2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

A Study on the Relative Pose Problem in an Active Vision System with varying Focal Lengths

B. Zhang and Y. F. Li*

Dept. of Manufacturing Engg and Engg Management, City University of Hong Kong, Kowloon, Hong Kong

**Author for correspondence, Email:* meyfli@cityu.edu.hk

***Abstract* - In this work, the relative pose problem is addressed in our structured light system. Assuming that there is an arbitrary planar structure in the scene, we suggest a method for estimating the rotation matrix and translation vector between the camera and the projector. In this system, the varying focal lengths of the camera are allowed and can be obtained without any further assumptions. Finally, we give some experimental results to validate this method.**

I. INTRODUCTION

Calibration of a vision system has always been an active topic in the computational robot vision. In general, this task involves obtaining the intrinsic parameters and extrinsic parameters of the system. During the past decades, many techniques have been proposed for solving this problem. These methods can be roughly classified into two categories, i.e. static calibration and dynamic calibration. In static calibration, a calibration target or device with precisely known structures is required to perform the system calibration [1, 2]. With a planar pattern, the target needs to be placed at several positions in front of the vision sensor. This process must be repeated again whenever the vision system is moved or adjusted. So it's very inconvenient to use this kind of techniques in dynamic environments with objects in motion. Therefore, it is desirable for a vision system to have the ability of recalibration without requiring external 3-D data as those provided by a precision calibration device. The vision system will automatically recalibrate itself whenever it's is changed. This kind of methods is generally named as dynamic calibration or self-calibration [3, 4].

In practice, the vision researchers find that an improved performance can be expected from dynamically calibrating the pose parameters while the intrinsic parameters are calibrated statically. This is known as the relative pose problem and many ingenious methods have been introduced to solve this problem. In [5], the authors proposed a geometric closed-form solution for the pose parameters based on volume measurement of tetrahedra composed of target points and the optical center of the vision system. When the point depths had been recovered from orthogonal decompositions, Paul [6] used two phases to solve the exterior orientation, i.e. first calculation of the projective parameters and then obtaining the motion parameters. In Nister's work [7], the relative pose problem was addressed using five non-coplanar points. However, a tenth degree polynomial was involved in his method. Some authors have also noticed that the relative pose problem can be solved from the correspondences between images of a planar structure. One of the earliest techniques is from Tsai [8] where the solution was provided by singular value decomposition of a plane-base Homography. Given the Homography matrix, Zhang [9] proposed a method for this problem from a case by case analysis of different geometric situations, where as many as six cases were considered. Since these methods are primarily for the noise-free images and solve for the exact solutions, they will be fragile in the presence of noise. For the projector-screen-camera system, Raij [10] and Okatani [11] investigated the auto-calibration methods by treating the projectors as virtual cameras and using the principles of planar auto-calibration. However, multiple projectors are required in their methods.

To overcome the mentioned limitations and improve the robustness of the vision system, we presented a calibration method using Homographic matrix where over-constrained systems were used for determining the relative pose automatically in [12]. In that method, algebraic tools, such as matrix decomposition, are employed to obtain a closed-form solution. In this paper, we will continue our work on this problem by exploring the geometric information in the structured light system. Assuming the internal parameters are known and there is an arbitrary planar structure in the scene, the translation vector is first obtained by constraints from the plane-induced parallax. Then the rotation matrix can be computed from the planed-based Homography. Besides, we show that some intrinsic parameters of the system, e.g. the focal lengths of the camera, can be calibrated without any further assumptions. Consequently, 3D reconstruction can be performed with traditional triangulation method. By dynamic calibration, we mean that only a single image is required to calibrate our system so that the variable parameters can be computed when the system is working. With its easy implementation, this method promises wide applications for active vision systems, such as robot hand-eye calibration and object recognition, etc.

II. PINHOLE MODEL AND THE HOMOGRAPHY

A. Pinhole Model

We take the classical pinhole model for the camera and the projector is regarded as a dual camera. Hence, they can be described by the following two matrices

$$\boldsymbol{K}_c = \begin{bmatrix} f_u & s & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \boldsymbol{K}_p = \begin{bmatrix} f_u^{'} & s^{'} & u_0^{'} \\ 0 & f_v^{'} & v_0^{'} \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

where f_u and f_v represent the focal lengths of the camera in terms of pixel dimensions along *U*-axis and *V*-axis respectively, and $(u_0 \quad v_0)^T$ is the principal point, *s* is a skew factor of the camera, representing cosine value of the subtending angle between *U*-axis and *V*-axis. Similar notations are defined for the projector. These parameters are generally named as the intrinsic parameters of the sensors.

For the camera and the projector, we define a right-handed coordinate system originated at their optical centers respectively. Let the world coordinate system coincides with that of the camera. Then the rotation matrix and translation vector between the camera and the projector, denoted by $\boldsymbol{R}$ and $\boldsymbol{t}$, are called the relative pose in the structured light system.

B. Homography

Given arbitrary plane π in 3D space, its images in the camera and the projector are related by a 3×3 transformation matrix according to the projective geometry. In general, this transformation is termed as plane-based Homography and denote by $\boldsymbol{H}$. Let $\boldsymbol{M}$ be an arbitrary point on π and its correspondent projections be $\boldsymbol{m}_c$ and $\boldsymbol{m}_p$ respectively. Then we have

$$\boldsymbol{m}_p = \lambda \boldsymbol{H} \boldsymbol{m}_c \quad (2)$$

where λ is a nonzero scale factor.

Let $\boldsymbol{H} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix}$, and its vector be

$$\boldsymbol{h} = (h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9)^{\mathrm{T}}.$$

For the *i*-th pair of projections, we denote $\boldsymbol{m}_{p,i} = [u_i \quad v_i \quad 1]^{\mathrm{T}}$ and $\boldsymbol{m}_{c,i} = [u_i^{'} \quad v_i^{'} \quad 1]^{\mathrm{T}}$.

According (2), each pair of corresponding points provides two constraints on the Homography. So given $n\,(n \geq 4)$ pairs of corresponding projection points of the scene, we have the following $2n$ equations:

$$\boldsymbol{A}\boldsymbol{h} = \boldsymbol{0} \quad (3)$$

where:

$$\boldsymbol{A} = \begin{bmatrix} u_1 & v_1 & 1 & 0 & 0 & 0 & -u_1^{'}u_1 & -u_1^{'}v_1 & -u_1^{'} \\ 0 & 0 & 0 & u_1 & v_1 & 1 & -v_1^{'}x_1 & -v_1^{'}v_1 & -v_1^{'} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ u_n & u_n & 1 & 0 & 0 & 0 & -u_n^{'}u_n & -u_n^{'}v_n & -u_n^{'} \\ 0 & 0 & 0 & u_n & v_n & 1 & -v_n^{'}u_n & -v_n^{'}v_n & -v_n^{'} \end{bmatrix}.$$

The solution for the vector $\boldsymbol{h}$ can be determined by singular value decomposition of matrix $\boldsymbol{A}$ in (3), and so is the Homography matrix $\boldsymbol{H}$.

III. CALIBRATING THE RELATIVE POSE

A. Calibration of the translation vector

For a point on plane π, it's well known that its projection in the camera can be transferred to the corresponding projection in the projector by the plane-based Homography. Therefore, the residual vector between the Homography transferred point and the projection point is a null vector. For a point that is not on the plane, its residual vector will not be null. This factor is illustrated in Fig. 1. Here, $\boldsymbol{m}_c$ and $\boldsymbol{m}_p$ represent a pair of corresponding projections of 3D point $\boldsymbol{M}$, while $\boldsymbol{m}_p^{'}$ and $\boldsymbol{m}_c$ are projections of $\boldsymbol{M}^{'}$. Then $\boldsymbol{H}\boldsymbol{m}_c$ and $\boldsymbol{m}_p$ will represent the same point since $\boldsymbol{M}$ belongs to the plane π , but not the case for $\boldsymbol{M}^{'}$ which is out of the plane. The line segment joining $\boldsymbol{m}_p$ and $\boldsymbol{m}_p^{'}$ is generally named as planar parallax or plane-induced parallax. The planar parallax is collinear with the epipolar denoted as $\boldsymbol{e}^{'}$. Many researchers use this property to establish correspondences between images and estimate depth information relative to the reference plane [13, 14, 15]. For example, Sawhney [14] and Irani [15] proposed two methods for 3D analysis based on the idea of planar parallax with respect to an arbitrary plane where the structure was solved directly from brightness measurements. The difference between them is that the former is applicable within two frames while the latter extends it to multiple frames. However, these methods depend on an accurate alignment of the images with respect to the viewed planar surface. In this paper, we will compute the translation vector and rotation matrix of the system from its geometric relationship. Thus, not only the depth information but also the three dimensional coordinates of the scene can be obtained by the triangulation method.

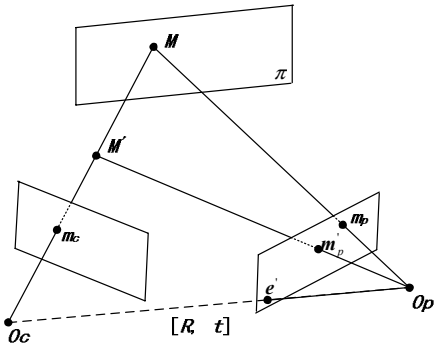


Fig. 1 Geometric relationship in the system

In Fig. 1, $\boldsymbol{O}_c$ and $\boldsymbol{O}_p$ represent the optical centers of the camera and the projector, and the translation vector $\boldsymbol{t}$ passes through them. Considering the plane defined by the three points $\boldsymbol{M}'$, $\boldsymbol{O}_c$ and $\boldsymbol{O}_p$, it's obvious that $\boldsymbol{m}'_p$ and $\boldsymbol{M}$ belong to this plane since they lie on the line segments $\overline{\boldsymbol{O}_p\boldsymbol{M}'}$ and $\overline{\boldsymbol{O}_c\boldsymbol{M}'}$, respectively. And $\boldsymbol{m}_p$ also lie on the plane since it belongs to the line segment $\overline{\boldsymbol{O}_p\boldsymbol{M}}$. Consequently, $\boldsymbol{m}_p$, $\boldsymbol{m}'_p$ and $\boldsymbol{t}$ are coplanar. Then we have

$$\left(\left(\boldsymbol{m}'_p - \boldsymbol{m}_p\right)\times \boldsymbol{m}_p\right)^{\mathrm{T}} \boldsymbol{t} = 0 \tag{4}$$

From (4), each projection pair provides one constraint on the translation vector. Thus two such pairs are sufficient to determine the translation vector since it has two degrees of freedom. In practice, we use more than two pairs to improve the reliability of the solution. So we have the following equations:

$$\boldsymbol{t}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{t} = 0 \tag{5}$$

where $\boldsymbol{Q} = \sum_i \left(\left(\boldsymbol{m}'_{p,i} - \boldsymbol{m}_{p,i}\right)\times \boldsymbol{m}_{p,i}\right)\left(\left(\boldsymbol{m}'_{p,i} - \boldsymbol{m}_{p,i}\right)\times \boldsymbol{m}_{p,i}\right)^{\mathrm{T}}$.

The solution for the translation vector is the eigenvector corresponding to the smallest singular value by singular value decomposition of $\boldsymbol{Q}$.

B. Recovering the Rotation Matrix

We assume that the equation of the plane π is $\boldsymbol{n}^{\mathrm{T}}\boldsymbol{M} = 1$. For the calibrated camera and projector, the Homography matrix can be expressed as

$$\lambda \boldsymbol{H} = \boldsymbol{R} + \boldsymbol{t}\boldsymbol{n}^{\mathrm{T}} \tag{6}$$

In (6), $\boldsymbol{H}$ and $\boldsymbol{t}$ have been obtained previously while the unknowns include λ, $\boldsymbol{R}$ and $\boldsymbol{n}$. We will first determine the scale factor λ and then the rotation matrix $\boldsymbol{R}$.

Let the translation vector be denoted as $\boldsymbol{t} = \begin{bmatrix} t_1 & t_2 & t_3 \end{bmatrix}^{\mathrm{T}}$ and its skew symmetric matrix be $[\boldsymbol{t}]_\times = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}$.

Multiplying matrix $[\boldsymbol{t}]_\times$ on both sides of (6), we have

$$\lambda[\boldsymbol{t}]_\times \boldsymbol{H} = [\boldsymbol{t}]_\times \boldsymbol{R} \tag{7}$$

As $\boldsymbol{R}$ is a rotation matrix, $\boldsymbol{R}\boldsymbol{R}^T = \boldsymbol{I}$. From (7), we obtain

$$\lambda^2[\boldsymbol{t}]_\times \boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}[\boldsymbol{t}]_\times = [\boldsymbol{t}]_\times[\boldsymbol{t}]_\times \tag{8}$$

In (8), the only unknown is λ^2. So we can solve it from (8) and the scale factor can be determined by the square root of λ^2. The sign of the scale factor should make both sides of (2) consistent. With i pairs of projections, if

$$\lambda \sum_i \boldsymbol{m}_{p,i}^{\mathrm{T}} \boldsymbol{H} \boldsymbol{m}_{c,i} < 0 \tag{9}$$

its sign should be reversed. One pair of projection points is sufficient to determine the sign. In practice, we use two or more pairs to increase its reliability as in (9).

Now, we can solve the rotation matrix $\boldsymbol{R}$ from (7). The method can be described as follows.

Let $\boldsymbol{C} = \lambda[\boldsymbol{t}]_\times \boldsymbol{H}$. We have

$$[\boldsymbol{t}]_\times \boldsymbol{R} = \boldsymbol{C} \tag{10}$$

Let $\boldsymbol{R}_i$ and $\boldsymbol{C}_i$ denote the *i*-th columns of matrices $\boldsymbol{R}$ and $\boldsymbol{C}$ respectively. From (10), we get

$$\boldsymbol{C}_i = \boldsymbol{t} \times \boldsymbol{R}_i \tag{11}$$

With *lagrange formulae* $(\boldsymbol{a}\times\boldsymbol{b})\times\boldsymbol{c} = (\boldsymbol{a},\boldsymbol{c})\boldsymbol{b} - (\boldsymbol{b},\boldsymbol{c})\boldsymbol{a}$, we obtain

$$\boldsymbol{R}_1 = \boldsymbol{C}_1 \times \boldsymbol{t} + \boldsymbol{C}_2 \times \boldsymbol{C}_3$$
$$\boldsymbol{R}_2 = \boldsymbol{C}_2 \times \boldsymbol{t} + \boldsymbol{C}_3 \times \boldsymbol{C}_1$$
$$\boldsymbol{R}_3 = \boldsymbol{C}_3 \times \boldsymbol{t} + \boldsymbol{C}_1 \times \boldsymbol{C}_2 .$$

Therefore, the solution for the rotation matrix is

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{C}_1 \times \boldsymbol{t} + \boldsymbol{C}_2 \times \boldsymbol{C}_3 & \boldsymbol{C}_2 \times \boldsymbol{t} + \boldsymbol{C}_3 \times \boldsymbol{C}_1 & \boldsymbol{C}_3 \times \boldsymbol{t} + \boldsymbol{C}_1 \times \boldsymbol{C}_2 \end{bmatrix} \tag{12}$$

IV. CALIBRATING THE FOCAL LENGTHS

In our system, the projector is used as a lighting projection device. Its intrinsic parameters are usually kept fixed once properly calibrated. The camera can be automatically self-adapted (focusing or zooming) for better performance. Recently, some researchers have studied this case by assuming a camera with vary focal length but zero skew, known aspect ratio and the principal point. For instance, Sturm [18] and Cao [19] addressed the calibration of focal length independently by setting the principal point to (0, 0), and the aspect ratio to one in their work.

In this section, we will show how to recover the possibly varying focal lengths of the camera with the remaining parameters known. In this case, the Homography matrix can be expressed as

$$\lambda \boldsymbol{H} = \left(\boldsymbol{R} + \boldsymbol{t}\boldsymbol{n}^{\mathrm{T}}\right)\boldsymbol{K}_c^{-1} \tag{13}$$

where $\boldsymbol{K}_c = \begin{bmatrix} f_u & 0 & 0 \\ 0 & f_v & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

From (13), we have

$$\lambda \boldsymbol{H}\boldsymbol{K}_c = \boldsymbol{R} + \boldsymbol{t}\boldsymbol{n}^{\mathrm{T}} \tag{14}$$

Multiplying matrix $[\boldsymbol{t}]_\times$ on both sides of (14), we get

$$\lambda [\boldsymbol{t}]_\times \boldsymbol{H}\boldsymbol{K}_c = [\boldsymbol{t}]_\times \boldsymbol{R} \tag{15}$$

Then we arrive at the following matrix equation:

$$\lambda^2 [\boldsymbol{t}]_\times \boldsymbol{H}\boldsymbol{K}_c \boldsymbol{K}_c \boldsymbol{H}^{\mathrm{T}} [\boldsymbol{t}]_\times = [\boldsymbol{t}]_\times [\boldsymbol{t}]_\times \tag{16}$$

In (16), the unknowns include λ and $\boldsymbol{K}_c$. Let $\boldsymbol{\varphi} = \begin{bmatrix} f_u^2 & f_v^2 & 1 \end{bmatrix}^{\mathrm{T}}$. Eliminating λ^2 in (16), we have the following linear equations for the intrinsic parameters in matrix $\boldsymbol{K}_c$

$$\boldsymbol{\Psi}\boldsymbol{\varphi} = 0 \tag{17}$$

where $\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{\alpha} & \boldsymbol{\beta} & \boldsymbol{\gamma} \end{bmatrix}$ is a 4×3matrix,

$$\boldsymbol{\alpha} = \begin{bmatrix} t_2h_4^2 - 2t_2^2h_4h_7 - 2t_1^2h_4h_7 + t_2^3h_7^2 + t_1^2t_2h_7^2 - t_2h_1^2 - t_2^3h_1^2 + 2t_1h_1h_4 + 2t_1t_2^2h_1h_4 - t_1^2t_2h_4^2 \\ -t_2^2h_1h_4 - t_1^2h_1h_4 + t_1t_2^2h_4h_7 + t_1^3h_4h_7 + t_2^3h_4h_7 + t_1^2t_2h_1h_7 - t_1t_2^3h_7^2 - t_1^3t_2h_7^2 + t_1t_2^3h_1^2 - 2t_1^2t_2^2h_1h_4 + t_1^3t_2h_4^2 \\ (t_1h_4 - t_2h_1)(t_2^2h_7 - t_1h_1 - t_2h_4 + t_1^2h_7) \\ t_1h_1^2 - 2t_2^2h_1h_7 - 2t_1^2h_1h_7 + t_1t_2^2h_7^2 + t_1^3h_7^2 - t_1t_2^2h_1^2 + 2t_2h_1h_4 + 2t_1^2t_2h_1h_4 - t_1h_4^2 - t_1^3h_4^2 \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} t_2h_5^2 - 2t_2^2h_5h_8 - 2t_1^2h_5h_8 + t_2^3h_8^2 + t_1^2t_2h_8^2 - t_2h_2^2 - t_2^3h_2^2 + 2t_1h_2h_5 + 2t_1t_2^2h_2h_5 - t_1^2t_2h_5^2 \\ -t_2^2h_2h_5 - t_1^2h_2h_5 + t_1t_2^2h_5h_8 + t_1^3h_5h_8 + t_2^3h_2h_8 + t_1^2t_2h_2h_8 - t_1t_2^3h_8^2 - t_1^3t_2h_8^2 + t_1t_2^3h_2^2 - 2t_1^2t_2^2h_2h_5 + t_1^3t_2h_5^2 \\ (t_1h_5 - t_2h_2)(t_2^2h_8 - t_1h_2 - t_2h_5 + t_1^2h_8) \\ t_1h_2^2 - 2t_2^2h_2h_8 - 2t_1^2h_2h_8 + t_1t_2^2h_8^2 + t_1^3h_8^2 - t_1t_2^2h_2^2 + 2t_2h_2h_5 + 2t_1^2t_2h_2h_5 - t_1h_5^2 - t_1^3h_5^2 \end{bmatrix}$$

and

$$\boldsymbol{\gamma} = \begin{bmatrix} -2t_2^2h_6h_9 + t_1^2t_2h_9^2 - t_1^2t_2h_6^2 - 2t_1^2h_6h_9 + 2t_1h_3h_6 + 2t_1t_2^2h_3h_6 + t_2h_3^2 + t_2^3h_6^2 - t_2h_3^2 - t_2^3h_3^2 \\ -t_1^2h_3h_6 - t_2^2h_3h_6 + t_1t_2^2h_6h_9 + t_1^3h_6h_9 - 2t_1^2t_2h_3h_9 + t_1^3h_3h_9 + t_2^3h_3h_9 - t_1t_2^3h_6^2 - t_1^3t_2h_6^2 + t_1t_2^3h_3^2 + t_1^3t_2h_6^2 \\ (t_1h_6 - t_2h_3)(t_2^2h_9 - t_1h_3 - t_2h_6 + t_2^2h_9) \\ -t_1t_2^2h_3^2 + t_1t_2^2h_6^2 + 2t_1^2t_2h_3h_6 + 2t_2h_3h_6 - 2t_2^2h_3h_9 - t_1h_6^2 + t_1h_3^2 + t_1^3h_6^2 - t_1^3h_3^2 - 2t_1^2h_3h_9 \end{bmatrix}.$$

Then we have the following proposition.

Proposition: The rank of matrix $\boldsymbol{\Psi}$ in (17) is 2, i.e.

$$rank(\boldsymbol{\Psi}) = 2 \tag{18}$$

The proof is briefly given here.

First, we set two independent vectors as

$$\boldsymbol{\lambda}_1 = \begin{bmatrix} 1 & 0 & -t_1 & -t_1/t_2 \end{bmatrix}^{\mathrm{T}}$$

$$\boldsymbol{\lambda}_2 = \begin{bmatrix} 0 & 1 & -1 & -2/t_2 \end{bmatrix}^{\mathrm{T}}.$$

It is easy to verify that $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ have the following relationship with the three columns in matrix $\boldsymbol{\Psi}$:

$$\boldsymbol{\alpha} = \alpha_1\boldsymbol{\lambda}_1 + \alpha_2\boldsymbol{\lambda}_2$$

$$\boldsymbol{\beta} = \beta_1\boldsymbol{\lambda}_1 + \beta_2\boldsymbol{\lambda}_2$$

$$\boldsymbol{\gamma} = \gamma_1\boldsymbol{\lambda}_1 + \gamma_2\boldsymbol{\lambda}_2$$

where α_i, β_i and γ_i are the i-th elements in $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ respectively.

Therefore, the rank of matrix $\boldsymbol{\Psi}$ must be 2.

By singular value decomposition of $\boldsymbol{\Psi}$, the solution for the vector $\boldsymbol{\varphi}$ is the singular vector corresponding to the smallest singular value. Considering the focal lengths should be positive, their solutions are given as

$$f_u = \sqrt{\varphi_1 / \varphi_3} \tag{19}$$

and

$$f_v = \sqrt{\varphi_2 / \varphi_3} \tag{20}$$

where φ_i is the i-th element of $\boldsymbol{\varphi}$.

Once $\boldsymbol{K}_c$ is known, the scalar factor λ in (16) is the only unknown parameter. So we can solve it using any one equation in (16) and again its sign is determined by (9).

Now, the only unknown in (15) is the rotation matrix $\boldsymbol{R}$, which can be computed by formula (12).

V. IMPLEMENTATION PROCEDURE

In summary, the procedure for calibrating the structured light system is given as follows:

Step 1: Computing the Homography matrix $\boldsymbol{H}$ between the camera plane and projector plane according to (3);

Step 2: Establishing constraints (5) and calculating the translation vector $\boldsymbol{t}$;

Step 3: If all the intrinsic parameters of the camera are fixed and known, determining the scale factor λ and its sign using (8) and (9) respectively;

If the focal lengths of the camera are varying, computing them first by (18) and (19), then the scale factor λ;

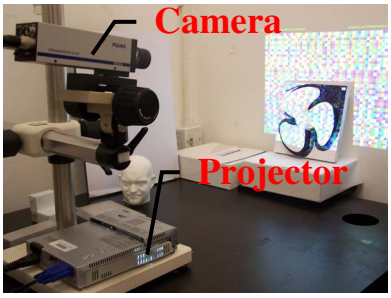
Step 4: Calculating the rotation matrix by formula (12);

Step 5: Optionally, the results can be improved by bundle adjustment, after having obtained the relative pose.

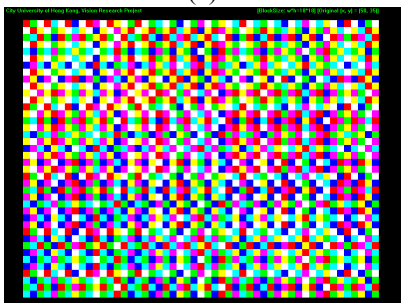
VI. EXPERIMENTS

Fig. 2a shows the system setup for real data experiments, which consists of a PULNIX TMC-9700 CCD camera and a PLUS V131 DLP projector. Fig. 2b gives the color-encoded light pattern for the projector, which can be used to uniquely identify the correspondences between the projector plane and the image plane. Here, seven different colors, i.e. red, green, blue; white, cyan, magenta; yellow, are used. It should be noted that rather than color dots in [16], we use color-encoded grid blocks. In practice, the grid blocks can be segmented more easily by edge detection. The encoded points are the intersection of these edges, so they can be found very accurately. When projecting dots, their mass centres must be located. In case of a dot appears only

partially in the image, its mass centre will be incorrect. 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 decreases the complexity of 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 look-up 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 items.



(a)



(b)

(a): The experimental setup
(b): A screen shot of the color-encoded light pattern
Fig. 2 The configuration of our structured light system

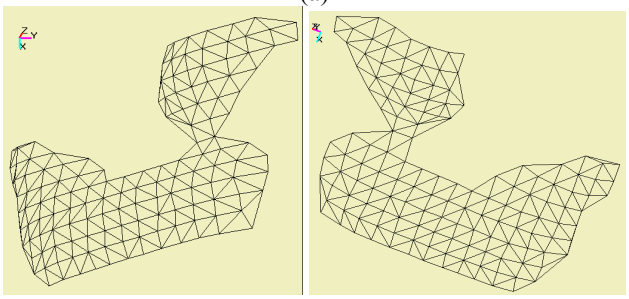
In static calibration, the intrinsic parameters of the camera and the projector were firstly determined by a planar pattern using Zhang's method [17]. Here, we placed the pattern at ten different positions to increase the calibration accuracy. In the current experiments, we assumed that the intrinsic parameters were all kept fixed and the wall in our lab used as the planar surface. When the system is working, the relative pose between the camera and the projector is arbitrarily changed. Then with more than four pairs of corresponding projections from the wall, we can calibrate the rotation matrix and the translation vector according to the proposed procedure. Consequently, 3D reconstruction is implemented to test the calibration results.

Fig. 3a gave a duck model used in the experiments. Here, a random frame from the video sequence was used to reconstruct the model. Fig. 3 b and Fig. 3c illustrated the polygonized results of the reconstructed point clouds in two different viewpoints. Totally 143 points from the model were reconstructed. It is observed that they are very similar to the original model. Since no attempt has been made to obtain the ground truth, we do not know real values of the variable parameters of the system and the point clouds. To evaluate their accuracy quantitatively, these point clouds were back-projected to the image plane and the projector plane respectively. Then the discrepancies between the real feature points and back-projected points were computed. In general, the more accurate the calibrated results, the smaller discrepancies we will have. To visualize the results, we plotted the feature points and back-projected points in Fig. 3d and Fig. 3e. Fig. 3f gave one zoomed part of the image. We can see that the feature points and back-projected points are very close to each other as expected. Quantitatively, the mean results of the discrepancies for the camera and the projector were given in Table 1. As very small discrepancies were obtained, this further demonstrated the validity and accuracy of our method.

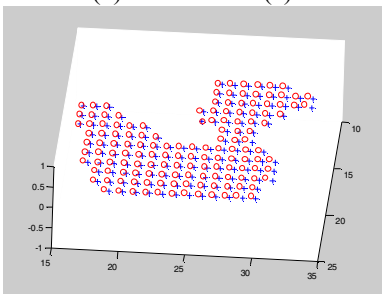
Table 1: Discrepancies between image points and re-projected points

The Mean Discrepancies	Image Plane	(0.1829, 0.3336)
	Projector Plane	(0.0244, 0.0142)

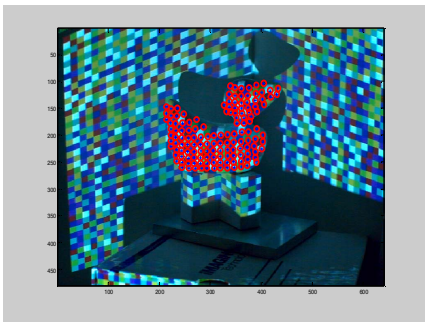
(a)



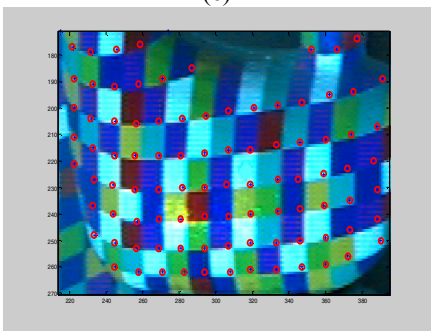
(b) (c)



(d)



(e)



(f)

(a) A duck model used for the experiment
(b, c) Polygonized results of the reconstructed points clouds from two viewpoints
(b, c, d) Original feature points and back-projected points: Blue '+' represents original feature points while red 'o' represents back-projected points from reconstructed 3D points. They should coincide with each other theoretically.
Fig. 3: An experiment on a duck model

VII. Conclusions

In this work, we have investigated the relative pose problem in the structured light system. Besides, we show that two of the intrinsic parameters, such as focal length and aspect ratio of the camera, can be computed without any further assumptions. In this method, the analytical solutions can be obtained from singular value decomposition and the reliability in implementation is easily improved via the use of redundant data. Compared with the classical eight-point algorithm, our method is more reliable and requires less points (six points are adequate) for a linear solution. Since a single image is sufficient for the calibration and 3D reconstruction, either the objects or the system can be moved or adjusted during the task. Besides, the scene plane can be a real or virtual one as long as it provides four or more coplanar points that will enable the estimation of the Homography matrix. Applications of this method include navigation of a mobile robot along ground plane, wall climbing robot for cleaning, inspection and maintenance of buildings and object recognition, etc. Future work includes studying the possibility of varying focal lengths of the projector and implementing the algorithm on a real mobile robot.

Acknowledgment

The work described in this paper was fully supported by a grant from the Research Grants Council of Hong Kong [Project No. CityU117605]

References

[1] R. Valkenburg, A. McIvor, Accurate 3D measurement using a structured light system, *Image and Vision Computing*, Vol. 16, No. 2, Feb. 1998, pp. 99-110.
[2] D. Q. Huynh, R. A. Owens and P. E. Hartmann, Calibrating a structured light stripe system: a novel approach, *International Journal of computer vision*, Vol. 33, No. 1, Sep. 1999, pp. 73-86.
[3] Y. F. Li and S. Y. Chen, Automatic recalibration of a structured light vision system, ***IEEE*** *Trans. on Robotics and Automation*, Vol. 19, No. 2, Apr. 2003, pp. 259 –268
[4] S. Y. Chen and Y. F. Li, Self-recalibration of a color-encoded light system for automated three-dimensional measurements, Measurement Science and Technology, Vol. 14, No. 1, Jan. 2003, pp. 33-40.
[5] M. Abidi and T. Chandra, A new efficient and direct solution for pose estimation using quadrangular targets, ***IEEE*** *Trans. on Pattern Analysis and. Machine Intelligence*, Vol. 17, No. 5, 1995, pp.534-538.
[6] Paul D. Fiore, Efficient Linear Solution of Exterior Orientation, IEEE Trans. on Pattern Analysis *and Machine Intelligence*, Vol. 23, No. 2, Feb. 2001, pp. 140-148.
[7] D. Nister, An efficient solution to the five-point relative pose problem, ***IEEE*** *Trans. on Pattern Analysis and. Machine Intelligence*, Vol. 26, No. 6, 2004, pp.756-770.
[8] R. Tsai, T. Huang, and W. Zhu, Estimating three dimensional motion parameters of a rigid planar patch, II: singular value decomposition, ***IEEE*** *Trans. Acoust. Speech, and Signal Process*, Vol. ASSP-30, 1982, pp. 525-534.
[9] Z. Zhang, A. R. Hanson, Scaled Euclidean 3D reconstruction based on externally uncalibrated cameras, ***IEEE*** *International Symposium on Computer Vision*, Coral Gables, Florida, Nov. 1995, pp. 37–42.
[10] A. Raij, M. Pollefeys, Auto-calibration of multi-projector display walls, *Proc. of the Ninth* ***IEEE*** *International Conference on Computer Vision and Pattern Recognition* (CVPR'04), Washington, DC, USA, Vol. 1, 23-26 Aug. 2004, pp. 14 – 17.
[11] T. Okatani, K. Deguchi, Autocalibration of a projector-camera system, ***IEEE*** *Trans. on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 12, Dec. 2005 pp. 1845 – 1855.
[12] B. Zhang, Y. F. Li and Y. H. Wu, Self-Recalibration of a Structured Light System via Plane-Based Homography, *Pattern Recognition*, Vol. 40, No. 4, April 2007, pp. 1368-1377.
[13] A. Criminisi, I. Reid and A. Zisserman, Duality, rigidity and planar parallax, *Proceedings of the European Conference on Computer Vision*, Freiburg, Germany, June 2-6, 1998, pp. 846-861.
[14] H. S. Sawhney, 3D Geometry from Planar Parallax, *Proc. of the* ***IEEE*** *International Conference on Computer Vision and Pattern Recognition* (CVPR'94), June 1994, pp. 929-934.
[15] M. Irani, P. Anandan and M. Cohen, Direct recovery of planar-parallax from multiple frames, ***IEEE*** *Trans. Pattern Analysis and Machine Intelligence*, Vol. 24, No. 11, 2002, pp. 1528–1534.
[16] P. M. Griffin, L. S. Narasimhan and S. R. Yee, Generation of uniquely encoded light patterns for range data acquisition, *Pattern Recognition*, Vol. 25, No.6, 1992, pp. 609--616.
[17] Z. Zhang, A flexible new technique for camera calibration, ***IEEE*** *Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 11, 2000, pp. 1330-1334.
[18] P. Sturm and Z. Cheng, *et al*, Focal length calibration from two views: method and analysis of singular cases, *Computer Vision and Image Understanding*, Vol. 99, No. 1, 2005, pp. 58-95.
[19] X. Cao and J. Xiao, et al, Self-calibration from turn-table sequences in presence of zoom and focus, Computer Vision and Image Understanding, Vol. 102, 2006, pp. 227-237.