Dense 3D reconstruction with an uncalibrated active stereo system

Hiroshi Kawasaki[1], Hiroshi Ohsawa[1], Ryo Furukawa[2], and Yasuaki Nakamura[2]

[1] Faculty of Engineering, Saitama University,
255, Shimo-okubo, Sakura-ku, Saitama, Japan
`{kawasaki,ohsawa}@mm.ics.saitama-u.ac.jp`
[2] Faculty of Information Sciences, Hiroshima City University,
3-4-1, Ozuka-higashi, Asaminami-ku, Hiroshima, Japan
`{ryo-f,nakamura}@cs.hiroshima-cu.ac.jp`

Abstract. In this paper, we describe a novel uncalibrated active stereo system using coded structured light. Structured-light-based active stereo systems generally consist of a camera and projector that require precise precalibration. Therefore, if we can eliminate the precalibration process from the system, the user can merely place the equipment arbitrarily and directly begin scanning the objects. This will greatly improve both the convenience and practicality of the system. In order to achieve this, we propose an original self-calibration method that can be considered as a camera-to-camera self-calibration method in which one of the cameras is replaced with a projector. We also propose a simultaneous 3D reconstruction method that utilizes multiple captured stereo pairs to increase the accuracy of the 3D estimation. Further, we suggest a simple solution to eliminate the ambiguity of scaling by attaching a laser pointer to the projector, which is important for the practical use of the 3D reconstruction.

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.

On the other hand, an active stereo system utilizes a light or laser projector for scanning and can thus retrieve high-precision correspondences with ease; therefore, the accuracy of the 3D points is relatively high. Another benefit of this system is that dense 3D points can be captured easily by controlling the light-projecting devices. Among the different types of active stereo systems, the structured-light-based system is widely used because of its several advantages such as scanning efficiency and simple, low-cost production.

One of the serious drawbacks of active stereo systems is that they essentially require precalibration between the camera and projector whenever the system conditions are changed. Since the precalibration is usually complicated and laborious, it significantly reduces 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 practicality of the system.

On the bases of these facts, we propose an active stereo system that does not require precalibration. Our proposed method is based on a self-calibration stereo method that can be considered as a camera-to-camera self-calibrated method in which one of the cameras is replaced with a projector. The relative position and parameters of the projector model are estimated from the epipolar constraints between the camera and projector. The estimation is performed by applying the Gauss-Newton method to the correspondence points obtained by the coded structured light. We also propose several methods to eliminate the ambiguity of scaling, which is inevitable in uncalibrated 3D reconstruction methods.

The contributions of our work are as follows: first, our system does not assume any limitations on the shapes of objects in the scenes (e.g., the inclusion of planar surfaces or calibration objects) or the camera models (e.g., assumption of an orthogonal camera model); second, we propose the simultaneous 3D reconstruction of multiple scans by which we can obtain constant scaling for multiple scenes and can greatly enhance accuracy; third, we propose a simple solution to eliminate the ambiguity of scaling by attaching a laser pointer to the projector, which is important for the practical use of 3D scanners.

2 Related works

Many active 3D scanning systems have been proposed to date. Among them, a projector-camera based system is commonly used because of several advantages. However, the system require a precise precalibration for installation and this is usually a laborious task.

In order to avoid the calibration problems mentioned above, many uncalibrated active stereo methods have been proposed [1–3]. Takatsuka [1] and Furukawa [3] 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 with regard to passive stereo systems (i.e., camera-to-camera systems)[4], and several researchers have attempted to apply these techniques to active stereo systems by substituting one of the paired stereo cameras with a projector. Fofi et al. [5] proposed an uncalibrated active vision method based on this concept. However, since they assumed an affine camera model, their technique requires capturing a plane in the scene and objects that have a large disparity in depth cannot be recovered. Further, they did not considered the problem of the inevitable ambiguity of scaling in their method.

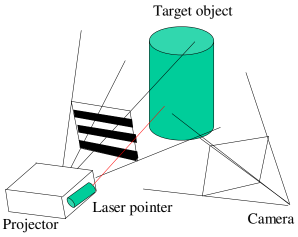


Fig. 1. Components of the 3D measurement system.

3 Uncalibrated active stereo system

3.1 System configuration

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 parameter, which cannot be estimated 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.

The camera and projector are oriented toward the object to measure the shape. A set of dense correspondence points is obtained by the structured light method. The camera-projector parameters are self calibrated using the points. The 3D locations of the correspondence points are then reconstructed by using the stereo method.

Our proposed system has the following features, which are highly desirable in a practical 3D measurement system. Firstly, the projector and the camera can be located arbitrarily. Secondly, there are no limitations imposed on the geometry of the measured scene.

3.2 Obtaining a set of correspondence points by structured light

To obtain correspondence points effectively by using video projector, coded structured light methods have been used and studied extensively[6, 7]. 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 the light patterns 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 light patterns, point-to-point correspondences between the directions from the projector and the pixels in the image are resolved.

3.3 Self-calibration and 3D reconstruction

Our aim is to construct an active stereo system that does not require any calibration process even if the camera or projector is moved arbitrarily; therefore, the camera parameters should be self-calibrated. We assume that the intrinsic parameters of the camera are known, except the focal length of the projector. This is because the intrinsic parameters of the camera can be obtained easily by existing methods, while those of the projector are more difficult to obtain. Further, while scanning zoom and focus are changed more frequently than in the case of cameras. Therefore, we estimate the focal length of the projector and the extrinsic parameters of the relative position between the camera and projector by our self-calibration method.

For the self-calibration, a nonlinear optimization, Gauss-Newton method, is applied. Recently, due to the improved computational capabilities of PCs, self-calibration and 3D reconstructions using only nonlinear optimizations have been studied by some researchers [8], and this approach is employed in our study.

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 $\mathbf{R}_p(\alpha_p, \beta_p, \gamma_p)$. Since the norm of the translation vector $\|\mathbf{t_p}\|$ cannot be resolved by a self-calibration, $\mathbf{t_p}$ is assumed to be an unit vector and is expressed by two parameters of polar coordinates. Thus, $\mathbf{t_p}$ is expressed as $\mathbf{t_p}(\rho_p, \phi_p)$.

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

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

where r is a parameter. The line from the camera is expressed as $s(u_{ci}, v_{ci}, -1)^t$ where s is a parameter.

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

$$\begin{aligned}
\mathbf{p}_{ci} &:= N\ (u_{ci},\ v_{ci},\ -1)^t, \\
\mathbf{q}_{ci}(\theta, f_p) &:= N\ \{\mathbf{R}_p(\alpha_p, \beta_p, \gamma_p)\}(u_{pi}/f_p, v_{pi}/f_p, -1)^t,
\end{aligned} \tag{2}$$

where N is an operator which normalizes a vector (i.e. $N\ \mathbf{x} := \mathbf{x}/\|\mathbf{x}\|$), and $\theta := (\alpha_p, \beta_p, \gamma_p)$ represents the parameters of rotation of the projector. Then, the signed distance between the lines is

$$E_i(\theta, \tau, f_p) := \mathbf{t_p}(\tau) \cdot N\ (\mathbf{p}_{ci} \times \mathbf{q}_{ci}(\theta, f_p)), \tag{3}$$

where "$\cdot$" indicates dot product, and $\tau := (\rho_p, \phi_p)$ represents the parameters of the translation.

$E_i(\theta, \tau, f_p)$ includes systematic errors whose variances change with the parameters (θ, τ, f_p) and the data index i. To compose an error evaluation function unbiased about the parameters (θ, τ, f_p), $E_i(\theta, \tau, f_p)$ 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

$$\begin{aligned}
D_{ci}(\theta, \tau, f_p) &:= \|\mathbf{t_p}(\tau) \times \mathbf{q}_{ci}(\theta, f_p)\| / \|\mathbf{p}_{ci} \times \mathbf{q}_{ci}(\theta, f_p)\|, \\
D_{pi}(\theta, \tau, f_p) &:= \|\mathbf{t_p}(\tau) \times \mathbf{p}_{ci}\| / \|\mathbf{p}_{ci} \times \mathbf{q}_{ci}(\theta, f_p)\|.
\end{aligned} \tag{4}$$

Using the distances, the signed distance normalized by the error level is expressed by $\tilde{E}_i(\theta, \tau, f_p)$ in the forms

$$\begin{aligned}
w_i(\theta, \tau, f_p) &:= \{\epsilon_c D_{ci}(\theta, \tau, f_p) + \epsilon_p D_{pi}(\theta, \tau, f_p)/f_p\}^{-1} \\
\tilde{E}_i(\theta, \tau, f_p) &:= w_i(\theta, \tau, f_p)\ E_i(\theta, \tau, f_p)
\end{aligned} \tag{5}$$

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

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

$$f(\theta, \tau, f_p) := \sum_{i=1}^{K} \{\tilde{E}_i(\theta, \tau, f_p)\}^2, \tag{6}$$

where K is the number of correspondences. The function is minimized using the Gauss-Newton method.

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

3.4 Simultaneous reconstruction of multiple scenes

Since the ambiguity of scaling inevitably exists in uncalibrated stereo methods, several problems occur in practical use. For example, when we scan an object from various view directions to capture its entire shape, different scaling parameters for each scan make it difficult to achieve correct registration and integration. In this paper, two solutions are proposed for this problem: the first solution is the simultaneous 3D reconstruction of multiple scans to estimate consistent camera parameters, and the second is a simple method that uses a laser pointer attached to the projector.

The simultaneous 3D reconstruction is performed as follows. First, multiple scenes are captured by keeping both the camera and projector fixed. The intrinsic and extrinsic parameters of the camera and projector are identical for all the scenes. Therefore, by merely joining the sets of correspondence points for all the scenes and applying the self-calibration algorithm described in section 3.3, consistent camera and projector parameters are obtained. The 3D reconstruction can then be performed for each scene using the estimated camera/projector parameters.

An advantage of this method is that the scalings of all the reconstructed scenes are the same because of the use of consistent camera and projector parameters for multiple scenes. This simplifies the problem of registration and integration while capturing the entire shape of an object.

3.5 Estimation of scaling parameter

A reconstructed 3D shape produced by our method is scaled by an unknown multiplier from the real shape. The simultaneous reconstruction described in the previous subsection is useful for obtaining multiple reconstructions with constant scaling, but we cannot estimate “real” factor of scaling for the scene with the method. For some applications, estimation of the real scaling factor is needed.

To achieve this, 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.

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. From the 3D point lit by the pointer in the image, the scaling parameter is calculated by triangulation method, prceesed for the line of the laser and the line of sight determined from the image.

The line formed by the laser pointer should be calibrated in the projector coordinate system. To do this, multiple points on the laser are obtained by measuring an object with a known shape lit by 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.

Table 1. Parameters estimated by calibration and from data.

	By calibration	From data
f_p	0.0338[m]	0.0329[m]
$(\alpha_p, \beta_p \gamma_p)$	(-9.3°, -31.6°, -13.0°)	(-8.2°, -30.9°, -12.7°)
$\mathbf{t_p}/\|\mathbf{t_p}\|$	(-0.610,0.446,-0.655)	(-0.581,0.441,-0.684)

4 Experiments

4.1 Evaluation of accuracy

To evaluate our proposed method, we scanned a scene of a cube (20cm × 20cm × 20cm) (Fig.2(a)and(b)), calculating the extrinsic parameters and the focal length of the projector. To evaluate the effectiveness of our simultaneous 3D reconstruction method, we performed simultaneous 3D reconstructions and single-scene 3D reconstructions and compared the results. For a comparison, 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 as the ground truth.

For the test data, we scanned a cube-shaped object 5 times changing the position of the object. During the scanning, we kept the camera and projector fixed. Then, 3D reconstruction is done for each scanned data, which is referred as a single-scene reconstruction. Simultaneous reconstruction is performed as follows. First, self-calibration is done using all the scanned data as a single input. Then, using the estimated parameters, each 3D shape is reconstructed by using the stereo method. Each of the self-calibrations is performed under 2 conditions: one is a self-calibration with fixed focal length of the projector and the other include the estimation of the focal length. For the fixed focal length condition, the focal length calculated by the explicit calibration were used. 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 estimated parameters are shown in Tab. 1.

We also evaluated the quality of the obtained 3D point set shown in Fig.2(c),(d). We applied a plane fitting algorithm for the faces in the scene, obtaining 3 planes (A,B,C) shown in Fig.2(a). The fitting of planes to the point sets were performed by principal component analysis. By using the estimated plane parameters, we calculated angles between the estimated planes. One of the measured angles of the 5 data sets are shown in Tab. 2. The averaged signed errors from the actual angles (90 °)and the roots of mean squared errors (RMS errors) of them are shown in Tab. 3.

Also, the residual RMS errors of the plane fitting algorithms were calculated, which are shown in Tab. 3.

From the results, we can see that the shapes of the cubes are correctly reconstructed for all of the conditions. From the tendency of the results, we can see that the results of the simultaneous reconstructions were better than the single reconstructions for both of the accuracy of angles and residuals of plane-fittings.

Table 2. Results of angles between estimated planes.

	Single Input		Simultaneous	
Focus of projector	Fixed	Selfcalib	Fixed	Selfcalib
between A-C	89.95°	90.23°	90.17°	90.06°
between B-C	89.87°	90.99°	89.88°	90.56°
between A-B	90.39°	92.09°	90.07°	91.33°

Table 3. Summary of 3D reconstruction.

	Single Input		Simultaneous	
Focus of projector	Fixed	Selfcalib	Fixed	Selfcalib
Ave. errors	0.02°	2.65°	-0.01°	-0.52°
RMSE (degree)	0.10 °	3.78°	0.07°	1.79°
RMSE of plane(mm)	0.72	0.77	0.62	0.65

Thus, these experimental results show the effectiveness of the simultaneous reconstructions.

Also, the results of the fixed focal lengths were better than those of self-calibrated focal lengths. One possible reason for the estimation errors would be considered inappropriate error model implementated in our system. If the error model was incorrect, estimation of the focal length would be biased. Further research is necessary for this area.

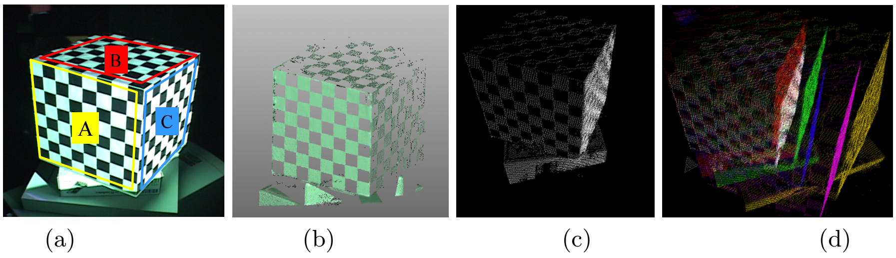
4.2 Scaling parameter evaluation

We evaluated our method for estimating the scaling parameter. First, the line formed by the laser light was calibrated by measuring 2 points lit by the laser. Then, we moved the camera and projector, scanned a cube 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 data. To evaluating the accuracy of the scaling estimation, we measured the length of 3 edges of the cube, a, b, and c. The results were 199.9mm, 199.2mm and 199.8mm, respectively. All estimated length of edges were close to the actual length 20cm, thus confirming that our estimation method is effective and practical.

4.3 Entire shape acquisition by simultaneous method

To demonstrate the effectiveness of our simultaneous 3D reconstruction method, we performed entire shape acquisitions for two objects, a china figurine and a helmet.

In order to construct entire shapes, first, we scanned each of the objects 8 times, rotating it by 45°. Then, all 3D shapes were recovered simultaneously



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

Fig. 2. Scanning of a cube with known size: (a) 3 faces used for accuracy estimations, (b) reconstructed 3D points, (d) 3D point set acquired by single-scene reconstruction, and (d) by simultaneous reconstruction.

by our simultaneous 3D reconstruction method. Finally, we apply an alignment algorithm for shape registration.

Results are shown in Fig.3. We can observe that multiple scanned shapes are integrated into a single shape without any gaps, although the system is uncalibrated. This is because all the scaling parameters for 8 scan data sets are the same, which is achieved by the simultaneous reconstruction method.

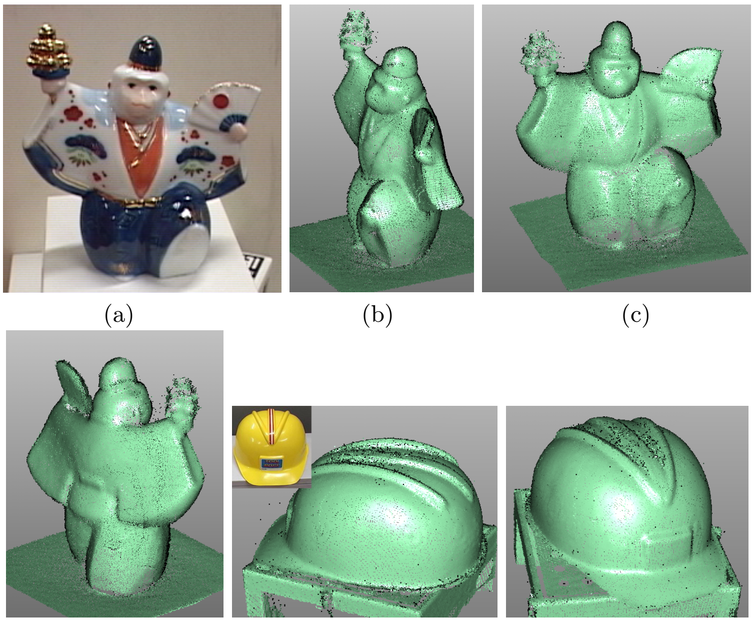
5 Conclusion

In this paper, we propose a novel uncalibrated active stereo system that 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 stereo system in which one of the cameras is replaced with a projector. We also propose the simultaneous 3D reconstruction method to increase accuracy and a simple method to eliminate the scaling ambiguity by 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. To verify the reliability and the effectiveness of our proposed method, we conducted several experiments with the proposed system and actual objects. The results of our experiments confirm the effectiveness of our proposed system.

References

1. Takatsuka, M., West, G.A., Venkatesh, S., Caelli, T.M.: Low-cost interactive active monocular range finder. In: CVPR. Volume 1. (1999) 444–449
2. Davis, J., Chen, X.: A laser range scanner designed for minimum calibration complexity. In: Third Int. Conf. on 3DIM. (2001) 91–98
3. Furukawa, R., Kawasaki, H.: Interactive shape acquisition using marker attached laser projector. In: Int. Conf. on 3DIM2003. (2003) 491–498



(a) (b) (c)

(d) (e) (f)

Fig. 3. Examples of the scanned objects:(a)(b)(c)(d) a china figurine, (e)(f) a helmet.

4. Faugeras., O.: Three-Dimensional Computer Vision - A Geometric Viewpoint. Artificial intelligence. M.I.T. Press Cambridge, MA (1993)
5. Fofi, D., Salvi, J., Mouaddib, E.M.: Uncalibrated vision based on structured light. In: ICRA. (2001) 3548–3553
6. Caspi, D., Kiryati, N., Shamir, J.: Range imaging with adaptive color structured light. IEEE Trans. on Patt. Anal. Machine Intell. **20** (1998) 470–480
7. Inokuchi, S., Sato, K., Matsuda, F.: Range imaging system for 3-D object recognition. In: ICPR. (1984) 806–808
8. Amano, A., Migita, T., Asada, N.: Stable recovery of shape and motion from partially tracked feature points with fast nonlinear optimization. In: 15th Vision Interface. (2002) 244–251