UNCALIBRATED ACTIVE STEREO FOR WIDE AND DENSE 3D DATA ACQUISITION

Hiroshi Kawasaki

Saitama University, Japan,
Dept. of Information and Computer Science
kawasaki@mm.ics.saitama-u.ac.jp

Ryo Furukawa

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

ABSTRACT

In this paper, we propose an uncalibrated, multi-image 3D reconstruction technique, 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 are difficult and sometimes impossible. In the proposed method, precalibration can be successfully omitted by applying the uncalibrated stereo technique, thereby multiple scanning while moving the camera or the projector is possible. As the result, users can freely move either the cameras or projectors to scan a wide range of objects.

1. INTRODUCTION

Efficient 3D data acquisition has been a important research topic for a long period. Recently, many digital archiving projects are proposed and 3D scanning technique become more popular and important. With respect to 3D scanning, 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 occulusions. 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.

In this paper, we propose two scanning methods. 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 of 3D reconstruction. Additionally, occlusions are reduced by multiple scans.

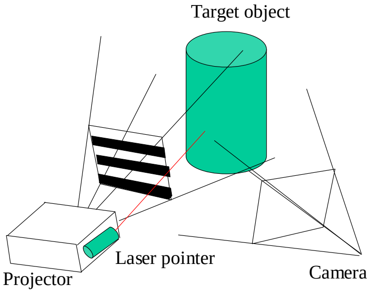


Fig. 1. Components of the 3D measurement system.

The second method is to scan while 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. 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 active stereo method [2]. 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. Figure 1 shows the configuration of the system.

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 in order to recon-

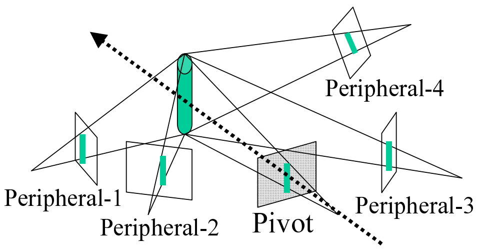
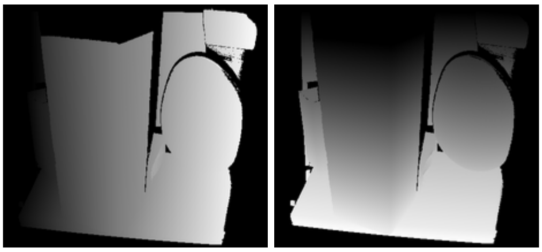


Fig. 2. A pivot scanning.



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

Fig. 3. Coded images by structured light.

struct a single 3D scene. However, it is possible to allow the movement either the camera or the projector.

Multiple scanning by moving one of the devices while the other is fixed is a practical implementation of the system. 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)

Another proposed scanning method involves keeping the camera and projector fixed, alternatively. We call this method as “an alternate scanning”, and this method can be 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.

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, even if occlusions for a single camera or a projector exist.
- Calibration of the extrinsic parameters is not necessary.
- 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, while those of the projector are more difficult to obtain.

2.1. 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, 3]. Since coded structured light method can fix only 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).

2.2. 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 a camera. Non-linear optimization (Newton method) of a squared error function on an image plane is used for the shape acquisition.

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 $\mathbf{R}_{c,j}(\alpha_{c,j}, \beta_{c,j}, \gamma_{c,j})$.

Let the normalized camera 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

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

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

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

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

$E_i(\theta_{\mathbf{j}}, f_p)$ includes systematic errors of the variances 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

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

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

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

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:

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

With our experience, the minimization stably converged from the rough initial values such as $\mathbf{t}_{\mathbf{p}} = (1, 0, 0)^t$ and $(\alpha, \beta, \gamma) = (0, 0, 0)$ in most cases.

By using converged parameter $\theta_{\mathbf{j}}$ and f_p, we can calculate initial 3D data by simply applying triangulation method.

2.3. Bundle adjustment of multiple stereo pairs

3D reconstruction by the uncalibrated stereo method, as described in section 2.2, is sometimes prone to unstable solutions, even if a sufficiently large number of correspondence points are provided. 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.

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 2.2.

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 as described in section 2.2.

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 the of extrinsic parameters of all the peripheral positions ($\theta_j (i = 1, 2, .., N_c)$).

Update of the extrinsic parameters $\theta_{\mathbf{j}} := \alpha_{c,j}, \beta_{c,j}, \gamma_{c,j}, \mathbf{t}_{\mathbf{c,j}}$ 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

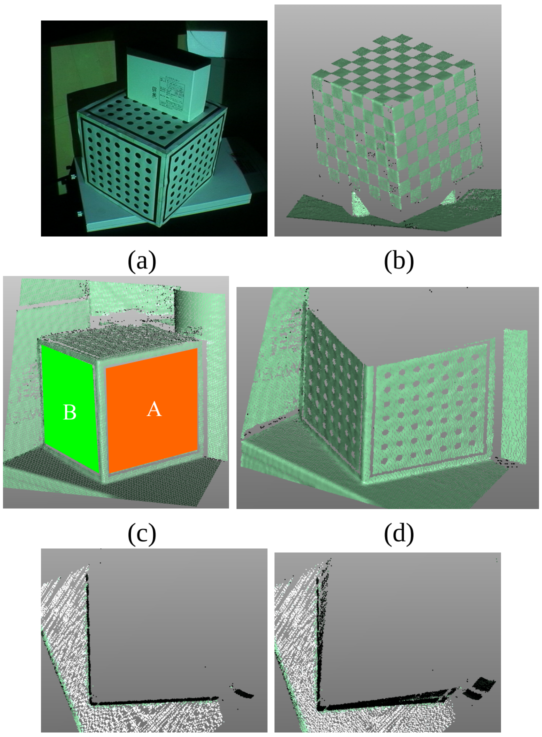
$$Trans(\theta_{\mathbf{j}}, \mathbf{x}) := \mathbf{R}_{c,j}(\alpha_{c,j}, \beta_{c,j}, \gamma_{c,j})\mathbf{x} + \mathbf{t}_{\mathbf{c,j}},$$

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

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

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.



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

Fig. 4. 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.

3. EXPERIMENTS

3.1. Evaluation of Accuracy

To evaluate the accuracy and effectiveness of our proposed method, we scanned a cube (20cm $\times$ 20cm $\times$ 20cm) as shown in Fig.4(a). For a comparison, we scan the cube under three different conditions; (i) estimate the 3D shape of the cube using a single pair of a camera and a projector, (ii) only camera moves and (iii) only projector moves. The results are shown in Tab. 1. We can observe that the accuracy of the 3D scan clearly increases when multiple stereo pairs are used for bundle adjustment.

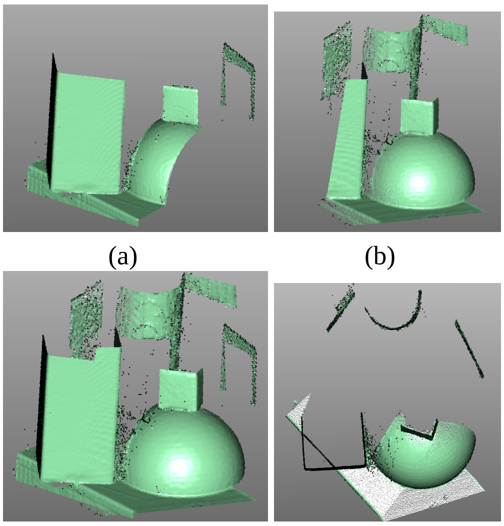
Table 1. Evaluation of shape estimation.

	$RMSD^*$ of A	Angle diff.
(i)without bundle adjust	0.0039 m	92.2 deg.
(ii)camera moves	0.0011 m	90.7 deg.
(iii)projector moves	0.0032 m	91.0 deg.

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

3.2. 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.5., which shows that all the scanning results are aligned correctly and that a large range of objects can be successfully retrieved.



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

Fig. 5. 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.

4. 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. 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.

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

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