

Three-dimensional shape measurement using a structured light system with dual cameras

Song Zhang, MEMBER SPIE

Shing-Tung Yau

Harvard University

Department of Mathematics

Cambridge, Massachusetts 02138

E-mail: szhang77@gmail.com

Abstract. A structured light system for three-dimensional shape measurement with single camera has the shortcoming of camera occlusion. To alleviate this problem, this paper introduces a structured light system with dual cameras for three-dimensional shape measurement. We discuss (1) system description, (2) system calibration, (3) three-dimensional data registration using the iterative closest-point (ICP) algorithm, and (4) three-dimensional data merging using holoimage. The principle of the system is introduced, and experiments are presented to verify its performance. © 2008 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.2835686]

Subject terms: calibration; structured light; dual cameras; holoimage; registration; phase shifting; merging.

Paper 070314R received Apr. 6, 2007; revised manuscript received Jul. 30, 2007; accepted for publication Aug. 8, 2007; published online Jan. 30, 2008.

1 Introduction

High-resolution accurate 3-D shape measurement is crucial for many applications, including medical imaging, inspection, and manufacturing.

Structured-light-based systems are more and more widely employed, due to their noncontact and noninvasive nature. However, for a structured light system using a single camera and a single projector, the field of view is limited. One of the typical problem of a single-camera-based structured light system is the occlusion caused by the camera. In certain areas, although the projector is able to illuminate the measured object, the camera cannot see it because the projector and the camera have to be at a certain distance and a certain angle to perform the measurement. This problem can be solved by turning the system or the object to another angle to cover the occlusion area of the camera. However, in certain applications, such as real-time measurement, that is not desirable. Another approach, which is adopted in this research, is to add another camera to the system to cover certain occlusions of the first camera. We are motivated to develop such a system by requests from different researchers.

We have been working on developing a system for real-time 3-D shape measurement with a single camera.¹ Our previous single-camera system is able to measure the absolute coordinates of dynamic 3-D shapes (such as facial expressions) in real time with very high quality.^{1,2} However, other researchers, especially those from industry, found that it is very difficult to use the data, since the rigid motion of the head is very difficult to obtain. This is called the *ear-to-ear* problem, because if the system can measure the surface area from left ear to right ear, the rigid motion of the head during experiment can be estimated easily. Another problem is to measure the nose. Due to occlusions of the camera, it is very difficult to measure two sides of the nose well. If a face is directly facing the projector, however, it

can be effectively illuminated from ear to ear.

In this research, we add another camera synchronized with the first camera to image the object from another angle. By this means, the two problems mentioned can be alleviated significantly. Since the second camera is synchronized with the first camera and acquires images independently, the measurement speed is not reduced. Our experiments demonstrate that the system is able to handle the occlusion problem near the nose, and the measurement areas are significantly larger. It can even solve the ear-to-ear problem for a human subject who can be illuminated by the projector from ear to ear.

On the other hand, adding another camera involves a number of tasks such as, (1) system calibration, (2) 3-D data registration, and (3) 3-D data merging. In this research, we extended the previously proposed calibration method for a structured-light system, due to Zhang and Huang,³ to the dual-camera system. The whole system is calibrated in the same world coordinate system, and corresponding pairs of 3-D data are well aligned. However, the 3-D data registration is still necessary for our data, due to the calibration errors of the system. In this research, we found that the iterative closest-point (ICP) algorithm is sufficient for registration.⁴⁻⁶ ICP is the dominant method for aligning three-dimensional models, based on the geometry, and sometimes the color, of the meshes; it is widely used for the output from 3-D shape measurement systems.⁷ The algorithm is very simple and is commonly used in real time systems. It iteratively estimates the transformation (translation, rotation) between two raw scans.

Once two patches of geometric surface are registered, the next step is to merge them into a single piece for future data analysis. Different software packages were tested to merge the data acquired by two cameras, including the volumetric range image processing (Vrip) package from Stanford (<http://graphics.stanford.edu/software/vrip/>), and a various software package from Geometry Systems Inc. (<http://www.geometrysystems.net/>). However, none of

these software packages can produce satisfactory results for our high-quality data. All of them smooth the geometry significantly and thereby lose details.

In order to preserve our data quality, we have proposed a technique called holoimage.⁸ Holoimage is a novel geometric representation; it encodes both shading and geometry information within the same image. Holoimage encodes the geometric information into fringe images similar to those in 3-D shape measurement techniques using phase-shifting algorithms. They are synthesized by the computer. Our experiments show that holoimage has much better performance than other software packages such as Vrip.

Section 2 introduces the phase-shifting algorithm used in this research. Section 3 discusses the system setup. Section 4 introduces the calibration of the whole system. Section 5 addresses the 3-D data registration problem. Section 6 introduces 3-D data merging using holoimage. Section 7 shows some experimental results, and Section 8 summarizes the work presented in this paper.

2 Three-Step Phase-Shifting Algorithm

Phase-shifting algorithms have been extensively employed for accurate and rapid 3-D shape measurement. Over the years, various phase-shifting algorithms have been developed, such as three-step, four-step, and five-step algorithms.⁹ In this research, we used a three-step phase-shifting algorithm for our system.

A three-step phase shifting algorithm with a phase shift of $2\pi/3$ can be written as

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

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

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

where $I'(x,y)$ is the average intensity, $I''(x,y)$ the intensity modulation, and $\Phi(x,y)$ the phase to be obtained. The phase $\Phi(x,y)$ can be solved for using the previous three equations:

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

3-D information is carried in the phase $\Phi(x,y)$. The value of $\Phi(x,y)$ obtained from Eq. (4) ranges from 0 to 2π . This step is called phase wrapping.

Phase unwrapping is necessary if multiple fringe stripes are used to obtain the continuous phase map.¹⁰ In this research, we utilize a rapid phase-unwrapping algorithm.¹¹ 3-D information can be obtained once the system is calibrated.^{3,12,13}

3 System Setup

A structured light system with a single camera and a single projector is typical for 3-D shape measurement. However, such a system setup has the problem of measurement shadow caused by the projector and occlusion caused by the camera. Figure 1 illustrates the problem of a structured light system with a single camera. In area A, although the projector is able to shine the structured pattern onto the

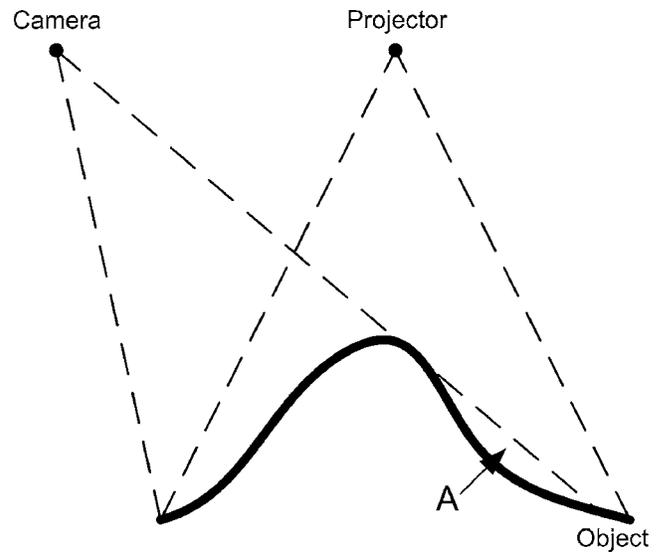


Fig. 1 Camera occlusion (area A) for a single-camera system.

measurement surface, the camera cannot see it. This is the occlusion caused by the camera. In this research, we add another camera from another viewing angle, as shown in Fig. 2, to alleviate this problem.

Figure 3 shows the system setup. The projector projects computer-generated fringe images onto the object; two cameras from different viewing angles capture deformed fringe images. Fringe images acquired by each camera are used to reconstruct one patch of the 3-D geometry through the phase-shifting algorithm. Finally a merged single set of 3-D mesh data is generated by software.

Figure 4 shows a photograph of the actual measurement system. In this system, the projector used is a digital light-processing (DLP) projector (PLUS U5-632h) with a resolution of 1024×768 . The focal range of the projector is $f = 18.4$ to 22.1 mm. The digital micromirror device (DMD) chip used for this projector is 0.7 in. The cameras are digital CCD cameras with an image resolution of 640×480 (Pulnix TM6740-CL). The camera sensor size is $7.4 \mu\text{m}$ (H) $\times 7.4 \mu\text{m}$ (V). Both cameras use Fujinon HF16HA-1B

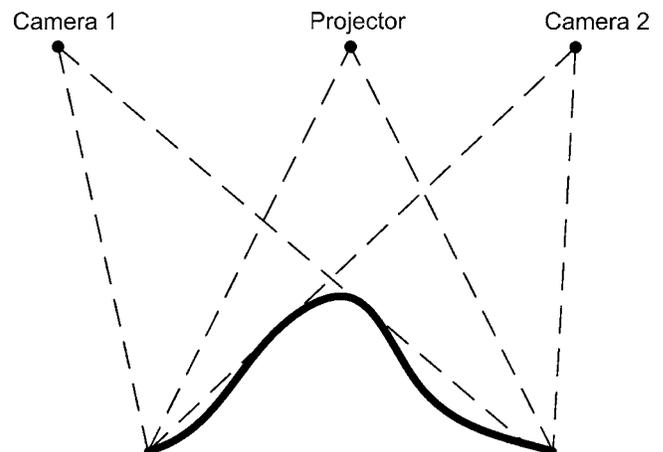


Fig. 2 A dual-camera system can alleviate the viewing problem of a single-camera system.

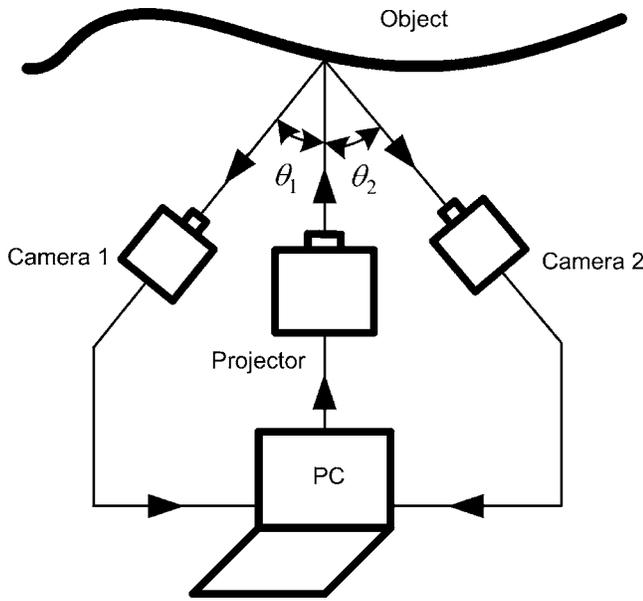


Fig. 3 Setup for a dual-camera system.

lenses with a focal length of 16 mm at $f/1.4$ to $f/16$. The exposure time used for the camera is approximately 2.78 ms. The frame grabber used in this system is a Matrox Solios XCL with CameraLink interface.

4 System Calibration

Calibration of a dual-camera system can be done by calibrating two single-camera structured light systems. However, if the two systems are calibrated independently, the 3-D data acquired by them will be difficult to align and merge. Since the relation between the two cameras is fixed, the calibration of the whole system can be performed in the same unique world coordinate system.

Figure 5 shows a diagram of the structured light system with dual cameras, where $(o_0^c; u^c, v^c)$ is the image coordinate system of camera 1, $(o^c; x^c, y^c, z^c)$ the camera 1 coordinate system, $(o^w; x^w, y^w, z^w)$ the world coordinate system, and f^c the focal length of camera 1; $(o_0^{c2}; u^{c2}, v^{c2})$

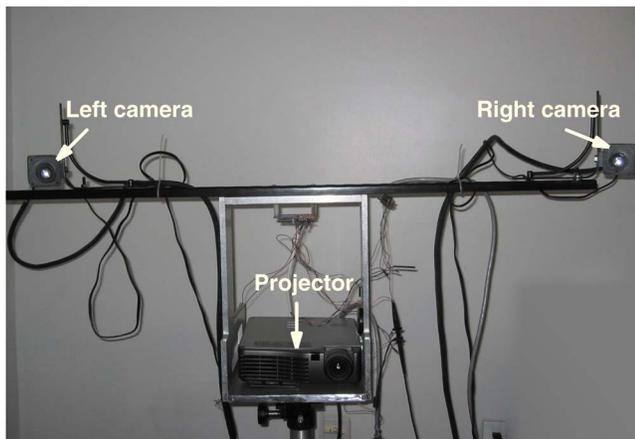


Fig. 4 Photograph of the actual system with dual cameras.

the image coordinate system of camera 2, $(o^{c2}; x^{c2}, y^{c2}, z^{c2})$ the camera 2 coordinate system, and f^{c2} the focal length of camera 2; $(o_0^p; u^p, v^p)$ the image coordinate system of the projector, $(o^p; x^p, y^p, z^p)$ the projector coordinate system, and f^p the focal length of the projector.

The calibration is to estimate the intrinsic parameters and the extrinsic parameters of the system. In this research, we implemented the system calibration method proposed by Zhang and Huang³ for each camera-projector pair. The camera model as well as the projector model used here is the linear model described by Zhang.¹⁴ For this model, the calibration is to find the relationship between the world coordinates and the image coordinates,

$$s\mathbf{I} = \mathbf{A}[\mathbf{R}, \mathbf{t}]\mathbf{X}^w, \tag{5}$$

where $\mathbf{I}=[u, v, 1]^T$ is the homogeneous image coordinates, $\mathbf{X}^w=[x^w, y^w, z^w, 1]^T$ is the homogeneous world coordinates, and s is a scale factor; $[\mathbf{R}, \mathbf{t}]$ is the extrinsic parameter matrix, where \mathbf{R} (a 3×3 matrix) and \mathbf{t} (a 3×1 vector) respectively represent the rotation and translation between the world coordinate system and the camera coordinate system; and \mathbf{A} is the camera intrinsic parameter matrix and can be expressed as

$$\mathbf{A} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{6}$$

where u_0, v_0 are the coordinates of the principal point, α and β are the focal lengths along the u and v axes of the image plane, and γ is the parameter that describes the skewness of the two image axes.

The unique world coordinates can be established by choosing the same pose of the calibration board to establish the world coordinates for all three devices (two cameras and one projector). Figure 6 shows the world coordinate selection from the images of the same calibration pose. The same physical point on the image is chosen as the origin of the coordinate system, the x and y axes have the same physical directions on the planar calibration board, and z is perpendicular to the calibration board. The projector calibration images can be generated by the method introduced in Ref. 3. Once a set of calibration images for each camera (37 images) and the calibration images (73 images) for the projector are generated, the intrinsic parameters of each sensor can be calibrated using the traditional camera calibration method and the MATLAB Toolbox.¹⁵ By selecting images in the same calibration pose, the extrinsic parameters can be calibrated. After calibration, the intrinsic parameters for the two cameras are

$$\mathbf{A}^{c1} = \begin{bmatrix} 16.22 \pm 0.09 & 0 & 2.05 \pm 0.03 \\ 0 & 16.29 \pm 0.09 & 1.75 \pm 0.04 \\ 0 & 0 & 1 \end{bmatrix} \text{ mm},$$

and

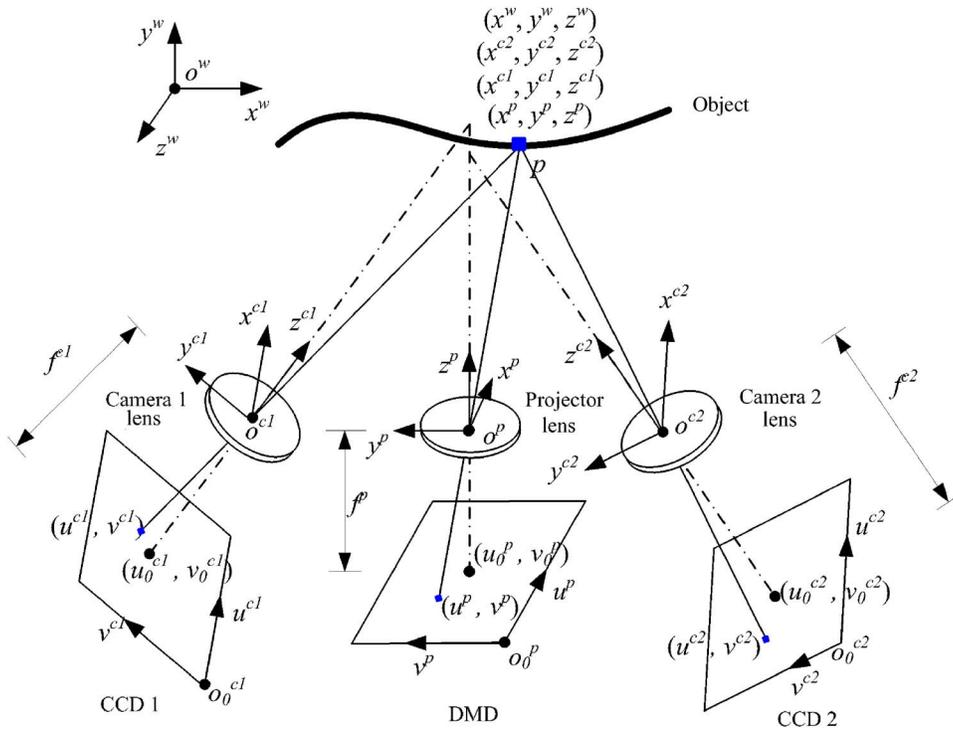


Fig. 5 Structured-light system with dual cameras using pinhole camera models.

$$\mathbf{A}^{c2} = \begin{bmatrix} 16.43 \pm 0.08 & 0 & 2.37 \pm 0.03 \\ 0 & 16.48 \pm 0.08 & 1.84 \pm 0.04 \\ 0 & 0 & 1 \end{bmatrix} \text{ mm},$$

respectively, and the projector intrinsic parameters are

$$\mathbf{A}^p = \begin{bmatrix} 22.22 \pm 0.14 & 0 & 13.36 \pm 0.09 \\ 0 & 22.22 \pm 0.14 & 5.09 \pm 0.11 \\ 0 & 0 & 1 \end{bmatrix} \text{ mm}.$$

The corresponding extrinsic parameters are

$$\mathbf{R}^{c1} = \begin{bmatrix} 0.006025 & 0.999981 & 0.001533 \\ 0.951931 & -0.006205 & 0.306250 \\ 0.306253 & -0.000386 & -0.951950 \end{bmatrix},$$

$$\mathbf{t}^{c1} = [-119.26, -75.83, 1059.92]^T \text{ mm};$$

$$\mathbf{R}^{c2} = \begin{bmatrix} 0.023355 & 0.999621 & 0.014566 \\ 0.849628 & -0.012168 & -0.527242 \\ -0.526865 & 0.024689 & -0.849591 \end{bmatrix},$$

$$\mathbf{t}^{c2} = [-124.84, -85.43, 1136.13]^T \text{ mm};$$

and

$$\mathbf{R}^p = \begin{bmatrix} 0.018930 & 0.998007 & -0.060205 \\ 0.993293 & -0.025642 & -0.112741 \\ -0.114060 & -0.057667 & -0.991799 \end{bmatrix},$$

$$\mathbf{t}^p = [-393.72, -100.14, 1053.32]^T \text{ mm},$$

respectively.

Once the intrinsic and extrinsic parameters of the system are calibrated, each camera-projector pair is able to generate one patch of the 3-D geometry in the same world coordinate system. Here we have equations

$$s^c \mathbf{I}^c = \mathbf{A}^c [\mathbf{R}^c, \mathbf{t}^c] \mathbf{X}^w, \quad (7)$$

$$s^p \mathbf{I}^p = \mathbf{A}^p [\mathbf{R}^p, \mathbf{t}^p] \mathbf{X}^w. \quad (8)$$

There are five unknowns, s^c, s^p, x^w, y^w , and z^w , in these equations. Since there are three equations in Eq. (7), and two additional equations in Eq. (8) using absolute phase, the world coordinates for each point can be solved for uniquely.

We found that due to the errors of the calibration, the two systems are not in exactly the same world coordinate system; they differ by a small translation (approximately 10 mm in the x direction, 2 mm in the y direction, and 24 mm in the z direction). This is understandable from the calibration procedures used in this research. As previously mentioned, each projector-camera pair is calibrated independently. As explained in our previous paper,³ the world coordinate system for each projector-camera pair is established by selecting the origin and the directions of the x and y axes from a pair of checkerboard images. In theory the projector image can be generated accurately pixel by pixel from the camera image; however, due to the sampling errors, the mapping is not 100% accurate, and it varies from position to position. Therefore, when we select the origin for the left projector-camera pair and that for the right

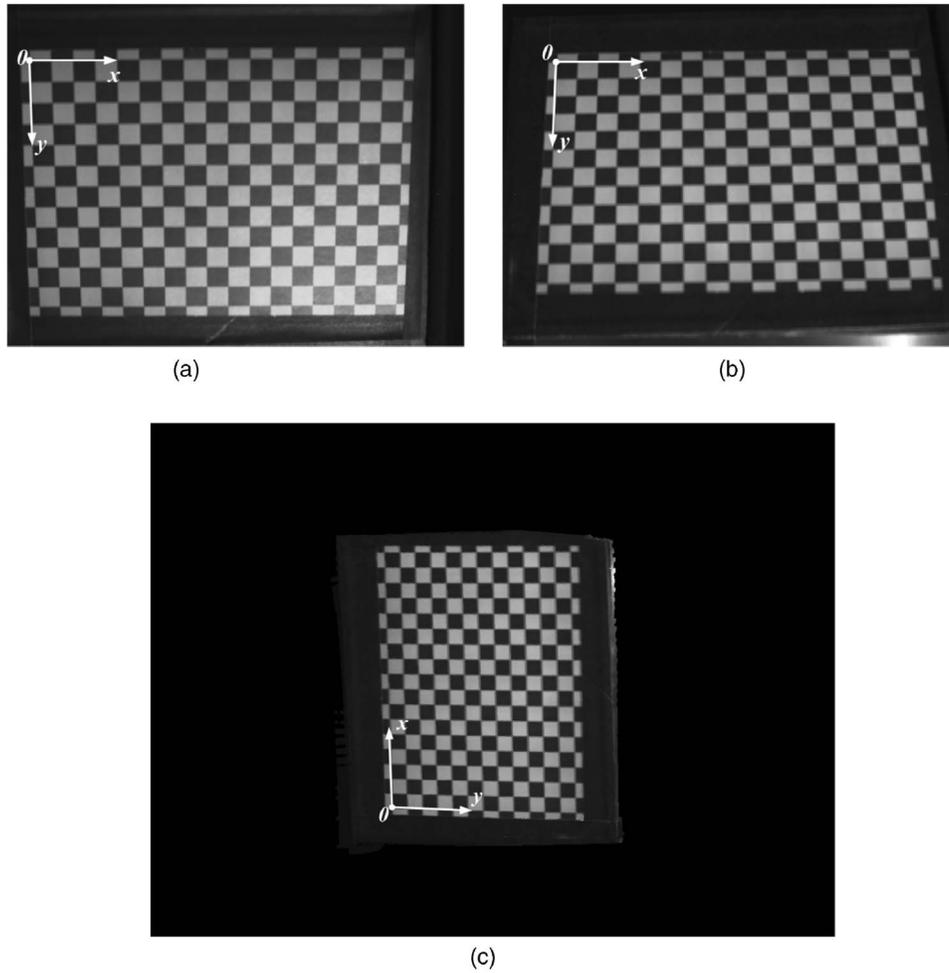


Fig. 6 System coordinate establishment: world coordinate system on (a) camera 1 calibration image, (b) camera 2 calibration image, (c) projector calibration image.

projector-camera pair, we cannot guarantee that the two are exactly the same. Hence, the two projector-camera pairs produce 3-D coordinates in different coordinate systems.

Our experiments found that there is no noticeable rotation between the two coordinate systems. Therefore, we believe that the coordinate inconsistency was introduced by the selection of the coordinate origins. Experiments revealed that this translation is fixed for the system; therefore, it can be determined by selecting a measurement point and finding the difference between its coordinates in the two systems.

5 3-D Registration

Ideally, after calibration, two patches acquired by two cameras should have the same coordinates for any common point. However, due to the calibration error and/or measurement error, the two patches are not aligned well. One way to solve this problem is to use a more complicated system calibration approach with an error lookup table (LUT), such as the approach introduced by Hu et al.¹³ Even so, the 3-D coordinates cannot be guaranteed to match well enough. Thus, 3-D data registration is necessary to obtain high-quality single-piece data. In this research, we used the iterative closest-point (ICP) algorithm to do 3-D registra-

tion. Since the two data patches from the two cameras are already roughly aligned, the 3-D data registration need only refine the alignment. The registration is basically to find the rigid motion between the two patches, which is a 3×4 matrix.

$$\mathbf{T} = [\mathbf{R}, \mathbf{t}], \tag{9}$$

where \mathbf{R} is a 3×3 rotation matrix and \mathbf{t} is the translation vector.

For our case, we have two geometry data to register, and the ICP algorithm is to find the matrix iteratively by finding the best match points. Assume two point-cloud surfaces S^1 and S^2 from the two surface patches, respectively, and let \mathbf{x}_i^1 and \mathbf{x}_j^2 be the points on S^1 and S^2 , respectively. The objective is to find the translation vector \mathbf{t} and rotation matrix \mathbf{R} such that the criterion

$$F(\mathbf{R}, \mathbf{t}) = \sum_{i=1}^{N^1} p_i d(\mathbf{R}\mathbf{x}_i^1 + \mathbf{t}, S^2) + \sum_{j=1}^{N^2} q_j d(\mathbf{R}^T \mathbf{x}_j^2 - \mathbf{R}^T \mathbf{t}, S^1) \tag{10}$$

is minimized, where $d(\mathbf{x}, S)$ denotes the distance from point \mathbf{x} to surface S , and where p_i takes value 1 if point \mathbf{x}_i^1 can be

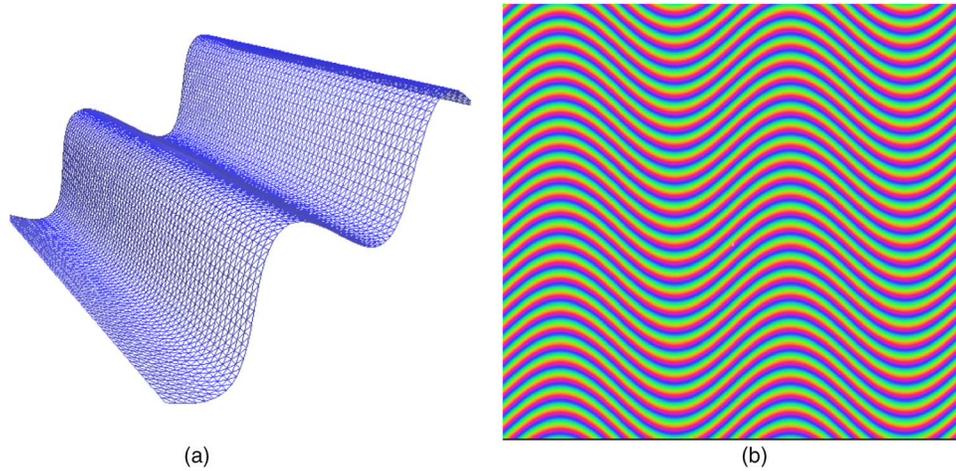


Fig. 7 Example of holoimage. (a) Sinusoidal profile, 3-D geometry. (b) The holoimage of the sinusoidal profile geometry (projection angle: 30 deg; fringe pitch: 16; holoimage size: 512×512).

matched to a point on surface S^2 in the second frame, and value 0 otherwise. The minimization of Eq. (10) must simultaneously satisfy the maximization of

$$\sum_{i=1}^{N^1} p_i + \sum_{j=1}^{N^2} q_j.$$

Otherwise, no matching point would be the optimal solution. To make the computation more efficient, we only find matching points on the surface patch S^2 for points on surface patch S^1 , with the maximization of $\sum_{i=1}^{N^1} p_i$. In other words, the ICP solution is to minimize the functional

$$F(\mathbf{R}, \mathbf{t}) = \frac{1}{\sum_{i=1}^{N^1} p_i} \sum_{i=1}^{N^1} p_i d(\mathbf{R}\mathbf{x}_i^1 + \mathbf{t}, S^2). \quad (11)$$

However, in practice, this is too slow, since the registration needs to be performed for every point on the surface patch S^1 to find its nearest point on the second patch S^2 . This is a very time-consuming procedure. In order to reduce the convergence time, the ICP algorithm is performed locally, which means that the registration is performed within a small region on S^2 . Assume the normal to an arbitrary point \mathbf{x}_i^1 on S^1 is \mathbf{n}^1 , and to an arbitrary point \mathbf{x}_j^2 on S^2 is \mathbf{n}^2 . If the angle between \mathbf{n}^1 and \mathbf{n}^2 is α , the search is performed if α is smaller than certain angle (45 deg for our case). Also, the search is performed only within a small area on the surface S^2 (within a sphere of radius 10 mm centered on the initial matching point). These two constraints enable the ICP algorithm to run efficiently. Our experiments demonstrated that this ICP algorithm works well if good initial starting matching points are given, which is the case for our system.

6 3-D Data Merging Using Holoimage

Once two 3-D data are registered, the next step is to merge them into one single data and resolve the overlapping areas. In an overlapping area, detecting the closest points between two patches and unifying them into one would be one so-

lution. However, it is very difficult for this approach to guarantee surface smoothness, because one point on surface S^2 might correspond to multiple points on surface S^1 . In this research, we used a technique called *holoimage* to merge the 3-D data.

6.1 Holoimage

The holoimage is a novel geometric representation introduced by Gu et al.⁸ It encodes both shading and geometry information within the same image. A holoimage encodes the geometric information into fringe images similarly to 3-D shape measurement techniques using phase-shifting algorithms. The holoimage is synthesized by the computer. However, there are fundamental differences between the synthesized holoimage and the captured fringe images in the real world. The projective texture mapping (corresponding to the projector in the real world) of a synthetic holoimage does not include shadows or self-occlusion. Another difference is related to color. Three monochromatic fringe projective textures can be combined into one color projective texture to generate one 24-bit color holoimage. However, using color fringe images in real measurement is undesirable, since the measurement accuracy is affected by the color of the object due to the problem of color coupling of the projector and the camera. Therefore, three monochromatic fringe images are usually needed to reconstruct the geometry.

It is very easy to synthesize a holoimage using a modern graphics pipeline. Three sinusoidal fringe patterns can be precomputed and stored as a three-channel 24-bit color texture image. In order to simplify the analysis of the holoimage, a canonical configuration is preferred, where both the projective texture and the camera use orthogonal projection, and the geometric object is normalized to be inside a unit cube. The color holoimage is encoded as

$$I_r(x, y) = \frac{255}{2} \left[1 + \cos \left(\frac{2\pi x}{P} - \frac{2\pi}{3} \right) \right], \quad (12)$$

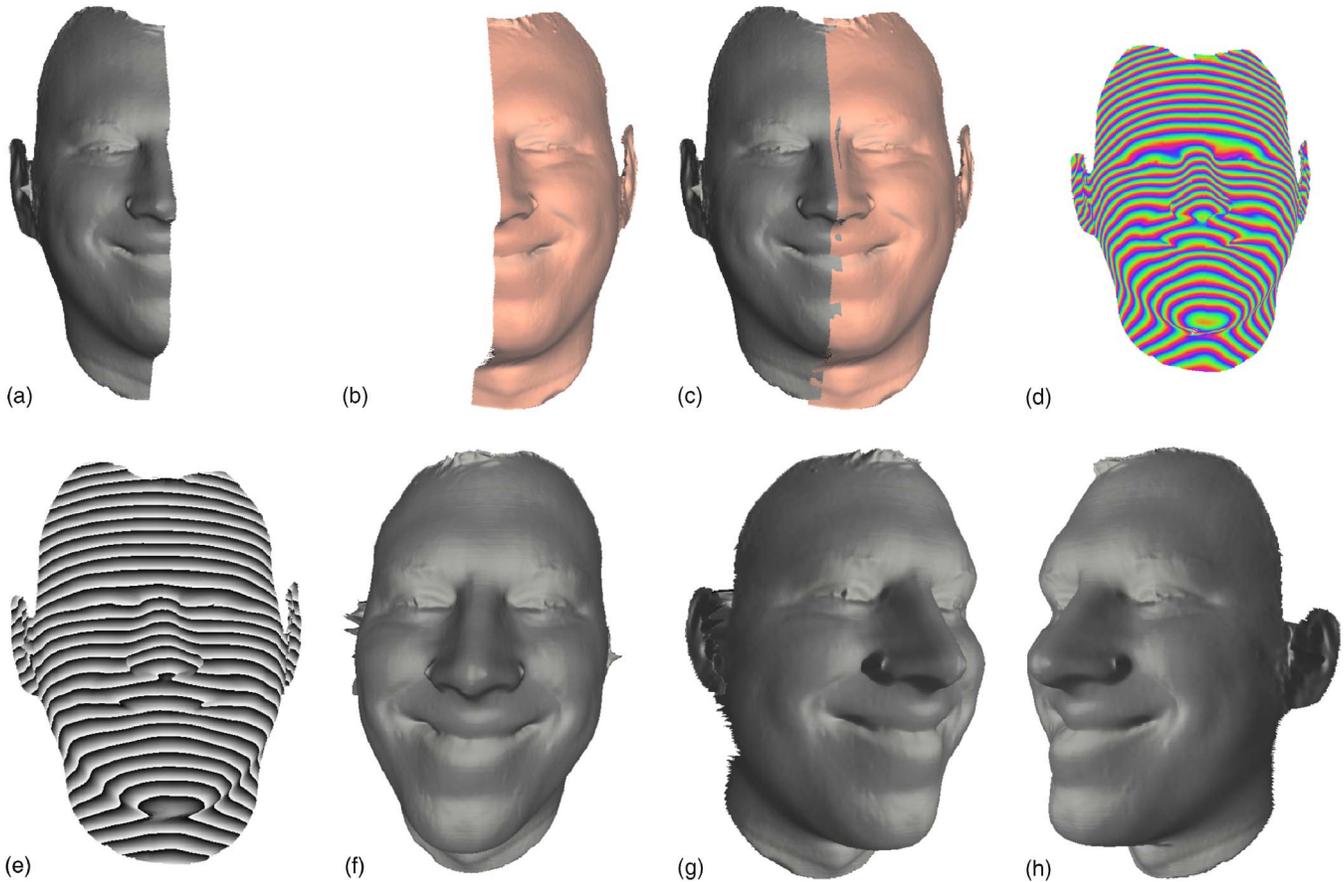


Fig. 8 3-D data merging using holography. (a) Left half face. (b) Right half face. (c) Two sides of faces rendered together in the same OpenGL scene. (d) Holoimage. (e) Phase computed from the holoimage. (f) 3-D geometry obtained from holoimage and rendered with perspective model. (g), (h) Two viewing angles of the resultant 3-D geometry after merging.

$$I_g(x,y) = \frac{255}{2} \left[1 + \cos\left(\frac{2\pi x}{P}\right) \right], \quad (13)$$

$$I_b(x,y) = \frac{255}{2} \left[1 + \cos\left(\frac{2\pi x}{P} + \frac{2\pi}{3}\right) \right], \quad (14)$$

where P is the fringe pitch (number of pixels per fringe period) used. The projected fringe images are distorted by the object virtually and rendered on the screen. Figure 7 shows an example of a sinusoidal-shaped object and its holoimage.

In the computation, if we only care about geometric information and do not wish to represent any shading or texture for the surface, we can set the OpenGL texture environment mode to *replace*. If we want to encode both geometry and shading information, we should set the texture environment mode to *modulate*. If a texture is also to be rendered on the surface, we need to use a multitexturing technique to generate the holoimage. Of course, if color texture is desirable, a color holoimage may bring errors and a monochromatic fringe image may have to be adopted.

6.2 3-D Data Merging

Two sets of 3-D data can be rendered together in the same scene using OpenGL as shown in Fig. 8(c), where the darker-color left half [as shown in Fig. 8(a)] represents the geometry from the left camera, and the brighter-color right

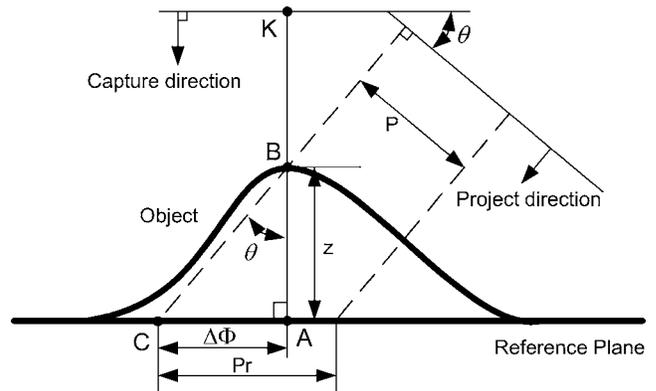


Fig. 9 Schematic diagram of a holography system using an orthogonal projection model.

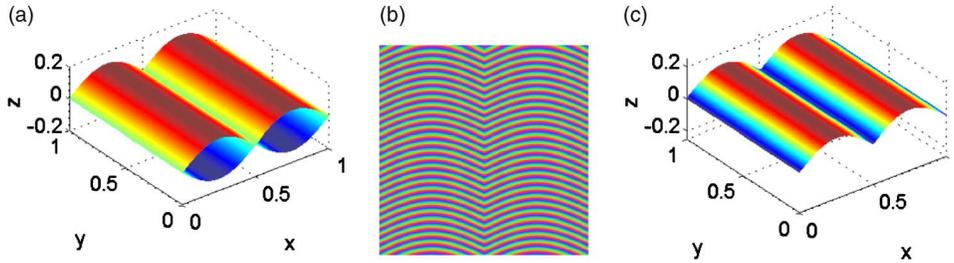


Fig. 10 Example of computing simulated ideal surfaces and merging them.

half [as shown in Fig. 8(b)] represents the geometry from the right camera. The color fringe images are projected onto the object virtually, using the orthogonal projection, and the rendered screen is recorded [image shown in Fig. 8(d)]. For this setup, only a single hologram is necessary for merging, since the captured geometry can be illuminated by a single projector. By adjusting the viewing angle of the rendering, the measurement data from two cameras can be seen completely. Since the angle between the virtual projector and the camera is precisely known and the projection is orthogonal, the 3-D geometry can be recovered easily, from the phase computed from the hologram, using a phase-to-height conversion algorithm, which is introduced in the next subsection.

6.3 Phase-to-Height Conversion

Figure 9 shows a diagram of the hologram system. Before any measurement, a reference plane is set. The reference plane is a plane with height 0 in the depth, or z , direction. Subsequent measurement is relative to the reference plane. An arbitrary point K in the captured image is assigned to point A on the reference plane and B on the object surface. The distance between A and B is the depth z of the measured object, $z=AB$. The corresponding phase for the reference plane is Φ_A if no object is placed and Φ_B when the object is there. From the projector point of view, the phase Φ_B on the object is the same as Φ_C on the reference plane, that is, $\Phi_C=\Phi_B$. The phase difference between points A and C is $\Delta\Phi=\Phi_A-\Phi_C$. Since the fringe image is uniformly distributed on the reference plane, the actual distance is

proportional to the phase difference of the fringe images. In other words, $AC=k \Delta\Phi$. Because ΔABC is a right triangle, we have

$$z = \overline{AC} / \tan \theta = k \Delta\Phi / \tan \theta$$

$$= k(\Phi_A - \Phi_C) / \tan \theta = k(\Phi_A - \Phi_B) / \tan \theta.$$

Assume the projection fringe has a fringe pitch of P . Once it is projected onto the reference plane from an angle of θ , it becomes $P_r = P / \cos \theta$. Assume the physical size of a pixel in millimeters is c , which is determined by the projection window size and the setup of the OpenGL scene. The constant k therefore becomes

$$k = \frac{cP_r}{2\pi} = \frac{cP}{2\pi \cos \theta}.$$

Hence,

$$z = \frac{cP(\Phi_A - \Phi_B)}{2\pi \sin \theta}. \tag{15}$$

The x and y values are proportional to their indices in the x and y directions with a constant of proportionality equal to the pixel size c .

6.4 Simulation Result

To show that the hologram technique can successfully merge two surfaces, we simulate two ideal surfaces and merge them using hologram. A typical surface profile, a sinusoidal one, is used for this purpose. Figure 10(a) shows

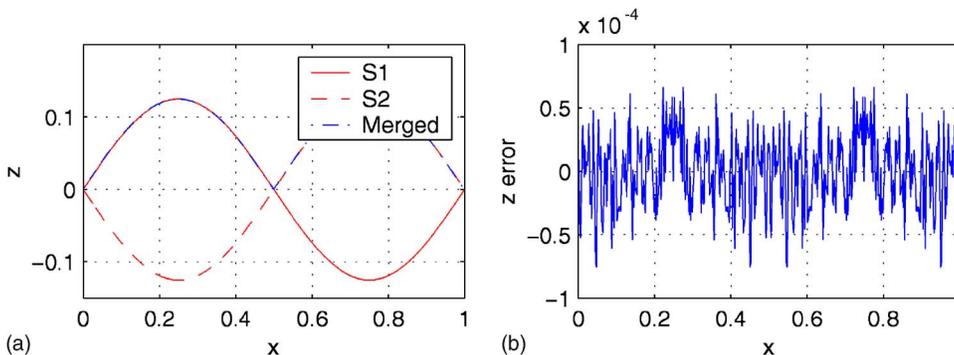


Fig. 11 Cross sections and the merging error. (a) The corresponding cross sections. (b) Merging error (approximately 2.9×10^{-5}).

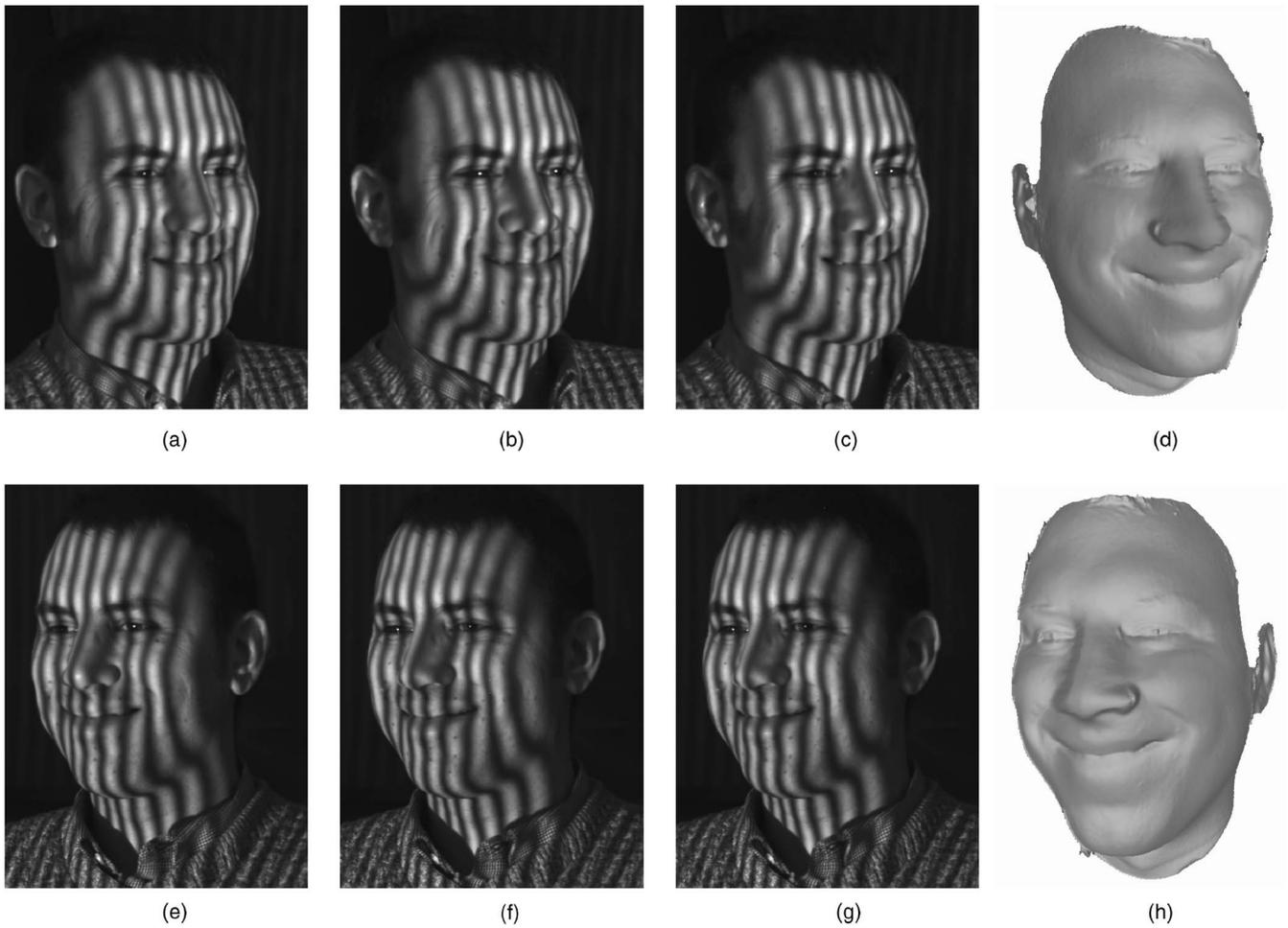


Fig. 12 3-D data acquired by two cameras. (a) to (c) are three fringe images acquired by the left camera. (d) shows the 3-D geometry acquired by the left camera. (e) to (g) are three fringe images acquired by the right camera, and (h) shows the 3-D geometry acquired by the right camera.

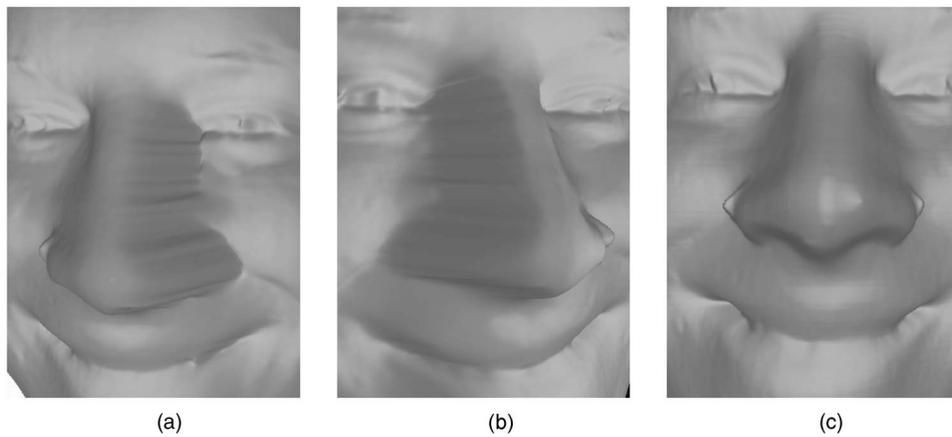


Fig. 13 Problem of occlusion. (a) The left camera cannot see the right side of the nose, where the profile is not correctly measured. (b) The right camera cannot see the left side of the nose, where the profile is not correctly measured. (c) After merging, both sides of the nose are properly measured.

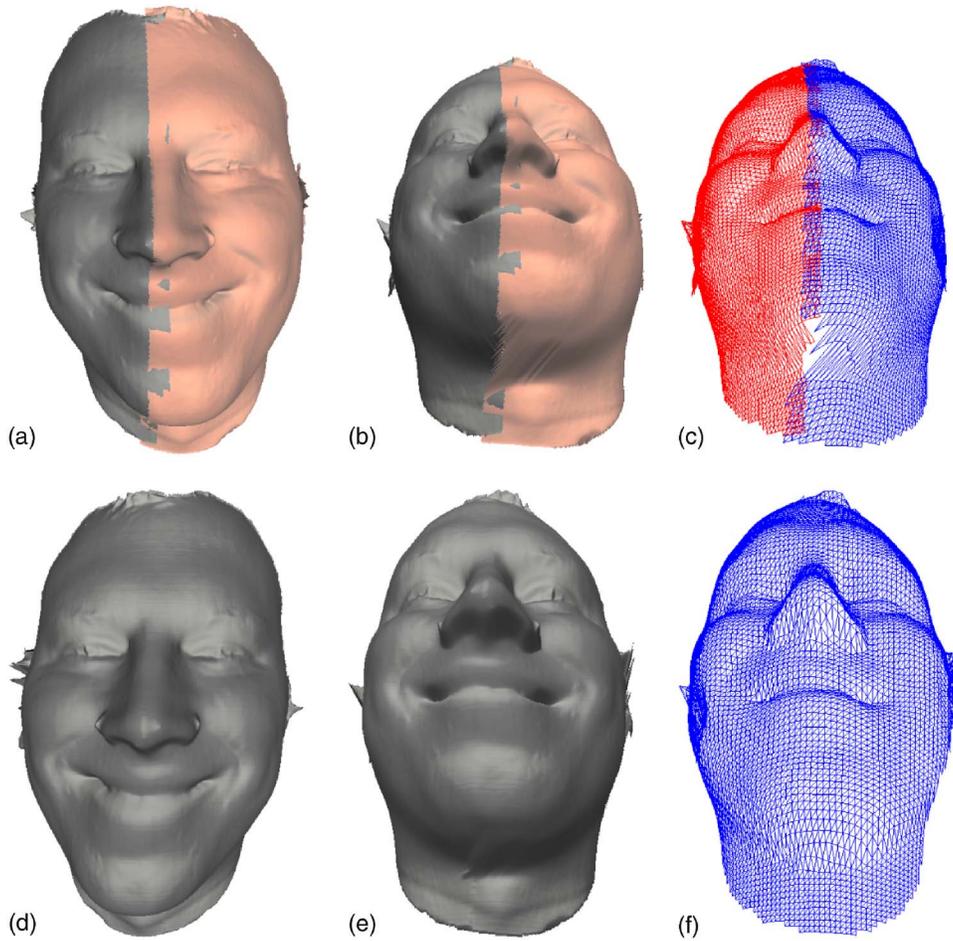


Fig. 14 Comparison between before and after merging. The first row shows the 3-D geometries before merging and rendered in the same scene. The second row shows the 3-D geometries after merging.

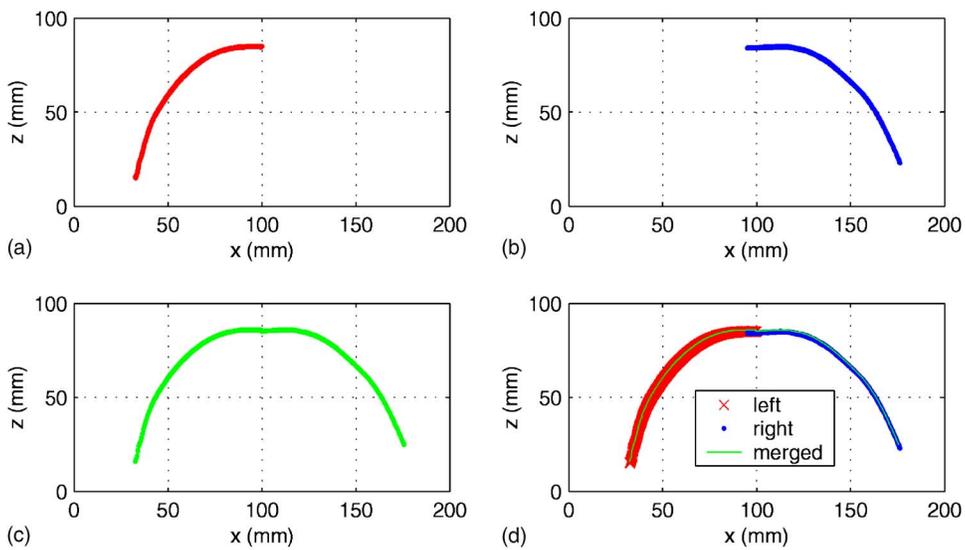


Fig. 15 Cross sections of the data before and after merging with $59.5 < y < 60.5$ mm: (a) from the 3-D data acquired from the left camera; (b) from the 3-D data acquired from the right camera; (c) from the 3-D data after merging. (d) Cross sections in (a) to (c) plotted on the same graph.

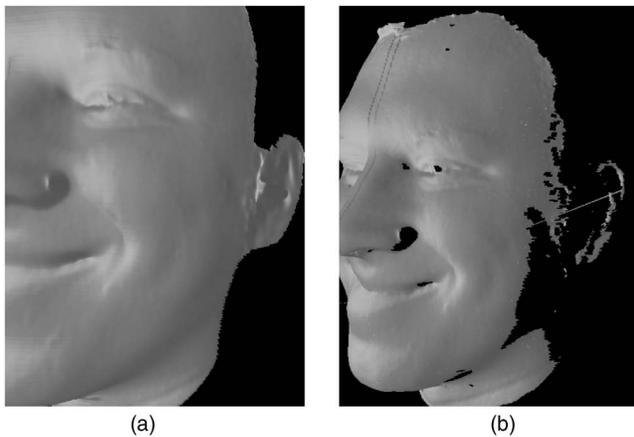


Fig. 16 Comparison between the results using (a) holoimage and (b) Stanford Vrip for 3-D geometry.

an example of merging two sinusoidal surfaces:

$$S1 : z = 0.125 \sin(2\pi y), \quad (16)$$

$$S2 : z = -0.125 \sin(2\pi y). \quad (17)$$

Figure 10(b) shows its holoimage, and Fig. 10(c) shows the merged result. This figure shows that only the front surface is generated after merging. Figure 11 shows their cross sections and the corresponding error. The absolute error is approximately 2.9×10^{-5} , and the relative error is approximately

$$2.9 \times 10^{-5} / 0.125 = 2.32 \times 10^{-4} = 0.02\%,$$

which is very small and can be neglected. The simulation result shows that the holoimage can successfully merge two surfaces without producing obvious errors.

7 Experiments

To verify the performance of the system, we measured a face using our system. Figure 12 shows two sides of the face acquired by two cameras independently. We can see here that due to camera occlusions, the profile of the nose was not well measured by either camera, as shown in Fig. 13(a) and 13(b). The results of measurement of the two face patches by the two cameras are refined using the ICP algorithm and then cut vertically near the center of the nose (this cutting is done automatically by detecting the normal directions of the surfaces). The geometry measured by the left camera keeps the left half, while the geometry measured by the right camera keeps the right half, which are shown in Fig. 8(a) and 8(b), respectively. Two half geometries are rendered using OpenGL in the same window whose holoimage is generated [Fig. 8(d)]. The 3-D geometry can be generated from the holoimage using phase wrapping, phase unwrapping, and the phase-to-depth conversion algorithm. Figure 8(f)–8(h) show the merged result.

Figure 13 shows a comparison near the region of the nose measured by each camera and the results after merging. It can be seen that before merging, the nose profile cannot be properly measured by either camera, while the result is satisfactory after merging. This experiment dem-

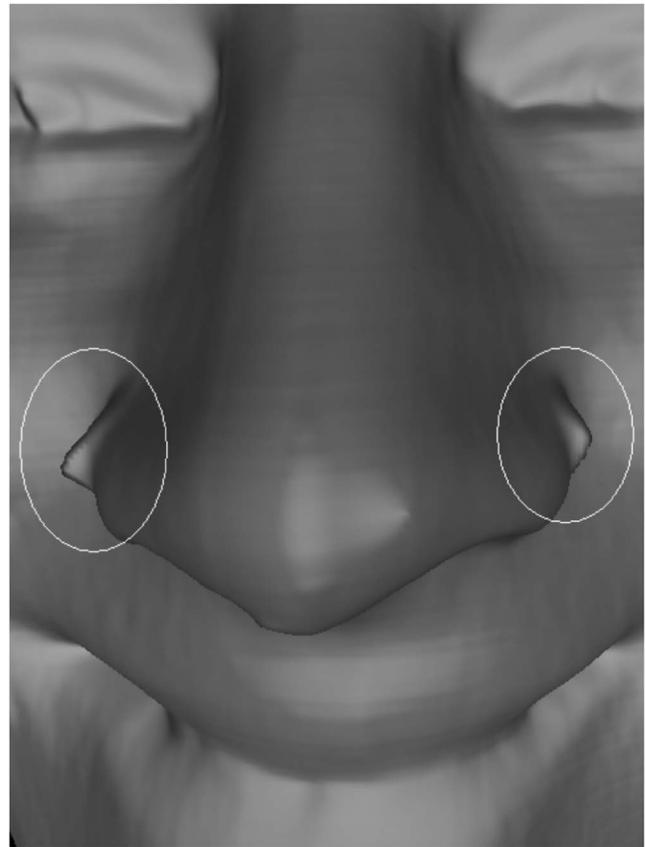


Fig. 17 Measurement error caused by the projector shadow.

onstrated that the system with dual cameras can alleviate the camera occlusion problem significantly. Figure 14 shows the comparison before and after merging in 3-D rendering mode. Figure 15 shows one cross section of the 3-D data before and after merging, with y ranging from 59.5 to 60.5 mm. The data before merging and after merging are overlapped well. Hence, the merging algorithm works successfully for real measured data.

Figure 16 shows the comparison between the merging results with holoimage and with the Stanford Vrip software package (GSI gives a similar result). It can be seen clearly that Vrip removed a lot of useful information from the ear region as well as the face, while holoimage preserved it. Vrip generated a number of holes in the data, while holoimage did not. The data quality is much higher for the data generated by holoimage, and the file size is only approximately 2/3 of that generated by Vrip.

It should be noted that the structured light system with dual cameras cannot solve the problem created by the projector. Figure 17 illustrates the problem, the geometric profiles enclosed in ellipses are not well measured. This is due to the fact that close to the region, the projection light is close to the tangent to the object surface, so that the phase is changing dramatically and cannot be extracted correctly.

8 Conclusions

We have presented a structured light system with dual cameras intended to obviate the camera occlusions of a single-camera system. We addressed the system calibration and

3-D data registration using the ICP algorithm and proposed a novel 3-D data merging using the holoimage technique. We demonstrated that a dual camera system would significantly alleviate the camera occlusion problem; for example, it could measure the profile of the nose properly. Moreover, the system could measure a larger area, even the whole face from ear to ear, which is good for estimating the rigid motion of the head for facial data analysis. In addition, because two cameras can work simultaneously and independently, the dual camera system is especially good for real-time applications where rotating the object to avoid the camera occlusion is not desirable. The data acquisition speed is not affected by the measurement, while the measurement area is much larger. Experiments were presented to demonstrate the performance of the proposed system.

Acknowledgments

The author would like to thank the team at Geometric Informatics Inc. (GI) for their assistance and discussions. This work was done at GI.

References

1. S. Zhang and P. S. Huang, "High-resolution, real-time three-dimensional shape measurement," *Opt. Eng.* **45**, 123601 (2006).
2. S. Zhang and S.-T. Yau, "High-resolution, real-time 3D absolute coordinate measurement based on a phase-shifting method," *Opt. Express* **45**, 2644–2649 (2006).
3. S. Zhang and P. S. Huang, "Novel method for structured light system calibration," *Opt. Eng.* **45**, 083601 (2006).
4. Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image Vis. Comput.* **10**, 145–155 (1992).
5. P. Besl and N. McKay, "A method for registration of 3-D shapes," *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 239–256 (1992).
6. Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *Int. J. Comput. Vis.* **13**, 119–152 (1994).
7. S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *Proc. 3rd Int. Conf. on 3-D Digital Imaging and Modeling*, pp. 145–152, IEEE (2001).
8. X. Gu, S. Zhang, P. Huang, L. Zhang, S.-T. Yau, and R. Martin, "Holoimages," in *SPM '06: Proc. 2006 ACM Symp. on Solid and Physical Modeling*, pp. 129–138 (2006).
9. D. Malacara, Ed., *Optical Shop Testing*, John Wiley and Sons, New York (1992).
10. D. C. Ghiglia and M. D. Pritt, *Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software*, John Wiley and Sons (1998).
11. S. Zhang, X. Li, and S.-T. Yau, "Multilevel quality-guided phase unwrapping algorithm for real-time three-dimensional shape reconstruction," *Appl. Opt.* **46**, 50–57 (2007).
12. R. Legarda-Sáenz, T. Bothe, and W. P. Jüptner, "Accurate procedure for the calibration of a structured light system," *Opt. Eng.* **43**(2), 464–471 (2004).
13. Q. Hu, P. S. Huang, Q. Fu, and F. P. Chiang, "Calibration of a three-dimensional shape measurement system," *Opt. Eng.* **42**, 487–493 (2003).
14. Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000).
15. J.-Y. Bouguet, "Camera calibration toolbox for MATLAB." Online. Available at: http://www.vision.caltech.edu/bouguetj/calib_doc.



Song Zhang is a research fellow at Harvard University. He received his doctoral degree in mechanical engineering from Stony Brook University in 2005. His major research interests include real-time 3-D optical metrology, 3-D machine and computer vision, and geometry processing.



Shing-Tung Yau is a professor of mathematics at Harvard University. He received his doctoral degree in mathematics from University of California–Berkeley in 1971. He has received a number of awards including a Fields Medal in 1982, a MacArthur Fellowship in 1984, the Crafoord Prize in 1994, and the (U.S.) National Medal of Science in 1997.