International Journal of Computer Vision 33(1), 73–86 (1999)

Calibrating a Structured Light Stripe System: A Novel Approach

D.Q. HUYNH
School of Information Technology, Murdoch University, Perth WA 6150 Australia
du@it.murdoch.edu.au; d.huynh@murdoch.edu.au

R.A. OWENS
Department of Computer Science, The University of Western Australia, Perth, WA 6907, Australia

P.E. HARTMANN
Department of Biochemistry, The University of Western Australia, Perth, WA 6907, Australia

Received February 27, 1998; Revised February 2, 1999; Accepted March 12, 1999

Abstract. The problem associated with calibrating a structured light stripe system is that known world points on the calibration target do not normally fall onto every light stripe plane illuminated from the projector. We present in this paper a novel calibration method that employs the invariance of the cross ratio to overcome this problem. Using 4 known non-coplanar sets of 3 collinear world points and with no prior knowledge of the perspective projection matrix of the camera, we show that world points lying on each light stripe plane can be computed. Furthermore, by incorporating the homography between the light stripe and image planes, the 4×3 image-to-world transformation matrix for each stripe plane can also be recovered. The experiments conducted suggest that this novel calibration method is robust, economical, and is applicable to many dense shape reconstruction tasks.

Keywords: calibration, projectivity, shape measurement, active stereo, triangulation

1. Introduction

Research on shape reconstruction and object recognition by projecting structured light stripes onto objects has been active since the early 1970s (Shirai and Suwa, 1971). The major difficulty involved in all structured light stripe systems is the ambiguity in identifying light stripes at regions where depth discontinuities occur. To avoid this ambiguity, most of the structured light stripe systems project one light stripe plane at a time, with either the light stripe plane panning through the scene (Shirai and Suwa, 1971; Lee and Wong, 1992) or the objects being placed on a computer controlled turntable (Sato et al., 1982; Jezouin et al., 1988). The overall time taken in such systems for reconstructing the imaged objects is therefore considerable. In an attempt to shorten the shape reconstruction time, other structured light stripe systems tackle the stripe identification problem by projecting light stripe planes encoded in some patterns, such as binary encoding (Sato et al., 1986; Alexander and Ng, 1987).

A basic structured light stripe system consists of one camera and one projector which together form an *active stereo* pair. The term *active* is commonly associated with such a system because light energy is emitted into the environment. Depending on the orientation of the structured light stripes, the projector is displaced either horizontally or vertically relative to the camera in space (see Fig. 1 for a system that projects horizontal structured light stripe planes). Unlike passive stereo which uses two cameras, a structured light stripe system generates dense world points by sampling image points on each light stripe in the image and totally avoids the so-called *correspondence problem*. The absence of

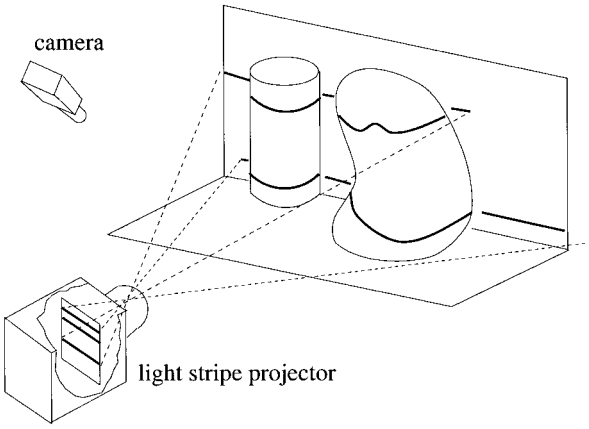


Figure 1. A structured light stripe system.

this difficult matching problem makes active stereo an attractive method for many shape measurement tasks (e.g., see (Jarvis, 1983) for a general review)—with the trade-off that the system must be fully calibrated prior to use.

The traditional approach to calibrating a structured light stripe system incorporates two separate stages: camera calibration and projector calibration.

In the camera calibration stage, the 3×4 world-to-image perspective transformation matrix P is estimated using at least 6 non-coplanar world points (relative to a world coordinate frame, say $\mathcal{F}_w$) and their corresponding image projection points (Bolles et al., 1981; Faugeras, 1993). If $\tilde{\mathbf{M}} = (M_1, M_2, M_3, 1)^\top$ is a world point and $\tilde{\mathbf{m}} = (m_1, m_2, 1)^\top$ is the corresponding image point in homogeneous coordinates then P, $\tilde{\mathbf{M}}$, and $\tilde{\mathbf{m}}$ are related by the following equation:

$$P\tilde{\mathbf{M}} = s\tilde{\mathbf{m}}, \tag{1}$$

where s is an unknown scalar. Since P is defined only up to a scale, 11 unknowns need to be estimated.

In the projector calibration stage, the coefficients of the equation of each light stripe plane relative to the same world coordinate frame $\mathcal{F}_w$ must be determined. Given that the equation of a plane is described by

$$aX_1 + bX_2 + cX_3 + d = 0, \tag{2}$$

where a, b, c, and d are the unknown coefficients, at least 3 non-collinear world points that fall onto the light stripe plane are required. While it is impractical to set up a large number of known world points for projector calibration, it is also difficult to ensure sufficient world points fall onto each light stripe plane. In the traditional calibration approach, this problem is overcome by using P obtained from the camera calibration stage. Firstly, two equations are available from (1) for each image point that lies on a given light stripe. Secondly, since calibration targets often consist of two or three mutually orthogonal surfaces whose plane equations in the world coordinate frame are known, if the image point is known to correspond to a world point that lies on one of these surfaces, then the plane equation of the surface provides one additional equation. By incorporating these 3 equations together, i.e., triangulation, the world coordinates of the given image point can be estimated. Repeating this process for a number of non-collinear world points yields an over-constrained system of equations in the form given in (2); solving for the plane coefficients is then a trivial procedure.

We propose in this paper a novel calibration method that uses 4 known non-coplanar sets of 3 collinear world points and computes, for each light stripe plane illuminated from the projector, a unique 4×3 image-to-world transformation matrix. These 12 calibration points can be either marked on a calibration target (as shown in Fig. 6) or marked on 4 non-coplanar vertical strips permanently fixed in the workspace.

The remainder of the paper is organized as follows. Section 2 presents a brief review on projective geometry. Section 3 details the proposed calibration method. Section 4 presents the completed experiments and a comparison of the traditional method and our proposed method. Section 5 presents conclusions.

2. Projective Geometry

An outline of notations is in order. We denote the known world coordinate frame by $\mathcal{F}_w$; the image plane by Π; the kth light stripe plane by Π_k; world points by bold upper case 3-vectors, e.g., $\mathbf{M} = (M_1, M_2, M_3)^\top$; world lines, each of which is internally represented as a 3-vector for the line direction and a world point that lies on it, are denoted by bold upper case letters with an overhead arrow, e.g., $\vec{\mathbf{L}}$; image points by bold lower case 2-vectors, e.g., $\mathbf{m} = (m_1, m_2)^\top$; matrices by upper case. In addition to the above notations, homogeneous or projective coordinates are denoted by adding a " ˜ " above the entities, e.g., $\tilde{\mathbf{M}} = (\mathbf{M}^\top, 1)^\top = (M_1, M_2, M_3, 1)^\top$ and $\tilde{\mathbf{m}} = (m_1, m_2, 1)^\top$ for the homogeneous coordinates of a world point and an image point. The reader is reminded that $\tilde{\mathbf{M}}$ and $\lambda\tilde{\mathbf{M}}$, for any non-zero scalar λ, denote the same point.

2.1. *Cross Ratio*

Given 4 points **p**, **q**, **r**, **m** lying on a line **l**, the number of parameterizations for **l** is infinite. One simple parameterization is to choose an arbitrary point, say **p**, as the origin and the distance between **p** and another point, say **q**, as unit length. Such a parameterization is given below:

$$(\mathbf{q}-\mathbf{p})\theta+\mathbf{p},$$

where θ is the parameter that defines points on the line. This parameterization gives $\theta_{\mathbf{p}}=0$, $\theta_{\mathbf{q}}=1$, $\theta_{\mathbf{r}}$ and $\theta_{\mathbf{m}}$ any real numbers depending on their positions relative to **p** on the line. The cross ratio r of the four collinear points **p**, **q**, **r**, and **m** (commonly referred to as {**p**, **q**; **r**, **m**}) is then defined as

$$r=\left(\frac{\theta_{\mathbf{p}}-\theta_{\mathbf{r}}}{\theta_{\mathbf{q}}-\theta_{\mathbf{r}}}\right)\Bigg/\left(\frac{\theta_{\mathbf{p}}-\theta_{\mathbf{m}}}{\theta_{\mathbf{q}}-\theta_{\mathbf{m}}}\right). \tag{3}$$

It can be verified that the cross ratio of four points on a line is independent of the choice of parameterization of the line.

Collinearity and cross ratio are known to be invariant under perspective projection (Semple and Kneebone, 1952) (see Fig. 2). This invariance property uniquely determines the coordinates of a point that is on the same line with three other points if the cross ratio of these four points is known. In particular, if 4 world points **P**, **Q**, **R**, and **M** are collinear in space then the 4 corresponding image points **p**, **q**, **r**, and **m** are also collinear and the cross ratios of the two sets of points are identical. So, given that the cross ratio is r and that the coordinates of the three world points **P**, **Q**, **R** are known, the coordinates of the fourth world point **M** immediately follows after computing the parameter $\theta_{\mathbf{M}}$ given below:

$$\theta_{\mathbf{M}}=\frac{r(\theta_{\mathbf{Q}}-\theta_{\mathbf{R}})\theta_{\mathbf{P}}-\theta_{\mathbf{Q}}(\theta_{\mathbf{P}}-\theta_{\mathbf{R}})}{r(\theta_{\mathbf{Q}}-\theta_{\mathbf{R}})-(\theta_{\mathbf{P}}-\theta_{\mathbf{R}})}. \tag{4}$$

This invariance property of the cross ratio will be applied later in the paper for determining world points that fall onto each light stripe plane.

2.2. *Homography*

In projective geometry, the term *homography* refers to the plane-to-plane transformation in the projective space (Fig. 3). Given 4 or more points $\tilde{\mathbf{m}}_i$, where $i=1,\ldots,n$ for $n\geq 4$, in general position (i.e., no three points are collinear) on a plane Π and their perspective projections $\tilde{\mathbf{m}}'_i$ onto another plane Π' under a perspective centre **C**, there exists a unique homography H, defined up to a scale, such that

$$H\tilde{\mathbf{m}}_i=\rho_i\tilde{\mathbf{m}}'_i,$$

where ρ_i's are unknown scalars. Matrix H can be solved as a linear system of equations or via the singular value decomposition of the data matrix.

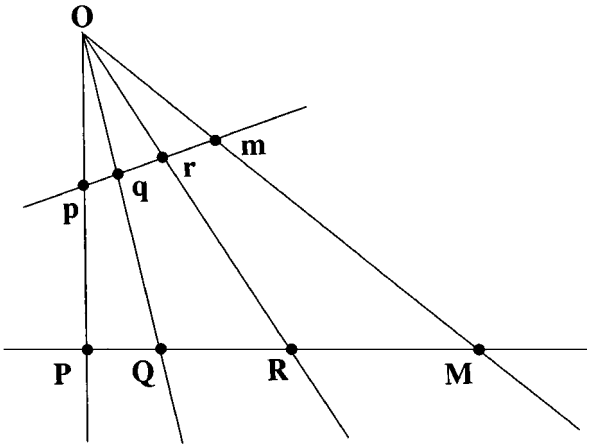


Figure 2. Invariance of collinearity and cross ratios under perspective projection. Here, {**p**, **q**; **r**, **m**} = {**P**, **Q**; **R**, **M**}.

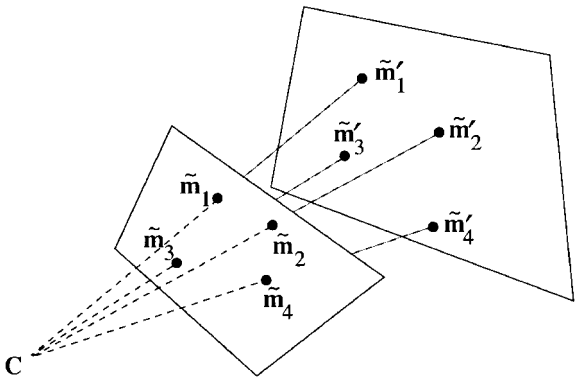


Figure 3. Plane-to-plane homography can be determined from 4 or more corresponding points in general position on two planes under a perspective centre.

3. A Novel Calibration Method

3.1. Related Work

With a laser range finder mounted at the wrist of a robot arm, Chen and Kak (1987) calibrated their vision system by moving the robot arm to several known positions. They proved that image coordinates can be directly transformed to world coordinates via a 4×3 image-to-world transformation matrix that is defined up to a scale. Since each world line provides 2 equations, to determine the 11 unknown elements of the matrix, the system requires at least 6 world lines.

Reid (1996) later extended the concept to known world plane and image point correspondences. His laser range finder was mounted on an AGV whose position relative to the world coordinate frame was known. Moreover, the transformations from sensor coordinate frame to vehicle coordinate frame, from the fixed light stripe coordinate frame to sensor coordinate frame, etc., were all known. With an appropriate choice of the stripe coordinate frame, the first row of the 4×3 image-to-stripe transformation matrix can be simplified to zero, leaving only 8 unknown elements to be solved. In his system, a number of known world planes are required.

3.2. Computing the Image-to-World Transformation

Inspired by the work of Chen and Kak (1987) and Reid (1996), we propose that the image-to-world transformation matrix for each light stripe plane illuminated from the projector can be computed from known world points. Our structured light stripe system, which consists of a camera and a projector is set up as depicted in Fig. 1. As mentioned before, 4 non-coplanar sets of 3 collinear points known relative to a world coordinate frame (Fig. 4) are required for calibrating the system.

We have shown in Section 2.1 how to compute the coordinates of an unknown world point that is collinear with 3 known world points. Here, we extend the idea to 4 known non-coplanar sets of 3 collinear world points. As shown in Fig. 4, the known collinear world points $\{\mathbf{P}_i, \mathbf{Q}_i, \mathbf{R}_i \mid i = 1, \ldots, 4\}$ do not fall onto the light stripe plane Π_k, but the 4 world points $\{\mathbf{M}_i \mid i = 1, \ldots, 4\}$ that do fall onto Π_k can be computed.

Let $\mathcal{F}_{\pi_{k2}}$ be a 2D coordinate frame on Π_k. Then, as mentioned in Section 2.2, any 4 or more points in general position defined in $\mathcal{F}_{\pi_{k2}}$ and their corresponding image projections define a unique homography H_k.

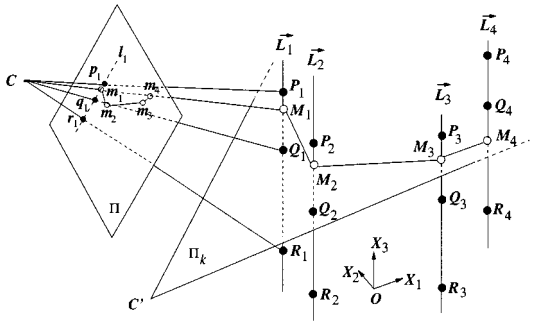


Figure 4. The proposed calibration method using 4 known non-coplanar sets of 3 collinear world points (dark circles $\{\mathbf{P}_i, \mathbf{Q}_i, \mathbf{R}_i \mid i = 1, \ldots, 4\}$). Open circles $\mathbf{M}_i$'s are unknown world points lying on Π_k; $\mathbf{C}$ and $\mathbf{C}'$ are the perspective centres of the camera and projector.

Since the homography exists regardless of the choice of the 2D coordinate frames in both projective planes, without loss of generality, $\mathcal{F}_{\pi_{k2}}$ can be arbitrarily defined. To relate the coordinate frame $\mathcal{F}_{\pi_{k2}}$, which has yet to be determined, with the world coordinate frame $\mathcal{F}_w$, we define a 3D coordinate frame $\mathcal{F}_{\pi_{k3}}$ such that the $\mathbf{X}_3$-axis of $\mathcal{F}_{\pi_{k3}}$ coincides with the normal vector $\mathbf{n}$ of Π_k (Fig. 5). It is apparent from this definition that both the $\mathbf{X}_1$- and $\mathbf{X}_2$-axes of $\mathcal{F}_{\pi_{k3}}$ would lie on Π_k, although their directions are still unknown at this stage. The arbitrary 2D coordinate frame $\mathcal{F}_{\pi_{k2}}$ is now defined as having its two axes coinciding with the $\mathbf{X}_1$- and $\mathbf{X}_2$-axes of $\mathcal{F}_{\pi_{k3}}$.

Given that the 4 world points $\mathbf{M}_i$, for $i = 1, \ldots, 4$ lying on Π_k have been determined, the following procedure shows how points can be transformed from $\mathcal{F}_w$ to $\mathcal{F}_{\pi_{k3}}$ and how the $\mathbf{X}_1$- and $\mathbf{X}_2$-axes of $\mathcal{F}_{\pi_{k3}}$ and $\mathcal{F}_{\pi_{k2}}$ are defined:

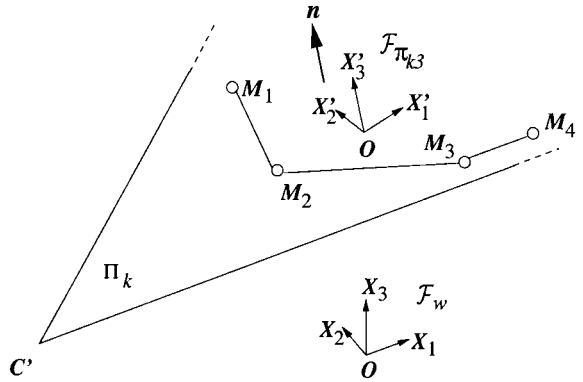


Figure 5. The relationship between $\mathcal{F}_{\pi_{k3}}$ and $\mathcal{F}_w$. The $\mathbf{X}_3$-axis of $\mathcal{F}_{\pi_{k3}}$ is parallel to the normal vector $\mathbf{n}$ of Π_k.

- compute the unit normal $\mathbf{n} = (n_1, n_2, n_3)^\top$ of Π_k using the 4 computed world points $\mathbf{M}_i$, for $i = 1, \ldots, 4$.
  In practice, all the $\mathbf{M}_i$'s may not be exactly coplanar in space, so a plane must be first fitted to the computed world points; vector $\mathbf{n}$ is then defined as the normal vector of the fitted plane.
- compute the angle ψ between $\mathbf{n}$ and the $\mathbf{X}_3$-axis ($= (0, 0, 1)^\top$) of $\mathcal{F}_w$. This is simply the inner product of the two vectors:

$$\psi = \cos^{-1}(\mathbf{n} \cdot (0, 0, 1)) = \cos^{-1}(n_3).$$

  It is straightforward from Fig. 5 that a rotation of angle ψ about an axis that is orthogonal to both $\mathbf{n}$ and the $\mathbf{X}_3$-axis of $\mathcal{F}_w$ will bring the two vectors to align. The next two steps of the procedure are therefore to determine the rotation axis $\mathbf{a}$ and then the rotation matrix R for the alignment.
- compute the unit vector $\mathbf{a}$:

$$\mathbf{a} = (\mathbf{n} \times \mathbf{X}_3)/\|\mathbf{n} \times \mathbf{X}_3\|.$$

- compute the rotation matrix R that brings $\mathbf{X}_3$ to align with $\mathbf{n}$. The matrix for this rotation can be derived as follows (e.g., see Bottema and Roth (1979) or Pavlidis (1982)):

$$R = \mathbf{a}\mathbf{a}^\top + \cos\psi(I - \mathbf{a}\mathbf{a}^\top) + \sin\psi(I \times \mathbf{a}),$$

  where I is the 3×3 identity matrix and $I \times \mathbf{a}$ is the skew-symmetric matrix formed by $\mathbf{a}$, i.e.,

$$I \times \mathbf{a} = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}.$$

  If $\mathbf{n}$ and $\mathbf{X}_3$ are parallel then $\psi = 0$ and $R = I$.
- since the origin of the coordinate frame $\mathcal{F}_{\pi_{k3}}$ can be arbitrarily chosen, without loss of generality the centroid $\bar{\mathbf{M}}$ of the four computed world points $\mathbf{M}_i$, for $i = 1, \ldots, 4$, is used as the coordinate frame origin. Our choice of the origin and the $\mathbf{X}_3$-axis of $\mathcal{F}_{\pi_{k3}}$ defines the 4×4 matrix T_k below:

$$T_k = \begin{bmatrix} R & -R\bar{\mathbf{M}} \\ \mathbf{0}^\top & 1 \end{bmatrix}, \tag{5}$$

  Ideally, matrix T_k would transform any world point $\tilde{\mathbf{M}} = (M_1, M_2, M_3, 1)^\top$ lying on Π_k to the form $\tilde{\mathbf{M}}' = (M_1', M_2', 0, 1)^\top$. That is,

$$T_k\tilde{\mathbf{M}} = \tilde{\mathbf{M}}'. \tag{6}$$

  However, it often occurs for real data that the computed $\mathbf{M}_i$'s are not perfectly coplanar in space, so the third component of each $\tilde{\mathbf{M}}_i'$ after the T_k transformation, may not be identically zero. Choosing the centroid $\bar{\mathbf{M}}$ as the origin of $\mathcal{F}_{\pi_{k3}}$ in fact helps average out these small non-zero values.
  Note that the transformation matrix T_k above has also defined the $\mathbf{X}_1$- and $\mathbf{X}_2$-axes of both $\mathcal{F}_{\pi_{k3}}$ and $\mathcal{F}_{\pi_{k2}}$.
- because the $\mathbf{X}_1$ and $\mathbf{X}_2$-axes of the coordinate frame $\mathcal{F}_{\pi_{k2}}$ are defined to be identical to those of the coordinate frame $\mathcal{F}_{\pi_{k3}}$, relative to the local 2D coordinate frame $\mathcal{F}_{\pi_{k2}}$ on Π_k, $\tilde{\mathbf{M}}'$ can be expressed in homogeneous coordinates as $\tilde{\mathbf{m}}' = (M_1', M_2', 1)^\top$. Points in $\mathcal{F}_{\pi_{k2}}$ can now be directly computed from world points lying on Π_k via the following 3×4 transformation:

$$T_{w\pi_k} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} T_k \equiv ST_k. \tag{7}$$

  Given that $\mathbf{M}_i \in \mathcal{F}_w$, $\forall i$ can be transformed to $\tilde{\mathbf{m}}_i' \in \mathcal{F}_{\pi_{k2}}$ via $T_{w\pi_k}$, the homography H_k that satisfies

$$H_k\tilde{\mathbf{m}}_i = \rho_i\tilde{\mathbf{m}}_i', \quad i = 1, \ldots, 4 \tag{8}$$

  can be computed. Here, $\tilde{\mathbf{m}}_i$'s are the intersection points of the kth light stripe with $\mathbf{l}_i$'s (Fig. 4) and ρ_i's are unknown scalars.
  Our aim now is to estimate the image-to-world transformation T_{iw_k} for every image point $\mathbf{m}$ that is known to be the projection of a world point $\mathbf{M}$ lying on Π_k. Since H_k and $T_{w\pi_k}$ have been estimated from the above procedure, given $\mathbf{m} \in \Pi$ and using (8) yields

$$H_k\tilde{\mathbf{m}} = \rho\tilde{\mathbf{m}}', \tag{9}$$

  where $\tilde{\mathbf{m}} = (\mathbf{m}^\top, 1)^\top$, $\tilde{\mathbf{m}}' \in \mathcal{F}_{\pi_{k2}}$, and ρ is a non-zero scalar.
  Suppose that $\rho\tilde{\mathbf{m}}' = (m_1', m_2', m_3')^\top$. Then relative to $\mathcal{F}_{\pi_{k3}}$, this is simply $(m_1', m_2', 0, m_3')^\top$. Let $\tilde{\mathbf{M}}'$ be this point in $\mathcal{F}_{\pi_{k3}}$ then it follows that

$$\tilde{\mathbf{M}}' = S^\top(\rho\tilde{\mathbf{m}}'),$$

  where S is defined in (7).

Since T_k is in the group of Euclidean transformations, T_k is invertible. Substituting (6) and (9) into the above gives

$$\tilde{\mathbf{M}} = T_k^{-1} S^\top H_k \tilde{\mathbf{m}}. \tag{10}$$

Define

$$T_{iw_k} = T_k^{-1} S^\top H_k. \tag{11}$$

This is the 4×3 image-to-world transformation matrix that maps any given image point lying on the kth light stripe directly to projective coordinates in $\mathcal{F}_w$. We may call it the "inverse projection matrix" or the "conversion matrix" as termed in Chen and Kak (1987).

3.3. *Summary and Discussion*

In summary, for each light stripe plane Π_k illuminated from the projector, the procedure for the proposed calibration method is as follows:

1. compute the 4 image points $\mathbf{m}_i$ that the kth light stripe intersects with $\mathbf{l}_i$, for $i = 1, \ldots, 4$ as described in Section 4.2 and the cross ratio r_i as given by (3) for each image line $\mathbf{l}_i$.
2. compute, based on the invariance of the cross ratios and (4), the 4 points $\mathbf{M}_i \in \mathcal{F}_w$ that Π_k intersects with $\vec{\mathbf{L}}_i$.
3. follow the procedure described above and compute the 4×4 transformation T_k defined in (5).
4. compute $T_{w\pi_k}$ given in (7), and then $\tilde{\mathbf{m}}'_i = T_{w\pi_k} \tilde{\mathbf{M}}_i$, for $i = 1, \ldots, 4$.
5. compute the homography H_k that maps $\tilde{\mathbf{m}}_i$ to $\tilde{\mathbf{m}}'_i$, $i = 1, \ldots, 4$.
6. compute T_{iw_k} defined in (11).

Our method is different from those given in Chen and Kak (1987) and Reid (1996) in the following respects:

- While world line to image point correspondences are used in (Chen and Kak, 1987) and world plane to image point correspondences are used in (Reid, 1996), world point to image point correspondences are used here.
- Because our method uses world point to image point correspondences, we are able to compute the world to stripe plane transformation T_k and homography H_k, and estimate the image-to-world transformation from explicit matrix multiplications. In Chen and Kak (1987) (or Reid (1996)), this matrix is estimated from world line (or world plane) to image point correspondences.

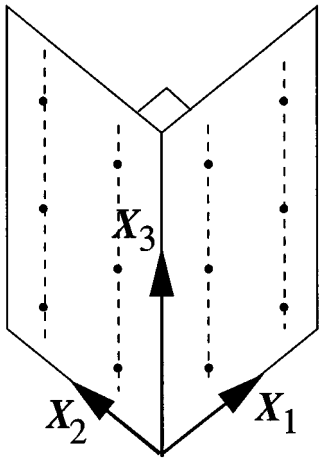


Figure 6. The calibration target and the coordinate system used for calibration.

The proposed calibration method can be easily extended to handle an arbitrarily large number of collinear triplets of points. Incorporating more collinear triplets of points into the system would yield more world points $\tilde{\mathbf{M}}_i$ for getting a better estimate of the normal vector $\mathbf{n}$ and the homography H of each light stripe plane. We have not incorporated more collinear triplets of world points into the system because of constraints on hardware. The calibration target that we used (see Fig. 6) only has 4 collinear triplets of points, which just meet the minimum requirement for the calibration method.

4. Experiments

Three of the experiments for testing the proposed calibration method are detailed here. To carry out the experiments, the output data produced by a modified version of the SHAPE measurement system as described in Alexander and Ng (1987) were used. Because the SHAPE system is protected under copyright, to completely avoid any involvement with the SHAPE software, we have written our own Matlab programs and we used only the output data of SHAPE in text format. For the completeness of this paper, the updated SHAPE system is briefly described below, and the

interested reader may refer to Alexander and Ng (1987) for details on the earlier version of the SHAPE system.

4.1. The SHAPE Measurement System

The SHAPE measurement system used in our experiments consists of a camera and a projector. The light stripe projector is a modified 35 mm slide projector with a liquid crystal light valve mounted in place. It is capable of projecting 128 horizontal light stripes, labelled 0 to 127. The calibration target that comes with the system is a heavy metal frame with two orthogonal surfaces corresponding to the $X_1 = 0$ and $X_2 = 0$ planes of $\mathcal{F}_w$ (Fig. 6). The 6 calibration points on each surface are arranged into 2 sets of 3 collinear points parallel to the $\mathbf{X}_3$-axis. These 12 points are known with high accuracy (to the order of 10^{-3} mm) and are ideal for testing the calibration method proposed in this paper.

Stripe identification in the images is prerequisite to the calibration and reconstruction processes. With the n light stripes encoded in binary, at least $\log_2(n) + 1$ images must be taken. The additional one image (with all light stripes turned on) is required so that light stripe number 0 can be identified. Stripe identification based on binary encoding of light stripe numbers works as follows: each light stripe number k has a binary representation, e.g., if $k = 3$ then the its binary representation is 00000011, having bits 0 and 1 on; in the $\log_2(n) + 1$ images taken, light stripe 3 in the projector will be therefore turned on in images 0 and 1 and off in all the other images. The overlaying of all the images taken will then enable all the light stripes to be correctly identified.

For 128 light stripes, SHAPE takes a total of $\log_2(128) + 4 = 11$ images. The four additional images in the formula are: Two images, with all light stripes turned off, are used for normalizing the intensity values across the remaining images; one image is taken with the even (bit 0 is off) stripe numbers turned on; one image with all the light stripes turned on. The total time required for taking the eleven 512×512 images was measured to be approximately 1 second.

SHAPE samples up to 128 points for each stripe detected in the image. With a total of 128 light stripes, the system can generate up to 16,384 data points for dense shape measurement applications. Each data point contains information such as the stripe number to which the data point belongs, the image coordinates, and the computed world coordinates.

4.2. Implementation and Results

To conduct our experiments, we input into Matlab the file containing the data points generated by SHAPE. Only the stripe number and the image coordinates of each data point were used. The world points from SHAPE were involved at the final stage for comparing our calibration method with the traditional structured light stripe calibration method.

An image of the calibration target, placed at approximately 1.5 m from the camera and illuminated by all the light stripe planes, is shown in Fig. 7. The coordinates of the 12 known world points $\{\mathbf{P}_i, \mathbf{Q}_i, \mathbf{R}_i \mid i = 1, \ldots, 4\}$ for calibration were firstly loaded into Matlab, the parameter θ (see Section 2.1) of each point was then estimated. The image coordinates of these world points were manually detected, the fitting of the image lines $\mathbf{l}_i, i = 1, \ldots, 4$ then followed.

To determine the intersection point of a light stripe with each $\mathbf{l}_i$ (the open circles $\mathbf{m}_i, i = 1, \ldots, 4$ shown in Fig. 4), a neighbourhood of 20 pixels on either side of each $\mathbf{l}_i$ was specified. Image points on the stripe falling into this neighbourhood were automatically selected, an image line $\mathbf{l}$ was then fitted to these points, the intersection point was finally computed by solving $\mathbf{l}_i$ and $\mathbf{l}$ simultaneously. On average, 6 image points were selected for this neighbourhood size. Note that this method of computing each $\tilde{\mathbf{m}}_i$ assumes that the corresponding $\tilde{\mathbf{M}}_i$ lies on a surface that is locally planar. If a different calibration target is used that violates this assumption then a different approach to computing each $\tilde{\mathbf{m}}_i$ will have to be taken.

After this, the world points $\mathbf{M}_i, \forall i$ were computed from the cross ratios. Following the summary of the calibration method given in the previous section, the image-to-world T_{iw_k} matrix for the kth stripe was computed. The final procedure was to transform all image points on each light stripe to projective coordinates using $T_{iw_k}, \forall k$.

The data file from SHAPE in the first experiment contains a total of 10,387 points, ranging from stripe 11 to stripe 127. Our method gave a total of 9897 points, ranging from stripe 20 to stripe 109. The reason that our method missed the first 9 and the last 18 stripes is due to the intersection point, $\mathbf{m}_i$ for some i, for these stripes being outside the image boundary and so the T_{iw_k} matrices for these light stripes could not be determined.

Since the computed world points are too dense for visualization, only alternate world points from alternate

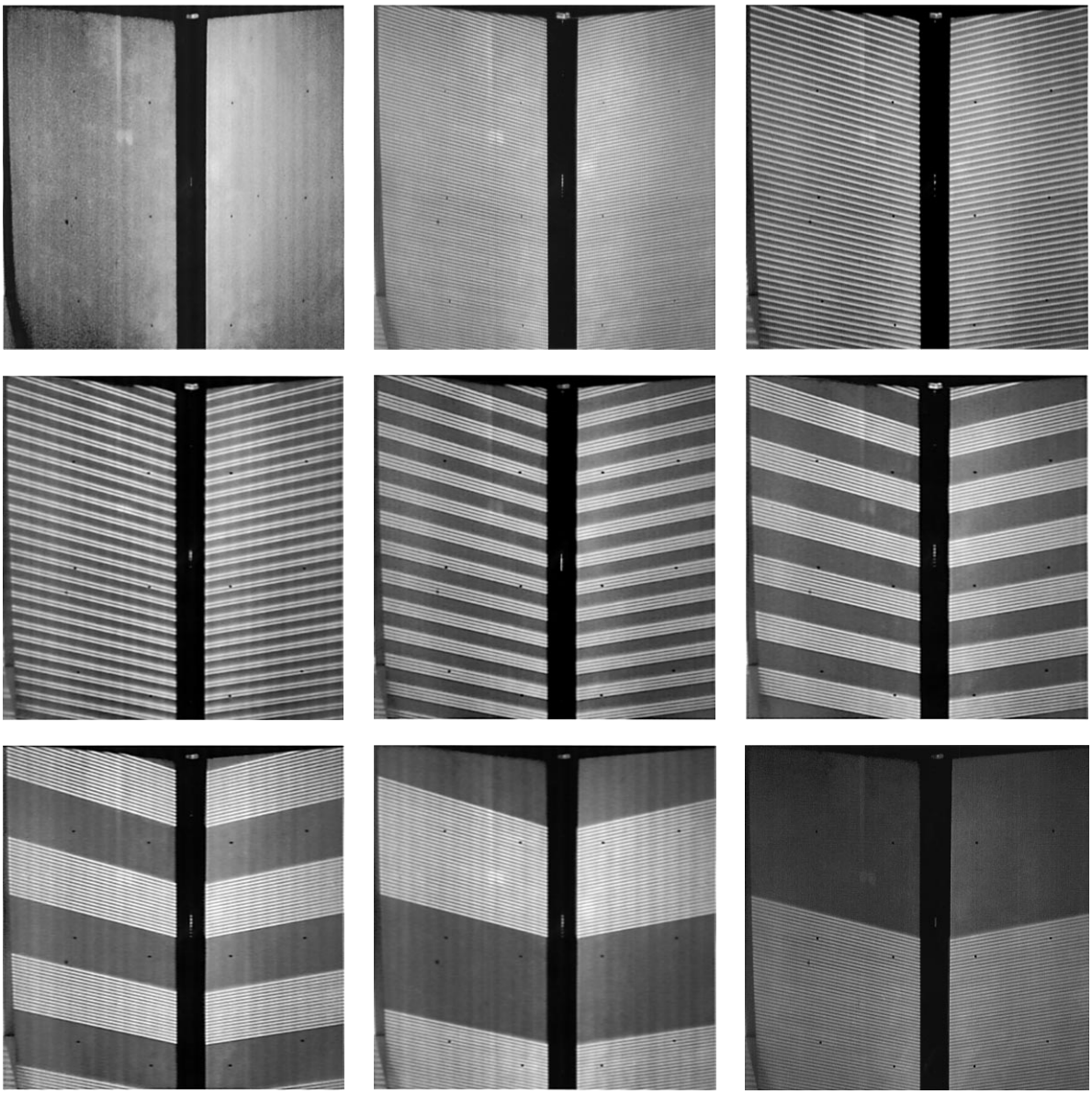


Figure 7. Light stripes are encoded in binary. Nine of the eleven images taken are shown here. First row from left to right: all stripes off, all stripes on, odd stripes on. Second row: two consecutive stripes on, four consecutive stripes on, eight consecutive stripes on. Third row: sixteen consecutive stripes on, thirty-two consecutive stripes on, sixty-four consecutive stripes on.

stripes are selected and displayed in Fig. 8. The plan view shows a very good coplanarity of the computed points on each surface of the calibration target; points on both surfaces also form a good right angle.

We should note that lens distortion has not been considered in our calibration method. However, for the type of lens used in our experiments, lens distortion has been minimized to a negligible level (observe the straightness of the line stripes in Fig. 7). The location of image points on each stripe is also a major factor that determines how well world points can be computed. In the SHAPE system, the centred line of each line stripe is detected to sub-pixel accuracy, so the image coordinates that we used in the experiments have been accurately determined.

SHAPE uses the Direct Linear Transform (DLT) described in Section 1 to calibrate the camera and projector and computes world points by triangulation.

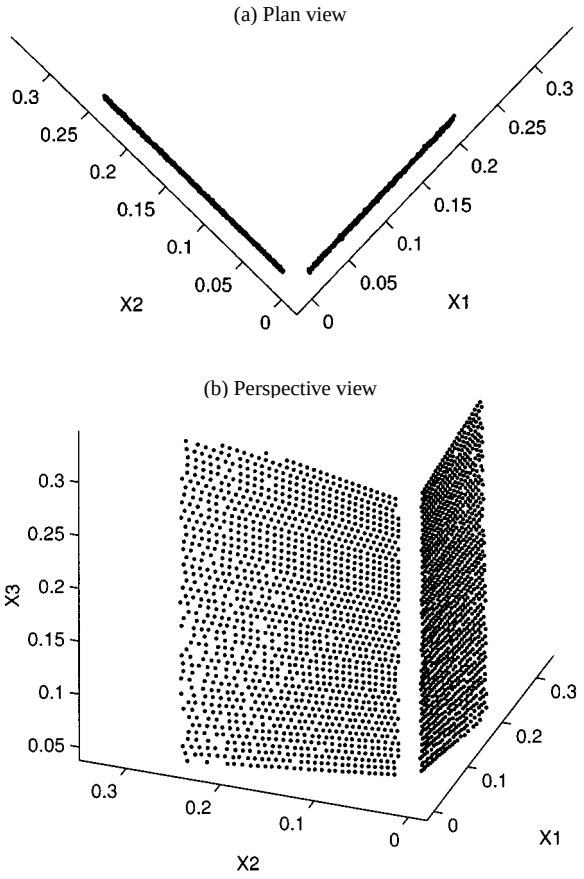


Figure 8. World points computed by our calibration method. The unit of each axis is metre.

The availability of the SHAPE world points makes it convenient for us to assess our calibration method. We compared the coordinates of world points obtained from the two methods by computing the Euclidean distance between each pair of corresponding world points produced by both methods. In the first experiment, the computed mean and standard deviation of all these distances were 1.136 and 1.181 mm. These small values confirm that the reconstructed world points using our calibration method are as good as those estimated from the traditional DLT and triangulation.

Figure 9 shows images of a mannequin, in our second experiment, with the light stripe patterns superimposed. Using the computed T_{iw_k} matrices from the previous experiment for stripes 20 to 109, the reconstructed world points are displayed in Fig. 10. In this experiment, the numbers of data points from SHAPE (stripes 33 to 127) and our method (stripes 33 to 109) were 6559 and 5557, respectively. The mean and standard deviation of the Euclidean distances of corresponding world points were computed to be 5.813 and 5.823 mm. Note that on the background plane of the mannequin there were more than 2000 world points whose image coordinates were not reliably detected due to the insufficient sharpness of the light stripes. The errors in these image coordinates rendered both the DLT and our methods to produce less accurate world points for the background plane. If these background world points were removed from the comparison procedure then, with the total number of common foreground points being 3190, the mean and standard deviation were estimated as 2.211 and 2.173 mm.

Our third experiment involves the reconstruction of a fan. Figure 11 shows an image of a fan with all the light stripes superimposed and Fig. 12 displays its reconstruction using our method. The original data from SHAPE contains 9187 points, ranging from stripes 25 to 125. Our method gave 7806 points, ranging from stripes 25 to 109. The mean and standard deviation for the corresponding world points produced by both methods were computed to be 3.415 and 4.491 mm. If background world points were excluded then the mean and standard deviation reduced to 1.778 and 1.220 mm.

The reconstructions from SHAPE for the three experiments show no visible differences from those figures displayed above. In fact, an examination of experiment 1 reveals that the world points from our proposed method have a better fit onto both surfaces of the calibration target, i.e., the absolute values of the X_1- or X_2-components are smaller. The small values of the computed means and standard deviations in the three experiments above show that our method performs as well as the traditional DLT.

4.3. A Note on the Cross Ratio and the Computation of $\theta_{\mathbf{M}}$

Let P, Q, R, M be four independent random variables having a uniform distribution over the interval $[a, b]$. Let Z and V be two random variables defined as

$$Z = (P - R)(Q - M)$$
$$V = (Q - R)(P - M).$$

Then $\mu_Z = \mathcal{E}(Z) = 0$ and $\mu_V = \mathcal{E}(V) = 0$, where $\mathcal{E}(.)$ denotes the expected value of the random variable concerned. Applying the Taylor series expansion on Z/V about the mean (μ_Z, μ_V) and then taking the

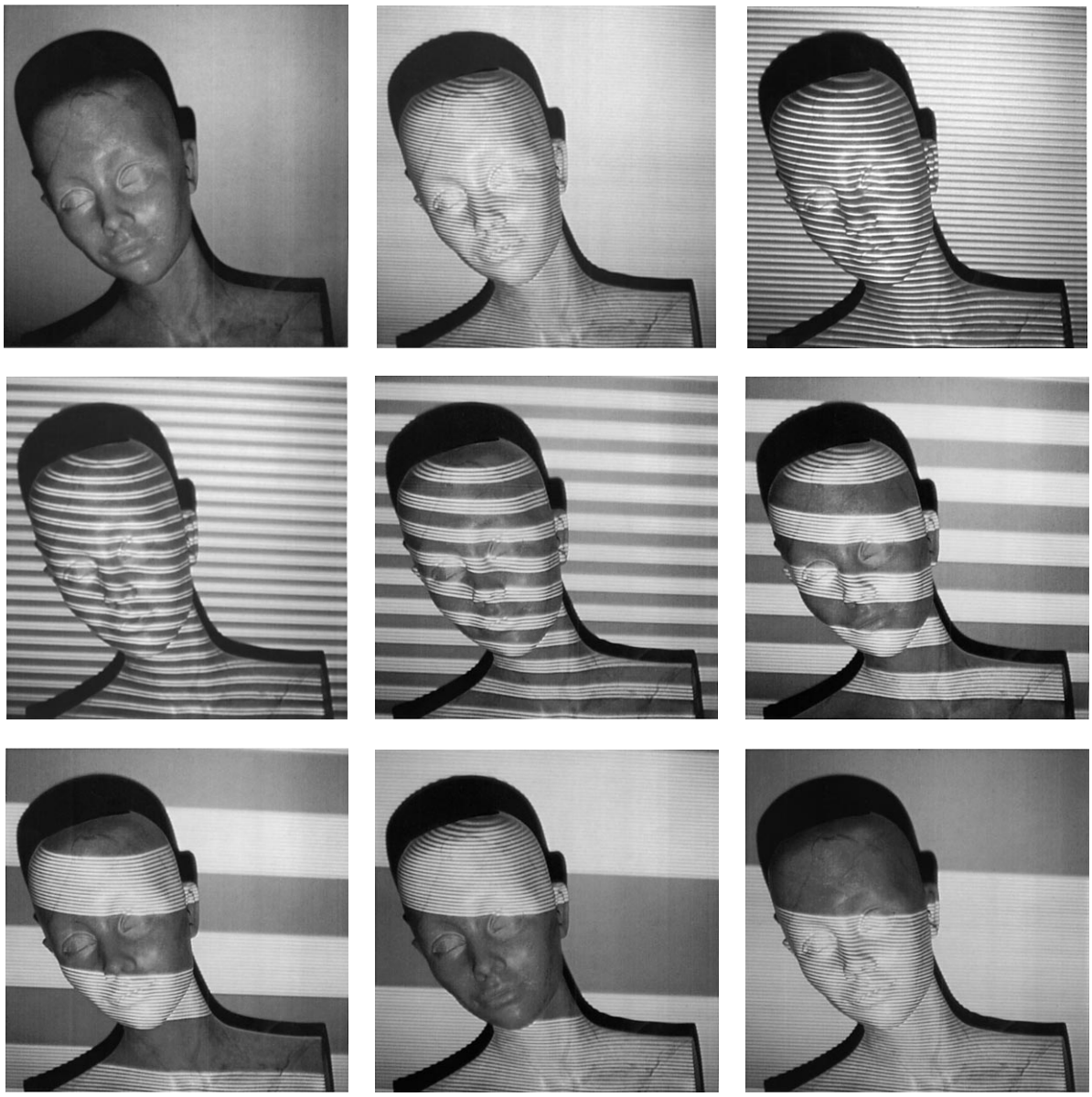


Figure 9. Nine of the eleven images of the mannequin. See Fig. 7 for explanations.

expectation yields (Mood et al., 1974)

$$\mathcal{E}(Z/V) \approx \frac{\mu_Z}{\mu_V} - \frac{\mathrm{Cov}(Z,V)}{\mu_V^2} + \frac{\mu_Z \mathrm{Var}(V)}{\mu_V^3}$$

where Cov(.,.) and Var(.) denote the covariance and the variance respectively. At $\mu_Z = \mu_V = 0$, the Taylor series does not exist and $\mathcal{E}(Z/V)$ is undefined. Thus, if one considers the four collinear points **P**, **Q**, **R**, **M**, whose parameter values $\theta_{\mathbf{P}}$, $\theta_{\mathbf{Q}}$, $\theta_{\mathbf{R}}$, $\theta_{\mathbf{M}}$ on the line are described by the random variables P, Q, R, M above, then the expected value of the random variable for the cross ratio, Z/V, is undefined. However, if at least 3 points (say **P**, **Q**, and **R**) on the line are known to be distinct then the intervals for the random variables P, Q, R, M can be modified to $[a, b)$, $[b, c)$, $[c, d)$, and $[a, d]$, where $a < b < c < d$. This gives $\mu_Z = \mathcal{E}(Z) = \frac{1}{4}((d-b)^2 - (c-a)^2)$ and $\mu_V = \mathcal{E}(V) = \frac{1}{4}(b-d)^2$, where $\mu_Z = 0$ if and only if $d - b = \pm(c - a)$ and $\mu_V = 0$ if and only if $b = d$.

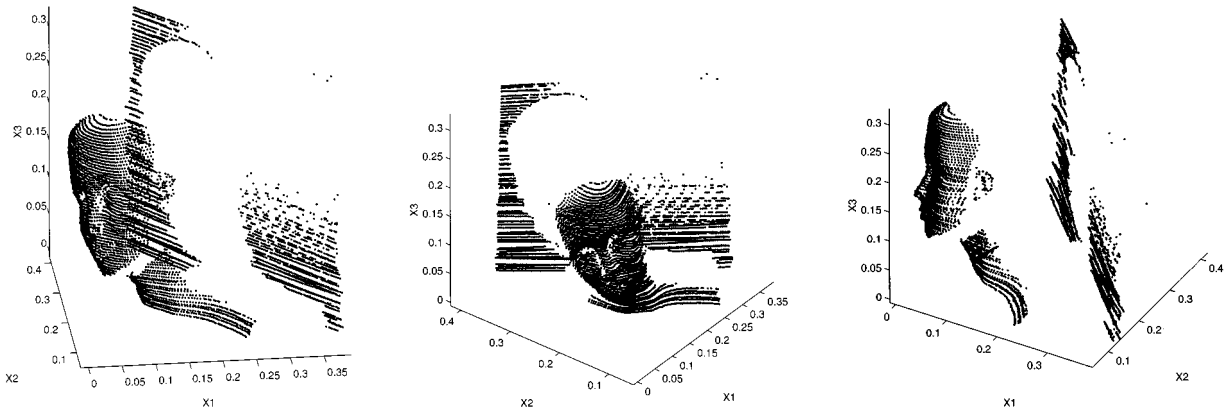


Figure 10. Reconstruction of the mannequin using the proposed calibration method.

In practice, when computing the cross ratio r, there are four cases that one may consider (see Section 2.1):

- *Case 1*: $\theta_{\mathbf{p}} = \theta_{\mathbf{r}}$, which results in $r = 0$;
- *Case 2*: $\theta_{\mathbf{q}} = \theta_{\mathbf{m}}$, which also results in $r = 0$;
- *Case 3*: $\theta_{\mathbf{q}} = \theta_{\mathbf{r}}$, which results in $r = -\infty$ or $r = +\infty$;
- *Case 4*: $\theta_{\mathbf{p}} = \theta_{\mathbf{m}}$, which also results in an infinite value of r.

Since the points $\mathbf{P}_i$, $\mathbf{Q}_i$, $\mathbf{R}_i$, for $i = 1, \ldots, 4$ on the calibration target are distinct in our experiments, Cases 1 and 3 are eliminated. If the image point $\mathbf{q}_i$ happens to fall onto a light stripe or in its vicinity then the cross ratio r estimated from the image for that light stripe will be zero or insignificant. The estimated $\mathbf{M}$ will be identical or close to $\mathbf{Q}$ as desired, as

$$\lim_{r\to 0} \theta_{\mathbf{M}} = \lim_{r\to 0} \frac{r(\theta_{\mathbf{Q}} - \theta_{\mathbf{R}})\theta_{\mathbf{P}} - \theta_{\mathbf{Q}}(\theta_{\mathbf{P}} - \theta_{\mathbf{R}})}{r(\theta_{\mathbf{Q}} - \theta_{\mathbf{R}}) - (\theta_{\mathbf{P}} - \theta_{\mathbf{R}})} = \theta_{\mathbf{Q}}.$$

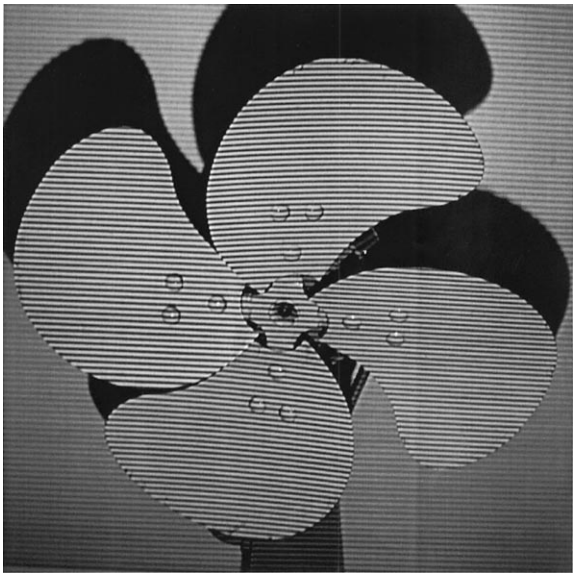


Figure 11. Image of a fan with all light stripes superimposed.

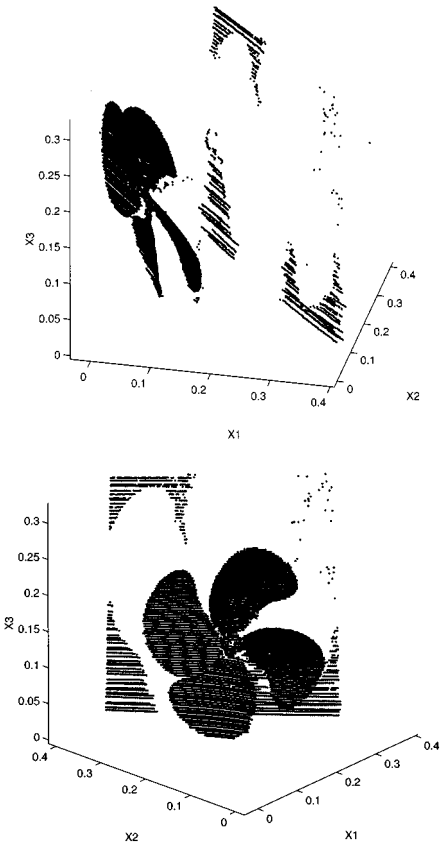
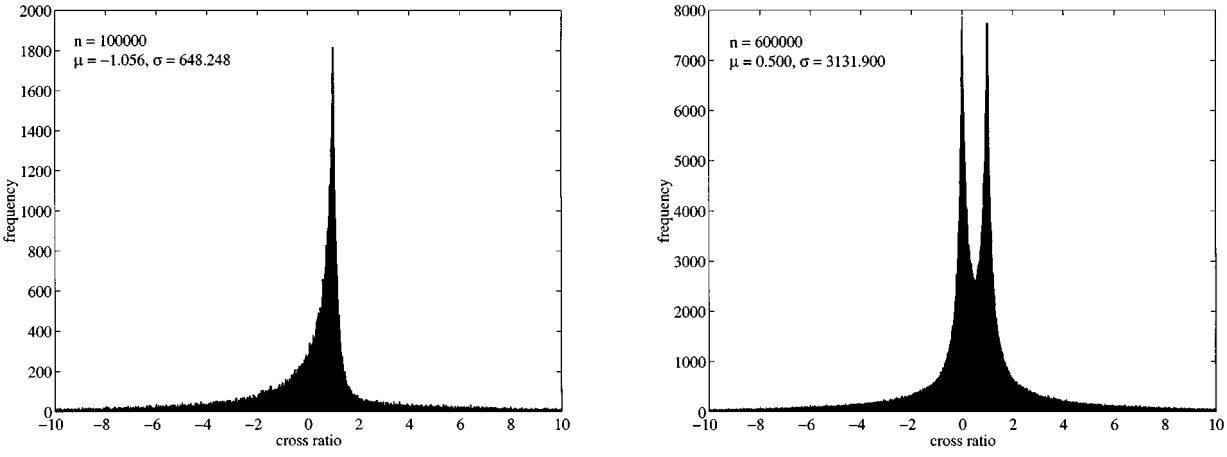


Figure 12. Reconstruction of the fan image using the proposed calibration method.



(a) (b)

Figure 13. (a) Histogram of 100,000 cross ratios computed from a simulation. (b) Histogram of the 600,000 cross ratios generated from (a) when the cross ratios for all the permutations of each tetrad are considered.

However, if the image point $\mathbf{p}_i$ falls onto a light stripe then the cross ratio r computed from the image will be very large or even infinite. Although

$$\lim_{r \to \pm\infty} \theta_{\mathbf{M}} = \theta_{\mathbf{P}}$$

and so $\mathbf{M}$ will approach $\mathbf{P}$ as required, to avoid the sensitivity in the computation of r, special care must be taken.

The cases where r is infinite can be avoided in practice by permuting the parameters in the tetrad. That is, instead of computing $\{\mathbf{p}, \mathbf{q}; \mathbf{r}, \mathbf{m}\}$, one may swap the positions of $\mathbf{p}$ and $\mathbf{q}$ and compute $\{\mathbf{q}, \mathbf{p}; \mathbf{r}, \mathbf{m}\}$. It is easy to verify that this gives a new cross ratio $r = 0$, which remains invariant under perspective projection since the positions of the corresponding world points $\mathbf{P}$ and $\mathbf{Q}$ are also swapped. Alternatively, swapping the positions of $\mathbf{p}$ and $\mathbf{r}$ will give $r = \{\mathbf{r}, \mathbf{q}; \mathbf{p}, \mathbf{m}\} = 1$. In general, the 24 permutations for each tetrad are known to yield 6 different cross ratios: r, $\frac{1}{r}$, $1-r$, $\frac{1}{1-r}$, $1-\frac{1}{r}$, $\frac{r}{r-1}$. If any two parameter values are identical in the tetrad then the number of different cross ratios is reduced to 3, namely 0, 1, and ∞ (Semple and Kneebone, 1952). In view of this, one can always restrict the cross ratio to within the interval $[-1, 1]$ by selecting an appropriate permutation for the parameters in each tetrad.

Figure 13(a) shows the histogram of the cross ratios computed for 100,000 randomly generated tetrads. The number of bins used for plotting the histogram is 1000. The parameters $\theta_{\mathbf{p}}$, $\theta_{\mathbf{q}}$, $\theta_{\mathbf{r}}$, $\theta_{\mathbf{m}}$ for the tetrads were independent, uniform random variables over the intervals $[-20, 0)$, $[0, 20)$, $[20, 40)$, and $[-20, 40]$, respectively. We note that for these chosen intervals, $\mu_Z = 0$ and $\mu_V = 400$. If all the six cross ratio values for each tetrad are taken into consideration, then the histogram will have the shape shown in Fig. 13(b) (see also the following paragraph). Figure 14 shows the new histogram of the same data in Fig. 13(a) by inverting all the cross ratios that are outside the interval $[-1, 1]$. The standard deviation σ was reduced to 0.532.

The probability density functions (pdf) of the cross ratio have been derived by Åström and Morin (1992) with the assumption that the four collinear points are independent random variables of identical uniform

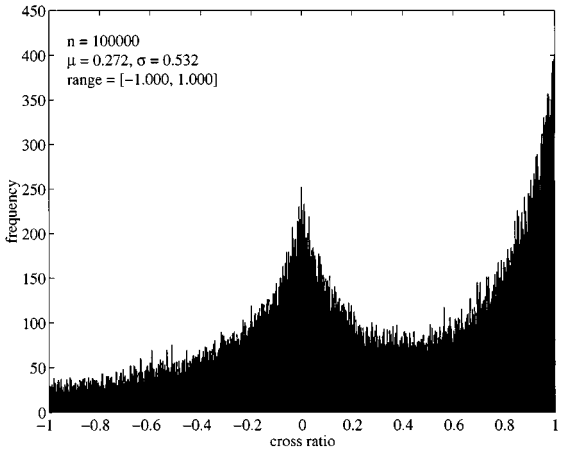


Figure 14. Histogram of the 100,000 cross ratios in Fig. 13(a) after restricting all the cross ratios to the interval $[-1, 1]$.

distribution, and by Maybank (1995) with the assumption that the independent random variables have identical Gaussian distribution. The pdf's reported in both papers have the same shape and are symmetric about the cross ratio $r = 0.5$ when all the permutations of tetrads are considered in the derivation. Since the main theme of this paper is on calibrating a structured light stripe system, only one cross ratio value was computed for each tetrad of collinear points detected in the image. If all the permutations of each tetrad are considered, then the histogram of the cross ratio, as shown in Fig. 13(b), corresponds well with the pdf's reported in Åström and Morin (1992) and Maybank (1995). However, if the four collinear points are distinct and if the permutations of tetrads are considered then the pdf of the cross ratio is symmetric about 0.5 but has 6 different peaks. Because of the scope of the paper, we will not discuss this analysis further.

4.4. Complexity Analysis

Analysis on the complexity of both methods indicates that the proposed calibration method is computationally cheaper in the shape reconstruction stage than the traditional calibration method.

To reconstruct each world point, our method needs to perform 12 multiplications to get the projective coordinates, this is then followed by 3 divisions (same complexity as multiplications assumed) to convert to inhomogeneous coordinates. So in total, 15 multiplications are required.

For the traditional method, after the recovery of the 3×4 perspective transformation matrix and the plane equation of each light stripe, the estimation of each world point requires the inversion of the 3×3 matrix in the following equation:

$$\begin{bmatrix} P_{11} - P_{31}x_1 & P_{12} - P_{32}x_1 & P_{13} - P_{33}x_1 \\ P_{21} - P_{31}x_2 & P_{22} - P_{32}x_2 & P_{23} - P_{33}x_2 \\ a_k & b_k & c_k \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} x_1 - P_{14} \\ x_2 - P_{24} \\ -d_k \end{bmatrix} \tag{12}$$

where P_{ij} is the ijth element of matrix P as described in (1), $(x_1, x_2)^\top$ is a detected image point on the kth light stripe, and a_k, b_k, c_k, d_k are the coefficients of the kth light stripe plane.

It is known that the most economical procedure to carry out from here is not to compute the matrix inverse but to apply Gaussian Elimination with partial pivoting and to think in terms of solving a system of linear equations (Golub and Van Loan, 1996) (p. 121). To find out the complexity of the traditional triangulation method for reconstructing a world point, we estimate that 6 multiplications are first required to construct the matrix in (12). Next, 8 multiplications (4 multiplications for each of the last two rows) are required to transform the matrix into the form

$$\begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & \times & \times \end{bmatrix}$$

and 4 multiplications are required to update the 3-vector on the right hand side of (12). To further convert the matrix into upper triangular form and to update the 3-vector accordingly, 4 additional multiplications are required. The number of multiplications required so far is already summed to 22 while the number of multiplications involved in the back substitution step for estimating the world point $(X_1, X_2, X_3)^\top$ has not yet been considered.

It is clear from the above analysis that the traditional triangulation method is more costly than our calibration method for world point reconstruction.

5. Conclusions

We have presented a novel calibration method for a structured light stripe system. The method uses 4 known non-coplanar sets of 3 collinear world points and, based on the invariance of cross ratios under perspective projection, computes world points that fall onto each light stripe plane illuminated from the projector. The computed world points and the corresponding image points determine a unique homography between the image plane and each light stripe plane. Finally, for each light stripe plane, the 4×3 image-to-world transformation is recovered from matrix multiplications. Experiments conducted for the proposed calibration method suggest that it is robust, economical, and can be applied to many shape measurement tasks.

Acknowledgments

We thank the anonymous referees for their comments and for their suggested improvements to the paper. This research was in part supported by a Japan Society for the Promotion of Science (JSPS) Fellowship ID No. P98162.

An earlier version of the article has been presented in (Huynh, 1997).

References

Alexander, B.F. and Ng, K.C. 1987. 3D Shape measurement by active triangulation using an array of coded light stripes. *SPIE: Optics, Illumination and Image Sensing for Machine Vision II*, 850:199–209.

Åström, K. and Morin, L. 1992. Random cross ratios. Rapport Technique RT88 IMAG-14, LIFIA, Institut Imag, Grenoble, France.

Bolles, R.C., Kremers, J.H., and Cain, R.A. 1981. A simple sensor to gather three-dimensional data. Tech. Report 249, SRI International, Stanford University.

Bottema, O. and Roth, B. 1979. *Theoretical Kinematics*. North-Holland: Amsterdam.

Chen, C.H. and Kak, A.C. 1987. Modeling and calibration of a structured light scanner for 3D robot vision. In *Proc. IEEE Conf. Robotics and Automation*, Vol. 2, pp. 807–815.

Faugeras, O. 1993. *Three-Dimensional Computer Vision—A Geometric Viewpoint*. The MIT Press: Cambridge.

Golub, G.H. and Van Loan, C.F. 1996. *Matrix Computations*. Johns Hopkins series in the mathematical sciences, 3rd edition, The Johns Hopkins University Press: Baltimore and London.

Huynh, D.Q. 1997. Calibration of a structured light system: A projective approach. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Puerto Rico, pp. 225–230.

Jarvis, R.A. 1983. A perspective on range finding techniques for computer vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 5(2):122–139.

Jezouin, J.L., Saint-Marc, P., and Medioni, G. 1988. Building an accurate range finder with off the shelf components. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 195–200.

Lee, T. and Wong, S.B. 1992. An active triangular range finder and reflectance sensing using scanning mirror. In *ICARCV'92*, Vol. 1, pp. CV–5.6.1–CV–5.6.5.

Maybank, S.J. 1995. Probabilistic analysis of the application of the cross ratio to model based vision. *International Journal of Computer Vision*, 16(1):5–33.

Mood, A.M., Graybill, F.A., and Boes, D.C. 1974. *Introduction to the Theory of Statistics*. 3rd edition, McGraw-Hill.

Pavlidis, T. 1982. *Algorithms for Graphics and Image Processing*. Springer-Verlag: Berlin-Heidelberg.

Reid, I.D. 1996. Projective calibration of a laser-stripe range finder. *Image and Vision Computing*, 14(9):659–666.

Sato, Y., Kitagawa, H., and Fujita, H. 1982. Shape measurement of curved objects using multiple slit-ray projections. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 4(6):641–646.

Sato, K., Yamamoto, H., and Inokuchi, S. 1986. Tuned range finder for high precision 3D data. In *Proc. International Conference on Pattern Recognition*, Paris, France, pp. 1168–1171.

Semple, J.G. and Kneebone, G.T. 1952. *Algebraic Projective Geometry*. Oxford University Press.

Shirai, Y. and Suwa, M. 1971. Recognition of polyhedrons with a range finder. In *2nd Int. Joint Conf. on AI*, London, pp. 80–87.