Building a 3D scanner system based on monocular vision

Zhiyi Zhang* and Lin Yuan

College of Information Engineering, Northwest A&F University, No. 22 XiNong Road, ShaanXi-YangLing, 712100, China

*Corresponding author: 815802490@qq.com

Received 4 November 2011; revised 26 December 2011; accepted 27 December 2011; posted 5 January 2012 (Doc. ID 157644); published 4 April 2012

This paper proposes a three-dimensional scanner system, which is built by using an ingenious geometric construction method based on monocular vision. The system is simple, low cost, and easy to use, and the measurement results are very precise. To build it, one web camera, one handheld linear laser, and one background calibration board are required. The experimental results show that the system is robust and effective, and the scanning precision can be satisfied for normal users.

OCIS codes: 100.0100, 100.6890.

1. Introduction

A three-dimensional (3D) scanner is a high-precision instrument system combining mechanical electro-optics and computer technology in one body. It can obtain the 3D surface digital information of an object or scene [1,2]. The information is usually expressed as a point cloud that shows the surface geometry shape and structure. A 3D shape model of an object can be reconstructed by processing the obtained point cloud.

According to the different ways of sensing, a 3D scanner can usually be divided into two types: contact and noncontact. Contact 3D scanners need physical touch to get the 3D information. They usually have high accuracy and reliability. However, they are usually expensive, need a long time to scan, and may damage the scanned object. Noncontact 3D scanners usually use laser or structured lighting to irradiate an object; the 3D coordinate points on its surface can be obtained by calculating its reflection information [3,4]. Compared to contact 3D scanners, noncontact scanning of is fast, its operation is simple and easy, and more importantly, it will not damage the scanned object. The disadvantage is that its precision may not be enough in some special applications.

The noncontact 3D measurement method can be divided into three categories: multiview stereo vision methods, shape from shading methods, and structured light projection triangle measurement methods. Multiview stereo vision methods encountered a similar problem: how to find the corresponding point in the multiple view images [5,6]. It is also difficult to apply the methods to scarcely textured areas. The shape from shading methods require prior knowledge of the reflectance of object surfaces. So they are difficult to control and end up being less accurate than the other methods. The structured light projection triangle measurement methods are a kind of active vision, because they can avoid the prior knowledge problem of the surface reflectance and can solve the problem of corresponding point matching by using temporal unwrapping [7], spatial unwrapping [8], and period coded [9] approaches.

In this paper, a low-cost noncontact 3D laser scanner system based on monocular vision is proposed [10,11]. In our system, in order to construct an ingenious Euclidean geometric space by transforming projective geometry, one background calibration board is used. One handheld linear laser and one web

1559-128X/12/111638-07$15.00/0

camera are used to obtain the specialty images, which can avoid the problem of corresponding point matching. This process will involve many technologies, such as camera-image display calibration, subpixel laser line extraction, environment light influence elimination, and so on. The contributions of the proposed system are that it is portable, easy to operate, and high precision and can work well in a more complex light environment, especially.

In Section 2, the Euclidean geometric space construction method from the projection image of monocular vision is described in detail. In Section 3, camera-image distance calibration, subpixel laser line extraction, and environment light influence elimination technologies are explained in detail. The software implementation and some points to note are briefly described in Section 4. Some experimental results are shown and discussed in Section 5. Conclusions and prospects are summarized in Section 6.

2. Principle of Geometric Construction

In general, monocular vision is usually considered the ideal pinhole camera. The pinhole camera model can better describe the mathematical relationship between the 3D coordinates of points and their corresponding two-dimensional projection points lying in the image plane.

A. Geometric Transform of a Projection

Now we consider the following circumstances as shown in Fig. 1. O is the origin of coordinates. x and z indicate the x-axis and z-axis directions, respectively. The y-axis direction can be expressed by the cross product of x and z. A, B, C are three space points lying on one straight line, and B is the midpoint of A and C. Corresponding to the space points A, B, and C, respectively, A', B', and C' are the projection points lying on the projection imaging plane P_{image} that is parallel to the plane xy. If d_z is the distance between O and P_{image}, the coordinates of A', B', C' can be expressed as A' $(x_{A'}, y_{A'}, d_z)$, B' $(x_{B'}, y_{B'}, d_z)$, C' $(x_{C'}, y_{C'}, d_z)$, respectively. By using the parametric expression equations of a straight line, the coordinates of A, B, C can be calculated as $A = t_1 * A'$, $B = t_2 * B'$, and $C = t_3 * C'$, respectively. Because B is the midpoint of A and C, then $t_2 = (t_1 + t_3)/2$ can be obtained. Other than simply knowing that t_1, t_2, and t_3 are greater than 0, we known nothing about the specific values.

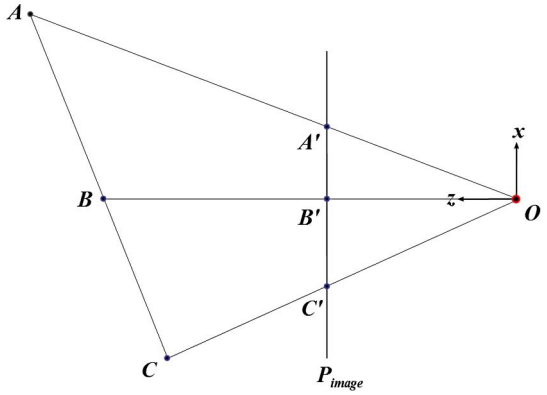

Fig. 1. (Color online) Geometric transform of a projection.

If d_{AB} is the distance between A and B and d_{BC} is the distance between B and C, when d_{AB}, d_{BC}, d_z, and the x, y coordinates of A', B', C' are already known, according to perspective projection geometry theory and the cosines theorem, the following equation can be obtained:

$$\begin{aligned} d_1^2 + d_2^2 - a^2 &= \frac{(t_1 d_1)^2 + (t_2 d_2)^2 - m^2}{t_1 t_2}, \\ d_2^2 + d_3^2 - b^2 &= \frac{(t_2 d_2)^2 + (t_3 d_3)^2 - m^2}{t_2 t_3}, \\ d_1^2 + d_3^2 - c^2 &= \frac{(t_1 d_1)^2 + (t_3 d_3)^2 - 4m^2}{t_1 t_3}. \end{aligned} \tag{1}$$

Here, $d_{AB} = d_{BC} = m$, and $d_i (i = 1, 2, 3)$ is defined as

$$\begin{aligned} d_1 &= \sqrt{x_{A'}^2 + y_{A'}^2 + d_z^2}, \\ d_2 &= \sqrt{x_{B'}^2 + y_{B'}^2 + d_z^2}, \\ d_3 &= \sqrt{x_{C'}^2 + y_{C'}^2 + d_z^2}, \end{aligned} \tag{2}$$

and a, b, c can be calculated as

$$\begin{aligned} a &= \sqrt{(x_{A'} - x_{B'})^2 + (y_{A'} - y_{B'})^2}, \\ b &= \sqrt{(x_{B'} - x_{C'})^2 + (y_{B'} - y_{C'})^2}, \\ c &= \sqrt{(x_{A'} - x_{C'})^2 + (y_{A'} - y_{C'})^2}, \end{aligned} \tag{3}$$

if we assume

$$\begin{aligned} U &= d_1^2 + d_2^2 - a^2, \\ V &= d_2^2 + d_3^2 - b^2, \\ W &= d_1^2 + d_3^2 - c^2, \end{aligned} \tag{4}$$

$$\begin{aligned} T_1 &= \frac{(U - d_2^2) d_1^2}{2W} - \left(d_1^2 - \frac{U}{2} + \frac{d_2^2}{4} \right), \\ T_2 &= \frac{(V - d_2^2) d_3^2}{2W} - \left(d_3^2 - \frac{V}{2} + \frac{d_2^2}{4} \right), \\ M_1 &= \frac{d_2^2}{4} - \frac{(U - d_2^2) d_3^2}{2W}, \\ M_2 &= \frac{d_2^2}{4} - \frac{(V - d_2^2) d_1^2}{2W}, \\ N_1 &= \frac{(U - d_2^2) 4m^2}{2W} - m^2, \\ N_2 &= \frac{(V - d_2)^2 4m^2}{2W} - m^2. \end{aligned} \tag{5}$$

Then, t_1, t_3, and t_2 can be calculated as

$$t_1 = \sqrt{\frac{M_1N_2 + T_2N_1}{T_1T_2 - M_1M_2}},$$

$$t_3 = \sqrt{\frac{M_2N_1 + T_1N_2}{T_1T_2 - M_1M_2}},$$

$$t_2 = \frac{t_1 + t_3}{2}. \quad (6)$$

Thus, through the above derivation, the 3D coordinates of three space points A, B, C can be obtained.

B. Construction of the Geometric Space

We consider the following circumstances as shown in Fig. 2. O is the origin of coordinates. A, B, C and D, E, F define two parallel straight lines in the plane P_l, where B is the midpoint of AC and E is the midpoint of DF. Similarly, G, H, I and J, K, L define two parallel straight lines in the plane P_r, where H is the midpoint of GI and K is the midpoint of JL. L_c is intersection line of P_l and P_r.

If a pinhole camera with known image distance d_z is placed on point O, then according to the previous derivation, when the distances d_{AB}, d_{BC}, d_{DE}, d_{EF}, d_{GH}, d_{HI}, d_{JK}, and d_{KL} and the relevant information projected on the image are known, the plane equations of P_l and P_r and the straight line equation of L_c can be calculated. So, the planes P_l and P_r with 12 calibration points as shown in Fig. 2 can be used as the background calibration board.

When the object is placed before the background calibration board, a linear laser can be used to irradiate it, as shown as in Fig. 3. Furthermore, when the left and right sides of background calibration board are extend beyond the image, the laser plane can be determined as shown in Fig. 4. The laser plane and P_l and P_r planes intersect in straight lines MN and QR, respectively. Here, points M and Q were limited to locations on the left and right of the image, and N and R were limited to locations on the straight lines AD and IL. Assume that MN and L_c intersect at point S, and QR and L_c intersect at point T. In theory,

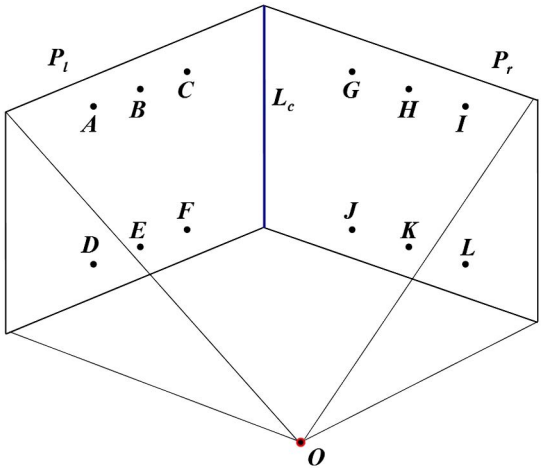

Fig. 2. (Color online) Construction of the geometric space by the background calibration board.

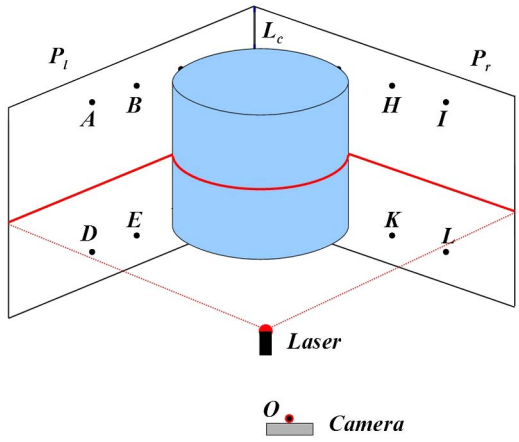

Fig. 3. (Color online) A linear laser can be used to irradiate the object that is placed before the background calibration board.

S and T are the same point. But because of error, S and T may not be the same point. In order to fix the error, the average value of S and T can be used to modify straight lines MN and QR. Then, the laser plane can be determined by modified straight lines MN and QR. Finally, the 3D point coordinates of the object surface can be obtained by calculating the intersection points of the laser plane and each straight line that connects an image laser point to the origin point. Here, the origin point is the center of camera lens.

3. Design of the 3D Scanner System

In Section 1, a prerequisite condition for all calculations was that d_z must be known. So, how to obtain an accurate d_z becomes the first problem that must solved. In addition, the extracted laser line from the image has a certain width that will seriously affect the accuracy of the results. The second problem to be solved is how to get a subpixel level laser line. Finally, when the environment light is complex, in order to ensure the laser line can be extracted from the image, it is necessary to take appropriate measures to eliminate environment light effects.

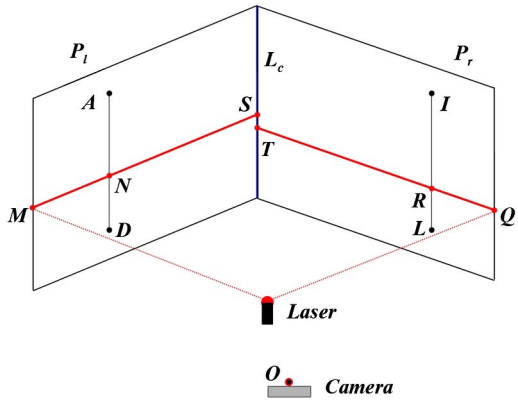

Fig. 4. (Color online) How to determine the laser plane when there is a little error.

A. Camera–Image Distance Calibration

In general, a 3D scanner system needs very detailed and complex calibration [12]. More importantly, this work needs to be completed by the user, but the user often cannot finish it. So for the convenience of users, a calibration method may be used to simplify this process as much as possible.

Figure 5 shows a simple background pattern for the image distance calibration where A and E are located in the vertical direction, B and D are located in the horizontal direction, and C is the midpoint of AE and BD. The distances between A and E and between B and D can be known in advance. The method of camera–image distance calibration is shown in Fig. 6.

The calibration procedure is divided into two operations. In the first step, we fix the calibration pattern board to a vertical plane, then appropriately adjust the distance and angle of the camera. When C', which is the projection point of C, is lying in the center of the image, the corresponding A' and E' and B' and D' are mutually symmetrical about point C', marking the position of the camera lens as O'. Then, to measure the distance from O' to C as $D_{O'C}$, record the distance from A' to E' that are lying in the image as $D_{A'E'}$. The second step is similar to the first step, but the difference is that we need to move and adjust the camera lens to another position, O''. That is to say, the distance $D_{O''C}$ and $D_{A''E''}$ can now be measured and calculated. We assume that ε is the distance error between the real camera lens center and the marker point position, and $D_{O'C'} = D_{O''C''} = d_z + \varepsilon$. According to the property of similar triangles, the following equations can be obtained:

$$\frac{d_z}{D_{A'E'}} = \frac{D_{O'C} - \varepsilon}{D_{AE}}, \frac{d_z}{D_{A''E''}} = \frac{D_{O''C} - \varepsilon}{D_{AE}}. \tag{7}$$

Then ε and the image distance d_z can be calculated as

$$\varepsilon = \frac{D_{A'E'}D_{O'C} - D_{A''E''}D_{O''C}}{D_{A'E'}D_{A''E''}}, \tag{8}$$

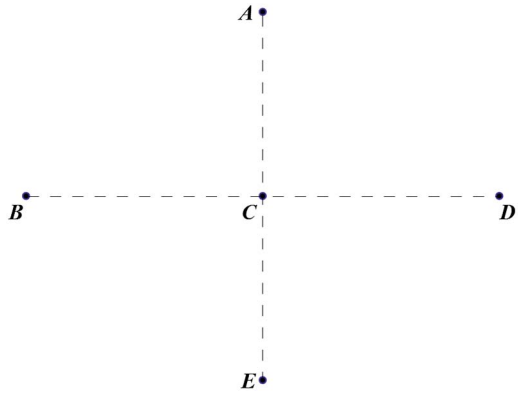

Fig. 5. (Color online) Simple background pattern for the image distance calibration.

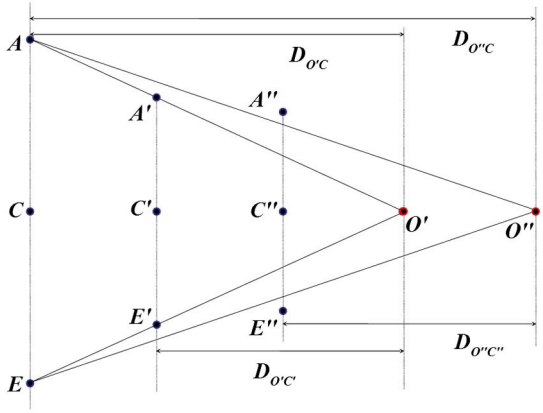

Fig. 6. (Color online) Camera–image distance calibration.

$$d_z = \frac{D_{A'E'}D_{A''E''}(D_{O''C} - D_{O'C})}{D_{AE}(D_{A'E'} - D_{A''E''})}. \tag{9}$$

B. Subpixel Laser Line Extraction

The images may include a variety of noise and distortion. In order to highlight useful information and limit useless information, it is necessary to process the original image by using some feature analysis [13]. Though the laser can provide sharp contrast to the environment, it is difficult to find the demarcation line of laser and background from the source color image. Therefore, in order to get the center of the laser line, it effective to convert the image into gray. In general, scanning distance does not exceed the range of the background calibration board, and the laser line in one image is always distributed in a certain width range, So by limiting the reasonable scan range, the scanning speed can be improved, and the adverse effects of reflected light from the environment can be effectively eliminated.

In the literature [14], two laser line center extraction methods have been provided. In our study, the weighted average method was used to obtain the subpixel level laser center line [15,16]. The method is briefly described as follows.

First, the approximate center of each column can be found by moving a 3×1 mask from top to bottom. When the sum of three pixel gray values is maximized, or it is greater than a limited threshold value, the middle of the location can be considered as the rough laser line center. Second, because the ideal light intensity of a laser line is a Gaussian distribution [17], the upper and lower boundaries of the laser line can be detected in the area adjacent to the rough center. In this paper, the Sobel operator is used to find the boundaries of the laser line. Finally, the precise laser line center G_c in one column can be calculated by the equation

$$G_c = \frac{\sum_{i=ub}^{lb} P_i N_i}{\sum_{i=ub}^{lb} N_i}. \tag{10}$$

Here, N_i is the gray value and P_i is the coordinate of pixel i that is lying in the area between the upper and lower laser boundaries. The method can be more precise, but this will slightly reduce the speed.

C. Elimination of Environment Light Influence

In general, the laser power in one system is fixed. Its light intensity will decay as the distance increases. Worse, in different lighting conditions, the diffused reflection signal intensity will be very different. When the environment light is stronger, the camera can capture a diffused reflection laser line signal that is weaker. In order to solve this problem and improve detection of the laser line, the difference image method was used to design effective algorithms. It is useful to take a reference image I_R without laser light and to use difference images $I_d = I_N - I_R$ in the following. Here, I_N is the current image with laser light. As the laser line will be either horizontal or vertical, we can reduce the problem to a one-dimensional detection of the laser line in each single column or row of the image.

For two images from different times, each pixel can produce a change in direction and intensity that will be called the characteristic change vector. When the background and the measurement object are stationary, we only need to focus on the intensity change. The light intensity change of time domain difference image $I_{(i,j)P}$ can be calculated as the following equation:

$$I_{(i,j)P} = \sum_{k=1}^{n} (V_{(i,j)R} - V_{((i,j))N})^2. \tag{11}$$

Here, $V_{(i,j)R}$ and $V_{(i,j)N}$ are the light intensity of pixel (i,j) corresponding to the reference image and current image, repsectively. n is the number of channels. When a color image is used, $n = 3$; if a gray image is used, $n = 1$.

4. Software Implementation

Because the projection geometric transformation is nonlinear, according to the method described in Section 2, for the instance of the projection of laser points lying on the x axis (or y axis), the calculation of the denominator as 0 will be encountered and lead to erroneous results. In order to avoid this problem, all the data points or pixels ought to be processed in the same quadrant of the coordinate system. Thus, the corresponding coordinate transformation is necessary.

If we only use light intensity, instead of the difference image method, to extract the laser line from the image, a light intensity threshold is essential. In our system, the light intensity threshold is fixed and does not require user configuration. It can be obtained through the advanced gray histogram analysis of the reference background image and the laser line image. The experimental results show that the threshold works very well over a wide range of light intensities and within relatively long distances. When the time domain difference image method was used, $I_{(i,j)P}$ values far greater than 0 would be treated as laser points.

Another interesting concern is that scan data will grow rapidly with the extension of scan time. These chaotic unstructured data may contain some noise and in some regions may be too intense. In order to get the proper amount of ordered structure data, a two-dimensional structured array may be used to store data. In a limited resolution space, only one point can be selected to constitute the OBJ format file.

5. Experimental Results

C/C++ Language and Windows API may be used for system development. The library used includes OpenCV and OpenGL. The camera is a Logitech QuickCam Pro 9000. The computer is Pentium(R) Dual-Core CPU T4300 operating at 2.10 GHz.

Table 1 shows the dependence test results for laser color, wavelength, and power, the illumination of environmental light, and the extraction method of the laser line. In this table, C & W are the color and wavelength of the laser, and EL is environmental light. "Max illumination of EL" means that when the illumination value of the natural environmental light is greater than the value shown in the table, the laser line most likely will not be detected because of destructive interference. It needs to be emphasized that the illuminance values were only the maximum values used in our experiments to facilitate measurement. They do not represent more stringent physical boundary values. In this column, the first number indicates the value when the light intensity method was used, and the second number indicates the value when the time domain difference image method was used.

Table 2 shows the different processing speeds for different image sizes. In this table, LIM is the light intensity method and DIM is the time domain difference image method. Because in one image only the points lying on one laser line can be computed, the processing speed depends on the camera's frame rate and computation time for 3D coordinates of points. It is obvious that processing speed will slow down with larger image sizes. The experimental results show that when the image resolution is not greater than

Table 1. Dependence Test Results for Laser Color, Wavelength, and Power and Illumination of Environmental Light for Different Laser Line Extraction Methods

C & W	Laser Power	Max Illumination of EL (lux)
Red, 635 nm	1 mW	0/50
	5 mW	80/200
	10 mW	200/500
	50 mW	200/500
Green, 532 nm	1 mW	20/50
	5 mW	100/250
	10 mW	200/500
	50 mW	500/1200

Table 2. Scanning Speeds for Different Image Sizes

Image Size (Pixels)	Frame Rate of LIM (fps[a])	Frame Rate of DIM (fps)
640 × 480	30	30
960 × 720	30	24
1280 × 960	24	12
1600 × 1200	16	8

[a]fps, frames per second.

1600 × 1200 pixels, the processing frame rate is greater than 12, except with the time domain difference image method. That mean that: in order to obtain better results, the moving speed of handheld lasers should match the processing speed. In fact, the processing speed will be faster with a lower image resolution, but the scanning accuracy will decrease.

Table 3 shows scanning accuracy situations for different distances when the image resolution 1600 × 1200 pixels is used. In the 280–3200 mm range, as the distance gets smaller, scanning will be more high precision. "Distance of Cam–BCB" means the distance from the camera to the deepest part of the background calibration board. "Distance of Adj-CP" means the shortest distance between two adjacent calibration points. From the results, although the maximum error and average error will increase as the scanning distance becomes larger, the relative error remains at around 0.08%. In a similar system described in the literature [10], the relative error is about 0.28%. That is because our system used a higher image resolution and simplified the background calibration. In addition, our background calibration board can be greater than or less than right angles and the error of camera–image distance calibration correction can be taken into account in advance, so our system can achieve precise and quick scans in a larger range.

Figures 7 and 8 show the scanned results of terracotta warriors and clay pots, respectively. Figure 7(a) shows the scanned point cloud of terracotta warriors, and Fig. 7(b) shows the depth map of the surface model obtained from Fig. 7(a). There are 39,460 vertices and 64,918 polygons in the model. Figure 8(a) shows the scanned point cloud of clay pots, and Fig. 8(b) shows the depth map of the surface model obtained from Fig. 8(a). There are 23,7823 vertices and 28,9861 polygons in the model. In the figure, depth information is expressed by pseudocolor, with far to near corresponding to blue-green-red.

Table 3. Scanning Accuracy Situations for Different Scanning Distances

Distance of Cam–BCB (mm)	280	400	800	1600	3200
Distance of Adj-CP (mm)	40	60	120	240	480
Maximum error (mm)	0.19	0.28	0.59	1.19	2.33
Average error (mm)	0.16	0.24	0.49	0.99	1.97

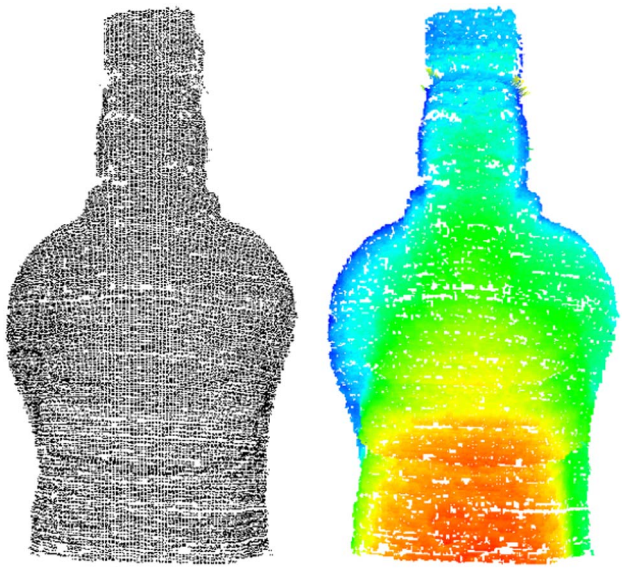

Fig. 7. (Color online) Scanned results of terracotta warriors.

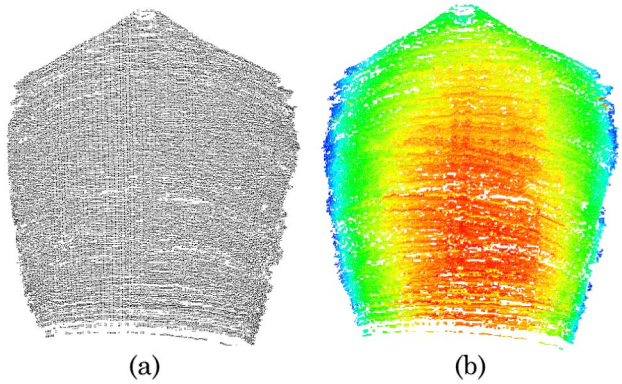

Fig. 8. (Color online) Scanned results of clay pots.

6. Conclusion

This paper presented a 3D scanning system based on geometric construction of monocular vision. According to the experimental results, when a 10 mW laser was used and the natural environmental light illumination was below 500 lx, for a variety of resolutions, at least 4000 valid vertex data can be obtained in 1 s. Within 3 m, the maximum error and average error are about 2 mm, and the relative error remained at around 0.08%. Thus, this method can basically reconstruct the 3D model, and the scanning precision and speed can be satisfied by normal users.

Because the linear laser is not a real plane, but a spindle cone, and diffuse reflectance and chaotic scattering of light will produce interference, some measurement data may be unstable. In order to solve this problem, laser shape calibration and reflected light analysis are essential. On this basis, figuring out how to build an omnidirectional scanning system with color and texture will be our future research direction.

This work was supported by the Special Fund from the Central Collegiate Basic Scientific Research Bursary (No. Z109021004) and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry (No. K314020901).

References

1. F. Chen, G. M. Brown, and M. Song, "Overview of three-dimensional shape measurement using optical methods," Opt. Eng. **39**, 10–22 (2000).
2. S. Zhang and P. S. Huang, "High-resolution, real-time three-dimensional shape measurement," Opt. Eng. **12**, 1–8 (2006).
3. D. Lanman and G. Taubin, "Build your own 3D scanner: 3D photography for beginners," in *Proceedings of SIGGRAPH '09* (ACM, 2009), article no. 8.
4. D. Lanman and G. Taubin, "Build your own 3D scanner: optical triangulation for beginners," in *Proceedings of SIGGRAPH Asia '09* (ACM, 2009), article no. 2.
5. S. Y. Park and M. Subbarao, "A multiview 3D modeling system based on stereo vision techniques," Mach. Vis. Appl. **3**, 148–156 (2004).
6. A. Censi, L. Iocchi, and G. Grisetti, "Scan matching in the Hough domain," *Proceedings of the IEEE International Conference on Robotics and Automation, 2005* (IEEE, 2005), pp. 2739–2744.
7. M. A. Merraez, J. G. Boticario, M. J. Labor, and D. R. Burton, "Agglomerative clustering-based approach for two-dimensional phase unwrapping," Appl. Opt. **44**, 1129–1140 (2005).
8. J. Guhring, "Dense 3-D surface acquisition by structured light using off-the-shelf components," Proc. SPIE **4309**, 220–231 (2001).
9. Y. Wang, K. Liu, D. L. Lau, and L. G. Hassebrook, "Period coded phase shifting strategy for 3-D real-time acquisition and reconstruction," IEEE Trans. Image Process. **20**, 3001–3013 (2011).
10. S. Winkelbach, S. Molkenstruck, and F. M. Wahl, "Low-cost laser range scanner and fast surface registration approach," Lect. Notes Comput. Sci. **4174**, 718–728 (2006).
11. S. Molkenstruck, S. Winkelbach, and F. M. Wahl, "3D body scanning in a mirror cabinet," Lect. Notes Comput. Sci. **5096**, 284–293 (2008).
12. M. Sheehan, A. Harrison, and P. Newman, "Automatic self-calibration of a full field-of-view 3D *n*-laser scanner," in *Proceedings of the International Symposium on Experimental Robotics (2010)* (Mobile Robotics Group, Oxford University, 2010), pp. 1–14.
13. E. Nasset and R. Nelson, "Using airborne laser scanning to monitor tree migration in the boreal-alpine transition zone," Remote Sens. Environ. **110**, 357–369, (2007).
14. Z. Lv and Z. Zhang, "Build 3D laser scanner based on binocular stereo vision," in *Proceedings of 2011 Fourth International Conference on Intelligent Computation Technology and Automation*, Vol. 1 (IEEE, 2011), pp. 600–603.
15. Z. Li, C. Wang, and Y. Shi, "An algorithm for detecting center of structured light stripe combining gradient sharpening with Barvcenter method," J. Image Graphics **13**(1), 64–68 (2008).
16. H. Xiong, Z. Zong, Q. Gao, and C. Chen, "Precise method for extracting center of structured light stripe," Comput. Eng. Appl. **45**(10), 235–237 (2009).
17. J. Wu, P. Wang, J. Chen, and M. Wu, "Method of linear structured light sub-pixel center position extracting based on gradient barycenter," J. Image Graphics **14**(7), 1354–1360 (2009).