Miniaturized three-dimensional endoscopic imaging system based on active stereovision

Manhong Chan, Wumei Lin, Changhe Zhou, and Jianan Y. Qu

A miniaturized three-dimensional endoscopic imaging system is presented. The system consists of two imaging channels that can be used to obtain an image from an object of interest and to project a structured light onto the imaged object to measure the surface topology. The structured light was generated with a collimated monochromatic light source and a holographic binary phase grating. The imaging and projection channels were calibrated by use of a modified pinhole camera. The surface profile was extracted by use of triangulation between the projected feature points and the two channels of the endoscope. The imaging system was evaluated in three-dimensional measurements of several objects with known geometries. The results show that surface profiles of the objects with different surfaces and dimensions can be obtained at high accuracy. The *in vivo* measurements at tissue sites of human skin and an oral cavity demonstrated the potential of the technique for clinical applications.

OCIS codes: 170.2150, 170.3890, 150.6910, 170.0110.

1. Introduction

Video-guided remote surgery provides minimally invasive treatment. In comparison with traditional open surgery, video-guided surgery has less adverse effects on the patients. In the procedure of video-guided treatment, an imaging device, such as a laparoscope or endoscope, is inserted through a small incision to visualize the surgical scene inside the patient's body. The commonly used endoscope system employs a single imaging channel with a wide-angle lens to obtain a large view of the field from an accessible organ site. The system provides a geometrically distorted two-dimensional (2-D) image of the surgical site. Moreover, valuable information about the lateral size and the depth of the imaged object is lost. These factors account for the limitations of current video-guided treatment. To guide the surgery with accurate information on the surface topology of the surgical site, a three-dimensional (3-D) and distortion-free imaging technique is desirable.

M. Chan, W. Lin, and J. Y. Qu (eequ@ust.hk) are with the Department of Electrical and Electronic Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China. C. Zhou is with the Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Science, P.O. Box 800-216, Shanghai 201800, China.

Received 27 August 2002.
0003-6935/03/101888-11$15.00/0

One of the most dominant features of the human visual system is that we have two eyes with which to perform stereovision. Because our two eyes are presented with a slightly different view of a scene, the brain is able to determine the depth of the scene by exploiting the slightly different inputs. This is the basic concept of stereovision technology that has been widely used in precision measurements and robotics.[1–4] For example, a binocular stereomatching technique captures two images with two cameras placed at different positions and uses triangulation to compute 3-D information of the imaged object.[5] To avoid difficulty in identification of the corresponding feature points in two images, a trinocular-matching algorithm has been proposed, in which a third camera is employed to overcome the problems inherent in binocular stereomatching.[6] Normally, the feature points, such as the edges or corners of the object at which the image intensity varies rapidly, are identified to match the images taken by multiple cameras that are placed at different positions and to calculate the 3-D coordinates. However, in an ambient lighting environment, there are three major obstacles that cause problems in securing feature point matching. First, false feature points can appear because of noise. Second, some feature points in one of the images might not appear in the other image (occlusion problem). More importantly, an object with a smooth surface, such as human tissue, produces plain characteristics in the image. It is impossible to

identify a sufficient number of feature points for accurate 3-D imaging.

To overcome the problems discussed above, an active method that projects feature points on an imaged object by use of an artificially controlled structured light has been developed.[2,7] In general, a pattern with a certain number of feature points can be projected onto the dark field of an imaged object, and the feature points can be easily extracted. This would significantly reduce the complexity of the matching images taken from different cameras and increase the accuracy of the 3-D measurement. To measure the surface profile of objects with plain spatial characteristics such as human tissue, the active 3-D vision method is particularly effective. In principle, the surface topology can be presented with a computerized graphic 3-D representation technique when surface information is known.

In clinical practice the accurate surface topology of a lesion area provides important diagnostic information for physicians to make objective and statistical judgments. The stereoscopic techniques were developed for 3-D endoscopy over the past two decades. A variety of 3-D endoscopic imaging systems have been evaluated clinically.[8–10] The stereoscopic endoscope can provide depth perception. However, it does not produce precise and quantitative information about the surface profile of an imaged object. Recently, a 3-D endoscope based on active vision was proposed.[11] A laser scanner was attached to the imaging channel of an endoscope to produce the structured light for 3-D measurements. It was found that the acquired surface topology of a lesion could be used in computer-aided diagnosis. However, the system with a fast laser scanner is complicated and makes it difficult to achieve real time 3-D imaging.

We present an endoscope system that can obtain 3-D information from an object with or without surface features based on active stereovision. The imaging system consists of a dual-channel endoscope. One channel projects structured light with a known pattern to the object of interest, and the other channel takes the image of the pattern to recover 3-D information about the imaged object. There is no scanning component in the system for the generation of structured light. The projection and imaging channels of the endoscope are calibrated independently. The calibration procedure involves determination of the external parameters (position and orientation) and the internal parameters (focal distance and lens parameters) of each channel. A ninth-order polynomial is employed to connect the image pixel coordinate with the true coordinate to correct the image distortion caused by the wide-angle lens. We identified the projected feature points by using the epipolar geometry together with a set of geometric and topological rules. After agreement of the feature points in the projection channel and the imaging channel was established, we computed the 3-D information of the imaged surface by using a simple triangulation relationship between feature points of the projection and imaging channels. We demonstrate that this technology can provide accurate information about the surface profile of various objects.

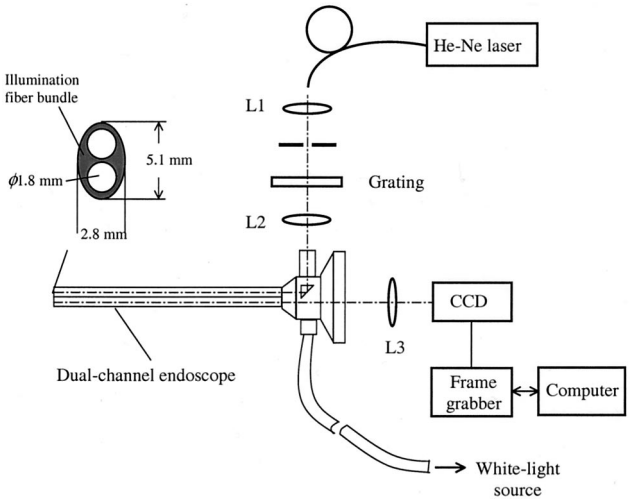


Fig. 1. Schematic of the dual-channel endoscopic imaging system.

2. Methods and Materials

A. Endoscopic Imaging System

The schematic of an active stereoimaging system is shown in Fig. 1. A custom-made rigid endoscope with dual-imaging channels was used to project structured light with a known pattern on the scene and to use the image of that pattern to recover 3-D information about the imaged surface. The optical parameters for both channels were identical. Each channel was built with a wide-angle lens at the tip and a series of rod lenses to relay the image collected by the wide-angle lens. Each imaging channel was 1.8 mm in diameter. The full view angle of each channel was approximately 50 deg. The separation of optical axes between the two channels was approximately 2 mm. The two channels were protected by a 5.1 mm $\times$ 2.8 mm stainless steel tubing as shown in Fig. 1. The space between the imaging channels and the protective tubing was filled with optical fiber to conduct white-light illumination during endoscopy. The image of the object was taken through the imaging channel and recorded with a CCD camera (Model WAT 502A, Watec America Corp., Las Vegas, Nev.). An image-processing board (Genesis, Matrox, Inc., Dorval, Canada) was used to capture the image with a revolution of 768 $\times$ 576 pixels at a frame rate of 25 frames/s. The captured images were stored in a computer for further processing.

The structured light was generated with a holographic binary phase grating and a collimated light source. The collimated light from a lower-power He–Ne laser was conducted to the imaging system by an optical fiber. Here, the holographic binary phase grating acted as a beam splitter to diffract the collimated illumination into multiple beams with a fixed separation angle between adjacent beams. An optimally designed and fabricated grating can distribute

most of the illumination light to the first-order diffracted beams with equal energy. With appropriate focusing, the diffracted beams can be projected to the imaged object and form a fine dot matrix that provides ideal feature points for the 3-D measurement of a surface profile. The number of diffracted beams is determined by the design and structure of the grating. By increasing the number of split beams, more feature points are available to sample the surface of the imaged object and higher accuracy of the 3-D measurement can be achieved.

Traditionally, a structured light is produced by projection of a mask with a known pattern to a scene.[2,7] Compared with the traditional technique, there are a few advantages to the use of holographic binary phase gratings to generate feature points. First, the dot matrix is projected into a dark field. The energy from the illumination source is equally distributed in each feature point, which minimizes the power level required from the illumination source. Second, the dot matrix is generated by splitting the illumination beams. It is not necessary to have a spatially uniform illumination source as in the traditional technique. Furthermore, the density of the feature points, which determines the spatial frequency of the sampling and the accuracy of the 3-D measurement, is not dependent on the size of the holographic grating when the fringe of the hologram is much smaller than the size of the grating. A grating designed for the generation of high density of a dot matrix can potentially be reduced to a small size. This is particularly important for integration of the structured light projection channel into a miniaturized system. However, it is difficult to make a small mask that produces high density of feature points because of diffraction limitation.

For this research, we designed and fabricated two gratings with optimal performance at 632.8 nm. One grating split the collimated incidence into 128 beams with a format of 8×16 dot matrix format. The other split the collimated incidence into 4096 beams of a 64×64 dot matrix. The detailed theory and technique for the design and fabrication of the binary phase grating have been discussed in a previous paper.[12] Briefly, the mask patterns for the gratings were designed based on numerical simulation. With ordinary microelectronic–lithography technology, the patterns were transferred to the photoresistance material onto pieces of glass. The glass pieces were then processed by the wet chemical-etching technique. The diluted HF was used as an etching solution. The etching speed was controlled by the addition of a certain amount of NH_4F. A white-light microscopic interferometer was used to inspect the surface shape and to detect the etching depth. The full size of the gratings was 10 mm $\times$ 10 mm.

The dot matrices produced by the gratings in the focal plane of a lens are shown in Fig. 2. For the 8×16 grating, the vertical and horizontal separation angles between adjacent beams are 0.29° and 0.145°, respectively. For the 64×64 grating, the vertical and horizontal separation angles between adjacent beams were the same: 0.145°. As can be seen, most of the energy from the illumination is distributed to the first-order diffraction beams by the 8×16 grating. The energy distributed to zero-order and higher-order beams is negligible. In the dot matrix generated by the 64×64 grating, the zero-order beam could not be eliminated. Also, two arrays appear of high-order diffractions across the zero-order beam. The exact reasons for the appearance of a zero-order beam and the two arrays of high-order beams have not yet been identified.

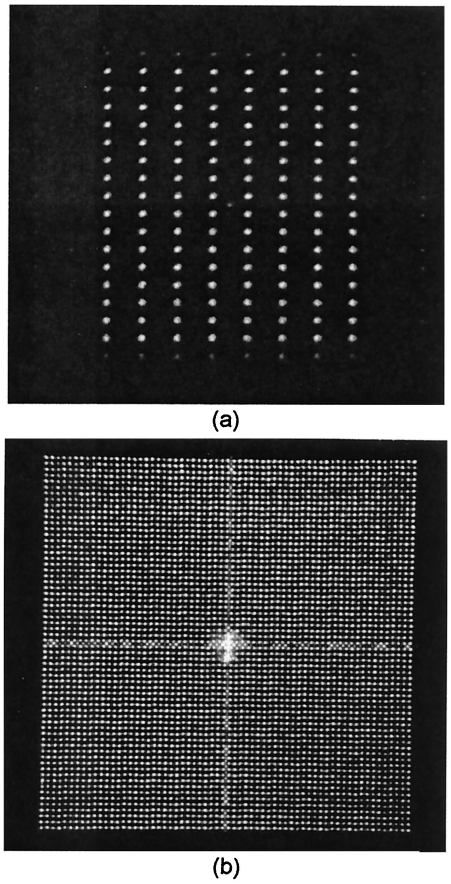


Fig. 2. Dot matrices produced by the gratings taken in the focal plane of a lens: (a) pattern generated by a 16×8 grating and (b) pattern generated by a 64×64 grating.

Although the zero-order and high-order beams took part of their energy from the illumination, the 64×64 grating can generate many more sample feature points than the 8×16 grating. Use of a quarter of the dot matrix generated by the 64×64 grating, can prevent interference from zero- and high-order beams. The maximal number of feature points from a quarter of a 64×64 dot matrix are then equal to $32 \times 32 = 1024$, which is a factor of 8 greater than that produced by the 8×16 grating. In this study, the 64×64 grating was chosen to generate feature points. A quarter of the dot matrix from the 64×64 grating was focused into the projection channel to create feature points. Figure 3 displays the dot matrix projected on a paper screen that was approximately 20 mm away from the tip of the endoscope.

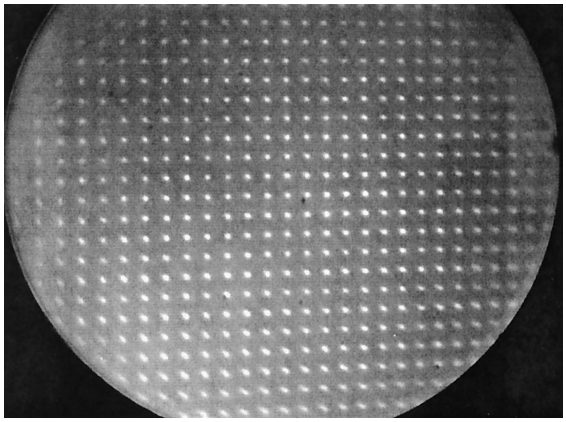


Fig. 3. Quarter of a dot matrix generated by a 64 × 64 grating taken through the imaging channel of an endoscope.

The image was captured from the imaging channel by the CCD camera shown in Fig. 1. Here, the object (a paper screen) was illuminated with the white light from the illumination channel of an endoscope for better illustration of the projection of a dot matrix on an imaged object. During 3-D measurements the white illumination light was turned off to maximize the contrast between dots and background, which ensures accurate recognition of feature points on the imaged object.

The holographic grating has an active area determined by the size of the laser beam. We varied the active area of the grating by changing the size of the pinhole in front of the grating from 1 to 6 mm as shown in Fig. 1. It was found that variation of the active area of the grating had no obvious effect on the pattern of structured light. This demonstrates that a grating can potentially be reduced to as small as a microlens (~1 mm in diameter) and integrated with optical fiber.

B. Calibration of the Endoscopic System

The calibration procedures establish the relationship between the coordinates of an imaging system and the real-world coordinates. Also, the distortion of the imaging system is revealed and then corrected. Our calibration procedures include two steps. First, a well-known pinhole camera model is used to calibrate the imaging channel.[13,14] Then, the projection channel is calibrated in a similar way by use of the calibration parameters of the imaging channel.

Existing camera calibration techniques can be classified into three categories: linear methods, direct nonlinear minimization methods, and multiple-stage methods.[14–16] The multiple-stage calibration methods can produce most of the calibration parameters by use of iterations or nonlinear optimization. Among these, Tsai's two-step calibration procedure is perhaps the most commonly used camera calibration algorithm.[14] We have used a modified Tsai algorithm to calibrate the dual-channel endoscopic imaging system.

An imaging system is treated as a pinhole camera in Tsai's algorithm. The pinhole camera model can be considered a good approximation for this study because the size of the imaging and the projection channels of the miniaturized endoscope is much smaller than the size of the imaged object. The pinhole camera model provides simple mathematical formulations. Briefly, the model consists of a rigid body transformation from a 3-D world coordinate system to a 2-D image coordinate system. It follows a projection with pinhole camera geometry from a 3-D real-world coordinate system to an ideal image coordinate in the plane of a CCD camera sensor. Next, the ideal image coordinates are transformed to an actual image coordinate to characterize the distortion of the imaging system. Finally, the actual image coordinate is related to an image pixel based coordinate with knowledge of the parameters for the CCD and the frame grabber.

Mathematically, these procedures can be represented as follows. If an object point in the world coordinate system $\mathbf{M}$ has a coordinate $(x_w, y_w, z_w)^T$, a perspective projection from the world coordinate system to the pinhole camera coordinate system can be formulated as

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = \mathbf{R}^T \begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix} + \mathbf{T}, \tag{1}$$

where $\mathbf{R} = \lfloor \hat{\mathbf{n}}_x \hat{\mathbf{n}}_y \hat{\mathbf{n}}_z \rfloor$ and $\mathbf{T} = [T_x T_y T_z]^T$ are the rotation matrix and the translation vector, respectively, which represent the alignment of the camera system.[14] Mathematically, $\mathbf{R}$ and $\mathbf{T}$ determine the transformation from the real-world coordinates to camera coordinates. We defined the optical axis of the imaging system as the z axis. The x and y axes are parallel to those of the image coordinate system. The origin of the world coordinate system is located at the tip of the endoscope. The origin of the pinhole camera coordinate system is defined as the optical center of the objective lens of the imaging channel.

Assuming a distortion-free projection, the image coordinate (x_u, y_u) of object point $\mathbf{M}$ is

$$x_u = f\frac{x_c}{z_c} = f\frac{\hat{n}_x{}^T\mathbf{M} + T_x}{\hat{n}_z{}^T\mathbf{M} + T_z},$$
$$y_u = f\frac{y_c}{z_c} = f\frac{\hat{n}_y{}^T\mathbf{M} + T_y}{\hat{n}_z{}^T\mathbf{M} + T_z}, \tag{2}$$

where f is the focal length of the pinhole camera that is the equivalent focal length of the endoscope and lens L3 shown in Fig. 1. The origin of the image coordinate system is defined at the intersection of the optical axis and the plane of the CCD sensor.

When we consider the distortion of the imaging system and the pixel size of the image grabbed by the computer, the camera model should include the parameters to represent distortion and image scaling. In general, there are three types of lens distortion: radial, decentering, and thin prism.[17] Of these, radial distortion is the dominant type. Including four terms of radial distortion coefficients, the distorted

real image coordinate (x_d, y_d) in the plane of the CCD sensor can be expressed as

$$x_d = (1 + \kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6 + \kappa_4 r^8) x_u,$$
$$y_d = (1 + \kappa_1 r^2 + \kappa_2 r^4 + \kappa_3 r^6 + \kappa_4 r^8) y_u, \quad (3)$$

where $r = ({x_d}^2 + {y_d}^2)^{1/2}$. The high-order polynomial terms are related mainly to the distortion at the edge of the image that is caused by the wide-angle lens.

Addition of the scaling factor to the camera model, the coordinate of the image grabbed by the computer (x_f, y_f), and that of the real image coordinate (x_d, y_d) are related as

$$x_f = s_x {d_x}^{-1} x_d + C_x,$$
$$y_f = s_x {d_y}^{-1} y_d + C_y, \quad (4)$$

where d_x and d_y are pixel sizes of the CCD camera, C_x and C_y are the pixel coordinates of the origin of the real image coordinate, and s_x is the scaling factor. In principle, with a series of known 3-D object points, all the intrinsic parameters of the imaging system, such as the focal length (f), the radial distortion coefficients $(\kappa_1, \kappa_2, \kappa_3, \kappa_4)$, the scaling factor (s_x), and the image center (C_x, C_y), can be calculated iteratively with Eqs. (1)–(4).

Calibration of the projection channel is similar to that of the imaging channel because it can be considered as a reverse imaging system. The pattern of the dot matrix generated by the grating in the 2-D image plane of the projection channel is projected to a 3-D space through the projection channel. It should be noted that the dot matrix remains constant at all times when viewed by the projection channel. The dot matrices projected into the planes at various distances from the tip of the projection channel can be treated as 3-D feature points for the calculation of the parameters of the projection channel. The mathematical formulations for the calibration procedures are the same as those for the calibration of the imaging channel except that the last step that connects the CCD coordinates to the pixel coordinates of the image captured by the computer is ignored.

C. Three-Dimensional Measurements

To use the feature points for the 3-D measurement of an imaged object, the same feature point needs to be identified from both the imaging and the projection channels. This lack of correspondence is an essential cornerstone of stereovision. When the lack of correspondence is solved, the triangulation relationship between a feature point and both channels can then be used to calculate the coordinates of the point.

In this research, the image of the dot matrix is recorded through the imaging channel. The dots are extracted and labeled in the image with a standard image-processing method. An algorithm based on epipolar geometry with certain geometric constraints is used to solve the lack of correspondence. A detailed description of the algorithm has been given in Refs. 18 and 19. Briefly, a camera ray, $\mathbf{R}_c$, and a projection ray, $\mathbf{R}_p$, are defined. A camera ray crosses the image of a feature point in the CCD plane, $\mathbf{M}_c$, the optical center of a pinhole camera model, and the physical feature point in the imaged object. This 3-D camera ray can be calculated precisely with the reverse projective matrix of the imaging channel obtained during the calibration procedures. Next, the part of the camera ray from the physical feature point to the endoscope tip is projected to the image plane of the projection channel to form a 2-D epipolar line. The 2-D epipolar line can be calculated accurately with the projective matrix of the projection channel. A dot generated by grating $\mathbf{M}_p$ with the closest distance to the epipolar line in the image plane of the projection channel is chosen as a possible candidate to correspond with image point $\mathbf{M}_c$ captured by the imaging channel. This procedure produces an epipolar constraint matrix to find a possible matched pair between the imaging and the projection channels.

We tested the epipolar constraint matrix by matching the accuracy of neighboring feature points. Here, it is assumed that, if a matched pair is found, the matched pairs for its four nearest-neighbor points can also be found by use of the same epipolar constraint matrix. This hypothesis is based on the assumption that the surface profile is a slowly varying function with respect to the spatial sampling frequency of the dot matrix. If there is a mismatched pair for one of the four nearest-neighbor points, the epipolar constraint matrix would not be considered as optimal. The next dot closest to the epipolar line would be chosen until the hypothesis were fulfilled. This procedure is repeated until the matched pairs all over the surface of the imaged object were found.

For an identified matched pair, a camera ray and a projection ray can be calculated in a similar way. The projection ray, $\mathbf{R}_p$, crosses the matching dot, $\mathbf{M}_p$, the optical center of the projection channel, and the physical feature point on the imaged object. It can also be calculated accurately with the projective matrix of the projection channel. The spatial intersection between $\mathbf{R}_c$ and $\mathbf{R}_p$ can define the 3-D coordinates of the feature point. In most cases, $\mathbf{R}_c$ and $\mathbf{R}_p$ do not intersect exactly with each other because of a calibration error. The position at which the distance between $\mathbf{R}_c$ and $\mathbf{R}_p$ is shortest is then defined as the spatial intersection point. Mathematically, the procedure to find a point that has the shortest distance to the camera ray and the projection ray involves solving the problem of a nonlinear system because of the optical system distortion. However, after both the imaging and the projection channels are calibrated, the solution can be obtained from a set of linear equations. The method to solve the problem is simple and standard.[20]

3. Results and Discussion

Calibration of an endoscopic imaging system involves calibration of the imaging channel and then

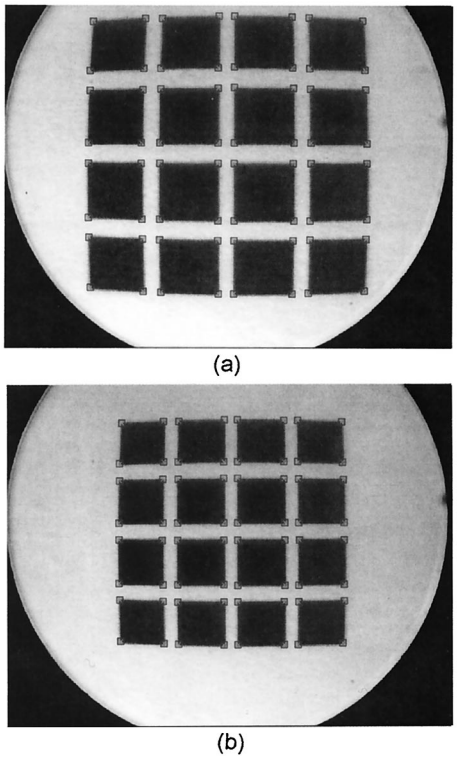


Fig. 4. Checkerboard pattern used for calibration of the imaging channel: (a) image captured at 19.75 mm from the endoscope and (b) image captured at 24.75 mm from the endoscope.

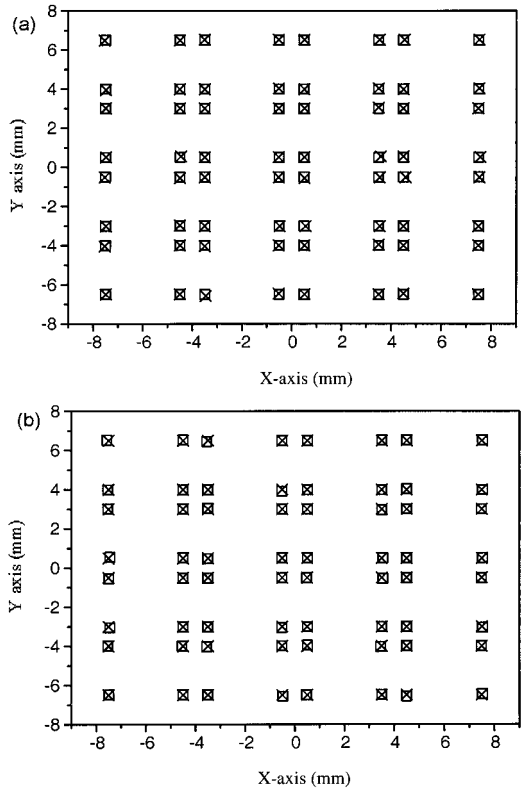


Fig. 5. Positions of true feature points and reprojected feature points in the calibration planes at calibration distances of (a) 19.75 mm and (b) 24.75 mm.

use of the calibrated imaging channel to calibrate the projection channel. A checkerboard pattern with 16 squares was made to generate object points with known 3-D coordinates for calibration of the imaging channel. The pattern was fixed on a one-dimensional translation stage that faced the imaging channel. The separation between the plane of the calibration pattern and the tip of the endoscope could be accurately controlled by the translation stage. The images of the calibration pattern taken at distances of 19.75 and 24.75 mm away from the tip of the endoscope are shown in Fig. 4. The corners of the squares in the pattern were extracted and used as the feature points for the calibration. The corners were accurately located by use of a corner detection algorithm.[20] [The extracted corners are marked with crosses in Fig. 4.] With all the corners detected, the total number of 128 feature points was then available for the calibration.

Calibration of the imaging channel was conducted following the procedure discussed in Subsection 2.B. We evaluated the accuracy of the calibration by calculating the 3-D reprojection error that was defined as the average percentage error between the true 3-D coordinates of the corners in the checkerboard pattern and their reprojected coordinates in the calibration planes.[17,21] The reprojected coordinates were intersection points between the camera rays of all the feature points and the calibration planes. Here, the calibration planes were set at 19.75 and 24.75 mm away from the tip of the endoscope. The 3-D reprojection error was calculated as

$$\epsilon\% = \frac{\sum_{i=1}^{N} \left| \frac{x_i - x_i'}{r_i} \right| + \left| \frac{y_i - y_i'}{r_i} \right|}{2N}, \qquad (5)$$

where (x, y) and (x', y') are the true coordinates and the reprojected coordinates, respectively; r_i is the distance from the ith feature point to the origin of the real-world coordinate; and N represents the number of corners used as feature points for the calibration. The positions of reprojected corners and true corners of the checkerboard pattern in the calibration planes are displayed in Fig. 5 to illustrate the evaluation of the calibration accuracy. The square symbols represent true feature points and crosses represent reprojected feature points. We found that the 3-D reprojection error was less than 2.1%. The distorted pattern seen in the images shown in Fig. 4 was well corrected in the reprojected corners shown in Fig. 5.

During the calibration procedure of the projection channel, the dot matrix was projected to the same calibration planes as those used for calibration of the imaging channel. The difference was that the checkerboard pattern was replaced with a plain screen. The images of the dot matrix on the screen

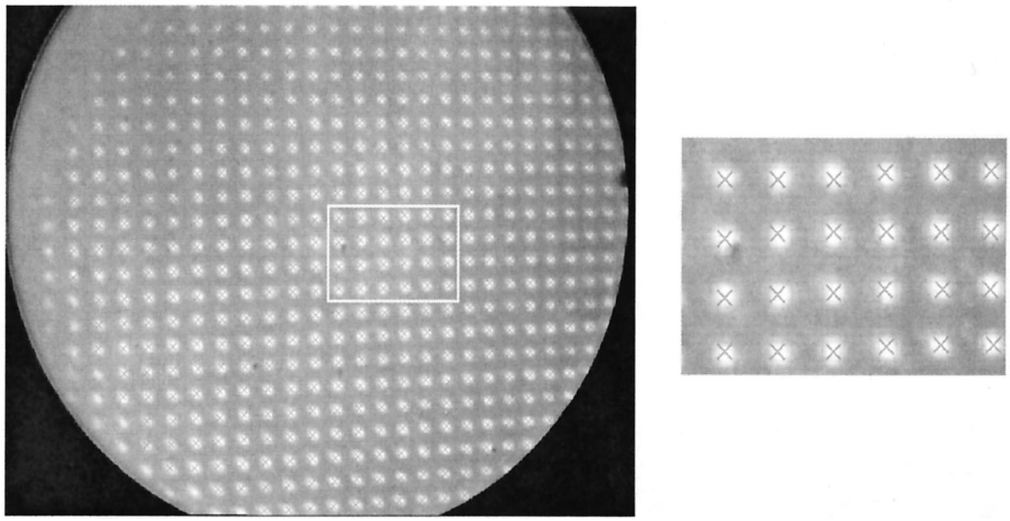


Fig. 6. Extraction of feature points by use of the blob detection algorithm. The cross points represent the gravity centers of the dots: left, image captured by the imaging channel; right, enlarged image from the frame shown at the right-hand side.

in the calibration planes were taken with the imaging channel. After binarization of the images, the dots in the matrix were extracted by use of the blob detection algorithm.[22] Recognition of the dot matrix is illustrated in Fig. 6. The 3-D coordinates of each dot in the calibration planes were then calculated with the reversed projective matrices obtained during calibration of the imaging channel. The dot matrices with known 3-D coordinates in the two calibration planes were then used as feature points for calibration of the projection channel.

In the image plane of the projection channel, the dots in the dot matrix were distributed uniformly as shown in Fig. 2 because the binary phase grating was designed to generate uniform separation angles between adjacent beams. The pattern remained constant as long as the projection optics were not changed. The feature points in the calibration planes always corresponded to the same dot matrix in the image plane of the projection channel. Following the same procedure as for calibration of the imaging channel, we calibrated the projection channel. We found that the 3-D reprojection error was 2.4%, which was comparable with the calibration error of the imaging channel.

As discussed in Subsection 2.C, the camera rays and the projection rays are used to calculate the 3-D coordinates of the feature points. Here, it should be emphasized that we determined all the projection rays after we calibrated the projection channel. The projection rays remain constant if the optical system of the projection channel is not changed. To obtain the 3-D coordinates of the feature points, only the camera rays need to be calculated with the captured image of the feature points and the projective matrix of the imaging channel.

Evaluation of the performance of the endoscopic imaging system for the 3-D measurements was conducted after the calibration of both channels. First, a target with the simplest geometric feature was used to evaluate the accuracy of the depth measurement. A rigid flat plane was fixed on a one-dimensional translation stage. The distance between the endoscope tip and the target surface could be accurately adjusted. The images of the dot matrices on the target were taken at a distance from the endoscope tip starting at 15 mm in increments of 5 mm up to 40 mm. The procedures described in Subsection 2.C were followed to calculate the 3-D coordinates of each feature point in the target. The typical results are shown in Fig. 7. The percentage of measurement errors of the separation between the endoscope and the plane of the target was less than 2% when the separation ranged from 15 to 45 mm.

In the second experiment, a step target as shown in Fig. 8(a) was used to evaluate the reconstruction of the flat surface at different angles. The step target was formed with one plane perpendicular to two parallel planes. The separation of the two parallel planes was 9.2 mm. The target was fixed to a rotator that permitted the angle between the optical axis of the endoscope and the orientation of the step target to be changed precisely. Here, the orientation of the step target was defined as the normal of planes A and C. Initially, the target orientation was set approximately parallel to the optical axis of the endoscope. The angle between the target and the optical axis was then changed in increments of 20 deg up to 60 deg. The distance of the rotation center from the endoscope was approximately 43.0 mm. The 3-D measurements were conducted at four angles between the orientation of the step target and the optical axis. Typical results are shown in Figs. 9 and 10. A quantitative evaluation of the 3-D measurements is summarized in Table 1. The measured coordinates of the feature points were fitted with three plane functions. Angles θ_{AB} and θ_{BC} were calculated by use of fitted planes A, B, and C, where distance d is the separation between fitted planes A and C. As can be

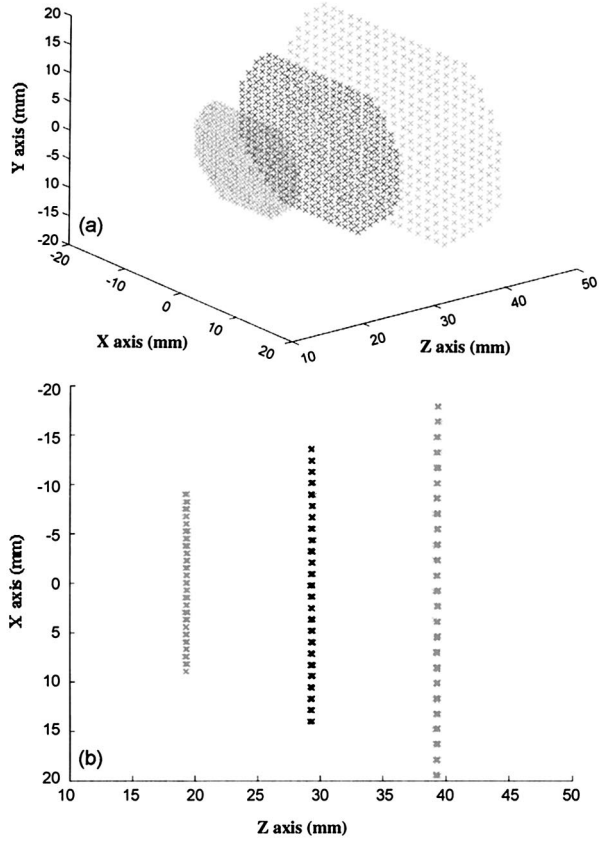


Fig. 7. Reconstruction of the surface of a flat target at different distances from the endoscope tip: (a) 3-D displayed results and (b) projection to the X–Z plane.

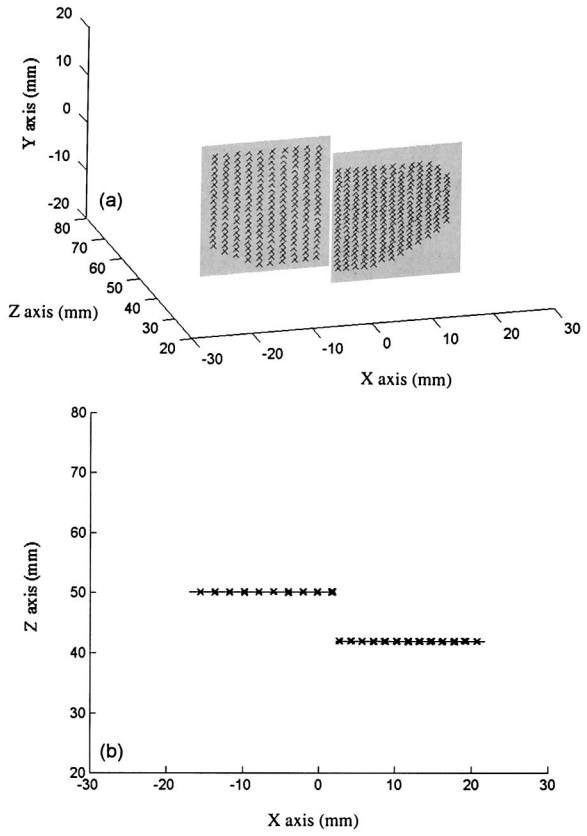


Fig. 9. (a) 3-D measurements taken at the initial angle and (b) projection of the reconstructed 3-D surface to the X–Z plane.

seen, the structure of the step target was recovered accurately at different view angles.

The last target to be used to evaluate the accuracy of the 3-D measurements was a curved object made with a pair of cylinders attached to each other. The diameters of the cylinders were 12.34 and 28.77 mm. The 3-D measurements were conducted from different angles to the target that was placed at distances of 30–40 mm away from the endoscope. The 3-D coordinates of the feature points on the surface of the target were fitted with two cylindrical functions. Typical results are shown in Fig. 11. We found that the diameters of the two fitted cylindrical functions were 12.67 ± 0.23 and 28.62 ± 0.36 mm in 3-D imaging from four different angles. The percentage error of the feature points to the fitted cylindrical surfaces was 2.7%, which demonstrates that the 3-D endoscopic system can recover a curved surface accurately.

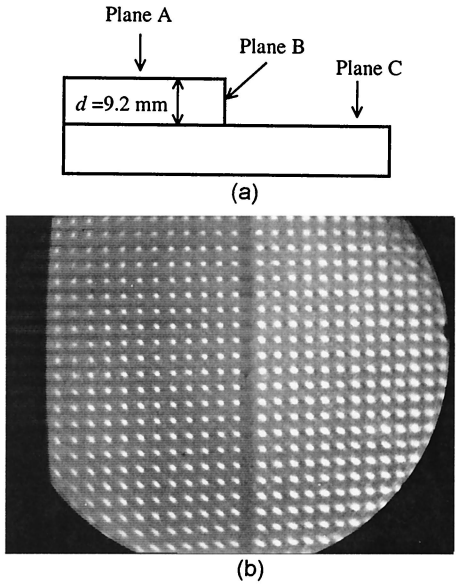


Fig. 8. (a) Structure of the step target that was used to evaluate the reconstruction of the flat surface at different view angles and (b) image of the step target taken at the initial angle.

For the first *in vivo* evaluation of the performance of the 3-D endoscope, we selected the tissue of a forearm. The inner side of the forearm was placed gently against a glass plate to form a flat tissue surface. The glass plate was attached to a precision rotator that was attached to a one-dimensional translation stage. The initial angle between the normal of the glass plate and the optical axis of the endoscope was set to approximately 30 deg to avoid specular reflection of the projection from the glass surface. The flat tissue surface was measured at different distances and angles. The results are shown in Fig. 12. We found that, although human tissue is highly

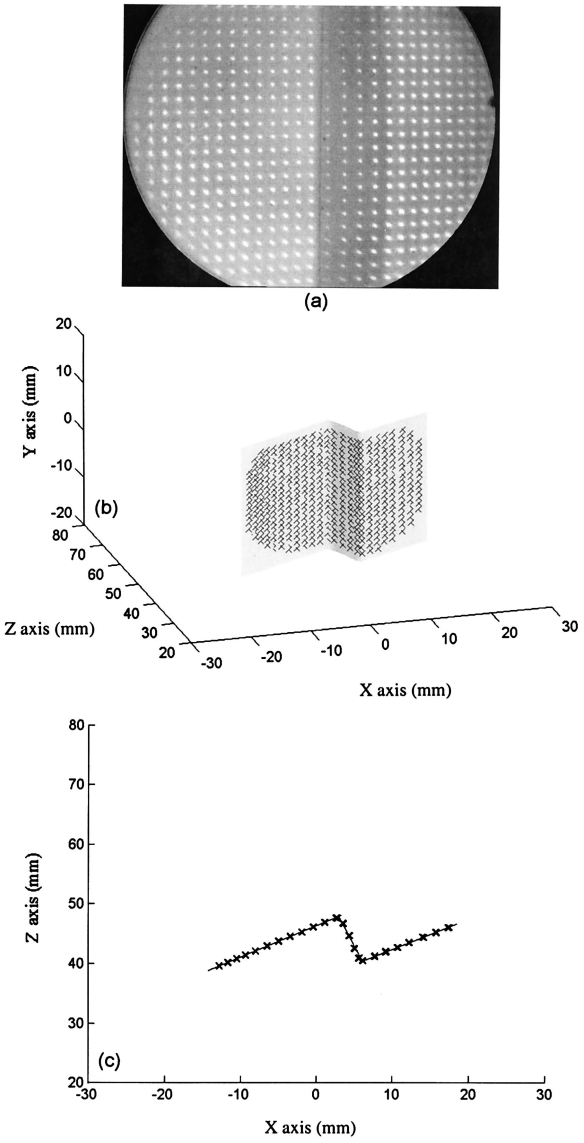


Fig. 10. (a) Image of the step target taken in increments of 20 deg, (b) 3-D measurements taken in increments of 20 deg with respect to the initial angle, (c) projection of the reconstructed 3-D surface to the *X*–*Z* plane.

turbid and the signals of the feature points are weak, the accuracy that we achieved with the 3-D measurement is comparable with that of test targets with a highly reflective surface.

Table 1. Summary of the 3-D Measurements with the Step Target[a]

$\Delta\theta$ (deg)	θ	θ_{AB}	θ_{BC}	d
0	1.6	—	—	9.11
20	23.3	91.04	90.85	9.38
40	41.5	91.22	90.38	9.32
60	62.4	90.98	91.11	9.01

[a] $\Delta\theta$ is the increment of the rotation angle with respect to the initially set angle; θ is the measured angle between the target and the optical axis; θ_{AB} and θ_{BC} are the measured angles between plane A, B and plane B, C, respectively; d is the measured separation distance between planes A and B.

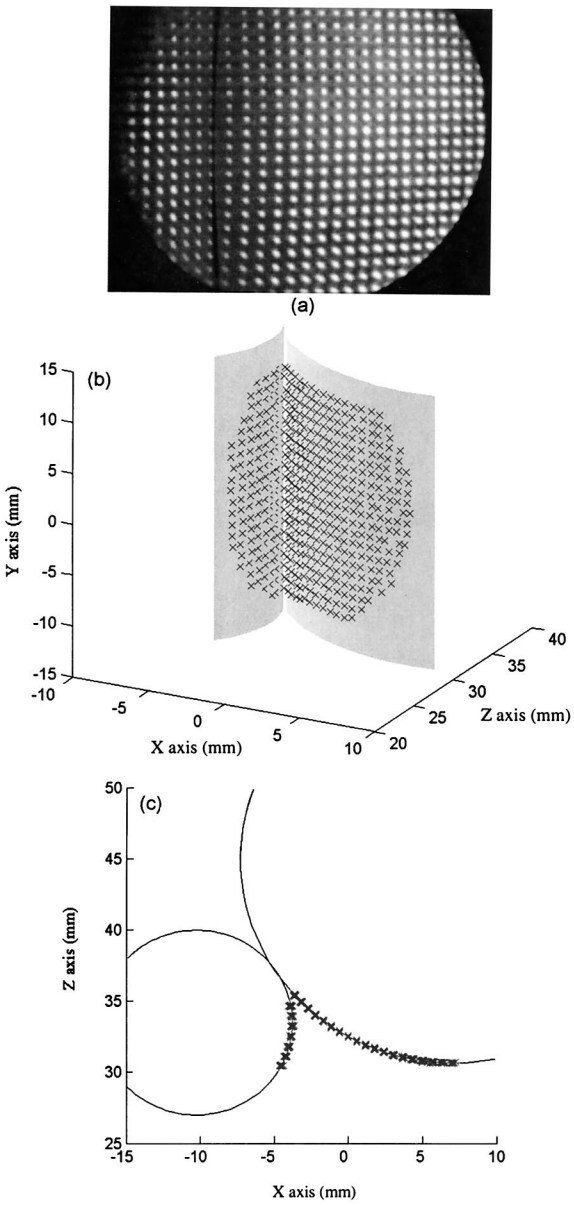


Fig. 11. Reconstruction of the curved surface of a target with two cylinders attached to each other: (a) image of the target taken with the endoscope close to the large cylinder, (b) 3-D measurement, (c) projection of the reconstructed 3-D surface to the *X*–*Z* plane.

To evaluate the potential of clinical application of the miniaturized 3-D endoscopic imaging system, the measurements were conducted on curved tissue surfaces. Here, the surface of the tips of the index and the middle fingers was first recovered by use of the 3-D endoscopic imaging system. Then, the surface of part of the oral cavity was examined with the endoscope. The reconstructed surfaces of the fingers and part of the oral cavity are shown in Figs. 13 and 14. We found that the recovered sizes of the fingers were consistent with the approximate measurements taken with a vernier caliper. The accuracy of the 3-D measurements of the oral cavity could not be evaluated quantitatively because the standard data

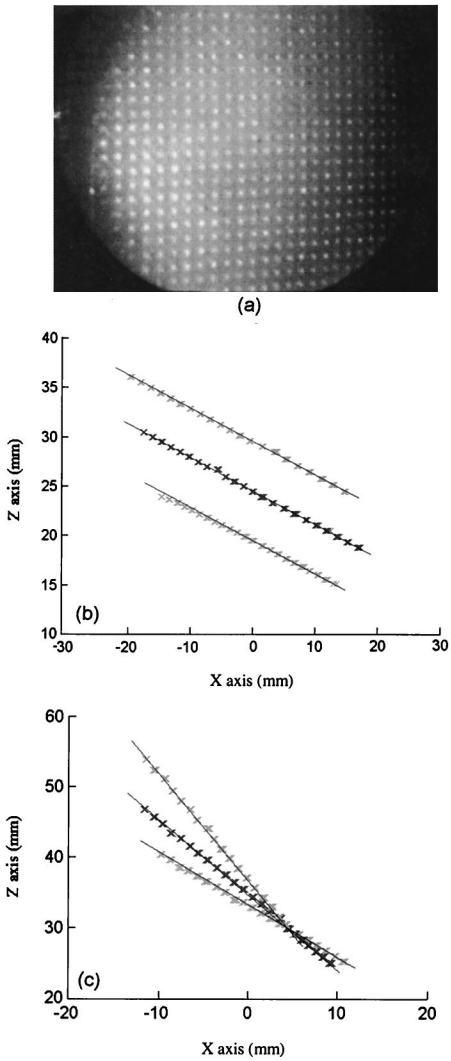


Fig. 12. 3-D measurements of skin tissue with a flat surface: (a) image taken at the initial angle and a distance away from the endoscope, (b) projections of the reconstructed surfaces to the *X–Z* plane at the initial distance and the distances in increments of 5 and 10 mm, (c) projections of the reconstructed surfaces to the *X–Z* plane at the initial angle and at angles in increments of 10 and 20 deg.

of the imaged tissue surfaces were not available. However, the projection of reconstructed 3-D surfaces to the *X–Z* plane in Fig. 15 demonstrates clearly that the depth information of the tissue surfaces was recovered.

4. Conclusion

We have presented a miniaturized endoscope system that can be used to obtain accurate 3-D information about an imaged object. The potential of this technology in clinical applications has been demonstrated. The imaging system takes advantage of a holographic binary phase grating to produce structured light for 3-D mapping of the surface profile of an imaged object. The grating can generate highly dense sampling points and distribute most of the energy from the illumination source uniformly to the sampling points. It has been demonstrated that the

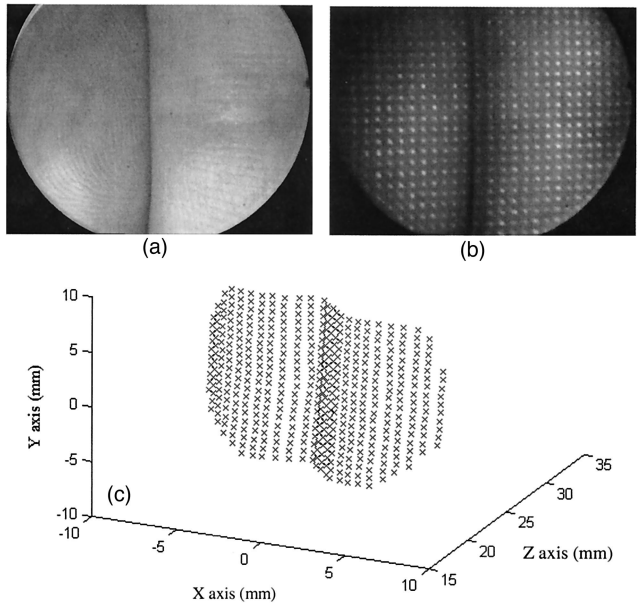
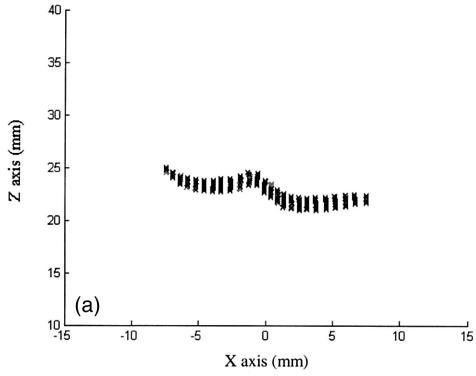


Fig. 13. Reconstructed surfaces of (a) fingers and (b) fingers projected by a dot matrix pattern. (c) Illustration of the reconstructed 3-D surface of fingers.

holographic grating can be made into a small size compatible with the optical fiber components (microlens, filter, etc.) without sacrificing the quality of the sampling pattern of the structured light. This makes it possible to miniaturize the 3-D imaging system by integration of a small grating, optical fiber, and a microlens into the projection channel of the endoscope.



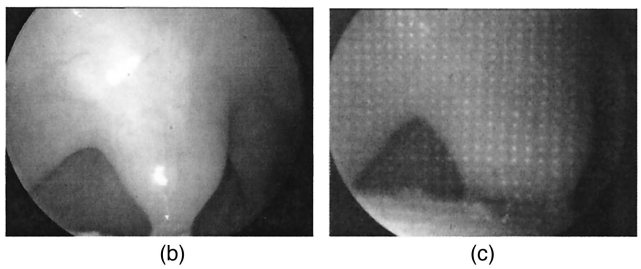


Fig. 14. (a) Projection of the 3-D surfaces to the *X–Z* plane. (b) Image of an oral cavity site for 3-D measurements and (c) image of the tissue site projected with feature points.

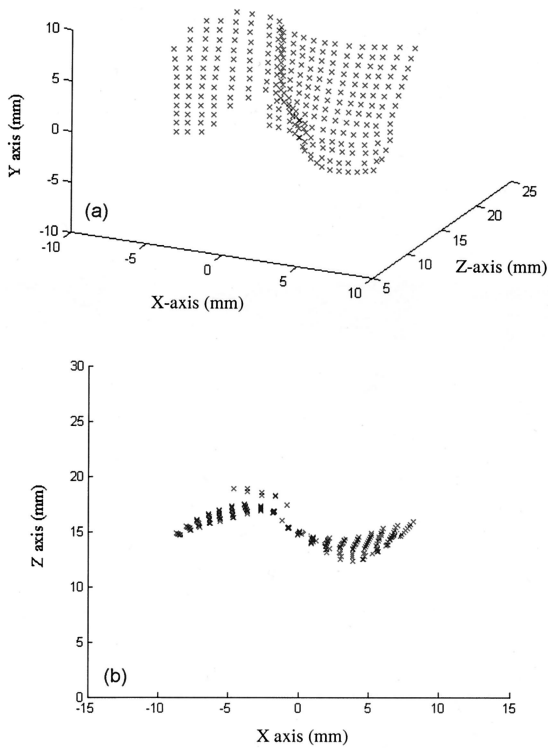


Fig. 15. (a) Illustration of the reconstructed 3-D surface of part of the oral cavity and (b) projection of the 3-D surfaces to the *X–Z* plane.

In future planned research, we will design and fabricate a small binary phase grating with minimized zero- and high-order diffractions. The projection channel will be made with an optical fiber based system. The structured light generated by the grating will be collimated and projected by use of microlenses and gradient-index lenses. The structured light will allow the projection channel to be integrated easily with a flexible endoscope system to access more organ sites and to perform 3-D imaging.

The authors acknowledge the support received from the Hong Kong Research Grants Council through grants HKUST6052/00M and HKUST6025/02M. J. Qu acknowledges the helpful discussions with P. Xi of the Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences.

References

1. S. T. Barnard and M. A. Fischler, "Computational stereo," ACM (Assoc. Comput. Math) Computing Surveys **14,** 553–572 (1982).
2. Y. F. Wang and J. K. Aggarwal, "An overview of geometric modeling using active sensing," IEEE Control Systems Mag. **8,** 5–13 (1988).
3. N. Ayache and P. T. Sander, *Artificial Vision for Mobile Robots: Stereo Vision and Multisensory Perception* (MIT, Cambridge, Mass., 1991).
4. F. Bernardini and H. Rushmeier, "The 3D model acquisition pipeline," Comput. Graph. Forum **21,** 149–172 (2002).
5. H. H. Baker, "Depth from edge and intensity based stereo," in *Stanford Artificial Intelligence Laboratory Tech. Rep. AIM-347* (Stanford University, Stanford, Calif., 1982).
6. A. Blake, D. McCowen, H. R. Lo, and P. J. Lindsey, "Trinocular active range-sensing," IEEE Trans. Pattern Anal. Mach. Intell. **15,** 477–483 (1993).
7. J. Batlle, E. Mouaddib, and J. Salvi, "Recent progress in coded structured light as a technique to solve the correspondence problem: a survey," Pattern Recogn. **31,** 963–982 (1998).
8. J. Strutz, "3-D endoscopy," HNO **41,** 128–130 (1993).
9. C. J. Calvano, M. E. Moran, L. D. Tackett, P. P. Reddy, and M. M. Pankratov, "New visualization techniques for in utero surgery: amnioscopy with a three-dimensional head-mounted display and a computer-controlled endoscope," J. Endourol. **12,** 407–410 (1998).
10. K. Yao, T. Matsui, H. Furukawa, T. Yao, T. Sakurai, and T. Mitsuyasu, "A new stereoscopic endoscopy system: accurate 3-dimensional measurement *in vitro* and *in vivo* with distortion-correction function," Gastrointest. Endosc. **55,** 412–420 (2002).
11. K. Hasegawa, K. Noda, and Y. Sato, "Electronic endoscope system for shape measurement," in *Proceedings of the 16th International Conference on Pattern Recognition* (Publications Office, IEEE Computer Society, P.O. Box 3014, Los Alamitos, Calif. 90720-1264, 2002), Vol. 2, pp. 761–764.
12. C. Zhou and L. Liu, "Numerical study of Dammann array illuminators," Appl. Opt. **34,** 5961–5969 (1995).
13. E. Hall, J. B. K. Tio, C. A. McPherson, and F. A. Sadjadi, "Measuring curved surfaces for robot vision," Computer **15,** 42–54 (1982).
14. R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV camera and lenses," IEEE J. Rob. Autom. **3,** 323–344 (1987).
15. J. Heikkila, "Geometric camera calibration using circular control points," IEEE Trans. Pattern Anal. Mach. Intell. **22,** 1066–1077 (2000).
16. J. Batista, H. Araujo, and A. T. de Almeida, "Iterative multistep explicit camera calibration," IEEE Trans. Rob. Autom. **15,** 897–917 (1999).
17. J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," IEEE Trans. Pattern Anal. Mach. Intell. **14,** 965–980 (1992).
18. G. Hu and G. Stockman, "3-D surface solution using structured light and constraint propagation," IEEE Trans. Pattern Anal. Mach. Intell. **11,** 390–402 (1989).
19. O. Faugeras, *Three-Dimensional Computer Vision: a Geometric Viewpoint* (MIT, Cambridge, Mass., 1993).
20. E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision* (Prentice-Hall, Upper Saddle River, N.J., 1998).
21. C. Chatterjee, V. P. Roychowdhury and E. K. P. Chong, "A nonlinear Gauss–Seidel algorithm for noncoplanar and coplanar camera calibration with convergence analysis," Comput. Vision Image Understand. **67,** 58–80 (1997).
22. R. C. Gonzalez and R. C. Woods, *Digital Image Processing* (Addison-Wesley, Reading, Mass., 1992).