surface measurements, figure; (100.2650) Fringe analysis; (120.3940) Metrology; (120.5050) Phase measurement.

References and links

1. J. Salvi, J. Pages, and J. Batlle, "Pattern codification strategies in structured light systems," *Pattern Recogn.* **37**, 827–849 (2004).
2. S. Zhang and P. Huang, "High-Resolution, Real-time 3D Shape Acquisition," in *IEEE Computer Vision and Pattern Recognition Workshop on Realtime 3D Sensors and Their Uses*, vol. 3, pp. 28–37 (2004).
3. K. G. Harding, "Phase Grating Use for Slope Discrimination in Moiré Contouring," in *Proc. SPIE*, vol. 1614, pp. 265–270 (1991).
4. Z. J. Geng, "Rainbow 3D Camera: New Concept of High-Speed Three Vision System," *Opt. Eng.* **35**, 376–383 (1996).
5. P. S. Huang, Q. Hu, F. Jin, and F. P. Chiang, "Color-Encoded Digital Fringe Projection Technique for High-Speed Three-Dimensional Surface Contouring," *Opt. Eng.* **38**, 1065–1071 (1999).
6. C. Guan, L. G. Hassebrook, and D. L. Lau, "Composite Structured Light Pattern for Three-Dimensional Video," *Opt. Express* **11**, 406–417 (2003).
7. S. Rusinkiewicz, O. Hall-Holt, and L. Marc, "Real-Time 3D Model Acquisition," in *SIGGRAPH*, pp. 438 – 446 (2002).
8. P. S. Huang, C. Zhang, and F. P. Chiang, "High-Speed 3D Shape Measurement Based on Digital Fringe Projection," *Opt. Eng.* **42**, 163–168 (2003).
9. P. S. Huang and S. Zhang, "A Fast Three-Step Phase Shifting Algorithm," *Appl. Opt.* (under press) (2006).
10. Q. Hu, P. S. Huang, Q. Fu, , and F. P. Chiang, "Calibration of a 3D Shape Measurement System," *Opt. Eng.* **42**, 487–493 (2003).
11. S. Zhang and P. S. Huang, "A Novel Structured Light System Calibration," *Opt. Eng.* (under press) (2006).
12. D. Malacara, ed., *Optical Shop Testing* (John Wiley and Sons, NY, 1992).
13. D. C. Ghiglia and M. D. Pritt, *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software* (John Wiley and Sons, Inc, 1998).
14. R. Legarda-Sáenz, T. Bothe, and W. P. Jüptner, "Accurate Procedure for the Calibration of a Structured Light System," *Opt. Eng.* **43**, 464–471 (2004).
15. D. A. Forsyth and J. Ponce, *Computer Vision-A Modern Approach* (Prentice-Hall, Inc., New Jersey, 2002).

1. Introduction

Accurate real-time 3D geometric shape measurement is increasingly important, with applications in manufacturing, inspection, entertainment, security, medical imaging, etc.

Optical 3D ranging methods [1], including stereovision, laser triangulation, and structured light, are extensively employed due to its non-contact and non-destructive nature. Among all existing ranging techniques, stereovision is probably the most studied method. However, the shortcoming of a stereo-based method is that the matching of stereo images is usually time-consuming. It is therefore difficult to realize real-time 3D shape reconstruction from stereo images. Unlike stereo-based methods, structured-light-based methods usually use much simpler processing algorithms. Therefore, it is more likely for them to achieve real-time performance.

There are essentially two approaches toward real-time 3D shape measurement. One approach is to use a single pattern, typically a color pattern [3, 4, 5]. Because they use color to code the patterns, the shape acquisition result is affected, to varying degrees, by the variations of the object surface color. Guan et al. proposed a method that uses a composite structured light pattern to realize real-time 3D shape measurement [6]. However, the resolution achieved is not high. The other approach is to use multiple patterns while switching them rapidly so that the number of patterns required to reconstruct one 3D shape can be captured in a short period of time. Rusinkiewicz et al. developed a real-time 3D model acquisition system that utilizes four patterns coded with stripe boundary codes [7]. The data acquisition speed achieved was 15 fps. However, like any binary coding method, the spatial resolution is relatively low in comparison with phase-shifting based methods. Huang et al. proposed a high-speed 3D shape measurement based on a rapid phase-shifting technique [8]. They use three phase-shifted, sinusoidal fringe patterns to achieve pixel-level resolution. Zhang and Huang developed a 3D shape acquisition system with a speed of up to 40 fps [2]. It is good enough to measure dynamic shapes within a small depth range. However, the system can only measure relative geometries. Therefore, it cannot measure rigid motion in depth beyond 2π . Moreover, because of the phase-to-height conversion algorithm used, the error is significant for large depth range measurement.

To our knowledge, it is very difficult for any existing 3D shape measurement system to obtain absolute coordinates in real-time with high resolution. This paper presents an improved version of the system developed by Zhang and Huang [2]. It is based on a digital fringe projection and fast phase-shifting method [9]. Three phase-shifted fringe patterns coded with three primary color channels (RGB) are sent to a DLP projector in B/W mode, and captured by a high-speed CCD camera synchronized with the projector. The data acquisition speed is 90 fps. Any successive three images can be used to reconstruct one 3D shape, therefore, the 3D data acquisition speed is 30 fps. This system can measure absolute coordinates of the object, hence, both geometric shapes and positions can be acquired. Since the absolute coordinates are measured, the system can significantly reduce the distortion error caused by the phase-to-height approximations. Therefore, the measurement depth range can be significantly larger, 700 mm for our system. The absolute coordinate measurement is possible only when absolute phase can be obtained. In order to obtain absolute phase, Hu et al. [10] and Zhang and Huang [11] used an additional centerline image. However, it is not desired for this research since our focus is to develop a high-speed system, while increasing the number of fringe images to reconstruct one 3D shape will decrease the measurement speed. In this research, a tiny cross marker is encoded in the projected fringe images, so that the 3D data acquisition speed is maintained but the absolute phase can be obtained. We propose a marker detection algorithm that takes advantage of the epipolar geometry, the image of one point from the projector is always imaged onto one line

in the camera image if no lens distortion exists. Our experiments demonstrate that more than 99% markers can be correctly detected using the proposed method. The absolute coordinates can be calculated from absolute phase if the system is calibrated.

Section 2 explains the principle. Section 3 shows experimental results. Section 4 concludes the paper.

2. Principle

2.1. Three-step phase-shifting algorithm

Phase-shifting methods are extensively employed in optical metrology, especially with the development of digital computers and digital display technologies. Three-step phase-shifting algorithms have the advantage of fast measurement because they require the minimum number of fringe images to reconstruct one 3D shape [12]. In this research, a three-step phase-shifting algorithm with a phase shift of $2\pi/3$ is used. The intensities of fringe images can be written as,

$$I_1(x,y) = I'(x,y) + I''(x,y) \cos[\phi(x,y) - 2\pi/3], \quad (1)$$

$$I_2(x,y) = I'(x,y) + I''(x,y) \cos[\phi(x,y)], \quad (2)$$

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

where $I'(x,y)$ is the average intensity, $I''(x,y)$ the intensity modulation, and $\phi(x,y)$ the phase. Solving Eqs. (1)-(3) simultaneously, we can obtain phase and data modulation,

$$\phi(x,y) = \tan^{-1} \left(\sqrt{3} \frac{I_1 - I_3}{2I_2 - I_1 - I_3} \right), \quad (4)$$

and

$$\gamma(x,y) = \frac{I''(x,y)}{I'(x,y)} = \frac{\sqrt{3(I_1 - I_3)^2 + (2I_2 - I_1 - I_3)^2}}{I_1 + I_2 + I_3}, \quad (5)$$

respectively. Where phase $\phi(x,y)$ in Eq. (4) is the so-called modulo 2π at each pixel, whose value ranges from 0 to 2π . If the fringe patterns contain multiple fringes, phase unwrapping is necessary to remove the sawtooth-like discontinuities and obtain a continuous phase map [13]. Once the continuous phase map is obtained, the phase at each pixel can be converted to xyz coordinates of the corresponding point on the object surface through calibration [14, 10, 11]. The average intensity $I'(x,y)$ represents a flat image of the measured object and can be used for texture mapping. Data modulation $\gamma(x,y)$ in Eq. (5) has a value between 0 and 1 and can be used to determine the quality of the phase data at each pixel with 1 being the best.

2.2. Marker detection

To reduce the measurement error caused by the marker, a tiny marker is used. However, locating the tiny marker from 2D image is generally difficult especially when the image is blurred. Figure 1(a) shows a typical example. The cross marker in the elliptical white window is barely visible. To detect the marker automatically, we generate a gamma map using Eq. (5), which is shown in Fig. 1(b), where the marker is more clearly stood out. However, in general, detecting the marker from the gamma map is not easy. In this research, a mask is created based on the quality of data, and is used to remove the background. The gamma map is also inverted: black to be white, and vice versa. Figure 1(c) shows the inverted gamma map after implementing the mask. Even though, our experiments showed that detecting cross accurately still failed sometimes. To improve the robustness of this procedure, a constraint is found to be helpful. From epipolar geometry [15], the ray of the cross center projected by the projector is imaged onto a line on the camera image if there is no camera lens distortion. Before any measurement, the

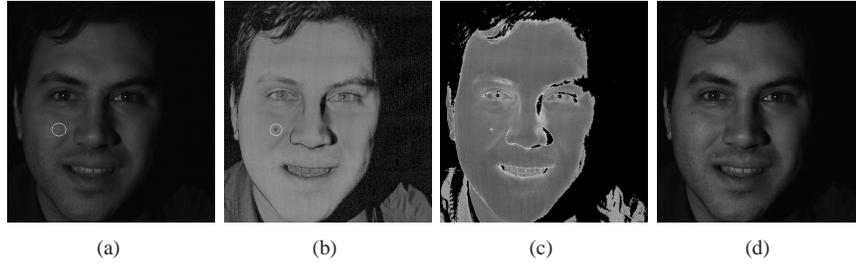


Fig. 1. Marker detection and removal. (a) 2D texture image with the marker. (b) Gamma map of the fringe images. (c) Inverted gamma map of the fringe images. (d) Texture image after the marker is removed.

cross epipolar line on the camera image is calibrated and stored for future use. To calibrate the line, a large white planar object posited at various distance, from extremely near to extremely far, from the scanner are imaged. The centers of the crosses on this series of images are found, which form the epipolar line. The marker searching only needs to be performed on the calibrated line, which makes the searching faster and more robust. We use a template-based method to find the marker center. The template we use is a 5×5 array,

$$f(x,y) = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 2 & 1 & 1 \\ 1 & 2 & 4 & 2 & 1 \\ 1 & 1 & 2 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

Assume the calibrated line is $y = h(x)$ for $x \in [1, w]$, where w is image width. Functional

$$g(x, h(x)) = \sum_{i=-2}^{i=2} \sum_{j=-2}^{j=2} \{ [w_g \times I_g(x+i, h(x)+j) + w_t \times I_t(x+i, h(x)+j)] f(i+2, j+2) \} \quad (6)$$

is defined to find the marker center. The center is located at the point where $g(x, h(x))$ reaches the maximum. Here, I_g is the inverted gamma map, I_t is the texture image, w_g is the weight for the gamma map, and w_t is the weight for the texture image. We consider both the texture and the gamma map to minimize the effect of noise and background. Our experiments show that more than 99% markers can be correctly detected using this method.

2.3. Absolute phase to coordinate conversion

Once the marker is detected, absolute phase can be obtained by shifting the whole phase map to make the marker point phase, ϕ_0 , to be absolute zero,

$$\phi_a(x, y) = \phi(x, y) - \phi_0(x_0, y_0). \quad (7)$$

Each point corresponds one line with the same absolute phase on the projected fringe image [11]. Therefore, we can establish a relationship between the captured fringe image and the projected fringe image,

$$\phi_a(u_c, v_c) = \phi_a(u_p). \quad (8)$$

We calibrate the system using the method proposed by Zhang and Huang [16, 11]. This method is essentially to make the projector capture images like a camera. The structured light system calibration therefore becomes a well-studied stereo system calibration. It is an elegant structured light system calibration method that significantly simplifies the system calibration

procedure. The intrinsic parameters matrices for the camera and the projector are A_c and A_p , respectively. The extrinsic parameter matrices for a fixed world coordinate system are M_c and M_p for the camera and the projector, respectively. Once the system is calibrated, the relationships between the world coordinate system and the camera and projector coordinate systems can be established, we obtain,

$$s_c[u_c, v_c, 1]^T = A_c M_c [X_w, Y_w, Z_w, 1]^T, \quad (9)$$

$$s_p[u_p, v_p, 1]^T = A_p M_p [X_w, Y_w, Z_w, 1]^T \quad (10)$$

s_c, s_p are camera and projector scaling factor, respectively, (u_c, v_c) and (u_p, v_p) are the camera and projector image coordinates, respectively, and (X_w, Y_w, Z_w) is the world coordinates.

In Eqs. (8)-(10), (X_w, Y_w, Z_w) , s_c, s_p, u_p , and v_p are unknowns, since there are seven equations, the world coordinate (X_w, Y_w, Z_w) can be uniquely determined.

2.4. Marker removal

When high quality 2D texture image is required, the marker on image needs to be removed. However, it is difficult to remove the marker without artifacts from the texture image. We found that from the properties of the phase-shifting algorithm, the marker can be completely removed.

Ideally, if no noises exist, data modulation ($\gamma(x, y)$ in Eq. (5)) should be always 1. On the other hand, image intensity on "cross" is I_0 for all images, $I_1 = I_2 = I_3 = I_0$. Hence $\gamma = 0$ for points on marker. Texture images can be obtained by averaging three fringe images,

$$t(x, y) = \begin{cases} (I_1 + I_2 + I_3)/3 = I' & (x, y) \in \text{fringe area} \\ I_0 & (x, y) \in \text{marker area} \end{cases} \quad (11)$$

We define a functional,

$$f(x, y) = (I_1 + I_2 + I_3)/3(1 + \gamma(x, y))/2 \quad (12)$$

$$= \begin{cases} I'(x, y) & (x, y) \in \text{fringe area} \\ I_0/2 & (x, y) \in \text{marker area} \end{cases} \quad (13)$$

if $I_0 = 2I'$, functional $f(x, y)$ is the same across image, and the marker is removed. The intensities of projected fringe images generated by the computer are,

$$I_1(x, y) = a + b(1 + \cos(\phi(x, y) - 2\pi/3)), \quad (14)$$

$$I_2(x, y) = a + b(1 + \cos(\phi(x, y))), \quad (15)$$

$$I_3(x, y) = a + b(1 + \cos(\phi(x, y) + 2\pi/3)). \quad (16)$$

To guarantee that the maker can be removed from captured fringe images, it has to satisfy $a + b = I_c$. Here I_c is the intensity value for the markers, which is 255 in our case. Figure 1(d) shows the result after removing the marker. It clearly shows that the marker is removed cleanly.

3. Experiments

To verify the performance of the system, we measured a human face. The data is acquired at 30 fps. Figure 2 shows six selected frames. During the experiment, the subject was asked to smile so that facial expressions are introduced. Movie in Fig. 3(a) and Fig. 3(b) show the feasibility of measuring geometric facial expressions. A heavily beard male face and a clean beautiful female face are measured. Movie in Fig. 3(c) shows the measurement results of a hand moving continuously over a depth range of approximately 700 mm. This video shows that the

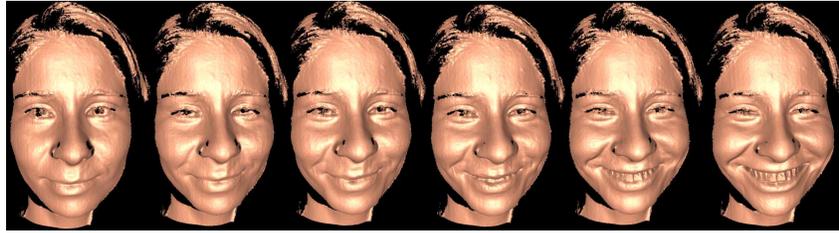


Fig. 2. Experimental result. Seven frames selected from a smiling sequence of a subject.



(a)

(b)

(c)

Fig. 3. (a) (2.1MB) Movie of facial expression. (b) (628KB) Movie of facial expression with texture. (c) (767KB) Movie of hand video moving over a depth range of 700 mm.

system can capture *absolute* coordinates of the object, namely, it can measure both geometric shapes and positions. It also shows that the feature details are clearly captured once the object is in focus and the fringe images are bright, while noisier when the object is far away from the focal plane, e.g., when the fringe images are defocused and darker. This is because when the fringe images are darker, the signal-to-noise is smaller. These experiments demonstrated that our system can measure both geometric shapes and positions for a relatively large depth range. More data is available at <http://math.harvard.edu/~songzhang>. Moreover, we measured a flat white board in focus and well illuminated, the measurement error was found to be RMS 0.10 mm. For a volume of 342(H) \times 376(V) \times 700(D) mm, the error is RMS 0.10-0.22 mm. It should be noted that we used a 3 \times 3 Gaussian filter for all data to smooth the most significantly random noises. If no filter is used, the measurement error is approximately RMS 0.15-0.30 mm over the same volume.

4. Conclusion

This paper presents a high-resolution, real-time 3D absolute coordinate measurement system based on a three-step phase-shifting method. The 3D shape measurement speed is 30 fps with 266K points per frame. Absolute 3D coordinates are obtained from three fringe images with a tiny marker. We propose a marker detection algorithm that takes advantage of epipolar geometry and utilizes the both the 2D texture image and the gamma map. Our experiments demonstrate that more than 99% markers can be correctly detected using the proposed method. Absolute coordinates are computed with the detected marker and the calibrated system parameters. We successfully demonstrate that our system can capture the details of facial expressions, both geometric shapes and positions of the hand moving over a depth range of 700 mm. To obtain high-quality texture images, we propose a method that can remove the markers cleanly. Although not relevant to metrology, it is important in applications such as computer graphics.

Acknowledgement

We thanks Dale Royer for helping generate video clips. The work was done in Geometric Informatics Inc. funded by the Advanced Technology Program of NIST.