Dense 3D Reconstruction with an Uncalibrated Stereo System using Coded Structured Light

Ryo Furukawa

Faculty of Information Sciences
Hiroshima City University
Hiroshima, Japan

Hiroshi Kawasaki

Institute of Industrial Science
Tokyo University
Tokyo, Japan

Abstract

In this paper, we describe a new uncalibrated active vision system. Normally, active vision systems need either precise calibrations or relatively fixed cameras and light sources, which require 3D measurement systems to make use of complicated measurement processes or measurement equipment. Recently, several uncalibrated active vision systems have been studied. Since calibrations of these systems are done either while or after shape data is obtained, the user can simply place the equipment arbitrarily and scan the objects. The proposed system is an uncalibrated active vision system which consists of a camera and a video projector. Geometrically, this is a camera-to-camera system with one camera replaced by a projector. Uncalibrated vision systems based on perspective projection devices are prone to unmeasured scaling parameters. To solve this problem, a laser pointer is attached to the projector. In the present work, coded structured light is used to obtain dense correspondence points, resulting in more detailed shape representations. Experiments using the implemented system show that it is able to reconstruct dense and precise 3D shapes.

1. Introduction

3D acquisition stereo systems can be categorized into two basic types: a passive stereo system and an active stereo system. The former can recover 3D shapes only from multiple images, therefore no special devices are necessary and the systems are usually easy to use. However, in order to recover 3D shapes from images by passive stereo, accurate correspondences between images are required, making this a difficult task. Furthermore, candidates for correspondent points are usually limited to a few feature points; thus, to recover dense 3D points, interpolation is necessary and data accuracy may be unreliable.

An active stereo system, on the other hand, utilizes light or laser projector for scanning, and thus has no difficulty in retrieving high precision correspondences; the accuracy of the 3D points is therefore relatively high. Another benefit of an active stereo system is that dense 3D points can easily be captured by controlling laser devices using structured light or other methods. Given the advantages of an active stereo system, such a system is usually adopted for practical use, especially for scanning intricately shaped objects.

One of the disadvantages of active stereo systems is that these systems usually require special and expensive devices, such as servo actuators for laser control, special structured light projectors, etc., which are usually heavy and costly. Recently, many types of low cost and handy-sized projectors have become readily available, and we therefore now have a good opportunity to develop a portable and low-cost active stereo system.

With regard to the disadvantages of active stereo systems, simplification of the calibration process is a critical issue. In an active stereo system whose projector and camera are not fixed to each other, precise precalibration is required before 3D scanning each time the conditions of the system are changed. Clearly, this significantly compromises the convenience of the system. If we can eliminate the precalibration process from an active stereo system, it will greatly improve both the convenience and the practicality of the system. Based on these facts, many active stereo systems which do not require precalibration have been proposed to date. In this paper, we propose an uncalibrated active stereo system which has three important advantages over other previously proposed uncalibrated system: first, our system can capture dense 3D points in a short period of time; second, the system does not require any calibration at any time during the entire scanning process; third, we provide a simple solution for eliminating ambiguity in the scaling of 3D data. The elimination of ambiguity is critical for the practical use of 3D scanners.

2. Related Works

Many active 3D scanning systems have been proposed to date. Among them, the active stereo system can be considered one of the most common techniques. One simple implementation of an active stereo system is to utilize a servo actuator to control the laser projector [16]. However, such a system is usually heavy, large and expensive because of the required high precision mechanical devices and the necessity for accurate calibrations. To avoid using precision mechanical devices, a structured light-based system may be used. However, while the structured light-based system has great advantages in scanning efficiency, i.e., it can retrieve dense 3D points in a short period of time, precise precalibration is necessary for installation and this is usually a laborious task.

To avoid the calibration problems mentioned above, many uncalibrated active stereo methods have been proposed [3, 9, 6, 17, 7, 11]. Takatsuka [17] and Furukawa [11] have proposed active stereo 3D scanners with online calibration methods. Each of their systems consists of a video camera and a laser projector attached with LED markers, and executes projector calibration for every frame. Thus, system configuration is relatively free and a real-time system is achieved. However, when calibration is performed for each frame, the system tends to have insufficient accuracy and low efficiency for practical use.

Uncalibrated stereo techniques have been studied extensively for passive stereo systems [8], and several researchers have proposed other practical methods based on these techniques, essentially substituting a projector for one camera of the stereo paired cameras. Fofi et al. [10] proposed a method in which a 3D shape is first reconstructed in a projective space and then is upgraded to Euclidean space. Their method does not require precalibration; to estimate the upgrade parameter, they assume an affine camera model, however the technique then cannot be applied to scenes which have a large disparity in depth. Because of their strong assumption of the affine camera model, they dismissed the problem of the inevitable ambiguity of scaling in an uncalibrated stereo method. This method also requires a plane in the scene which should be captured.

Li and Lu's method [15] executes a projector calibration for light planes projected to two planes in the scene. After calibration, the camera can be freely moved and zooming is also allowed during the 3D acquisition process. However, under this method, new calibration is required whenever the projector position is changed. Additionally, this method requires calibration for each plane and the authors do not mention how dense 3D points are acquired. Li and Lu's method adopts an off line calibration to eliminate scaling ambiguity.

Note that ambiguity of scaling inevitably exists for uncalibrated stereo methods and this usually causes many problems in practical use. For example, when we scan an object from varying view directions to capture the entire shape of the object, different scaling parameters for each scan make it difficult to achieve correct registration and integration. To avoid such problems, the following two techniques can be considered:

- Fix both the projector and the camera to keep the scaling value constant.
- Scan an object of known size and pre-determine the scaling value.

Both of these techniques usually take a long time and greatly reduce the flexibility of the system.

The system we propose here is essentially based on an uncalibrated stereo method for a passive stereo system, which we extend in order to be able to apply it to a large number of correspondences retrieved by a space coded structured light method. With our system, dense 3D points are simultaneously retrieved with online calibration, which is done by minimizing a normalized error function. We also propose a simple method of eliminating ambiguity of scaling by attaching a laser pointer to the projector.

If we can use multiple scanning data taken moving the location of the projector (or the camera), we can improve the results of 3D reconstructions obtained by the proposed method. This technique is described in another paper[12].

3. System Configuration and 3D Reconstruction

The 3D reconstruction system developed in this work consists of a video projector and a camera. A laser pointer is attached to the projector and is used for determining the scaling parameters, which cannot be fixed with uncalibrated stereo methods. If the ambiguity of the scaling parameter can be left unsolved, the laser pointer can be omitted. Fig. 1 shows the configuration of the system.

When the shape of an object is measured, the camera and the projector are oriented toward the object. With the structured light method, a set of dense correspondence points is obtained. The 3D locations of the correspondence points are reconstructed with an uncalibrated stereo method.

Our 3D reconstruction system has the following features, which are highly desirable in a practical 3D measurement system:

- The projector and the camera can be located arbitrarily.
- No limitations are imposed on the geometry of the measured scene except that, if the scaling parameters must be measured, the point lit by the laser pointer should be observable from the camera.

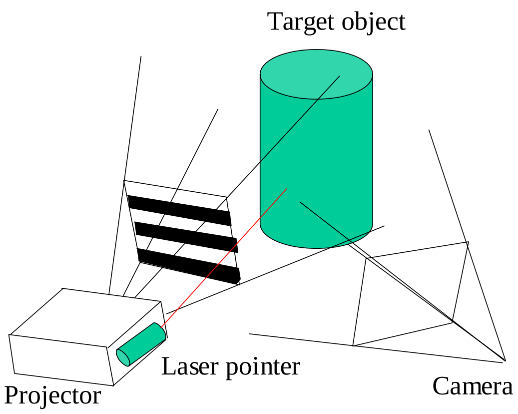


Figure 1: Components of the 3D measurement system.

- Recalibration is not needed even if the locations of the camera and the projector are changed.
- The measurement process is completely automatic without any need for user intervention.
- A dense depth image is obtained with the correct scaling parameter.

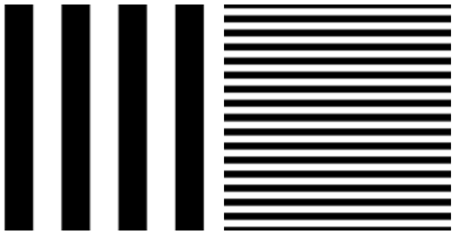
A video projector can be thought of as a reversed camera, and we can define intrinsic parameters of a projector such as a focal length like cameras.

In the present work, the intrinsic parameters of the camera are assumed to be known, while the focal length of the projector is assumed to be unknown. This is because the intrinsic parameters of the camera can be obtained by existing methods, such as taking pictures of a calibration pattern, while those of the projector are more difficult to obtain. For example, to estimate the parameters of the projector, a certain pattern should be projected onto a special object, like a calibration pattern on a plane, the projected pattern should be captured by a camera, and the pictures should be analyzed.

3.1. Obtaining a set of correspondence points by structured light

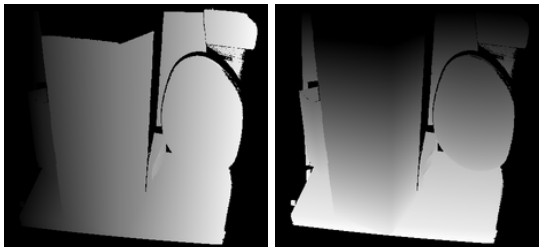
For active stereo systems, structured light is often used to obtain correspondence points. To resolve the correspondence effectively, coded structured light methods have been used and studied [2, 5, 4, 14, 13]. In the present method, directions from the projector are encoded into the light patterns, which are projected onto the target surface. The light patterns projected to each pixel are decoded from the obtained images, and the mapping from each pixel in the images to directions from the projector is obtained.

Since dense correspondence points are desired to obtain detailed shape data, one of the previously proposed coded



(a) x-coordinate (b) y-coordinate

Figure 2: Example of binary patterns for coded structured light.



(a) x-coordinate (b) y-coordinate

Figure 3: Coded images by structured light.

structured light methods is used in this work, specially the gray code method proposed by Inokuchi [14](Fig.2). Since the gray code encodes 1D locations in the projected patterns, we applied the code twice, once for the x-coordinate of the projected pattern and once for the y-coordinate. Based on the compound gray codes, point-to-point correspondences between the directions from the projector and the pixels in the image are resolved (Fig.3).

3.2. 3D reconstruction

From the set of correspondence points, 3D reconstruction can be carried out using uncalibrated stereo methods. One such method involves reconstructing the 3D shape in the projection space and upgrading it into a Euclidean space. Another method is to solve the epipolar constraints in Euclidean space. In the present work, non-linear optimization is applied from the first.

We chose to proceed in this way because both reconstruction in the projection space and solving the epipolar constraints allow more freedom than the real 3D geometry has because they do not take the constraints of rigid transformation or pre-measured camera parameters into account. Thus, non-linear optimization has often been applied to improve the solution and refine the results. Recently, given the improved computational capabilities of PCs, 3D reconstructions using only non-linear optimizations have been studied by some researchers [1]. In the present study, we have chosen to follow this approach.

We call a coordinate system which is fixed with the projector (or the camera) the projector (camera) coordinate system. Coordinate values expressed in this system are the projector (camera) coordinate. The origin of the projector (camera) coordinate system is the optical center of the projector (camera). The forward direction of the projector (camera) is the minus direction of the z-axis of the projector (camera) coordinate system. The x and y-axis of the projector (camera) coordinate system are parallel with the vertical and horizontal directions of the image coordinate system of the screen.

Let the focal length of the projector be f_p, and the direction vector of the ith correspondence point expressed in the projector coordinates be

$$(u_{pi},\ v_{pi},\ -f_p)^t.$$

Here, we express the rigid transformation from the projector coordinates to the camera coordinates as the rotation matrix $\mathbf{R}_p$ and the translation vector $\mathbf{t_p}$. The rotation is expressed by the parameters of Euler angles α_p, β_p and γ_p, and the rotation matrix is thus expressed as $\mathbf{R}_p(\alpha_p, \beta_p, \gamma_p)$.

The direction of the correspondence points observed by the camera is converted to the screen coordinates of a normalized camera, with corrected effects of the lens distortions. Let the converted coordinates be

$$(u_{ci},\ v_{ci},\ -1)^t.$$

If the epipolar constraints are met, the lines of sights from the camera and the projector intersect in the 3D space. The line from the projector in the camera coordinates is

$$r\{\mathbf{R}_p(\alpha_p, \beta_p, \gamma_p)\}(u_{pi}/f_p,\ v_{pi}/f_p,\ -1)^t + \mathbf{t_p}, \qquad (1)$$

where r is a parameter. The line from the camera is

$$s(u_{ci},\ v_{ci},\ -1)^t, \qquad (2)$$

where s is a parameter.

To achieve the epipolar constraints, the distance between the two lines (1) and (2) should be minimized. Let the direction vectors of the lines be expressed as

$$\begin{aligned} &\mathbf{p}_{ci} := N\ (u_{ci},\ v_{ci},\ -1)^t, \\ &\mathbf{q}_{ci}(\theta, f_p) := \\ &\quad N\ \{\mathbf{R}_p(\alpha_p, \beta_p, \gamma_p)\}(u_{pi}/f_p, v_{pi}/f_p, -1)^t, \end{aligned} \qquad (3)$$

where N is an operator which normalizes a vector (i.e. converts a vector into a unit vector with the same direction), and $\theta := (\alpha_p, \beta_p, \gamma_p, \mathbf{t_p})$ represents the tuple of the extrinsic parameters of the projector. Then, the distance between the lines is

$$E_i(\theta, f_p) := |\mathbf{t_p} \cdot N\ (\mathbf{p}_{ci} \times \mathbf{q}_{ci}(\theta, f_p))|, \qquad (4)$$

where "·" indicates dot product.

$E_i(\theta, f_p)$ includes systematic errors whose variances change with the parameters (θ, f_p) and the data index i. To compose an error evaluation function unbiased about the parameters (θ, f_p), $E_i(\theta, f_p)$ should be normalized by the expected error level. Assuming the epipolar constraints are met, the distance from the intersection of the lines to the camera and the projector are

$$\begin{aligned} &D_{ci}(\theta, f_p) := \|\mathbf{t_p} \times \mathbf{q}_{ci}(\theta, f_p)\| / \|\mathbf{p}_{ci} \times \mathbf{q}_{ci}(\theta, f_p)\|, \\ &D_{pi}(\theta, f) := \|\mathbf{t_p} \times \mathbf{p}_{ci}\| / \|\mathbf{p}_{ci} \times \mathbf{q}_{ci}(\theta, f_p)\|, \end{aligned} \qquad (5)$$

respectively. Using the distances, the distance normalized by the error level is

$$\tilde{E}_i(\theta, f_p) := \frac{E_i(\theta, f_p)}{\epsilon_c D_{ci}(\theta, f_p) + \epsilon_p D_{pi}(\theta, f_p)/f_p} \qquad (6)$$

where ϵ_c and ϵ_p are the errors intrinsic to the camera and the projector expressed as lengths in the normalized screen planes. In our experiments, we used pixel sizes for ϵ_c and ϵ_p.

Then, the function $f(\theta, f_p)$, which is minimized with the non-linear optimization is expressed as the following form:

$$f(\theta, f_p) := \sum_i \tilde{E}_i(\theta, f_p) + (\|\mathbf{t_p}\| - 1)^2 \qquad (7)$$

The last term gives the constraint to reduce the freedom of the scaling. Without this term, one of the globally optimal solution is $\mathbf{t_p} = \mathbf{0}$, which is meaningless.

3.3. Estimation of Scaling Parameter

The measured 3D shape is scaled by an unknown multiplier from the real shape. The following methods can be applied to determine the multiplier: which are

1. measuring the length of the two points on the real shape,
2. measuring an object with a known shape (a calibration object) and the target object successively without moving the camera nor the projector,
3. or measuring a calibration object and the target object simultaneously.

However, all of these techniques normally require some human intervention such as measuring or specifying the calibration object, making it difficult to develop a completely automatic measuring process.

In order to determine the scaling parameters more easily, we attach a laser pointer to the projector and project a mark onto the measured surface, which is then observed by the camera. The projected laser light forms a fixed line in the 3D space expressed by the projector coordinates. The 3D

Table 1: Parameters estimated by calibration and from data.

	By calibration	From data
f_p	0.043[m]	0.043[m]
$(\alpha_p, \beta_p \gamma_p)$	(4.58˚, 15.6˚, 4.31˚)	(7.19˚, 17.0˚, 2.95˚)
$\mathbf{t_p}/\|\mathbf{t_p}\|$	(0.49, -0.63, 0.60)	(0.50, -0.67, 0.54)

point lit by the pointer is determined from the image, and the scaling parameter is calculated from this data.

The line formed by the laser pointer should be calibrated when it is attached to the projector. To do this, multiple points on the laser are obtained by measuring an object with a known shape placed such that it crosses the laser line. The points are fitted to a line to obtain the parameters of the laser line. Calibration is needed only when the laser pointer is attached to the projector.

The calculation to evaluate the scaling of the measured shape with a 3D point which is known to be on the laser line is as follows. Let the detected 3D location be $(x_{pm}\ y_{pm}\ z_{pm})^t$ in the projector coordinates. The real 3D coordinates of the point can be expressed as $\lambda(x_{pm}\ y_{pm}\ z_{pm})^t$, where λ is the scaling multiplier. Since λ is unknown, the point is on the line which passes both the optical center of the projector and the point $(x_{pm}\ y_{pm}\ z_{pm})^t$. The point marked by the pointer is on the calibrated line; thus, by taking the cross point of the line $\lambda(x_{pm}\ y_{pm}\ z_{pm})^t$ and the calibrated laser line, we can decide the scaling parameter λ.

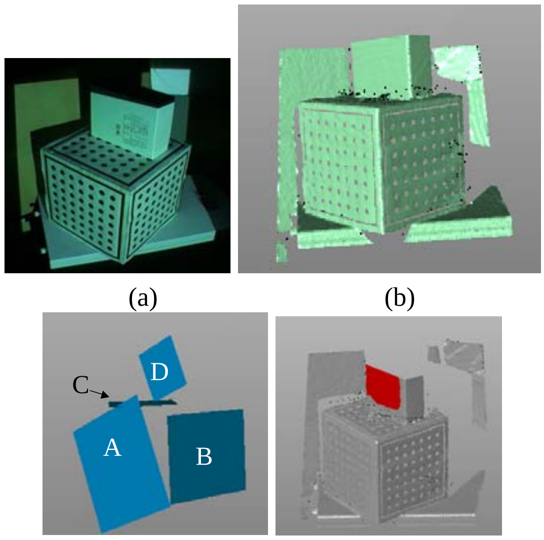
4. Experiments

4.1. Evaluation of Accuracy

To evaluate the accuracy and effectiveness of our proposed method, we scanned a scene of a cube (20cm $\times$ 20cm $\times$ 20cm) and a box (Fig.4(a)), calculating the extrinsic parameters and the focal length of the projector.

For a comparison, We also performed a calibration of the extrinsic parameters and the focal length using the known 3D positions of the vertices of the cube. The initial values of the position and direction of the projector were α_p=0˚, β_p=20˚, γ_p=0˚, $\mathbf{t_p}$=(1,0,0), $f_p = 0.05$. The results are shown in Tab. 1.

We also evaluated the quality of the obtained 3D point set shown in Fig.4(b). We applied a plane fitting algorithm for the faces in the scene, obtaining 4 planes (A,B,C,D) shown in Fig.4(c). By using the estimated plane parameters, we calculated angles between the estimated planes. The angles between plane A and B is 93.0˚, A and C is 91.7˚, B and C is 87.3˚, and A and D is 1.8˚. We also calculated the deviations of the points shown in Fig.4(d) from the plane D. The root of the mean squared deviations from the plane was 0.46mm in real scale.



(a) (b) (c) (d)

Figure 4: Scanning of a cube with known size: (a) the scanned scene, (b) the scan result, (c) planes fitted to the point sets, and (d) points compared to plane D.

The deviations of the points from the plane D indicate the intrinsic error level of our system, taking noise and quantization effects into account. Our experimental results show that the residuals are small enough to ensure the high accuracy of 3D scanning using our system. The angle between planes indicates the correctness of the calibration method, and should show 90˚ for planes A-B,A-C,B-C, or 0˚ for A-D, if the parameters are correctly estimated. The measured angles show that the errors of the angles varied from 1˚ to 3˚.

Finally, we evaluated our method of estimating the scaling parameter. First, the line formed by the laser light was calibrated by measuring 2 points lit by the laser. The laser line was expressed in the projector coordinates using the real scale. Then, we moved the camera and projector, scanned the cube shown in Fig.5(a) with the initial values $(\alpha_p, \beta_p, \gamma_p)$=(0˚, 20˚, 0˚), $\mathbf{t_p}$=(1,0,0), $f_p = 0.05$, and calculated the scale of the measured point set from the detected point lit by the laser. We fitted planes to the scaled point set at the top and the bottom of the cube, and measured the length between the planes at the nearest edge of the cube (Fig.5(b)). The estimated length was 19.9cm, which is close to the actual length 20cm, thus confirming that our estimation method is effective and practical.

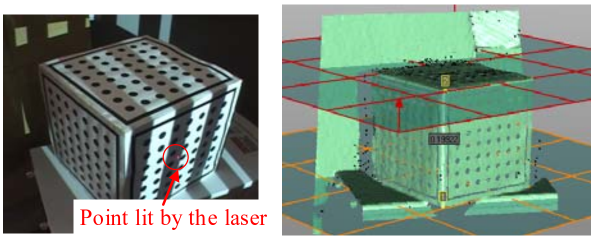


Figure 5: Scaling using a known laser line: (a) the scanned scene and the point lit by the laser, and (b) the measured edge

4.2. Large range scanning by moving the camera

We scanned a scene repeatedly by moving the camera. Since our proposed method does not require camera-projector calibration for scanning, we can continuously move the camera while performing a sequence of scans. With this feature of scanning, we can scan a wide range of objects without any special registration method. Our scanning results are shown in Fig.6, which shows that all scanning results are aligned correctly from the beginning and that a large range of objects is successfully retrieved.

4.3. Examples

Finally, we scanned several objects with intricate shapes and curved surfaces to verify the reliability and the effectiveness of our proposed method. Results are shown in Fig.7.

5. Conclusion

In this paper, we propose a new uncalibrated active vision system which enables dense 3D scanning with a single scanning process and without any precise calibrations or special devices. Our proposed method is based on an uncalibrated stereo technique for a passive vision system with one of the cameras replaced by a projector. We also propose an efficient method of estimating the scaling parameter of the system by simply attaching a laser pointer to the projector.

By using our proposed method, the camera and projector can be arbitrarily installed and it is possible to start 3D scanning immediately and without any precalibrations or complicated preparations. Another advantage of our method is that, since the calibration process is eliminated, we can freely move the projector while a sequence of scans is being carried out, and thus, we can scan a wider range of the object more accurately and efficiently than with other methods. To verify the reliability and the effectiveness of our proposed method, we conducted several experiments with

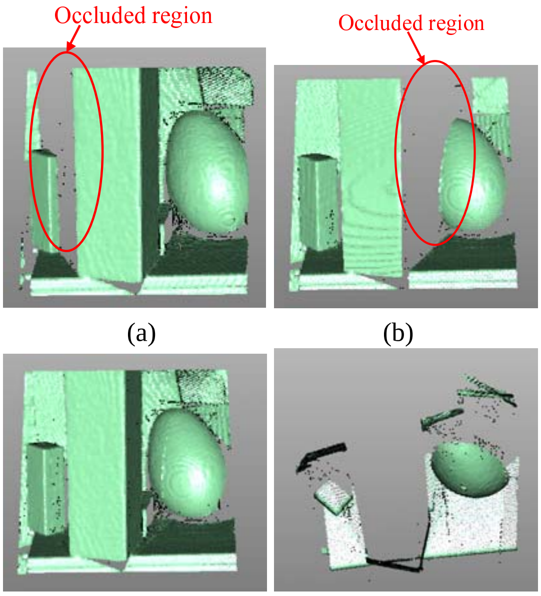


Figure 6: Scanning a intricate scene from various view directions: (a) the scanned point set with the first camera position, (b) the scanned point set with the second camera position, (c) the integrated point set, and (d) the integrated point set shown from the top.

the proposed system and actual objects. The results of our experiments confirm the effectiveness of our proposed system.

References

[1] A. Amano, T. Migita, and N. Asada. Stable recovery of shape and motion from partially tracked feature points with fast nonlinear optimization. In *15th Vision Interface*, pages 244–251, 2002.

[2] J. Batlle, E. Mouaddib, and J. Salvi. Recent progress in coded structured light as a technique to solve the correspondence problem: a survey. *Pattern Recognition*, 31(7):963–982, 1998.

[3] J. Y. Bouguet and P. Perona. 3D photography on your desk. In *Int. Conf. Computer Vision*, pages 129–149, 1998.

[4] K. L. Boyer and A. C. Kak. Color-encoded structured light for rapid active ranging. *IEEE Trans. on Patt. Anal. Machine Intell.*, 9(1):14–28, 1987.

[5] D. Caspi, N. Kiryati, and J. Shamir. Range imaging with adaptive color structured light. *IEEE Trans. on Patt. Anal. Machine Intell.*, 20(5):470–480, 1998.

[6] Chang Woo Chu, Sungjoo Hwang, and Soon Ki Jung. Calibration-free approach to 3D reconstruction using light

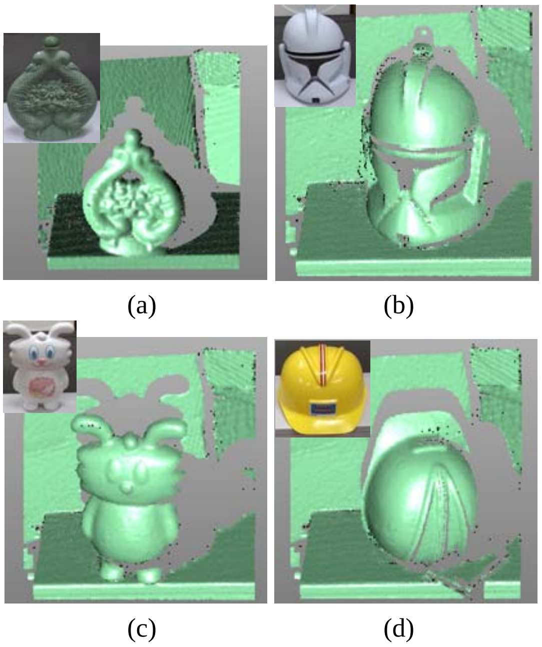


Figure 7: Examples of the scanned objects:(a) An ornament, (b) a mask, (c) a doll, (d) a helmet.

stripe projections on a cube frame. In *Third Int. Conf. on 3D Digital Imaging and Modeling*, pages 13–19, 2001.

[7] James Davis and Xing Chen. A laser range scanner designed for minimum calibration complexity. In *Third Int. Conf. on 3D Digital Imaging and Modeling*, pages 91–98, 2001.

[8] O. Faugeras. *Three-Dimensional Computer Vision - A Geometric Viewpoint*. Artificial intelligence. M.I.T. Press Cambridge, MA, 1993.

[9] R. B. Fisher, A. P. Ashbrook, C. Robertson, and N. Werghi. A low-cost range finder using a visually located, structured light source. In *Second Int. Conf. on 3D Digital Imaging and Modeling*, pages 24–33, 1999.

[10] David Fofi, Joaquim Salvi, and El Mustapha Mouaddib. Uncalibrated vision based on structured light. In *ICRA*, pages 3548–3553, 2001.

[11] Ryo Furukawa and Hiroshi Kawasaki. Interactive shape acquisition using marker attached laser projector. In *Int. Conf. on 3D Digital Imaging and Modeling2003*, pages 491–498, 2003.

[12] Ryo Furukawa and Hiroshi Kawasaki. Uncalibrated multiple image stereo system with arbitrarily movable camera and projector for wide range scanning. In *Proceedings of the 5th international conference on 3-D digital imaging and modeling (3DIM 2005) (to appear)*, 2005.

[13] O. Hall-Holt and S. Rusinkiewicz. Stripe boundary codes for real-time structured-light range scanning of moving objects. In *Int. Conf. Computer Vision*, volume 2, pages 359–366, 2001.

[14] S. Inokuchi, K. Sato, and F. Matsuda. Range imaging system for 3-D object recognition. In *ICPR*, pages 806–808, 1984.

[15] Y. F. Li and R. S. Lu. Uncalibrated euclidean 3d reconstruction using an active vision system. *IEEE Transactions on Robotics and Automation*, 20(1):15–25, 2004.

[16] MINOLTA Co. Ltd. Vivid 900 non-contact digitizer. In *http://www.minoltausa.com/vivid*.

[17] Masahiro Takatsuka, Geoff A.W. West, Svetha Venkatesh, and Terry M. Caelli. Low-cost interactive active monocular range finder. In *CVPR*, volume 1, pages 444–449, 1999.