Toward dynamic recalibration and three-dimensional reconstruction in a structured light system

Y. F. Li and B. Zhang

Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, Kowloon, Hong Kong

Received April 28, 2006; revised August 11, 2006; accepted September 11, 2006; posted October 5, 2006 (Doc. ID 70447); published February 14, 2007

We present a method for dynamic recalibration and 3D reconstruction via a structured light system. Assuming that the light planes cast from the digital light projector have been calibrated off-line, we show that the focal length, aspect ratio, and all motion parameters of the camera can be determined on-line. Then the 3D reconstruction can be carried out by either a traditional triangulation method or a more efficient transformation-based method. In the latter method, a single image is sufficient for the whole process of calibration and reconstruction. Thus a hand-held camera can be used. Computer simulation and real data experiments were carried out to validate the method. © 2007 Optical Society of America

OCIS codes: 000.3110, 110.0110, 120.0120, 120.4640, 150.0150, 330.0330.

1. INTRODUCTION

Camera calibration and 3D reconstruction have been widely investigated, and many methods have been proposed in recent years. These methods can be classified into passive and active methods. In general, the passive methods make use of special camera motion[1] or a structured scene[2,3] for camera calibration. Since multiple images are generally involved, the passive methods suffer from the ambiguity of correspondences between feature points in the images. This is a difficult problem to solve as there is a trade-off between the disparity and overlapped area size, especially when dealing with free-form textureless surfaces.

To avoid such ambiguity, an active vision system consisting of a projector and a camera can be adopted. To precisely reconstruct a 3D model with such an active vision system, it needs to be carefully calibrated by some special calibration device.[4,5] The traditional calibration procedures are usually carried out off-line, and they have to be repeated each time the setting is changed. Frequent recalibration in using such a system is a laborious and tedious task. As a result, the applications of such a system are limited, since the configuration and parameters must be kept unchanged during the entire measurement process. When working in an unknown environment, e.g., in underwater exploration by a vision-guided robot, changes of the position and configuration of the vision sensor become necessary. It is thus desirable for a vision system to have the ability of self-recalibration[6,7] or uncalibrated reconstruction[8] without requiring a special calibration device.

By defining an image-to-world transformation between a 3D point on the stripe light plane and its corresponding pixel in the image, researchers have proposed some methods for calibrating their structured light systems. Chen and Kak[9] proposed a line-to-point method to calibrate this transformation for the vision system of a laser range finder given six or more known world lines. Reid[10] used the plane-to-point method to determine this transformation where the laser range finder was mounted on a free-ranging automated guided vehicle whose position relative to the world coordinate frame was known. Based on cross-ratio invariance and point-to-point calibration, Huynh *et al.*[11] also proposed a calibration method using four known noncoplanar sets of three collinear world points in the scene. These methods can be considered as static calibration since any adjustment of their systems, e.g., autofocus or movement of the camera, will not be allowed when performing the vision tasks. This causes inconvenience in some practical applications. To allow on-line adjustment of the parameters of the vision system, we proposed in our previous work a way to dynamically compute the different components of the image-to-world transformation by assuming two known planes in the scene.[12] Once all the components had been recovered, the transformation matrix can be obtained by simple matrix multiplication. In this paper, we follow a similar framework for calibrating the structured light system by assuming only one known plane in the scene. Here, the generic formula for the plane-to-image homography is derived first in the system. Then the variable parameters, i.e., the focal length, aspect ratio, and all motion parameters of the camera, are computed by analyzing the scene plane-to-image homography. Finally, the stripe light plane-to-image homography is computed from the generic formula, which, in turn, enables the computation of the on-line image-to-world transformation. Both the previous and this method use a single image for the 3D reconstruction. However, they differ in two important aspects: (1) The method here provides calibration for the variable parameters of the camera while the previous one does not. (2) The relaxed requirement of only one known plane with our method is

much easier to meet, promising wider practical applications. With a closed-form solution, our method avoids the time-consuming nonlinear optimization. Hence, a handheld camera can be used in a dynamic environment.

The remainder of the paper is organized as follows: Section 2 introduces the system models. Section 3 presents the reconstruction principle. Section 4 gives the procedure for dynamic calibration and 3D reconstruction. Section 5 reports the results of computer simulations and real data experiments. Finally, we conclude this paper in Section 6.

Throughout this paper, scalars are denoted by normal face symbols while vectors and matrices are represented by lowercase and uppercase boldface letters. A roman superscript T represents the transpose of a vector or matrix. The subscripts c, p, and s, respectively, denote the camera-related, projector-related, and scene-plane-related elements. The notations $\mathbf{M}$ and $\mathbf{m}$ denote the 3D and 2D points, respectively.

2. SYSTEM MODEL

A. Structure of the Vision System

Figure 1 illustrates the layout of the proposed vision system, which mainly consists of a CCD camera and a digital light projector (DLP). Suppose there is a blank plane in the scene, e.g., the ground plane or desktop. By blank plane, we mean that no special structure or feature points regarding the plane are required. The projector is used to illuminate the scene with a light pattern that will produce two sets of light planes, in the horizontal and vertical directions, respectively. Each light plane produces a deformed curve on the object's surface and one or more line segments on the plane (see the bold line segments in Fig. 1). These line segments and curves, together with other light planes, will intersect at many feature points. Some of these feature points, which are used for measuring the object, will be on the surface of the object, while some will lie on the plane that will be used for calibrating the system. The camera captures the illuminated scene in its image plane.

B. Camera Model

We use the classical pinhole model for our camera. Let $\mathbf{K}_c$ be its calibration matrix. The principal point of the camera is assumed to be unchanged. This assumption can be satisfied to a sufficient extent in practice and has already been used by many researchers.[13,14] For example, in recent research,[14] a method was proposed for calibrating the focal length assuming all other parameters are known and fixed. Therefore, we can ignore the effect of the principal point by simple point translation. Then the calibration matrix becomes

$$\mathbf{K}_c = \begin{bmatrix} \tau f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{1}$$

Assuming that the rotation matrix and translation vector of the camera are $\mathbf{R}_c$ and $\mathbf{t}_c$, then a 3D arbitrary point $\tilde{\mathbf{M}} = (M_1 \quad M_2 \quad M_3 \quad 1)^T$ and its corresponding image point $\tilde{\mathbf{m}} = (m_1 \quad m_2 \quad 1)^T$ in the camera are related by the following projection equation:

$$\lambda \tilde{\mathbf{m}} = \mathbf{K}_c [\mathbf{R}_c \quad \mathbf{t}_c] \tilde{\mathbf{M}}, \tag{2}$$

where λ is a nonzero scale factor.

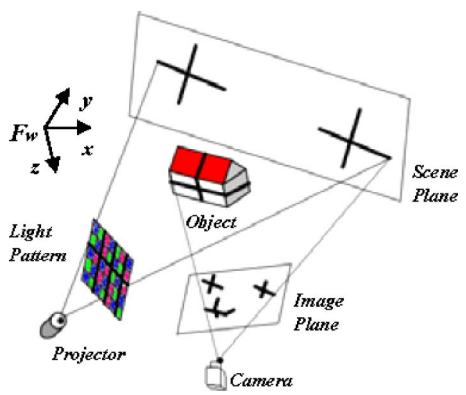


Fig. 1. (Color online) Layout of the proposed vision system.

3. RECONSTRUCTION PRINCIPLE

A. Image-to-World Transformation

Let the world coordinate frame be F_w with (x, y, z) representing its three coordinate axes in the scene. We treat the light planes emitted from the projector and the scene plane as ordinary planes (without any special features required) and assume that their equations have been calculated by a previous method.[12] Let Π be any one plane involved. Assuming $\overline{\mathbf{M}} = (\overline{M_1} \quad \overline{M_2} \quad \overline{M_3})^T$ is a known point on Π, we define a local frame $F_{3\Pi}$ with its origin at $\overline{\mathbf{M}}$, whose XY axes lie on the plane and whose Z axis aligns with its normal vector. In what follows, we discuss the principle of the 3D reconstruction.

Let $\mathbf{n} = (n_1 \quad n_2 \quad n_3)^T$ be the vector of the plane Π in the world frame F_w. Since the direction of the z axis of the frame F_w is $\mathbf{z} = (0 \quad 0 \quad 1)^T$, the angle between $\mathbf{n}$ and the z axis is $\theta = \cos^{-1}(\mathbf{n}^T z) = \cos^{-1}(n_3)$.

The unit vector that is perpendicular to both the z axis of the frame F_w and the normal vector $\mathbf{n}$ is $\mathbf{r} = (\mathbf{n} \times \mathbf{z})/\|\mathbf{n} \times \mathbf{z}\|$.

Then the rotation matrix from the world frame F_w to $F_{3\Pi}$ can be given by the Rodrigues formula

$$\mathbf{R} = \cos\theta \cdot \mathbf{I} + (1 - \cos\theta)\mathbf{r}\mathbf{r}^{\mathrm{T}} + \sin\theta \cdot [\mathbf{r}]_\times, \tag{3}$$

where $\mathbf{I}$ represents a 3×3 identity matrix, and $[\mathbf{r}]_\times$ is the skew matrix of vector $\mathbf{r}$ given by

$$[\mathbf{r}]_\times = \begin{bmatrix} 0 & -r_3 & r_2 \\ r_3 & 0 & -r_1 \\ -r_2 & r_1 & 0 \end{bmatrix}.$$

Let $\mathbf{M}_i = (M_{i1} \quad M_{i2} \quad M_{i3})^T$ be an arbitrary point on Π and $\tilde{\mathbf{M}}_i = (M_{i1} \quad M_{i2} \quad M_{i3} \quad 1)^T$ be its homogeneous form. Expressed in $F_{3\Pi}$, this point will be

$$\mathbf{M}_{pi} = \mathbf{R}\mathbf{M}_i - \mathbf{R}\overline{\mathbf{M}}. \tag{4}$$

From Eq. (4), we have

$$\tilde{\mathbf{M}}_{pi} = \mathbf{T}\tilde{\mathbf{M}}_i, \tag{5}$$

where

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & -\mathbf{R}\bar{\mathbf{M}} \\ \mathbf{0}^T & 1 \end{bmatrix}$$

is a 4×4 matrix. It transforms a point from the world frame F_w to the specific plane frame $F_{3\Pi}$ (Fig. 2). It should be noted that $\tilde{\mathbf{M}}_{pi}$ has the following coordinate form:

$$\tilde{\mathbf{M}}_{pi} = (m_{p1} \quad m_{p2} \quad 0 \quad 1)^T. \tag{6}$$

We can rewrite Eq. (5) as follows:

$$\tilde{\mathbf{m}}_{pi} = \mathbf{S}\mathbf{T}\tilde{\mathbf{M}}_i, \tag{7}$$

where $\tilde{\mathbf{m}}_{pi} = (m_{p1} \quad m_{p2} \quad 1)^T$ and

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The above equation describes the relationship for a point between the world frame F_w and the planar frame $F_{3\Pi}$. In this paper, it is used when computing the homography between the scene and the image planes.

According to projective geometry, there exists a unique homography transformation between the plane and its image. Let it be $\mathbf{H}$ and let $\tilde{\mathbf{m}}_{ci}$ be the correspondent point of $\tilde{\mathbf{m}}_{pi}$ in the camera image. Then we have

$$\tilde{\mathbf{m}}_{ci} = \lambda_i \mathbf{H}\tilde{\mathbf{m}}_{pi}. \tag{8}$$

From Eqs. (7) and (8), we have

$$\tilde{\mathbf{M}}_i = \frac{1}{\lambda_i}\mathbf{T}_{wi}\tilde{\mathbf{m}}_{ci}, \tag{9}$$

where

$$\mathbf{T}_{wi} = \mathbf{T}^{-1}\mathbf{S}^{\mathrm{T}}\mathbf{H}^{-1}. \tag{10}$$

Here, $\mathbf{T}_{wi}$ is called the image-to-world transformation. It transforms a pixel in the image to its corresponding 3D point on the space plane relative to the world frame F_w. It is a 4×3 matrix and can be determined up to a nonzero scale. Therefore, it has 11 degrees of freedom.

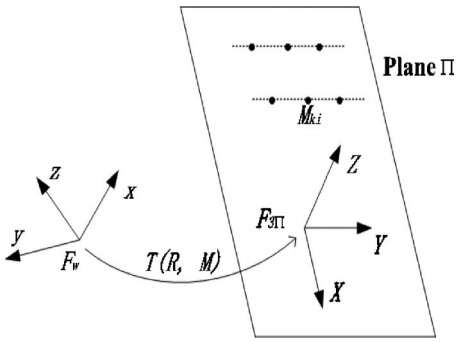


Fig. 2. Relationship between F_w and $F_{3\Pi}$ where $\mathbf{T}$ is a function of $\mathbf{R}$ and $\bar{\mathbf{M}}$.

B. Analysis of the System

From Eq. (9), we can see that, once the transformation $\mathbf{T}_{wi}$ is recovered, the 3D reconstruction is straightforward, i.e., simply by matrix multiplication. From Eq. (10), we can see that $\mathbf{T}_{wi}$ has three components: the transformation $\mathbf{T}$, constant matrix $\mathbf{S}$, and homography $\mathbf{H}$. From Eqs. (3) and (5), we can see that the transformation $\mathbf{T}$ depends only on the equation of the plane. In our system configuration, we assume that the scene plane and the light planes of the projector are calibrated and kept fixed. Therefore, the transformation $\mathbf{T}$ will remain constant once it has been recovered in the initial calibration. We assume that some intrinsic or extrinsic parameters of the camera should be adjusted to achieve a better performance in different cases. As a result, the homography $\mathbf{H}$ will vary in practical applications and needs to be calibrated on-line. In what follows, we will discuss the way for dynamically calibrating the homography and image-to-world transformation.

4. DYNAMIC RECALIBRATION

A. Plane-Image Homography

From Eq. (7), for an arbitrary point $\tilde{\mathbf{M}}_i$ on the plane Π, we have

$$\tilde{\mathbf{M}}_i = \mathbf{T}^{-1}\mathbf{S}^{\mathrm{T}}\tilde{\mathbf{m}}_{pi}. \tag{11}$$

Substituting Eq. (11) into Eq. (2), we get

$$\lambda\tilde{\mathbf{m}}_{ci} = \mathbf{K}_c[\mathbf{R}_c \quad \mathbf{t}_c]\mathbf{T}^{-1}\mathbf{S}^T\tilde{\mathbf{m}}_{pi}. \tag{12}$$

Considering Eqs. (8) and (13), we have

$$\sigma\mathbf{H} = \mathbf{K}_c[\mathbf{R}_c \quad \mathbf{t}_c]\mathbf{T}^{-1}\mathbf{S}^T, \tag{13}$$

where σ is a scale factor.

The above formula is an explicit expression of the plane-to-image homography. It is a 3×3 matrix and can be determined up to a nonzero scale factor. Since the plane is arbitrary, the formula is applicable for both the stripe light plane-to-image and scene plane-to-image homography. Therefore, we call it a generic homography.

Now, let the homography between the scene plane and the image be $\mathbf{H}_s$. As illustrated in Fig. 1, the light planes produce many 3D feature points on the scene plane. Since the equations of the light planes and the scene plane are determined at the static calibration stage, the coordinates of these points can be computed from these equations. For example, when a feature point $\tilde{\mathbf{M}}_i$ is obtained, we can transform it by Eq. (7) to get $\tilde{\mathbf{m}}_{pi}$. The correspondent pixel $\tilde{\mathbf{m}}_{ci}$ of $\tilde{\mathbf{M}}_i$ in the image can be extracted from the image. Then the pair of points, i.e., $\tilde{\mathbf{m}}_{pi}$ and $\tilde{\mathbf{m}}_{ci}$, gives two constraints on the homography $\mathbf{H}_s$. If four or more pairs of points are considered, this homography can be determined in a least-squares sense. In what follows, we present our method for dynamic calibration assuming that the homography $\mathbf{H}_s$ has been obtained.

B. Dynamic Recalibration of Camera Parameters

According to the formula of the transformation for the scene plane in Eq. (5), we have

$$\mathbf{T}_s^{-1} = \begin{bmatrix} \mathbf{R}_s^{\mathrm{T}} & \bar{\mathbf{M}}_s \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix},$$

where $\mathbf{R}_s$ denotes the rotation matrix from the world frame to that of the scene plane, and $\bar{\mathbf{M}}_s$ is any known point on the plane.

Let

$$\mathbf{R}_c = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{bmatrix} \quad \mathbf{R}_s = \begin{bmatrix} \mathbf{r}_{1s} \\ \mathbf{r}_{2s} \\ \mathbf{r}_{3s} \end{bmatrix},$$

where $\mathbf{r}_i$ and $\mathbf{r}_{is}$ are the ith row of the matrix $\mathbf{R}_c$ and $\mathbf{R}_s$, respectively. Let $\mathbf{t}_c = [t_1 \quad t_2 \quad t_3]^T$.

From Eq. (13), we have

$$\sigma\mathbf{H}_s = \begin{bmatrix} \tau f\mathbf{r}_1\mathbf{r}_{1s}^T & \tau f\mathbf{r}_1\mathbf{r}_{2s}^T & \tau f(\mathbf{r}_1\bar{\mathbf{M}}_s + t_1) \\ f\mathbf{r}_2\mathbf{r}_{1s}^T & f\mathbf{r}_2\mathbf{r}_{2s}^T & f(\mathbf{r}_2\bar{\mathbf{M}}_s + t_2) \\ \mathbf{r}_3\mathbf{r}_{1s}^T & \mathbf{r}_3\mathbf{r}_{2s}^T & \mathbf{r}_3\bar{\mathbf{M}}_s + t_3 \end{bmatrix}. \quad (14)$$

From the first two columns of Eq. (14), we obtain

$$\mathbf{R}_c\mathbf{r}_{1s}^T = \begin{bmatrix} \frac{\sigma}{\tau f}\mathbf{H}_{11} \\ \frac{\sigma}{f}\mathbf{H}_{21} \\ \sigma\mathbf{H}_{31} \end{bmatrix}, \quad (15)$$

$$\mathbf{R}_c\mathbf{r}_{2s}^T = \begin{bmatrix} \frac{\sigma}{\tau f}\mathbf{H}_{12} \\ \frac{\sigma}{f}\mathbf{H}_{22} \\ \sigma\mathbf{H}_{32} \end{bmatrix}, \quad (16)$$

where $\mathbf{H}_{ij}$ means the element on the ith row and jth column of the matrix $\mathbf{H}_s$.

From Eqs. (15) and (16), we will get the following three equations:

$$\left[\frac{\sigma}{\tau f}\mathbf{H}_{11}\right]^2 + \left[\frac{\sigma}{f}\mathbf{H}_{21}\right]^2 + (\sigma\mathbf{H}_{31})^2 = 1, \quad (17)$$

$$\left[\frac{\sigma}{\tau f}\mathbf{H}_{12}\right]^2 + \left[\frac{\sigma}{f}\mathbf{H}_{22}\right]^2 + (\sigma\mathbf{H}_{32})^2 = 1, \quad (18)$$

$$\left[\frac{\sigma}{\tau f}\right]^2\mathbf{H}_{11}\mathbf{H}_{12} + \left[\frac{\sigma}{f}\right]^2\mathbf{H}_{21}\mathbf{H}_{22} + \sigma^2\mathbf{H}_{31}\mathbf{H}_{32} = 0. \quad (19)$$

In Eqs. (17)–(19), there are three unknowns, i.e., f, τ, and σ. From the three equations, we can obtain the solutions,

$$f = \sqrt{\frac{a_2c_1 - a_1c_2}{b_1c_2 - b_2c_1}}, \quad (20)$$

$$\tau = \sqrt{\frac{b_1c_2 - b_2c_1}{a_1b_2 - a_2b_1}}, \quad (21)$$

$$\sigma = \pm\sqrt{\frac{\tau^2f^2}{\tau^2f^2\mathbf{H}_{31}^2 + \tau^2\mathbf{H}_{21}^2 + \mathbf{H}_{11}^2}}, \quad (22)$$

where

$$a_1 = \mathbf{H}_{21}^2 - \mathbf{H}_{22}^2, \quad b_1 = \mathbf{H}_{31}^2 - \mathbf{H}_{32}^2, \quad c_1 = \mathbf{H}_{11}^2 - \mathbf{H}_{12}^2,$$

$$a_2 = \mathbf{H}_{21}\mathbf{H}_{22}, \quad b_2 = \mathbf{H}_{31}\mathbf{H}_{32}, \quad c_2 = \mathbf{H}_{11}\mathbf{H}_{12}.$$

Since both f and τ should be positive, unique solutions are obtained. As for σ, its sign should make both sides of Eq. (8) consistent, which means if $\Sigma_i\mathbf{m}_{c,i}^{\mathrm{T}}\mathbf{H}\mathbf{m}_{p,i} > 0$, the sign is positive, otherwise, it is negative.

Then, we can scale the homography by σ. Let $\hat{\mathbf{A}} = \sigma\mathbf{H}_s$. From the first row of Eq. (14), we have

$$\mathbf{r}_1\mathbf{r}_{1s}^T = \frac{1}{\tau f}A_{11},$$

$$\mathbf{r}_1\mathbf{r}_{2s}^T = \frac{1}{f}A_{12},$$

$$\|\mathbf{r}_1\|_2 = 1. \quad (23)$$

We can obtain the solution for $\mathbf{r}_1$ from Eq. (23). Similarly, we have the solution for $\mathbf{r}_2$ by

$$\mathbf{r}_2\mathbf{r}_{1s}^T = \frac{1}{f}A_{21},$$

$$\mathbf{r}_2\mathbf{r}_{2s}^T = \frac{1}{f}A_{22},$$

$$\|\mathbf{r}_2\|_2 = 1. \quad (24)$$

The third row of the rotation matrix is $\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2$.

For the translation vector of the camera, we can obtain it by comparing the third column of both sides in Eq. (14), and give as

$$t_1 = \frac{1}{\tau f}\tilde{\mathbf{A}}_{13} - \mathbf{r}_1\bar{\mathbf{M}}_s,$$

$$t_2 = \frac{1}{f}\tilde{\mathbf{A}}_{23} - \mathbf{r}_2\bar{\mathbf{M}}_s,$$

$$t_3 = \mathbf{A}_{33} - \mathbf{r}_3\bar{\mathbf{M}}_s. \quad (25)$$

To now, we have obtained all the variable parameters of the camera. Then, the stripe light plane-to-image homography can be computed according to its explicit formula, so is the on-line image-to-world transformation matrix.

C. Implementation Procedure

The implementation of the structured light system mainly consists of two steps: static initial calibration and Euclidean reconstruction with dynamic recalibration of the system. Assuming that the static calibration has been

accomplished, the procedure of dynamic recalibration and Euclidean 3D reconstruction consists of the following seven steps:

1. The object to be reconstructed is placed in front of the scene plane and illuminated by the projector. According to the scene geometry, the camera may undergo a rigid motion, zooming and focusing to enable the sensor to fully capture the illuminated scene. Then the scene image is acquired.
2. Select four or more illuminated points on the scene plane and transform them by Eq. (7). Then extract their correspondences in the image and computing the scene plane-to-image homography by Eq. (8).
3. Calculate the focal length and aspect ratio of the camera and the scale factor of the homography according to Eqs. (20)–(22), then scale the homography.
4. Compute the first and second rows of the rotation matrix, i.e., $\mathbf{r}_1$ and $\mathbf{r}_2$, by Eqs. (23) and (24) and the third row $\mathbf{r}_3$. Then we have the rotation matrix $\mathbf{R}_c$.
5. Solve Eq. (25) for the translation vector $\mathbf{t}_c$ of the camera.
6. Compute the stripe light plane-to-image homography by formula (13).
7. Compute the image-to-world transformation by formula (10) and implement the Euclidean reconstruction by Eq. (9).

D. Discussion

Some researchers have discussed the camera calibration method from plane-to-image homography using a structured planar pattern.[2,3] They mainly took two steps to solve this problem, i.e., first determining the image of absolute conic, followed by computing the variable parameters of the camera. In this paper, with an ordinary plane, the variable parameters are found by analyzing the scene plane-to-image homography directly in the structured light system. Then the on-line image-to-world transformation is determined, and the 3D reconstruction can be implemented immediately.

5. EXPERIMENTS

In the first group of experiments, we show that acceptable results can be obtained for dynamic calibration with our method. Then the remaining experiments are focused on the validity of the calibration and 3D reconstruction. These experiments were carried out on both numerical simulation and real image data.

A. Experiments with Simulation Data

We first carried out some numerical simulations to verify the validity of our method. The simulated system has the following parameters. The focal length, aspect ratio, and principal point of the camera were 950, 1.1, and (300, 200), respectively. The resolution of the image was 740 $\times$480. The orientation and position of the camera were set to be $(\pi/4, \pi/5, \pi/6)$ and (100, 200, 300). Two sets of light planes from the projector, i.e., eight in the horizontal direction and eight in the vertical direction, were assumed to be known. The object was represented by a set of 3D points intersected by the light planes and some other random planes. The equation of the known scene plane was denoted as $Z=X+2Y+250$. In the experiments, we defined the residual errors as the discrepancies between the computed results and their theoretical values, where the absolute errors were taken for scalars, and the norm of the errors were taken for vectors or matrices.

In the simulation study, we find that the more accurate the scene plane-to-image homography is, the better calibration results we will have. It is known that a more accurate estimation of the homography can be obtained if more point pairs are used. In general, the larger the area of the plane in the image, the more feature points can be extracted. In practice, however, the available points are limited in number. Therefore, we conducted some tests on the effects of the number of points on the performance of the algorithm. Here, the number of points ranged from 10 to 60. For each case, 100 trials were run. A constant level of Gaussian noise, $N(0, 0.5)$, was added to the image pixels. The results on the calibration and reconstruction are shown in Fig. 3, where data1, data2, data3, and data4 present the average residual errors for the focal length, rotation angles, translation vector, and 3D reconstruction, respectively. This figure shows that when the number of points increases from 10 to 60, the residual errors in the calibrated parameters and 3D reconstruction results decrease sharply. Acceptable results can be obtained when the number is 40 or more. For example, less than 1 mm errors are obtained in the 3D reconstruction in this range. Therefore, we will use 40 points for the subsequent experiments.

Next, we tested the robustness of our method against different levels of noise on the image pixel. Here, Gaussian noise whose level ranges from 0 to 1 pixel was added to the image pixels. For each noise level, we performed 100 trials and computed the average residual error. The residual errors in the calibration and reconstruction are shown in Fig. 4(a). It can be seen that when there is no noise in the data, the errors are very small. The residual errors in the calibration and reconstruction increase with the noise level. To alleviate the effect of noise, two methods should be adopted: (1) extracting feature points with

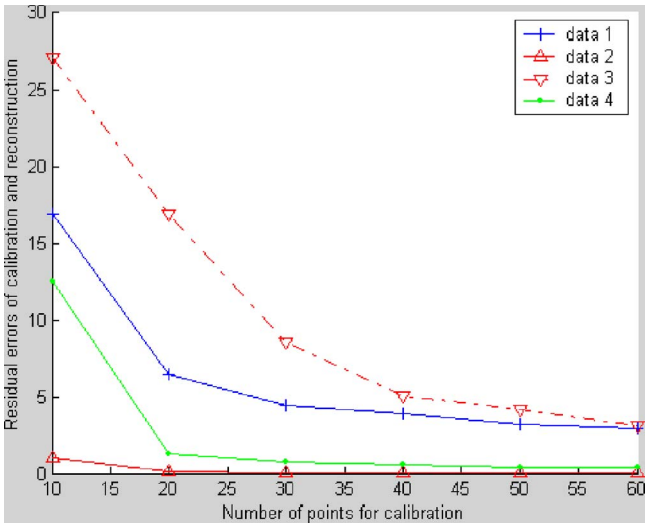
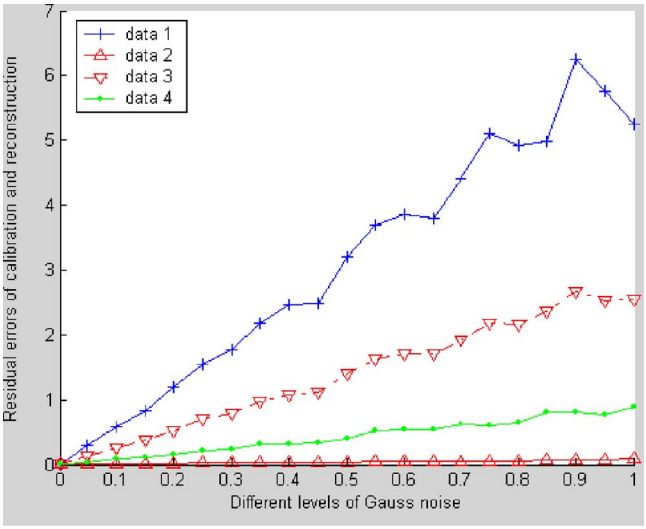
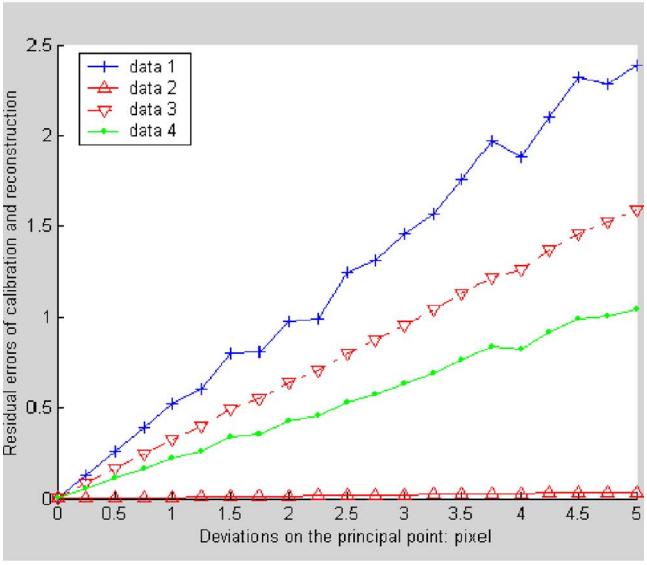


Fig. 3. (Color online) Residual errors as a function of the number of used points.



(a) Residual Errors as a function of noise levels



(b) Effects of deviation of the principal point

Fig. 4. (Color online) Residual errors in the dynamic calibration and 3D reconstruction. (a) Residual errors as a function of noise levels, (b) effects of deviation of the principal point.

the subpixel accuracy method, e.g., by the Harris corner finder method; (2) extracting as many points as possible and computing the scene plane-to-image homography in the least-squares sense.

In this paper, we assume that the principal point of the camera is fixed. However, the principal point may deviate a bit from its original position when the camera's focus is changed. Some researchers have studied the effect of this deviation.[15,16] In this work, we have tested this effect by numerical simulations. Here, we assumed that the principal point was perturbed by Gaussian noise ranging from 0 to 5 pixels. For each noise level, 100 trials were run, and the average residual errors in the calibration and reconstruction were computed as shown in Fig. 4(b). From this figure, we can see that the residual errors of calibration and reconstruction increase with the increased deviation of the principal point but still within a reasonable range. For example, when the focal length deviated by 2.3879 against its theoretical value of 950, the average error in the 3D reconstruction was at most 1 mm. Hence, small deviation of the principal point is tolerable in our method.

B. Real Data Experiments

1. System Configuration

In the real data experiments, the vision system consists of two major components, i.e., a PULNIX TMC-9700 CCD camera and a PLUS V131 DLP projector (see Fig. 5). The object to be reconstructed is placed in front of the scene plane. The projector is controlled by a computer to project a bundle of structured light onto the scene. The camera captures the illuminated scene by an image. Assuming that the projection points have been segmented into the points on the scene plane and those on the object's surface from the image, the procedure for system calibration and 3D reconstruction of the object can be implemented. As the 3D reconstruction is performed using homographic transformation rather than the triangulation method, the camera's orientation and position do not affect the accuracy significantly. Hence, the camera can be held by hand and moved freely in the experiments.

2. Design of Light Pattern

For generating the light pattern, there are many codification methods, including the time-multiplexing method, spatial neighborhood method, and direct codification method.[17] Since dynamic scene or moving objects may be involved in this system, we implemented a spatial neighborhood method where the light pattern formed a special case of perfect maps, and all the pattern information was encoded into a single shot.[18] The algorithm can be summarized as follows.

Let $\omega=\{1,2,\ldots,\gamma\}$ be a set of color primitives (for example, 1=red, 2=green, 3=blue, etc.). Given these color primitives, a 1D string is constructed such that each triplet of adjacent colors is distinct from every other triplet. This string is considered as the first row in the light pattern, and its size is γ^3+2. Similarly, a second string is constructed such that each pair of adjacent colors is distinct from every other pair, and the size is γ^2+1. For the other rows in the pattern, modulo operation is iteratively performed with the first row and each element of the second string. Then we have the light pattern whose size is $(\gamma^3+2)\times(\gamma^2+1)$.

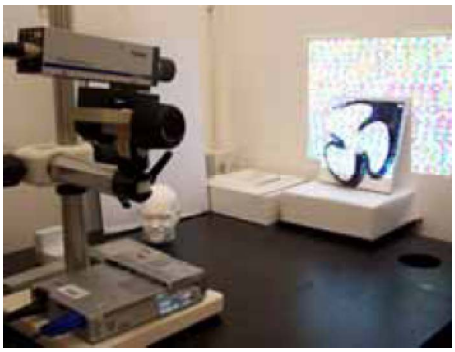


Fig. 5. (Color online) Configuration of our structured light system.

In practice, a trade-off needs to be made between the resolution of the pattern, i.e., the number of color primitives used and the complexity in the image processing. In general, the larger the number of color primitives used, the more complex the image processing and the more sensitive it is to noise. To preserve a good resolution and facilitate image processing, seven different colors, i.e., red, green, blue, white, cyan, magenta, and yellow were used. With the coding method, we can have a matrix of 50×345. Considering the requirement of our vision system, a 40×51 submatrix, in which any 3×3 neighbor is unique and any two consecutive blocks are distinct, is selected. Figure 6 gives a screen shot of the light pattern.

It should be noted that rather than color dots,[18] we use color-encoded grid blocks. The grid blocks can be segmented more easily by edge detection. The encoded points are the intersection of these edges, so that they can be found very accurately. When projecting dots, their centroids must be located. When a dot appears only partially in the image, its centroid will be erroneous. Moreover, the grid techniques allow adjacent cross points to be located by tracking the edges, but this is not the case for dot representation. This feature not only helps reduce the complexity in the image processing but also simplifies the decoding process. To decode the light pattern, a code word is defined by the color value of a grid block and its eight neighbors (north, south, west, east, northwest, southwest, northeast, and southeast). Then, a lookup table can be constructed containing the code words and their row's and column's indexes in the pattern. Since the first and last rows and columns in the pattern need not be considered, the table has 1862 elements.

3. Experimental Results

In the first experiment, with the variable parameters calibrated by our method, an industrial workpiece with known dimensions was used for testing [Fig. 7(a)]. The computer-aided design (CAD) model of the reconstructed result is given in Fig. 7(b), From this figure, we can see that the accuracy of the reconstruction is acceptable. For example, the neighboring surfaces of the model are nearly perpendicular to each other as in the real workpiece. Here, two geometrical elements, which were labeled as height and width, are evaluated. The measured quantities and the relative errors against their ground truth are listed in Table 1. It can be seen from the result that the accuracy of the reconstruction is acceptable.

In the second experiment, we tested on a general object with free-form surfaces, i.e., a fan model in Fig. 8(a). The object with the light pattern cast is shown in Fig. 8(b). Here, a total of 650 feature points from the fan's surface were extracted for reconstruction. Figure 8(c) shows the 3D point cloud reconstructed using parameters obtained from the dynamic calibration. Figure 8(d) is the CAD model of the clouds. Here the ground truth of these points was not known. To evaluate the accuracy, we reconstructed them using parameters obtained by a static calibration method and then computed the error distances between these two sets of points. Here, the mean and standard deviations of all these error distances were evaluated and are given in Table 2. For comparison, we listed our result together with that from Huynh and co-workers'[11] third experiment, since it was obtained un-

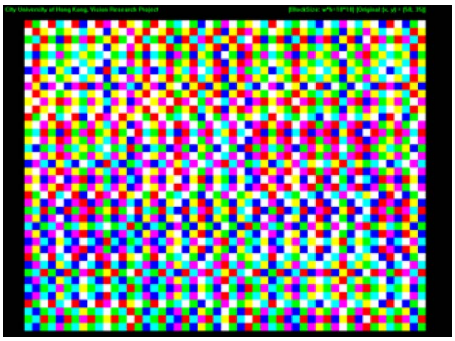
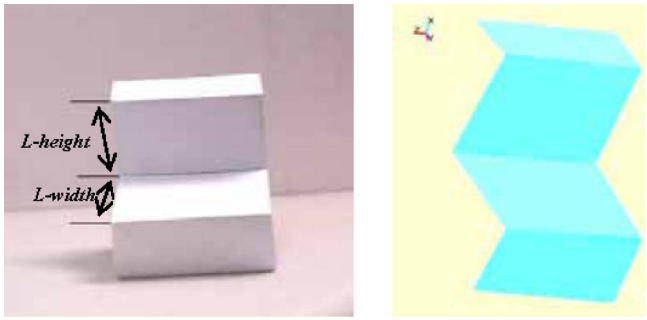


Fig. 6. (Color online) Screen shot of a color-encoded light pattern.

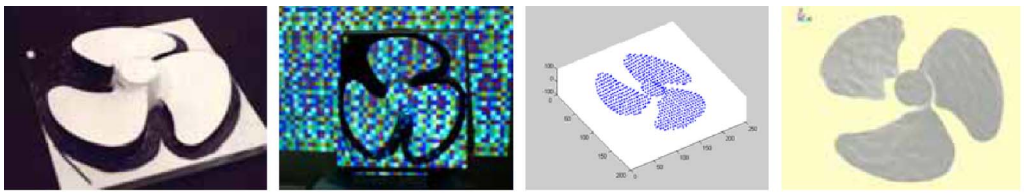


(a) The workpiece model (b) CAD model of the reconstructed result

Fig. 7. (Color online) Experiment on a workpiece. (a) Workpiece model, (b) CAD model of the restructured result.

Table 1. Measurements and Relative Errors

Elements	Actual Value (mm)	Measured Value (mm)	Relative Error (%)
L height	35	35.6501	1.86
L width	45	45.8913	1.98



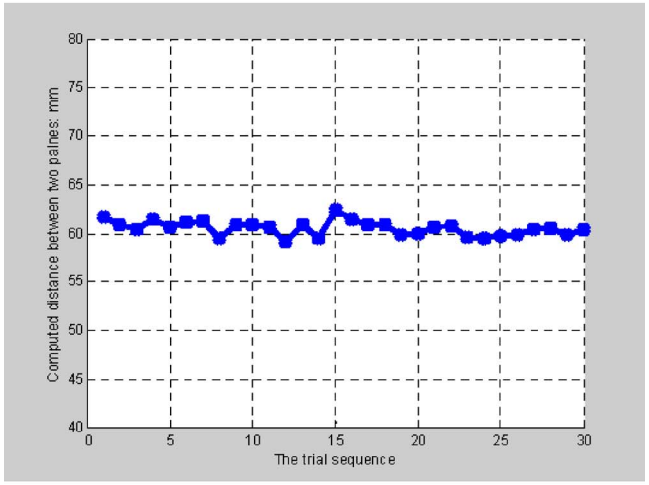
(a) The Fan model (b) An image captured (c) Reconstructed 3D point clouds (d) CAD model of the coulds

Fig. 8. (Color online) Experiment on a fan model. (a) fan model, (b) image captured, (c) reconstructed 3D point clouds, (d) CAD model of the clouds.

Table 2. Comparison of the Error Distance

Different Method	Mean Value (mm)	Standard Deviation (mm)
Huynh and co-workers' results	1.778	1.220
Our results	1.9878	0.6077

(a) A profile of the system setup



(b) The measured distance between the two planes

Fig. 9. (Color online) Example of the demonstration system. (a) Profile of the system setup, (b) measured distance between the two planes.

der a similar condition. From this table, we can see that the mean errors by the two methods are consistent (although ours is a bit larger), while the standard deviation of our method is significantly smaller than that of Huynh *et al.* This demonstrates the validity and accuracy of our scheme.

We also conducted another experiment to demonstrate the use of the system in dynamically changing configurations [Fig. 9(a)]. Two parallel planes were reconstructed and their distance measured. In this test, the two planes were static (thus with a constant distance between them), while the relative pose between the camera and the projector was changed all the time via the handheld camera. The results of 30 samples were recorded and shown in Fig. 9(b). Despite the changes in the camera pose, the measured distance between the two planes turned out to be relatively constant with a mean of 60.52 mm and the standard deviation of 0.76 mm. This result is quite satisfactory, considering the fact that no special effort was made in the initial calibration of the projector and the scene plane.

6. CONCLUSIONS

In this paper, we present a method for dynamic recalibration and 3D reconstruction for a structured light vision system. The advantage of our method is that the dynamic recalibration and 3D reconstruction are formulated in a closed-form solution, which avoids the possible convergence problem with an iterative algorithm. In addition, our method requires only a single image for the whole process of dynamic calibration and reconstruction. Thus a handheld camera can be used in this system. These features are especially important for some time critical tasks in a dynamic scene.

With the relaxed requirement of only one known plane, the proposed method is potentially useful for many practical applications. For example, in mobile robot navigation, we can assume that the ground plane of the floor or the road surface always exists and remains unchanged relative to the projector mounted on the robot during its navigation, which satisfies the requirement here. Besides, it is particularly applicable for a desktop vision system in which the desktop can be taken as the known scene plane.

ACKNOWLEDGMENT

The work described in this paper was fully supported by grants from the Research Grants Council of Hong Kong (projects CityU1206/04E and CityU117605).

REFERENCES

1. T. Moons, L. van Gool, M. Proesmans, and E. Pauwels, "Affine reconstruction from perspective image pairs with a relative object camera translation in between," IEEE Trans. Pattern Anal. Mach. Intell. **18**, 77–83 (1996).
2. Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1330–1334 (2000).
3. P. Sturm and S. Maybank, "On plane-based camera calibration: a general algorithm, singularities, applications," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, June 23–25 (IEEE, 1999), pp. 432–437.
4. A. M. McIvor, "Nonlinear calibration of a laser stripe profiler," Opt. Eng. (Bellingham) **41**, 205–212 (2002).
5. J. Y. Bouguet and P. Perona, "3D photography on your desk," Tech. Rep. (California Institute of Technology, 1997).
6. Y. F. Li and S. Y. Chen, "Automatic recalibration of a structured light vision system," IEEE Trans. Rob. Autom. **19**, 259–268 (2003).
7. S. Y. Chen and Y. F. Li, "Self-recalibration of a color-encoded light system for automated three-dimensional measurements," Meas. Sci. Technol. **14**, 33–40 (2003).
8. D. Fofi, J. Salvi, and E. Mouaddib, "Uncalibrated reconstruction: an adaptation to structured light vision," Pattern Recogn. **36**, 1631–1644 (2003).

9. C. H. Chen and A. C. Kak, "Modeling and calibration of a structured light scanner for 3-D robot vision," in *Proceedings of the IEEE Conference on Robotics and Automation* (IEEE, 1987), pp. 807–815.
10. I. D. Reid, "Projective calibration of a laser-stripe range finder," Image Vis. Comput. **14**, 659–666 (1996).
11. D. Q. Huynh, R. A. Owens, and P. E. Hartmann, "Calibrating a structured light stripe system: a novel approach," Int. J. Comput. Vis. **33**, 73–86 (1999).
12. B. Zhang and Y. F. Li, "An efficient method for dynamic calibration and 3D reconstruction using homographic transformation," Sens. Actuators A **119**, 349–357 (2005).
13. M. Li and J. Lavest, "Some aspects of zoom lens camera calibration," IEEE Trans. Pattern Anal. Mach. Intell. **8**, 1105–1110 (1996).
14. P. Sturm, Z. Cheng, C. Chen, and A. Poo, "Focal length calibration from two views: method and analysis of singular cases," Comput. Vis. Image Underst. **99**, 58–95 (2005).
15. R. G. Wilson, "Modeling and calibration of automated zoom lenses," Ph.D. dissertation (Carnegie Mellon University, 1994).
16. A. G. Wiley and K. W. Wong, "Geometric calibration of zoom lenses for computer vision metrology," Photogramm. Eng. Remote Sens. **61**, 69–74 (1995).
17. J. Salvi, J. Pages, and J. Batlle, "Pattern codification strategies in structured light systems," Pattern Recogn. **37**, 827–849 (2004).
18. P. M. Griffin, L. S. Narasimhan, and S. R. Yee, "Generation of uniquely encoded light patterns for range data acquisition," Pattern Recogn. **25**, 609–616 (1992).