Uncalibrated multiple image stereo system with arbitrarily movable camera and projector for wide range scanning

Ryo Furukawa
Faculty of Information Sciences,
Hiroshima City University, Japan
ryo-f@cs.hiroshima-cu.ac.jp

Hiroshi Kawasaki
Faculty of Information Engineering,
Saitama University, Japan
kawasaki@mm.ics.saitama-u.ac.jp

Abstract

In this paper, we propose an uncalibrated, multi-image 3D reconstruction, using coded structured light. Normally, a conventional coded structured light system consists of a camera and a projector and needs precalibration before scanning. Since the camera and the projector have to be fixed after calibration, reconstruction of a wide area of the scene or reducing occlusions by multiple scanning are difficult and sometimes impossible. In the proposed method, multiple scanning while moving the camera or the projector is possible by applying the uncalibrated stereo method, thereby achieving a multi-image 3D reconstruction. As compared to the conventional coded structured light method, our system does not require calibration of extrinsic camera parameters, occlusions are reduced, and a wide area of the scene can be acquired. As compared to image-based multi-image reconstruction, the proposed system can obtain dense shape data with higher precision. As a result of these advantages, users can freely move either the cameras or projectors to scan a wide range of objects, but not if both the camera and the projector are moved at the same time.

1 Introduction

For practical use of a 3D scanner, an active system is usually adopted because of its high accuracy, robustness, and efficiency. With respect to 3D scanners, a coded structured light stereo system can capture dense 3D data efficiently. The system has therefore been intensively studied. A coded structured light system usually consists of a camera and projector and these need to be precalibrated. Once the camera and the projector are precalibrated, they are fixed at their positions. This leads to several drawbacks in the system. For example, it is impossible to fix occlusions. Additionally, the portability of the system is seriously compromised.

In contrast, we propose an uncalibrated, coded structured light system, which exploits multi-image information. Since the system does not require precalibration of the extrinsic parameters, scanning can be repeated while moving the camera and the projector. As compared to the conventional coded structured light method, this system needs no calibration of extrinsic camera parameters, occlusions are reduced, and a wide area of the scene can be acquired.

With regards to a projector being equivalent to a camera, the proposed method has an apparent similarity with the uncalibrated stereo 3D reconstruction that uses multiple images. As compared to a feature-based, multi-image reconstruction, the proposed system can obtain dense shape data. In addition, since the correspondences errors are much smaller when coded structured light is used, the 3D reconstruction is more stable than feature-based systems. Thus, this system can produce better results with fewer images.

Since this system needs no calibration of the extrinsic parameters, the camera and the projector can be moved, except that both of the devices shouldn't be moved at the same time, so that it is possible to keep track of the consistent extrinsic parameters. This is because correspondences are only acquired between the positions of the camera and the projector during the scanning, using the coded structured light method.

Based on this consideration, two scanning methods are used in this paper. The first method is to scan moving only one device(i.e., either the camera or the projector), while the other device is fixed. Since this scanning method utilizes information of multiple images, the precision in estimating the camera positions improves, producing a better quality 3D reconstruction. Additionally, occlusions are reduced by multiple scans.

The second method is to scan by alternatively keeping one device fixed. For example, a scene can be scanned by moving the camera while the projector is stationary. The camera is then fixed and the scene is scanned moving the projector. The steps can be repeated. With this type of scanning, the system can cover a wider area of the scene, even

up to the entire model acquisition.

2 Related Work

Many active 3D scanning systems have been proposed to date; of these, the active stereo system can be considered as 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 [13]. However, such a system is usually heavy, large, and expensive because of the high precision mechanical devices required 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 [2, 8, 5, 14, 6, 10]. Takatsuka [14] and Furukawa [10] have proposed active stereo 3D scanners with on-line 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, users have the freedom to decide on the system configuration, 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 to be of practical use.

Uncalibrated stereo techniques have been studied extensively for passive stereo systems [7], and several researchers have proposed other practical methods based on these techniques, essentially substituting one camera of the stereo paired cameras with a projector. Fofi et.al. [9] proposed a method in which a 3D shape is first reconstructed in a projective space and is then upgraded to Euclidean space. In order to estimate the upgrade parameter, they assume an affine camera model, however, in this case, the technique cannot be applied to scenes that have a large disparity in depth. Also, this method requires a plane to be captured in the scene and cannot resolve the scaling ambiguity. Thus, the system is not practical for actual use.

With regard to multiple scannings, the methods proposed above do not assume multiple cameras and projectors. Therefore, to reconstruct a large scene, which cannot be obtained in a single scanning, merging several reconstruction results from multiple scanning while moving the camera or the projector is required. Since estimated values of the scaling parameters or the focal length may differ for each scanning, it is often very difficult to achieve correct registration and integration. In order to resolve such problems, application of multi-image uncalibrated stereo methods appears to be a promising idea, because these methods take the coherency of the scaling and the focal length into account. In this paper, we propose an uncalibrated stereo method that uses multiple cameras and projectors, to make a single consistent shape, using the projector as a reversed camera. With this method, scaling ambiguity can be minimized and wide range of geometry can be efficiently retrieved.

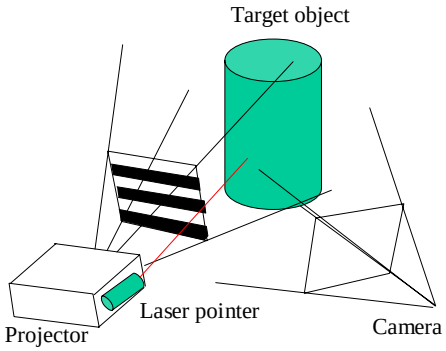


Figure 1. Components of the 3D measurement system.

3 System Configuration and 3D Reconstruction

The 3D reconstruction system developed in this work consists of a video projector and a camera. The video projector is used as a substitute for the camera, which is used in the uncalibrated stereo method. Since a video projector can be thought of as a reversed camera, we can define intrinsic parameters of a projector, such as focal length, as with cameras. A laser pointer is attached to the projector and is used for determining the scaling parameters that cannot be fixed with uncalibrated stereo methods. If the ambiguity of the scaling parameter can be left unsolved, the laser pointer can be omitted. Figure 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.

In this work, a multi-image uncalibrated stereo method is achieved by moving the camera and the projector. Since the correspondences are only obtained between the position of the camera and the projector on scanning, moving both the devices simultaneously is not allowed while reconstructing a single 3D scene. However, it is possible to allow the movement either the camera or the projector.

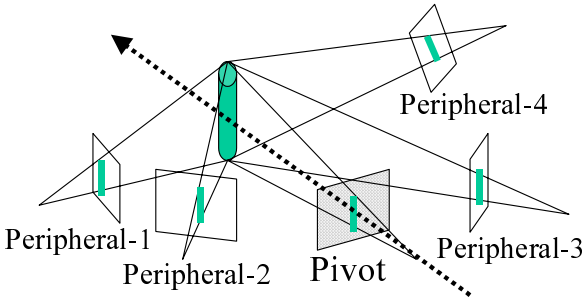
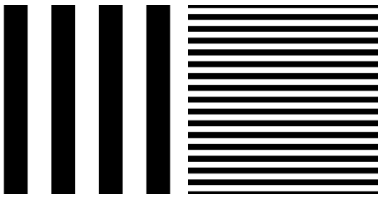


Figure 2. A pivot scanning.



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

Figure 3. Example of binary patterns for coded structured light.

Multiple scanning by moving one of the devices while the other is fixed is an efficient method to improve precision of the results. In this work, this type of scanning is referred as "a pivot scanning." Additionally, the position of the fixed device is called "the pivot position," and the positions of the other device are called "the peripheral positions."(Fig.2)

With a pivot scanning reconstruction, a relatively simple algorithm makes it possible to use information of multiple images as is described in section 3.3.

Another successful scanning method involves keeping the camera and projector fixed, alternatively. The following is one example of the scanning. First, "a pivot scanning" is conducted by moving the camera while the projector is fixed. The camera is then fixed and "a pivot scanning" is conducted by moving the projector. This process can be repeated as many times as necessary. Since the extrinsic parameter is calculated on scanning by the uncalibrated stereo method for the projector-camera pair, it is possible to calculate extrinsic parameters in-between arbitrary scannings by exploiting the sequence of alternative movement of the camera and the projector. We call this type of scanning "an alternate scanning." Using this technique, we can reconstruct a wider area of the scene. Obtaining the object's shape as observed from all the directions is also possible.

Since, "an alternate scanning" is divided into several pivot scannings, each of the pivot scannings which forms an alternate scanning is resolved separately by multi-image stereo, and the results are merged into a single scene, which gives the resolution of the alternate scanning.

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

- No limitations are imposed on the geometry of the measured scene, except that, if the scaling parameters are to be measured, the point lit by the laser pointer should be observable from the camera.

- Calibration of the extrinsic parameters is not needed.

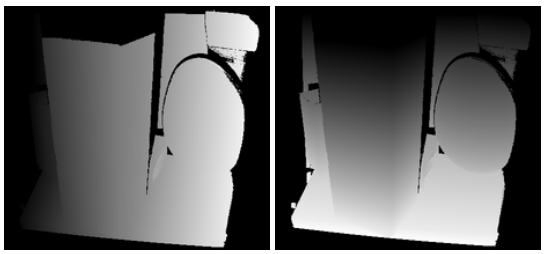
- Dense 3D reconstruction with the correct scaling parameter is achieved.

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 Correspondences between images 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 [1, 4, 3, 12, 11]. 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 onto 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 structured light methods is used in this work, particularly the gray code method proposed by Inokuchi [12] (Fig. 3). 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.4).



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

Figure 4. Coded images by structured light.

3.2 Initial 3D reconstruction

To obtain an initial 3D shape to resolve a pivot scanning, an uncalibrated stereo method is applied to a set of correspondence points between the pivot position and one of the peripheral positions. The devices of the image pair should include a projector and camera. Non-linear optimization of a squared error function on an image plane is used for the shape acquisition.

A coordinate system that is fixed with the projector or the camera is referred to as the device coordinate system. Coordinate values expressed in this system are the device coordinates. The origin of the device coordinate system is the optical center of the projector or the camera. The forward direction of the projector or the camera is represented by the negative values on the z-axis of the device coordinate system. The x-axis and y-axis of the device coordinate system are parallel to the vertical and horizontal directions of the image coordinate system of the screen. The device coordinates of the camera are also referred to as the camera coordinates, and those of the projector are referred to as the projector coordinates.

Here, we assume that the projector is fixed and placed at the pivot position. The discussion remains almost the same if the camera were to be fixed and placed at the pivot position.

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 of the jth camera as the rotation matrix $\mathbf{R}_{c,j}$ and the translation vector $\mathbf{t_{c,j}}$. The rotation is expressed by the parameters of Euler angles $\alpha_{c,j}$, $\beta_{c,j}$ and $\gamma_{c,j}$ and the rotation matrix is thus expressed as $\mathbf{R}_{c,j}(\alpha_{c,j}, \beta_{c,j}, \gamma_{c,j})$.

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

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

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 of the jth camera is

$$r\{\mathbf{R}_{c,j}(\alpha_{c,j}, \beta_{c,j}, \gamma_{c,j})\}(u_{pi}/f_p,\ v_{pi}/f_p,\ -1)^t + \mathbf{t_p}, \tag{1}$$

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

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

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

$$\begin{aligned}&\mathbf{p}_{ci} := N\ (u_{ci,j},\ v_{ci,j},\ -1)^t,\\ &\mathbf{q}_{ci}(\theta, f_p) :=\\ &\ N\ \{\mathbf{R}_{c,j}(\alpha_{c,j}, \beta_{c,j}, \gamma_{c,j})\}(u_{pi}/f_p, v_{pi}/f_p, -1)^t,\end{aligned} \tag{3}$$

where N is an operator that normalizes a vector (i.e., converts a vector into a unit vector with the same direction), and $\theta_{\mathbf{j}} := (\alpha_{c,j}, \beta_{c,j}, \gamma_{c,j}, \mathbf{t_{c,j}})$ represents the tuple of the extrinsic parameters of the projector. Then, the distance between the lines is

$$E_i(\theta_{\mathbf{j}}, f_p) := |\mathbf{t_{c,j}} \cdot N\ (\mathbf{p}_{ci} \times \mathbf{q}_{ci}(\theta_{\mathbf{j}}, f_p))|, \tag{4}$$

where "$\cdot$" indicates a dot product.

$E_i(\theta_{\mathbf{j}}, f_p)$ includes systematic errors the variances of which change with the parameters $(\theta_{\mathbf{j}}, f_p)$ and the data index i. To compose an error evaluation function unbiased with respect to the parameters $(\theta_{\mathbf{j}}, f_p)$, $E_i(\theta_{\mathbf{j}}, f_p)$ should be normalized by the expected error level. Assuming that the epipolar constraints are met, the distance from the intersection of the lines to the camera and the projector is

$$\begin{aligned}&D_{ci}(\theta_{\mathbf{j}}, f_p) := \|\mathbf{t_{c,j}} \times \mathbf{q}_{ci}(\theta_{\mathbf{j}}, f_p)\| / \|\mathbf{p}_{ci} \times \mathbf{q}_{ci}(\theta_{\mathbf{j}}, f_p)\|,\\ &D_{pi}(\theta_{\mathbf{j}}, f) := \|\mathbf{t_{c,j}} \times \mathbf{p}_{ci}\| / \|\mathbf{p}_{ci} \times \mathbf{q}_{ci}(\theta_{\mathbf{j}}, f_p)\|,\end{aligned} \tag{5}$$

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

$$\tilde{E}_i(\theta_{\mathbf{j}}, f_p) := \frac{E_i(\theta_{\mathbf{j}}, f_p)}{\epsilon_c D_{ci}(\theta_{\mathbf{j}}, f_p) + \epsilon_p D_{pi}(\theta_{\mathbf{j}}, f_p)/f_p} \tag{6}$$

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

Then, the function $f(\theta_{\mathbf{j}}, f_p)$, which is minimized with the non-linear optimization is expressed in the following form:

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

The last term gives the constraint to reduce the freedom of the scaling. Without this term, one of the globally optimal solutions $\mathbf{t_{c,j}} = \mathbf{0}$ becomes meaningless.

In this study, the Newton method was applied for the nonlinear optimization. With our experience, if the focal length of the projector f_p was fixed to the certain value, the extrinsic parameters stably converged from the rough initial values such as $\mathbf{t_p} = (1, 0, 0)^t$ and $(\alpha, \beta, \gamma) = (0, 0, 0)$ in most case.

If we estimate both f_p and the extrinsic parameters simultaneously, they sometimes do not converge from the initial values $\mathbf{t_p} = (1, 0, 0)^t$ and $(\alpha, \beta, \gamma) = (0, 0, 0)$. At this case, first, we estimate the extrinsic parameters by fixing f_p the constant value and then, we optimize both f_p and the extrinsic parameters simultaneously by using the estimated extrinsic parameters as the initial inputs.

In our experiments described in sec.4, we set the initial values of extrinsic parameters as $\mathbf{t_p} = (1, 0, 0)^t$ and $(\alpha, \beta, \gamma) = (0, 0, 0)$ when the camera was placed at the right-hand side of the projector and $\mathbf{t_p} = (-1, 0, 0)^t$ and $(\alpha, \beta, \gamma) = (0, 0, 0)$ when the camera was placed at the left-hand side of the projector. In terms of f_p, 1/2 to 3 times the actual focal length was set for initial value. Under these conditions, f_p and the extrinsic parameters were successfully converged for all case.

Once we obtain the parameters t_p and R_p , we can directly recover the 3D shapes by the stereo method.

3.3 Bundle adjustment of multiple stereo pairs

3D reconstruction by the uncalibrated stereo method, as described in section 3.2, is sometimes prone to unstable solutions, even if a sufficiently large number of correspondence points are provided. One example is the case when the target objects are placed at relatively far away from the camera and projector, another example is when the focal lengths of both the camera and the projector are very long. In these cases, the projections of the devices appear like para-perspective projections. Assuming para-perspective projections, 3D reconstruction using only epipolar geometries is impossible. Thus, as the problem tends to become more like a case of para-perspective-projections, the solution gets unstable, even with a sufficient number of correspondence points. Using three or more camera images is one solution for this situation. Since 3D reconstruction with three images is possible even while assuming para-perspective projections (i.e. factorization [15] based methods), the unstableness of the solution is alleviated.

In the 3D reconstruction of a pivot scanning, information of correspondences between 3 or more images is used. Using the information of multiple correspondences for each point, more precise reconstruction can be performed than by using the uncalibrated stereo method with two images.

For the 3D reconstruction, we select a sufficient number of reference points. The direction from the devices to each reference point for all the images, including the pivot image, should be known. The reference points are sampled from the points that comply with this condition. The 3D positions of the reference points and all the extrinsic parameters for the peripheral positions are simultaneously estimated.

Let the coordinates of the ith reference points in the device coordinates of the pivot position be expressed as $r_i(i = 1, 2, .., N_r)$, and the extrinsic parameters of the jth peripheral positions be expressed as θ_j. These values are updated iteratively. For the initial values of r_i, we use the result of the 3D reconstruction using the uncalibrated stereo method described in section 3.2.

The algorithm of the reconstruction is as follows.

1. Sample the reference points that are observable in the pivot position and the peripheral positions.

2. Calculate the 3D positions of the reference points from correspondences between the pivot position and one of the peripheral positions by using the uncalibrated stereo method.

3. Repeat the following steps until all θ_j converge.

   (a) Repeat the following steps for all the indexes of the peripheral positions ($j = 1, 2, ..., N_c$).

      i. Update the extrinsic parameters θ_j for the jth peripheral image, by using the current estimations of the positions of the reference points $r_i(i = 1, 2, .., N_r)$.

   (b) Update the positions of the reference points $r_i(i = 1, 2, .., N_r)$ from the current estimations of the extrinsic parameters of all the peripheral positions ($\theta_j(i = 1, 2, .., N_c)$).

Update of the extrinsic parameters $\theta_{\mathbf{j}} := \alpha_{c,j}, \beta_{c,j}, \gamma_{c,j}, \mathbf{t_{c,j}}$ is performed by minimizing errors on the image plane for each peripheral position. Let the coordinate transformation from the device coordinates of the pivot position to those of the ith peripheral position be

$$Trans(\theta_{\mathbf{j}}, \mathbf{x}) := \mathbf{R}_{c,j}(\alpha_{c,j}, \beta_{c,j}, \gamma_{c,j})\mathbf{x} + \mathbf{t_{c,j}},$$

the mapping of the projection by the standard camera be $Proj$, and the depth of the ith reference point measured by the device coordinates of the pivot position be d_{pi}. By

minimizing

$$Q(\theta_{\mathbf{i}}) := \\ ||Proj(Trans(\theta_{\mathbf{i}}, u_{pi}d_{pi}/f_p\ v_{pi}d_{pi}/f_p\ -d_{pi})^t) \\ -(u_{ci,j}\ v_{ci,j})^t||^2,$$

the extrinsic parameters $\theta_{\mathbf{j}}$ are estimated.

Update of the depth values of the reference points $d_{k,j}$ is done by averaging the depth values calculated by triangulation between the device at pivot position and each of the devices at peripheral positions, this is carried out as follows.

1. Repeat the following steps for each camera index $j = 1, 2, ..., N_c$.
   (a) Calculate the depth image of the projector by triangulation between the projector and the jth camera, treating the extrinsic parameters $\theta_{\mathbf{j}}$ as the input and the depth of the kth pixel of the projector $d_{k,j}$ as the output.
2. Calculate the averaged depth image of the projector by $d_k = \sum_{j=1}^{N_c} d_{k,j}/N_c$, where d_k is the kth depth value.

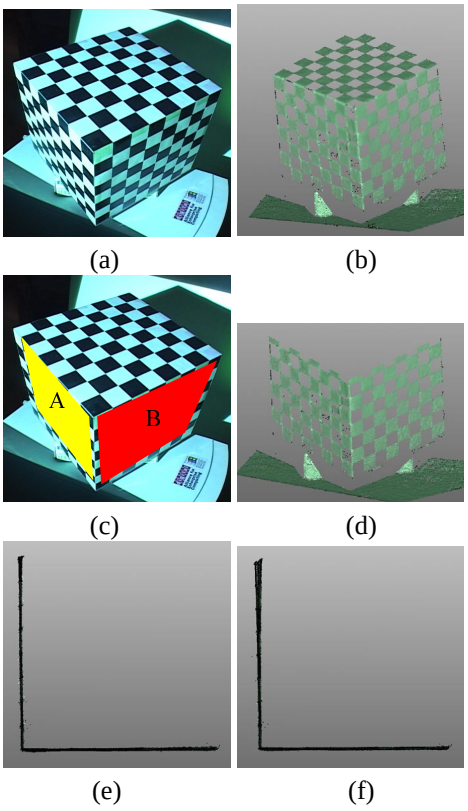
4 Experiments

4.1 Evaluation of Accuracy

To evaluate the accuracy and effectiveness of our proposed method, we performed 3D reconstruction of a cube (20cm $\times$ 20cm $\times$ 20cm) shown in Fig.5(a) using our proposed method. We scanned the cube three times and processed the scanned data with various conditions. The experimented conditions are 3D reconstructions with (condition iv) and without the bundle adjustment process, and the experiments without the bundle adjustment process were performed for both fixed focal length (condition ii) and self-calibrated focal length (condition iii). The scaling factors of the scanned data are incoherent for conditions ii and iii, so we scaled the reconstructed data so that distances from the pivot device to one selected point becomes the same length for those conditions. We also performed an explicit calibration of the extrinsic parameters and the focal length of the projector using the known 3D positions of the markers on the cube (condition i). The resolution of the devices were 720 $\times$ 480 for the camera, and 1024 $\times$ 768 for the projector.

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 of estimated parameters are shown in Tab. 1.

We tried several initial values of focal length for the condition of self-calibrated focal length (iii), and we could get convergence from between 1/2 *to* 3 times the focal length obtained by the explicit calibration. By repeating optimization to refine the parameters, we could get practically the same result. For the condition of fixed focal length (ii and iv), we used the explicitly calibrated focal length as the constant.



(a) (b) (c) (d) (e) (f)

Figure 5. Scanning of a cube with known size: (a) the scanned cube, (b) the scan result, (c) planes fitted to the point sets, (d) cross section of the cube, (e) cross section of our proposed method, and (f)cross section without bundle adjustment.

To evaluate the quality of the obtained 3D point set, we applied a plane fitting algorithm for the faces of the cube, planes A and B shown in Fig.5(c) were obtained. The planes were a collection of number of scannings without any alignment algorithm; Note that mis-alignment naturally occurs and RMSD is usually high under such situation. By using the estimated plane parameters, we calculated the angles between the estimated planes. The actual angle is 90˚. The results are shown in Tab. 2.

From the results, we can see the accuracy of the condition with bundle adjustment was the next to the explicit calibration for both the RMSD of planes and the angles. Thus, the bundle adjustment actually refined the accuracy of the 3D reconstructions. Also, we can see that the results without the bundle adjustments (conditions ii and iii) were

Table 1. Parameters estimated by calibration and from data.

	(i)	(ii)	(iii)	(iv)
	Explicit calibration	without bundle adjustment		with bundle adjustment
Focus		Fixed	Self-calibrated	Fixed
f_p	0.0338[m]	0.0338[m]	0.0334[m]	0.0338[m]
$(\alpha_p, \beta_p \gamma_p)$	(-4.9˚, -20.5˚, -22.0˚)	(-4.9˚, -20.5˚, -22.0˚)	(-5.2˚, -20.6˚, -22.2˚)	(-5.3˚, -20.5˚, -22.3˚)
$\mathbf{t_p}/\|\mathbf{t_p}\|$	(0.436,0.534,-0.725)	(0.435,0.533,-0.725)	(0.442,0.538,-0.718)	(0.434,0.536,-0.724)

Table 2. Evaluation of shape estimation.

	(i)	(ii)	(iii)	(iv)
	Explicit calibration	without bundle adjustment		with bundle adjustment
Focus		Fixed	Self-calibrated	Fixed
$RMSD^*$ of plane	0.0004[m]	0.0011[m]	0.0026[m]	0.0008[m]
Angle between planes(actual 90˚)	90.1˚	90.8˚	89.5˚	90.4˚

*The root of the mean squared deviations from the plane.

accurate enough for many purposes.

4.2 Integration of the shape

We scanned an object repeatedly by moving the camera and the projector. 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, it is possible to scan a wide range of objects without any special registration method. Our scanning results are shown in Fig.6, which shows that all the scanning results are aligned correctly without any other alignment algorithm and that a large range of objects can be successfully retrieved in an efficient manner.

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 and Fig.8. Fig.8 shows that our method can scan an entire shape of the object with consistent scaling parameter even if the system is uncalibrated.

5 Conclusion

In this paper, we propose a coded structured light system that does not require extrinsic camera and projector calibration. Since our method does not require any calibration while scanning, users can freely move either the camera or the projector to scan a wide range of the scene without occlusions. To achieve this, the bundle adjustment technique is applied to the multiple scannings, and it has also resulted in improvements in the accuracy of 3D estimation. With our proposed method, several typical drawbacks of the coded structured light system, such as, narrow range of scanning area, occlusions of the object, and difficulty in the maneuvering the system can be solved. To verify the reliability and the effectiveness of the proposed method, we conducted several experiments with the proposed system and actual objects. The results of our experiments confirm the effectiveness of the proposed system.

References

[1] 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.

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

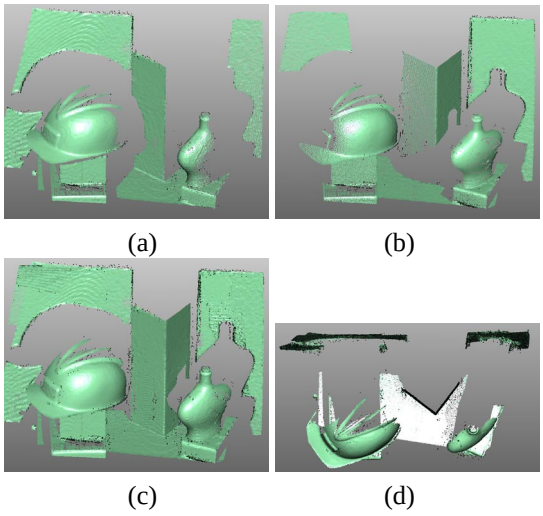
[3] K. L. Boyer and A. C. Kak. Color-encoded structured light for rapid active ranging. *IEEE Trans. on Patt. Anal. Machine Intell.*, 9(1):14–28, 1987.

[4] 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.

[5] C. W. Chu, S. Hwang, and S. K. Jung. Calibration-free approach to 3D reconstruction using light stripe projections on a cube frame. In *Third Int. Conf. on 3D Digital Imaging and Modeling*, pages 13–19, 2001.

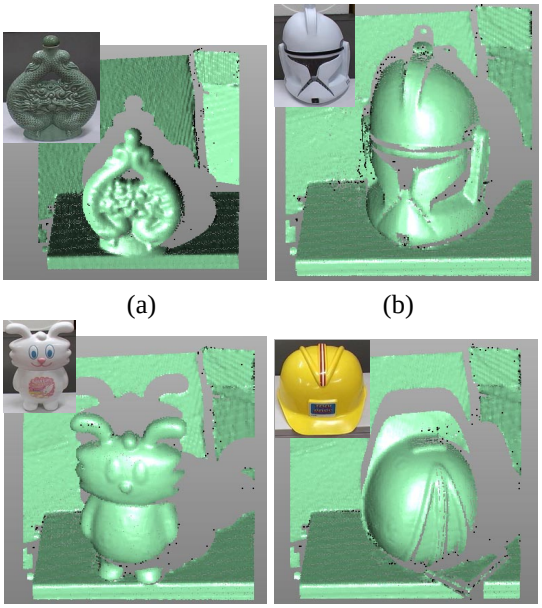
[6] J. Davis and X. Chen. A laser range scanner designed for minimum calibration complexity. In *Third Int. Conf. on 3D Digital Imaging and Modeling*, pages 91–98, 2001.

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



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

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



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

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

[8] 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.

[9] D. Fofi, J. Salvi, and E. M. Mouaddib. Uncalibrated vision based on structured light. In *ICRA*, pages 3548–3553, 2001.

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

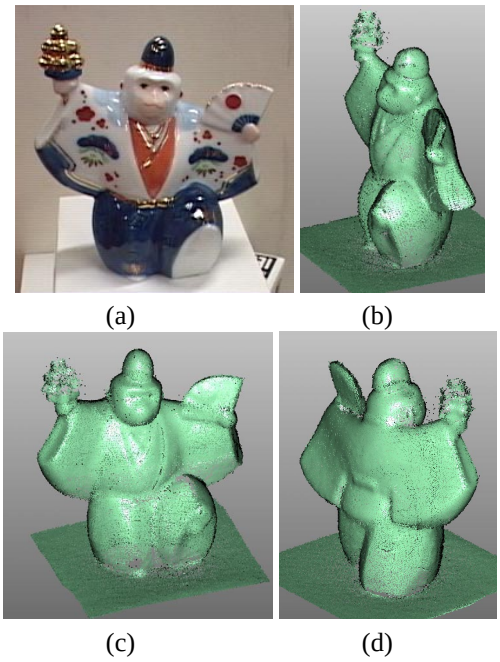
[11] O. Hall-Holt and S. Rusinkiewicz. Stripe boundary codes for real-time structured-light range scanning of moving objects. In *Int. Conf. Computer Vision*, volume 2, pages 359–366, 2001.

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

[13] M. C. Ltd. Vivid 900 non-contact digitizer. In *http://www.minoltausa.com/vivid*.

[14] M. Takatsuka, G. A. West, S. Venkatesh, and T. M. Caelli. Low-cost interactive active monocular range finder. In *CVPR*, volume 1, pages 444–449, 1999.

[15] C. Tomasi and T. Kanade. Shape and motion from image stream under orthography: A factorization method. *Int.J.of Computer Vision*, 9:137–189, 1992.



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

Figure 8. Result of the entire shape scanning:(a) picture of the object: a china figurine (b)(c)(d) 3D shapes from various view directions.