ELSEVIER

Image and Vision Computing 16 (1998) 99–110

image AND VISION COMPUTING

Accurate 3D measurement using a structured light system

R.J. Valkenburg*, A.M. McIvor

Industrial Research Limited, P.O. Box 2225, Auckland, New Zealand

Received 3 July 1996; revised 18 March 1997; accepted 22 April 1997

Abstract

This paper discusses a method for obtaining accurate 3D measurements using a temporally encoded structured light system. An objective of the work was to have a balance in the accuracy of all components in the system. This was achieved by including lens distortion in the models for both the camera and projector which comprise the structured light system. In addition, substripe estimation was used to estimate projector stripe values as a complement to subpixel estimators used for locating image features. Experimental evaluation shows that it is important to use substripe estimation and incorporate lens distortion in the projector model. © 1998 Elsevier Science B.V.

Keywords: Structured light; Triangulation; Spatial intersection; Calibration; Subpixel; Substripe

1. Introduction

Structured light is one of a variety of methods for acquiring 3D information about the world. This paper describes a method for obtaining accurate 3D measurements using a structured light system (SLS).

Structured light is a general concept and there are a number of ways of exploiting it to obtain 3D information e.g. [1–7]. As such, a brief description of the specific SLS considered in this paper follows. Refer to Fig. 1. The SLS consists of a projector together with a camera. The projector projects a coded stripe pattern on the scene and the camera captures an image. Hence for each visible point in the world there is a corresponding stripe number (stripe value) and image location (pixel coordinates). Given the parameters of the SLS each stripe value defines a plane in the world and each pixel defines a ray in the world. The intersection of this ray and plane defines a unique 3D location in the world. The parameters of the SLS are obtained during system calibration. This is achieved by presenting a calibration reference in the field of view which features a set of fiducial marks whose spatial location is known to high accuracy. The SLS is operated normally and the pixel coordinates and stripe value for each of the fiducial marks are obtained. These triples of reference, pixel and stripe coordinates are used to estimate the unknown parameters of a mathematical model of the SLS. In this way calibration can be regarded as a parameter estimation problem.

In practice the projector projects a sequence of nine banded light patterns (actually nine differential pairs to make the system more robust to varying background illumination and surface reflectivity) on the scene controlled by a liquid crystal shutter (LCS) as shown in Fig. 2. The first pattern is with the LCS fully open (full illumination). Each of the remaining patterns represents a bit in an 8-bit binary grey code for the stripe value (often referred to as temporal encoding) [3,4]. In general, for a b-bit code, the kth bit ($k = 0 \ldots b - 1$) of the code for the nth value ($n = 0 \ldots 2^b - 1$) is given by $\lfloor (1/2)((n/2^k) + 1) \rfloor \bmod 2$. This temporal encoding scheme enables 256 stripes to be differentiated in the scene. An important property of the code is that no two edges (level changes) are coincident throughout the sequence. As the most likely place to erroneously read a bit is at a level change the code helps reduce the possibility of detecting more than 1 bit of the code erroneously. In addition, detecting the bit incorrectly at a level change will only affect the value of the stripe by one. It should be noted that this stripe encoding completely eliminates the need for resolving any correspondence between points in the image and stripes.

The above description is a simplification. Projector lens distortion results in each stripe value corresponding to a (slightly) curved surface in the world. Similarly, camera lens distortion results in the image point associated with a ray slightly displaced from its nominal location. Furthermore, the discrete nature of the CCD and LCS gives rise to uncertainty in the true pixel coordinates and stripe value.

Experience with camera calibration suggests that addressing these issues will improve the accuracy of 3D

* Corresponding author. Tel.: +64 9 3034116; fax: +64 9 3028196; e-mail: b.valkenburg@irl.cri.nz

PII S0262-8856(97)00053-X

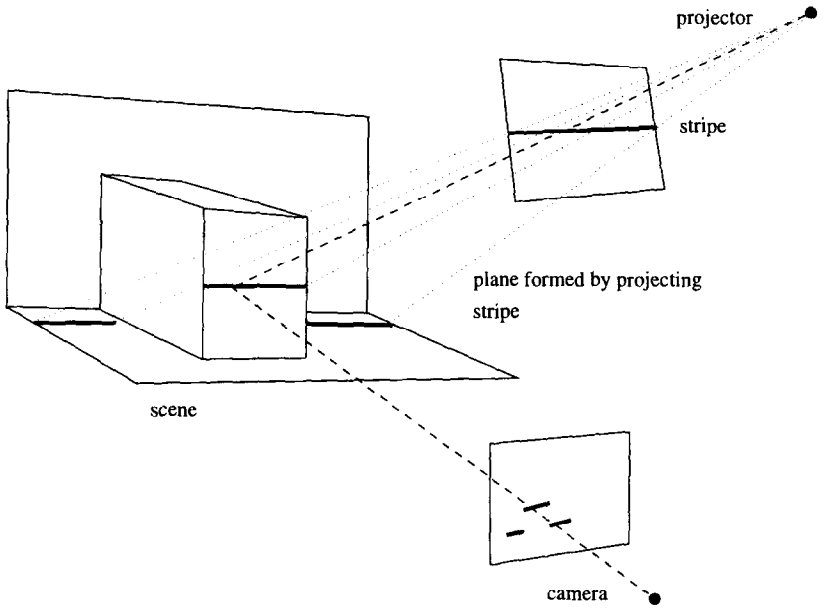

Fig. 1. Idealised structured light system.

data obtained with the SLS. The first problem is addressed by including lens distortion into the models of the camera and projector. The introduction of lens distortion into the projector model is more difficult than for the camera because distortion is a 2D phenomenon but only one coordinate of the position of a point on the projector emitter is available. The second problem is addressed through the use of subpixel and substripe operators.

The format for the remainder of this paper is as follows. In Section 2 models for the camera and projector are introduced. The algorithms used for subpixel and substripe estimation are described in Section 3. These algorithms do not take perspective distortion into account and the effect of not doing so is analysed. The method for system calibration is discussed in Section 4 and the procedure used for spatial intersection is described in Section 5. In Section 6 experiments are described and results presented to establish the performance of the system. Finally, in Section 7 the paper is

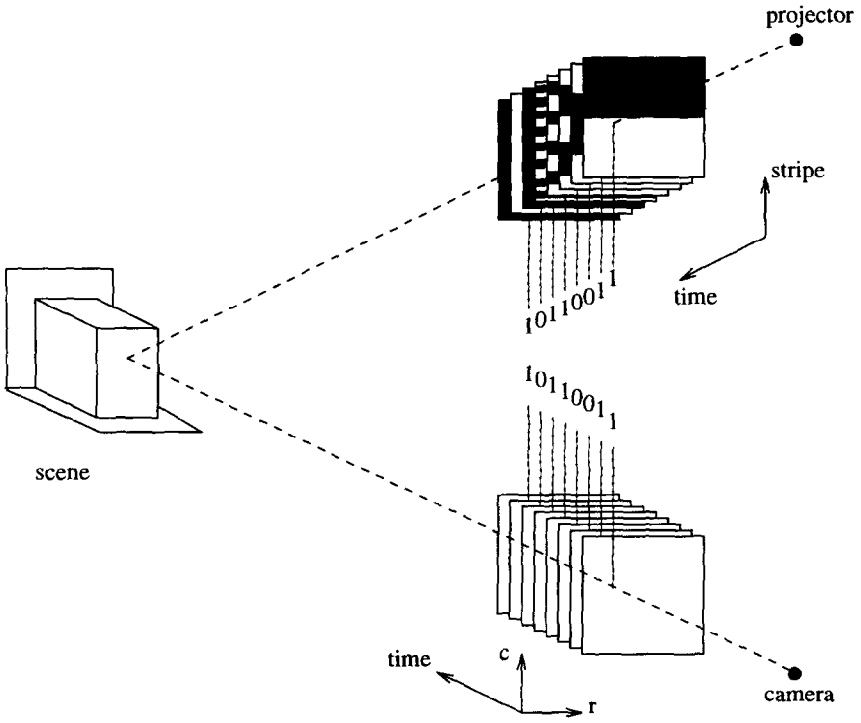

Fig. 2. Temporal encoding scheme.

concluded with a discussion regarding possible improvements to the system.

2. Model of SLS

This section describes models for the CCD camera and projector which comprise the SLS.

2.1. Camera model

Refer to Fig. 3. Let P be a point in the world, $\mathbf{C}_w$ a *world coordinate system* and $(XYZ)^t$ be the coordinates of P in $\mathbf{C}_w$. Define the *camera coordinate system*, $\mathbf{C}_c$, to have its W-axis parallel with the *optical axis* of the camera lens, its U-axis parallel with the u-axis of $\mathbf{C}_i$ (defined later) and origin located at the *perspective centre*. Let $(UVW)^t$ be the coordinates of P in $\mathbf{C}_c$. The coordinates $(UVW)^t$ are related to $(XYZ)^t$ by a rigid body coordinate transformation,

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = R_c \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + T_c \tag{1}$$

where R_c is a 3×3 rotation matrix and T_c is a 3×1 translation vector.

The *principal point* is the intersection of the imaging plane with the optical axis. Define the 2D *image coordinate* system, $\mathbf{C}_i$, to be in the image plane with its origin located at the principal point, u-axis in the fast scan direction and v-axis in the slow scan direction of the CCD. Let p be the projection of P onto the image plane and let $(\overline{uv})^t$ be the coordinates of p in $\mathbf{C}_i$. Then $(\overline{uv})^t$ are given by,

$$\begin{pmatrix} \overline{u} \\ \overline{v} \end{pmatrix} = \frac{f_c}{W} \begin{pmatrix} U \\ V \end{pmatrix} \tag{2}$$

where f_c is the principal distance.

Unfortunately the lens does not provide a perfect perspective projection. Radial lens distortion is incorporated into the model in the following way. Let $(uv)^t$ be the actual observed image point after being subject to lens distortion. Then $(uv)^t$ is related to $(\overline{uv})^t$ by,

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \overline{u} \\ \overline{v} \end{pmatrix} [1 + K_c(\overline{u}^2 + \overline{v}^2)] \tag{3}$$

where K_c is a coefficient which controls the amount of radial distortion.

Finally, it is necessary to model the image sampling performed by the CCD and framegrabber. Let C_p be the *pixel coordinate system* associated with the digital image. The pixel coordinates are related to the image coordinates by,

$$\begin{pmatrix} x_p \\ y_p \end{pmatrix} = \begin{pmatrix} s_c^x & k_c \\ 0 & s_c^y \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} c_c^x \\ c_c^y \end{pmatrix} \tag{4}$$

where s_c^x, s_c^y are scale factors (pixel/mm), c_c^x, c_c^y are the pixel coordinates of the principal point and k_c is a shear coefficient (pixel/mm).

2.2. Projector model

The model for the projector is similar to that for the camera as the projector can conceptually be regarded as a camera with a 1D image, acting in reverse.

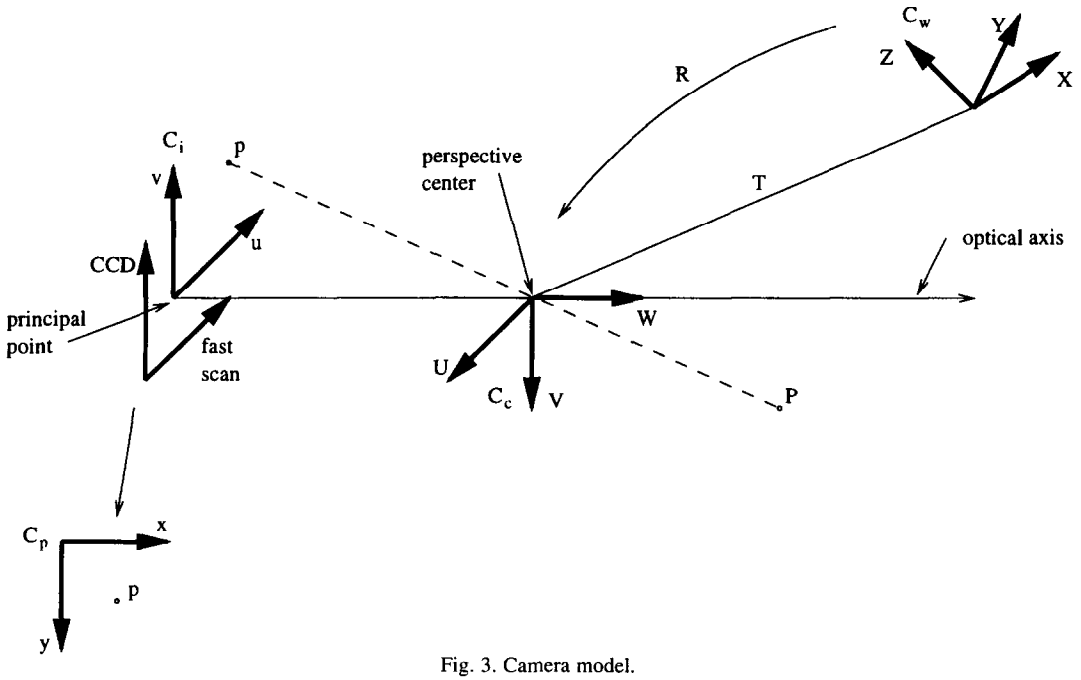

Fig. 3. Camera model.

Explicitly, the projector model differs in that the *stripe coordinate system* $\mathbf{C}_s$, the counterpart of $\mathbf{C}_p$, is 1D. This gives rise to the following set of equations,

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = R_p \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + T_p \tag{5}$$

$$\begin{pmatrix} \bar{u} \\ \bar{v} \end{pmatrix} = \frac{f_p}{W} \begin{pmatrix} U \\ V \end{pmatrix} \tag{6}$$

$$u = \bar{u}[1 + K_p(\bar{u}^2 + \bar{v}^2)] \tag{7}$$

$$x_s = s_p^x u + c_p^x \tag{8}$$

Physically, each stripe value x_s gives rise to a unique line on the projector emitter given by $\{(u, v): u = (x_s - c_p^x)/s_p^x, v \in \mathbb{R}\}$. The line is projected and distorted by the lens giving rise to a slightly curved surface in the world.

2.3. Discussion about models

The above models describe the geometric behaviour of the camera and projector and are based on a physical understanding of these devices. Any camera model used for high accuracy work would need to include most of the components of the above model. When $K_c = 0$ and $k_c = 0$ the camera described by the resulting model is referred to as a *perfect camera*. The perfect camera model is regarded as the most basic camera model and additional terms can be treated as refinements. Similarly, a *perfect projector* is the device described by the projector model when $K_p = 0$.

The shear, k_c, is used to model instability in the PLL line-synchronisation when the camera uses composite video output [8]. The scale factors, s_c^x, s_c^y, are governed by the inter-pixel spacing. In addition, when the sampling process is not pixel synchronous, s_c^x is used to absorb discrepancies in sampling rates in the camera and framegrabber.

It is clear that there is a redundant parameter within the set of camera parameters $\{s_c^x, s_c^y, f_c\}$. This problem can be removed by arbitrarily fixing one of them. For the calibrations considered here f_c is set to unity. While this affects the interpretation of some of the parameters in the model it does not affect the range of the model. Another approach is to set the value of s_c^y from the inter-pixel spacing given in the camera data sheet [8]. Similar considerations are made for the projector.

Expressions to include lens distortion are usually more complex than Eq. (3) and Eq. (7) and often include more radial distortion coefficients and some decentering distortion coefficients [8–10]. More importantly, the lens distortion is usually regarded as a function from the measured image coordinates $(uv)^t$ to the projected image coordinates $(\bar{u}\bar{v})^t$.

Some justification for selecting the lens distortion models used is required. Firstly, an effort has been made to achieve balance between the complexity of the camera and projector models as the system performance will be dictated by the least accurate component.

The distortion was incorporated as a function of $(\bar{u}\bar{v})^t$ to $(uv)^t$ as it is more difficult to include in the usual way for the projector. This is because distortion is a 2D phenomenon, whereas only one coordinate of the position of a point on the projector emitter is available. However, the ranges of both formulations are almost equivalent for typical system configurations [11]. Therefore this modification makes little difference to the solution.

One motivation for incorporating distortion in the usual way in a camera model is that it gives rise to a linear spatial intersection algorithm for multiple camera configurations (e.g. stereo pair). However, this is not true for a system containing a projector. The resulting spatial intersection problem is non-linear when distortion is included irrespective of how the distortion is incorporated into the model.

Ignoring lens distortion and the finite extent of the stripes, it is clear that the projector can be translated in a direction parallel to the line of intersection of the stripe planes without any effect on the image. It follows that this translation component (element T_p^2 in the model) will not be estimable from a given set of data. This observation is also made in Ref. [1]. Introducing distortion allows T_p^2 to be estimated but it cannot be recovered with any reasonable degree of confidence. Similar problems are encountered with the optic centre parameters of a camera when using the method of *analytic plumb-line calibration* [10,12].

The orientation of the axes of $\mathbf{C}_p$, $\mathbf{C}_i$, $\mathbf{C}_c$ as depicted in Fig. 3 are natural. They are chosen to be consistent with the CCD scan directions so that s_c^x, s_c^y are both positive. A similar choice can be made for the projector model.

The camera and projector models can be summarised as follows. Let

$$\Theta_c = (s_c^x\ s_c^y\ C_c^x\ c_c^y\ k_c\ K_c\ \omega_c^1\ \omega_c^2\ \omega_c^3\ T_c^1\ T_c^2\ T_c^3)^t \tag{9}$$

$$\Theta_p = (s_p^x\ c_p^x\ K_p\ \omega_p^1\ \omega_p^2\ \omega_p^3\ T_p^1\ T_p^2\ T_p^3)^t \tag{10}$$

where $\omega_c^k, k = 1 \ldots 3$ and $\omega_p^k, k = 1 \ldots 3$ parameterise R_c and R_p respectively (e.g. Euler angles [13]). The first six parameters of the camera model and the first three of the projector model are referred to as intrinsic parameters because they are independent of the world coordinate frame.

Using the above, the transformation from world coordinates, p_w, to pixel coordinates, p_p, resulting from the composition of Eqs. (1)–(4) can be expressed more succinctly as follows,

$$p_p = F_c(p_w;\ \Theta_c) \tag{11}$$

Similarly, the transformation from world coordinates to the stripe value, p_s, resulting from the composition of Eqs. (5)–(8) can be expressed,

$$p_s = F_p(p_w;\ \Theta_p) \tag{12}$$

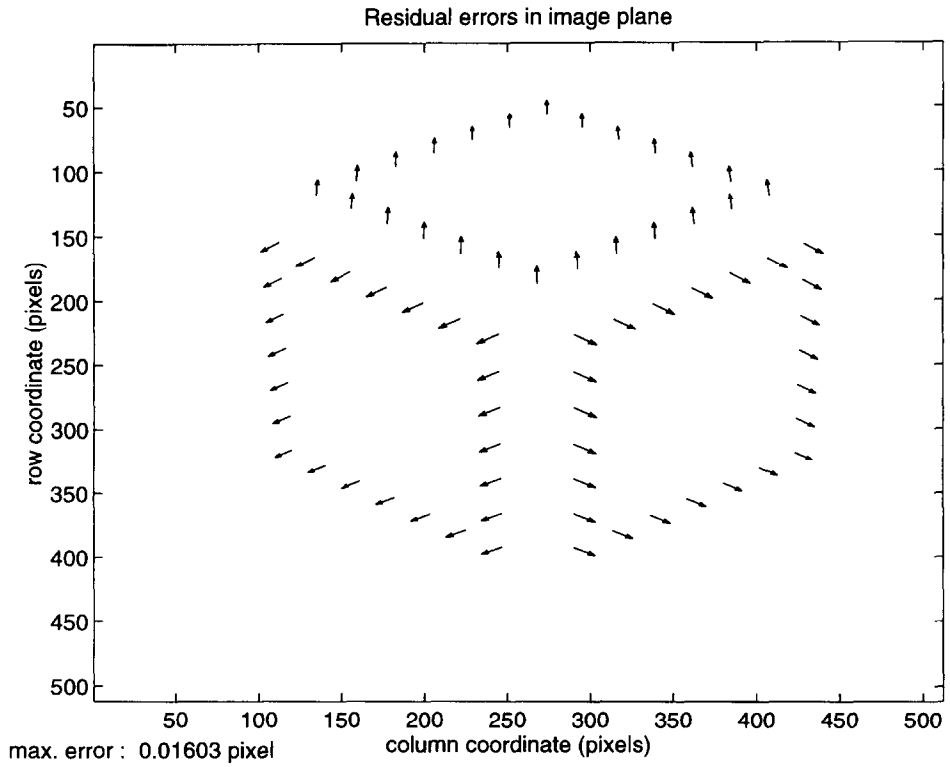

Fig. 4. Centroid error due to perspective distortion.

3. Localisation of coordinates

Both the calibration and spatial intersection procedures require the pixel coordinates and stripe value of points of interest in the world. Obtaining good estimates for the value of these quantities is important in order to obtain accurate estimates of the parameters of the SLS during calibration and accurate estimates of 3D point locations during spatial intersection.

3.1. Subpixel estimation

During calibration, the subpixel location of the centroids of the fiducial marks on the calibration reference are required. The greyscale centroid operator, described in Ref. [14], is used to estimate the centroids of the fiducial marks in the image. Such centroid-based techniques have a number of advantages. They are: (a) robust to noise (based on integration rather than differentiation, as are many edge detectors); (b) robust to small amounts of aliasing (theoretically exact if the sampling frequency is greater than the Nyquist frequency) [15]; (c) computationally inexpensive.

Unfortunately, the centroid of the image of the fiducial mark does not in general coincide with the image of the centroid, due to perspective distortion. This introduces a systematic error in the centroid estimate and is a fundamental limitation of centroiding techniques.

Ignoring lens distortion, an estimate of this error can be obtained as follows. Consider a flat circular region Δ of radius r. Let n be the normal to Δ and t be the location of the centroid of Δ, both in camera coordinates. If c denotes the image of the centroid and $\hat{c}$ denotes the centroid of the image of Δ then [11],

$$c = \frac{f_c}{t_3}\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \tag{13}$$

$$\hat{c} = \frac{f_c}{t_3^2 + r^2(n_3^2 - 1)}\begin{pmatrix} t_1 t_3 + r^2 n_1 n_3 \\ t_2 t_3 + r^2 n_2 n_3 \end{pmatrix} \tag{14}$$

As required, $\hat{c}$ reduces to c when $n = e_3$ (Δ parallel with image plane). The error in the image plane is given by $\epsilon_{\hat{p}_i} = c - \hat{c}$. This error can be transformed into pixel coordinates using the scale factors and shear coefficient.

Fig. 4 shows a plot of the errors (in pixel coordinates) for a typical calibration configuration[1]. Each error is represented as a vector with its base at the associated fiducial mark. The vector magnitudes have been scaled by a factor of 1000 for clarity. In the current system these errors are considered acceptable given the advantages of the centroid operator.

During spatial intersection two situations can arise, one where the image location is freely chosen and the other where it is dictated by a feature in the world. The need to use subpixel estimation does not arise in the first situation,

[1] This and other examples are taken from the data for trial 4 described in Section 6.

e.g. during the generation of a dense range map with image locations at the pixel lattice points. The type of subpixel operator used during the second situation depends on the nature of the feature e.g. edges, ridges, blobs [14,16,17].

3.2. Substripe estimation

During calibration and spatial intersection the substripe values of specified world points are required.

Substripe estimation is more difficult than subpixel estimation. One reason for this is that the stripe information is not directly available from the projector in the sense that the pixel location information is available from the camera. It is the images of the stripes cast on the scene which are used to recover the stripe values. In this way the stripes are effectively sampled twice, once by the LCS of the projector and once by the CCD sensor of the camera. Another difficulty, specific to temporally encoded systems, is that the stripe values are encoded in a sequence of images rather than available in a single image as for pixel locations.

It is assumed that the 8-bit stripe value is available for each pixel in the image. The image formed from these values is referred to as the *stripe image*. The discrete nature and coarse quantisation of the stripe image gives rise to poor accuracy if used directly. The algorithm adopted for substripe estimation is simple. A region of interest, Ω, is established with its centre at the subpixel location, (x_p,y_p), where the substripe estimate, x_s, is to be determined. A least-squares polynomial facet is fitted to the stripe image in Ω and the substripe estimate made by interpolating this facet.

Initially, consider the case where the underlying surface is flat. In this situation a planar facet is fitted to the stripe image in Ω. Assuming no lens distortion in the camera and projector, the stripe values and the pixel coordinates for a planar patch in the world will be related by a rational function because the composition of two projective maps is a projective map. Hence a planar facet does not model perspective effects precisely. The errors introduced by using a planar facet can be estimated as follows.

Let C_c and C_p denote the 3×4 and 2×4 perspective transformation matrices (PTMs) [21] for the (distortionless) camera and projector, respectively. Let T be a 4×3 homogeneous transformation matrix describing the coordinate transformation between a 2D coordinate system associated with the planar patch and the world coordinate system. It follows that,

$$\alpha \begin{pmatrix} x_s \\ 1 \end{pmatrix} = C_p T (C_c T)^{-1} \begin{pmatrix} x_p \\ y_p \\ 1 \end{pmatrix} \tag{15}$$

This clearly shows that the stripe values and pixel coordinates are related by an equation of the form

$$x_s = \frac{n_1 x_p + n_2 y_p + n_3}{d_1 x_p + d_2 y_p + d_3} \tag{16}$$

Let x_s be the true substripe value at subpixel location (x_p,y_p)

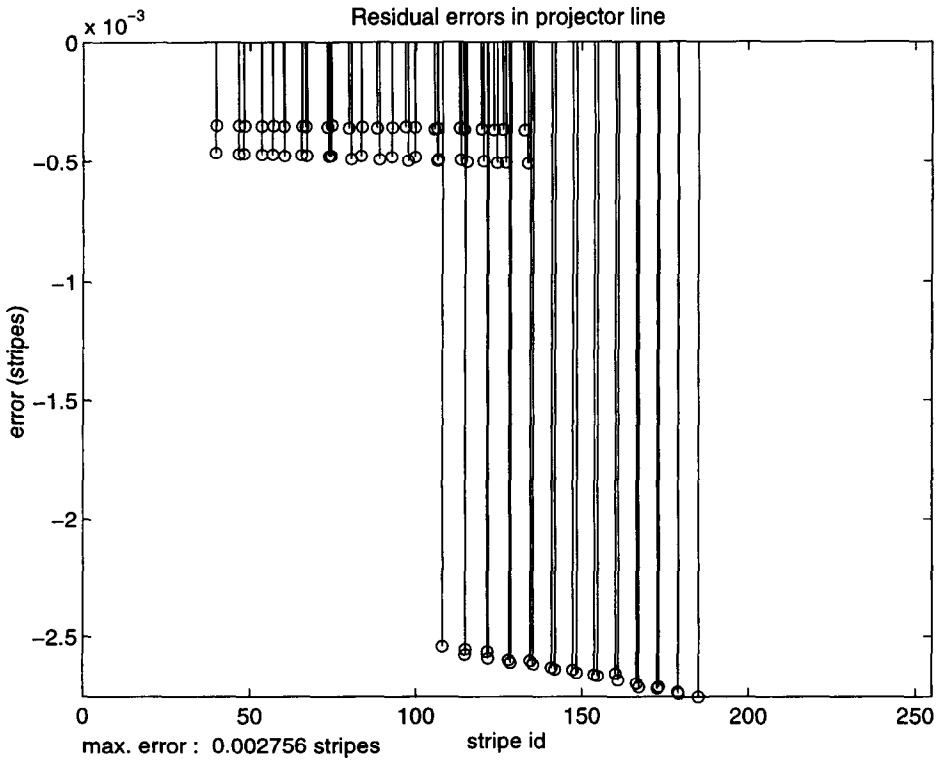

Fig. 5. Stripe error due to perspective distortion.

calculated using Eq. (15). Let (X_p,Y_p) be the $n \times 2$ matrix of pixel coordinates in Ω ($n = 17 \times 17$ pixels) and X_s be the n-vector of substripe values at coordinates (X_p,Y_p). The 3×1 vector of coefficients of the least-squares plane fitted to X_s on Ω is given by,

$$c = (A^tA)^{-1}A^tX_s \tag{17}$$

where $A = [X_pY_p1_n]$ is the $n \times 3$ design matrix associated with the plane. The estimate of x_s is given by

$$\hat{x}_s = c_1x_p + c_2y_p + c_3 \tag{18}$$

and the error is given by $\epsilon_{\hat{x}_s} = x_s - \hat{x}_s$.

Fig. 5 shows a stem plot of these errors for a typical calibration configuration. The horizontal axis gives the stripe value calculated at the centroids of the fiducial marks and the vertical lines give the error in stripes at these locations. The errors are insignificant compared with the effects which result from the 8-bit quantisation of the stripe value and can be ignored in practice. Indeed, any effort to estimate the parameters of the rational function on Ω results in over parameterisation problems.

Now consider the case of more general curved surfaces. In this situation higher order 2D polynomial facets are fitted to the stripe image. The order of the facet model depends on the size of the region of interest, Ω, which must cover a sufficient number of stripes. In practice, a region of interest of size 17×17 pixels and 2D third-order polynomials are used. Note that polynomial fitting does not introduce blurring in the sense of low-pass filtering (Ref. [18], Section 14.8). For example, a polynomial of degree n (which can have considerable oscillation) is invariant to a polynomial filter of order n (i.e. the filter which is defined by fitting and interpolating a polynomial of degree n).

Planar facets are only used when the surface under consideration is known to be flat, such as the faces of the calibration reference. In all other cases a polynomial facet is used. The planar facet was introduced because it provides superior estimates when the surface has very low curvature. More sophisticated criterion for selecting between these models are under development as are more complex substripe estimators.

4. SLS calibration

As described in Section 1, calibration involves estimating the unknown SLS parameters from a number of known world points and their corresponding pixel coordinates and stripe values.

Assume there are n fiducial marks on the calibration reference. Let $p_w(j)$, $p_p(j)$ and $p_s(j)$ be the world coordinates, pixel coordinates and stripe coordinates respectively, of the jth fiducial mark. In addition, let P_w, P_p and P_s be vectors of sizes $3n$, $2n$ and n, respectively, formed by stacking the coordinate vectors. The vectors P_p and P_s are measured from the image sequence as described in Section 3 and will consequently suffer from measurement noise.

If μ_{P_p} and μ_{P_s} are the true values of P_p and P_s, respectively, then the observations are given by,

$$P_p = \mu_{P_p} + \epsilon_p \tag{19}$$

$$P_s = \mu_{P_s} + \epsilon_s \tag{20}$$

The SLS model can then be written,

$$\mu_{P_p} - F_c(P_w;\ \Theta_c) = 0$$

$$\mu_{P_s} - F_p(P_w;\ \Theta_p) = 0 \tag{21}$$

The following assumptions are made: P_w is measured without error; $P_p \sim N(\mu_{P_p},\ \sigma_p^2 I_{2n})$; $P_s \sim N(\mu_{P_s},\ \sigma_s^2 I_n)$.

Maximum likelihood estimation of $\{\Theta_c,\Theta_p\}$ gives rise to the following non-linear least-squares (NLLS) problems [19],

$$\min_{\Theta_c} \|P_p - F_c(P_w;\ \Theta_c)\|^2 \tag{22}$$

$$\min_{\Theta_p} \|P_s - F_p(P_w;\ \Theta_p)\|^2 \tag{23}$$

These problems can be solved using any general technique for NLLS problems such as Gauss-Newton or Levenberg-Marquardt [18].

Obtaining good initial estimates for parameters is desirable as it helps ensure that the global minimum is obtained and reduces computation by reducing the number of iterations required. Initial estimates can be obtained from a variety of sources depending on the nature of the problem. The approach adopted here is to estimate and decompose the PTMs for the camera [20] and projector [21].

5. Spatial intersection

Spatial intersection is used to find an estimate of the location of a point in the world coordinate system given its pixel coordinates and stripe value. The lens distortion in the projector results in slightly curved surfaces being projected onto the scene rather than planes. Consequently the spatial intersection procedure involves solving a homogeneous non-linear system in three variables. However, linear methods can be used to calculate a good initial approximation and physical considerations preclude the presence of any other solutions in a large neighbourhood of the true solution.

Let $p_p = (x_py_p)^t$, and $p_s = (x_s)$ be the pixel coordinates and stripe value of a point P. Then the world coordinates of P, denoted p_w, are given by the solution of the non-linear system,

$$\begin{pmatrix} p_p \\ p_s \end{pmatrix} - \begin{pmatrix} F_c(p_w;\ \Theta_c) \\ F_p(p_w;\ \Theta_p) \end{pmatrix} = 0 \tag{24}$$

This problem can be solved using any general technique such as a quasi-Newton strategy [18].

The initial linear estimate of p_w is found as follows. Let C_c^k be the kth row of the 3×4 camera PTM and C_p^k the kth row of the 2×4 projector PTM. Then the PTM equations relating p_w, p_p and p_s, i.e.

$$\beta_p \begin{pmatrix} p_p \\ 1 \end{pmatrix} = C_c \begin{pmatrix} p_w \\ 1 \end{pmatrix} \quad \beta_s \begin{pmatrix} p_s \\ 1 \end{pmatrix} = C_p \begin{pmatrix} p_w \\ 1 \end{pmatrix} \tag{25}$$

can be rearranged into the 3×4 linear homogeneous system,

$$\begin{pmatrix} C_c^1 - x_p C_c^3 \\ C_c^2 - y_p C_c^3 \\ C_p^1 - x_s C_p^2 \end{pmatrix} \begin{pmatrix} \alpha p_w \\ \alpha \end{pmatrix} = 0 \tag{26}$$

A solution to Eq. (26) can be obtained (e.g. using SVD) and the estimate of p_w obtained from this homogeneous vector.

When there are a large number of points to process computational savings can be made. A generalisation of the vector cross product for $n-1$ vectors in an n-dimension linear space can be defined implicitly by $\langle w, v_1 \times \ldots \times v_{n-1} \rangle = \det(v_1, \ldots v_{n-1}, w)$ [22]. It is clear that $v_1 \times \ldots \times v_{n-1}$ can be written $\Sigma_k \det(v_1, \ldots v_{n-1}, e_k) e_k$ and is orthogonal to v_k where $\{e_k\}_{k=1}^n$ are the standard basis vectors of $\mathbb{R}^n$. Therefore a solution to Eq. (26) is given by the (generalised) cross product of the rows of the 3×4 matrix,

$$\begin{pmatrix} \alpha p_w \\ \alpha \end{pmatrix} = \sum_{k=1}^{4} \det(C_c^1 - x_p C_c^3, C_c^2 - y_p C_c^3, C_p^1 - x_s C_p^2, e_k) e_k \tag{27}$$

Dividing by α and using the linearity and antisymmetry of the determinant tensor, the kth component of p_w is given by,

$$p_w^k = \frac{C_{1,2,1}^k - x_p C_{3,2,1}^k - y_p C_{1,3,1}^k - x_s C_{1,2,2}^k + x_s x_p C_{3,2,2}^k + x_s y_p C_{1,3,2}^k}{C_{1,2,1}^4 - x_p C_{3,2,1}^4 - y_p C_{1,3,1}^4 - x_s C_{1,2,2}^4 + x_s x_p C_{3,2,2}^4 + x_s y_p C_{1,3,2}^4} \tag{28}$$

where $C_{i,j,l}^k = \det(C_c^i, C_c^j, C_p^l, e_k)$ are constants which depend only on the camera and projector PTMs and can be precomputed.

6. Experimental evaluation

The structured light system used for the experiments comprised a K2T LCS and controller (housed in a Kodak projector), TM-6CN Pulnix camera and S2200 Data Cell framegrabber. The LCS shutter has 256 stripes and 512×512 images were captured. The calibration reference is a 150 mm cube with 72 fiducial marks (of 5 mm radius) arranged over three faces as shown in Fig. 6. The 3D location of each fiducial mark centroid has been measured to an accuracy of 0.1 mm in a coordinate system attached to the cube. During calibration this coordinate system is taken to be the world coordinate system. Calibration thus fixes the world coordinate system, and all subsequent measurements are expressed in this world coordinate system until the system is recalibrated. The experimental setup has the projector above the camera with the angle between their optic axes being approximately 12° to give a working volume of 250 mm diameter, at 1600 mm in front of the camera.

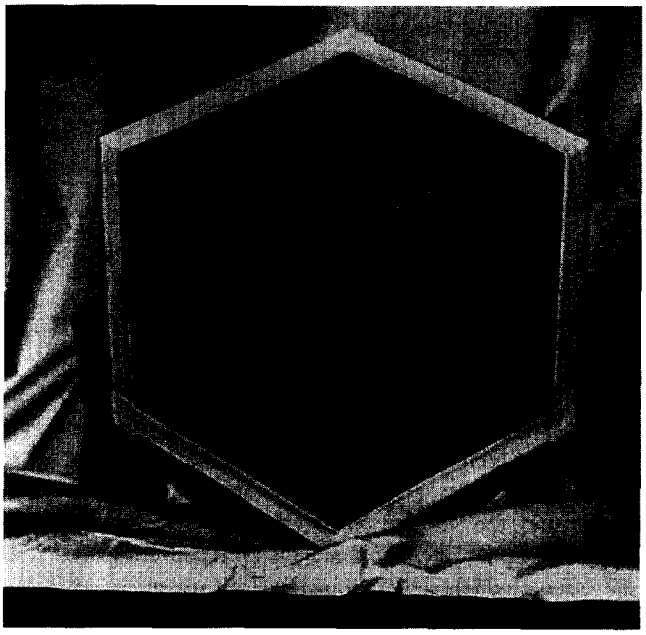

Fig. 6. Calibration reference.

To evaluate the performance of the system, seven trials were used with the calibration reference in a different position in each. For each trial the *observed data* consist of the pixel coordinates and stripe values of the fiducial marks. The calibration parameters for each trial were extracted from the observed data using the procedure described in Section 4.

The estimated pixel coordinates and stripe values are obtained by projecting the world reference coordinates using the calibrated system model. The *pixel residuals* and *stripe residuals* are the difference between the measured and estimated values. The *spatial intersection errors* are the difference between the reference coordinates of the fiducial marks and their coordinates estimated by spatial intersection. Figs 7–9 show the pixel residuals, stripe residuals and spatial intersection errors, respectively, for trial 4. The fiducial mark numbers in Fig. 9 were obtained from an arbitrary ordering of the fiducial marks on the cube.

Table 1 shows the average magnitude of the spatial intersection errors for each of the calibration trials. Each average

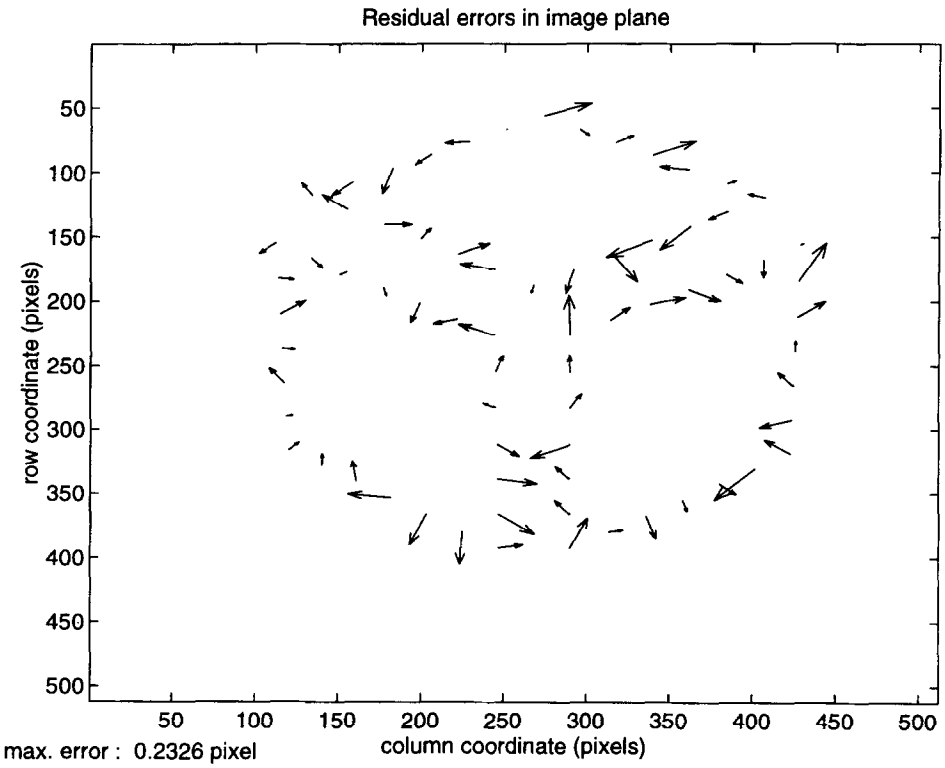

Fig. 7. Pixel residual errors.

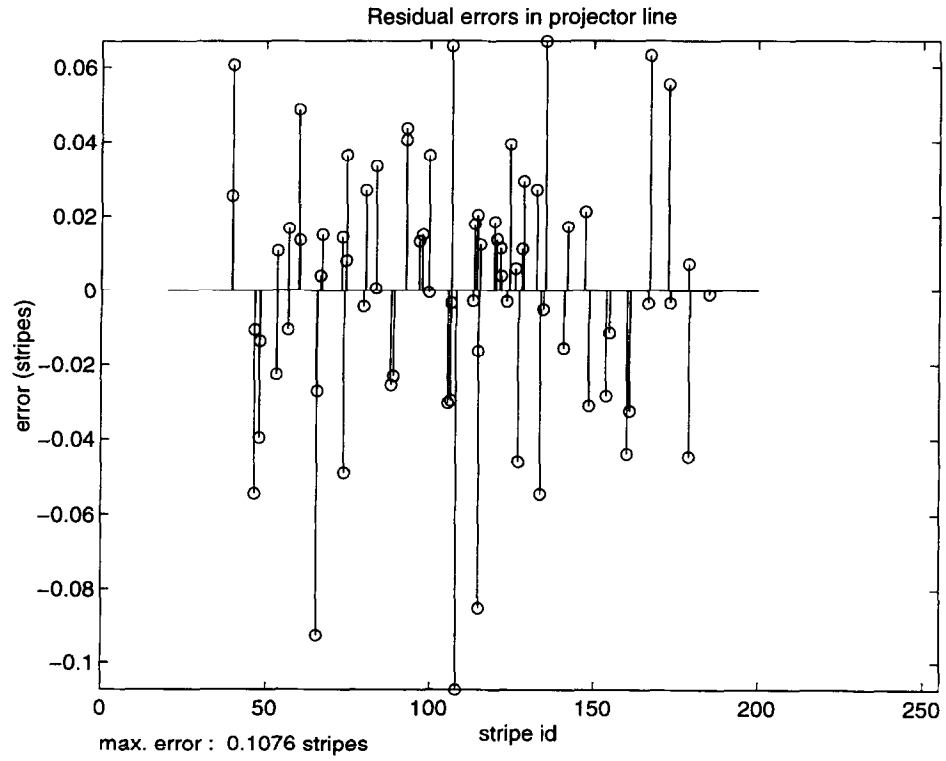

Fig. 8. Stripe residual errors.

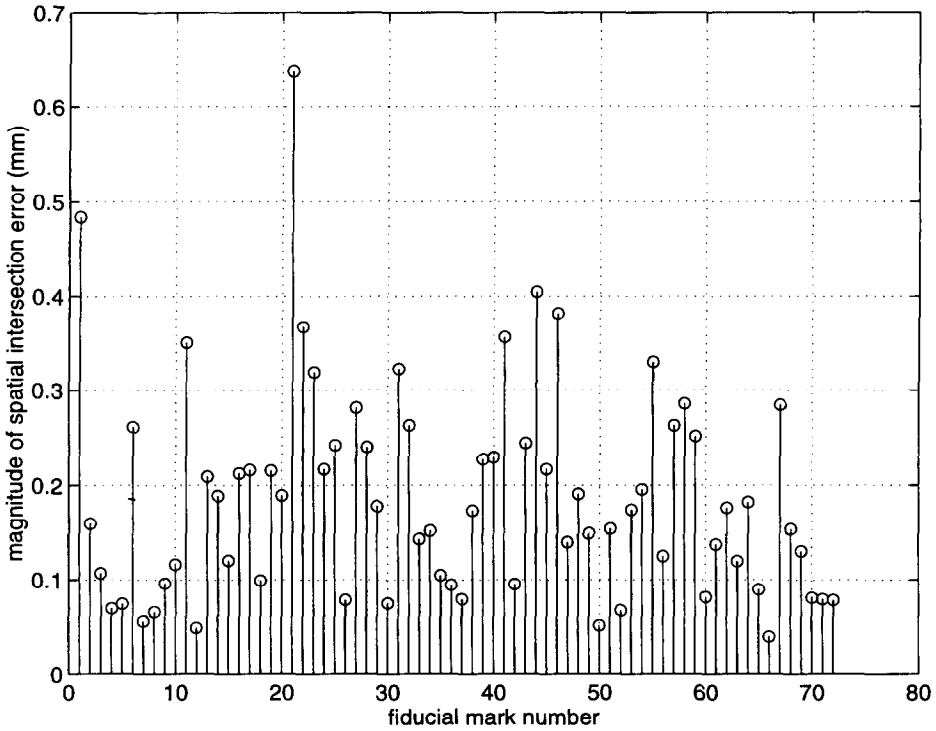

Fig. 9. Spatial intersection errors.

was obtained using all 72 fiducial marks. For comparison, Table 2 shows the average errors when the camera and projector models include no radial distortion coefficients. Furthermore, Table 3 shows the average errors when no substripe estimation is used. Excluding either radial distortion or substripe estimation increases the error in all trials. On average, over all trials, adding substripe estimation improves the performance by 82% while adding lens distortion improves the performance by only 13%. This shows that substripe estimation has a more significant effect.

Given the observed data for a trial and the calibration parameters for that trial, or any other, spatial intersection can be used to estimate the coordinates of the fiducial marks in the world coordinate system established by the calibration. These can be compared with the reference coordinates by model fitting, i.e. estimating the 3D rigid body transformation between them using a least-squares error criterion [23]. The *model fitting error* is the difference between the estimated world coordinates and the transformed reference coordinates. Table 4 shows the average magnitude of the model fitting error for all of the calibration trials and observed data sets. The diagonal entries of Table 4 are very close to, and less than, the values in Table 1. This indicates that the spatial intersection procedure is accurately recovering the coordinates in the world coordinate system. The model fitting errors are largest when the spatial intersection data set is farthest from the calibration data set.

Table 5 shows the intrinsic parameters for the SLS obtained from each trial. The results offer clear support to the manufacturer's claim that the framegrabber aspect ratio is unity. Recall (Section 2.3) that the camera has been normalised so $f_c = 1$, thus the camera scale factors are the ratio of the true principal distance and the pixel spacing. The camera scale factors obtained during calibration agree with the nominal value of 2200 pixels/mm. In addition the camera principal point is consistent with the nominal value of (255, 255) pixels. The shear coefficient is small with respect to s_c^x (0.14%) suggesting a very small contribution. The scale factor for the projector is close to the nominal value of 1020 stripes/mm.

However there are significant variations in many of the parameters. An examination of the estimated dispersion matrix for the camera shows some parameters have high

Table 1
Average magnitude of the spatial intersection errors for calibration trials

	Trial						
	1	2	3	4	5	6	7
Average error (mm)	0.154	0.245	0.296	0.187	0.280	0.246	0.233

Table 2
Average magnitude of the spatial intersection errors for calibration trials with no radial distortion in the SLS model

	Trial						
	1	2	3	4	5	6	7
Average error (mm)	0.226	0.283	0.311	0.233	0.292	0.261	0.264

Table 3
Average magnitude of the spatial intersection errors for calibration trials with no substripe estimation

	Trial						
	1	2	3	4	5	6	7
Average error (mm)	1.348	1.312	1.371	1.521	1.233	1.170	1.444

Table 4
Average magnitude of the model fitting error (mm) for each calibration trial (horizontal) and each measured data set (vertical)

Trial	1	2	3	4	5	6	7
1	0.154	0.405	0.271	0.207	0.322	0.285	0.298
2	0.389	0.245	0.445	0.337	0.246	0.428	0.392
3	0.335	0.613	0.296	0.367	0.490	0.307	0.437
4	0.205	0.379	0.312	0.187	0.298	0.297	0.248
5	0.408	0.301	0.476	0.349	0.280	0.451	0.374
6	0.324	0.564	0.273	0.355	0.455	0.245	0.424
7	0.295	0.475	0.308	0.247	0.366	0.292	0.233

variances and there are high correlations between parameters. This can be partially explained by the experimental configuration of the structured light system. With the distance between the camera and world reference much larger than the diameter of the world reference, a weak perspective camera model provides a reasonable explanation of the observed data. Consequently many parameters are correlated. In particular, T_c^3 has a relatively large variance and is highly correlated with s_c^x and s_c^y.

The above remarks are also applicable to the projector parameters. However, the situation for the projector is inherently worse than that for the camera as the projector has nine parameters (compared with 12 for the camera) but only half as many data points are available. As mentioned previously, introducing K_p allows T_p^2 to be estimated but it has a large variance as expected. Furthermore, when the distance between the projector and world reference is much larger than the diameter of the world reference the stripe planes are nearly parallel in the working volume. Hence a component of the rotation is nearly unidentifiable. This gives rise to large variances in the rotation parameters.

It is important to note that the actual values of the camera and projector parameters are incidental in this situation and it is the values of the measured 3D coordinates that are of real interest. Consequently, the quality of the model is measured by its ability to provide accurate 3D data through spatial intersection rather than good camera/projector parameter estimates. For example, including T_p^2 and K_p in the projector model is regarded as an improvement because it improves the accuracy of the spatial data regardless of the fact that T_p^2 cannot be recovered to any useful accuracy.

7. Conclusion

This paper discussed a method for obtaining accurate 3D measurements using a temporarily encoded structured light system. Lens distortion was included in the models for both the camera and projector. In addition, a substripe estimator was used to estimate projector stripe values as well as using a subpixel estimator for locating image features.

Experimental evaluation shows that it is important to use substripe estimation and incorporate lens distortion in the projector model. Of these two, the improvement due to using substripe estimation is more significant than that resulting from including radial distortion in the projector and camera models.

The performance achieved by the system is in the order of 0.3 mm for the geometry and conditions described in Section 6. The geometry of the SLS configuration (particularly the triangulation angle) substantially affects the system performance. This must be taken into account when comparing the performance with other systems. As discussed in Section 6, while the spatial intersection accuracy is high, not all camera and projector model parameters are reliably estimated.

The performance could be improved by using a better calibration reference. In particular, one having more fiducial marks and having them better distributed. Also the reference needs to be measured more accurately than at present because the spatial intersection errors are of the same order as the measurement tolerance.

Further improvements could be obtained by using

Table 5
The six intrinsic parameters for the camera followed by the three intrinsic parameters for the projector for each calibration trial

Trial	1	2	3	4	5	6	7
s_c^x	2411.062	2399.667	2425.255	2418.480	2403.901	2443.054	2408.316
s_c^y	2412.349	2402.982	2425.084	2421.023	2407.875	2445.911	2410.448
c_c^x	289.389	283.201	301.069	274.053	271.295	314.980	272.129
c_c^y	271.802	303.288	258.478	275.559	309.616	261.080	259.293
k_c	4.214	2.316	4.308	3.285	1.938	4.006	3.136
K_c	− 0.164	− 0.176	− 0.149	− 0.139	− 0.200	− 0.219	0.035
s_p^x	899.157	1020.565	1151.871	1015.006	1061.155	1162.614	1059.352
c_p^x	159.906	240.831	130.128	183.441	233.480	132.600	181.811
K_p	− 0.085	− 0.123	0.317	− 0.116	− 0.138	0.384	− 0.109

'multiple frame methods' where a single calibration is performed on the coordinate data obtained with the calibration reference in a variety of different positions. The intrinsic parameters are common to all positions, whereas the pose parameters vary for each position. However this would be time consuming for this SLS because 18 images need to be taken at each position to obtain the stripe information.

The performance of the centroid operator used for estimating the centroid of the fiducial marks can be improved by better identification of the support region and background removal. Substantial improvements to the sub-stripe estimator should also be possible by recognising the step-like structure of a stripe image. Alternatively a sub-stripe algorithm could be designed based on the original 18 banded light patterns. Other avenues for improving the performance include using a digital output pixel synchronous large format camera and a projector with more stripes.

Acknowledgements

This work has been supported by the New Zealand Foundation of Research Science, and Technology Contract CO8410, Objective 4.

References

[1] T. Stahs, F. Wahl, Fast and versatile range data acquisition in a robot work cell, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, July 7–10, 1992, pp. 1169–1174.

[2] Y. Shirai, 3D computer vision and applications, in: 11th International Conference on Pattern Recognition, International Association for Pattern Recognition, IEEE Computer Society Press, 1992, pp. 236–245.

[3] K. Sato, H. Yamamoto, S. Inokuchi, Tuned range finder for high precision 3D data, in: International Conference on Pattern Recognition, Paris, France, October 1986, IAPR, IEEE Computer Society, pp. 1168–1171.

[4] K. Sato, S. Inokuchi, Range-imaging system utilizing nematic liquid crystal mask, in: International Conference on Computer Vision, IEEE Computer Society Press, 1987, pp. 657–661.

[5] R. Jarvis, Range sensing for computer vision, in: A.K. Jain, P.J. Flynn (Eds.), Three-Dimensional Object Recognition Systems, vol. 1 of Advances in Image Communications, Elsevier Science Publishers, Amsterdam, 1993, pp. 17–56.

[6] E. Trucco, R.B. Fisher, A.W. Fitzgibbon, Direct calibration and data consistency in 3-D laser scanning, in: E. Hancock (Ed.), BMVC94: Proceedings of the 5th British Machine Vision Conference, Sheffield, BMVA Press, September 1994, pp. 489–498.

[7] A. Busboom, R.J. Schalkoff, Direct surface parameter estimation using structured light: A predictor-corrector based approach, Image and vision Computing 14 (5) (June 1996) 311–321.

[8] H.A. Beyer, An introduction to photogrammetric camera calibration, in: Invited paper, Seminaire Orasis, St. Malo, France, September 23–27, 1991.

[9] C.C. Slama (Ed.), Manual of Photogrammetry, American Society of Photogrammetry, 1980.

[10] D.C. Brown, Close range camera calibration, Photogrammetric Engineering 37 (8) (August 1971) 855–866.

[11] R.J. Valkenburg, A.M. McIvor, Accurate 3D measurement using a Structured Light System, Report 576, Industrial Research Limited, June 1996.

[12] H.M. Karara (Ed.), Non-Topographic Photogrammetry, American Society for Photogrammetry and Remote Sensing (ASPRS), 2nd ed., 1989.

[13] K.S. Fu, R.C. Gonzalez, C.S.G. Lee, Robotics: Control, Sensing, Vision and Intelligence, McGraw-Hill, 1987.

[14] R.J. Valkenburg, A.M. McIvor, P.W. Power, An evaluation of subpixel feature localisation methods for precision measurement, in: S.F. El-Halim (Ed.), Videometrics III, vol. 2350, SPIE, 1994, pp. 229–238.

[15] B.F. Alexander, K.C. Ng, Elimination of systematic error in subpixel accuracy centroid estimation, Optical Engineering 30 (9) (September 1991).

[16] P. Seitz, Optical superresolution using solid-state cameras and digital signal processing, Optical Engineering 27 (7) (July 1988).

[17] E.P. Lyvers, O.R. Mitchell, M.L. Akey, A.P. Reeves, Subpixel measurement using a moment-based edge operator, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (12) (1989).

[18] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes in C: the Art of Scientific Computing, Cambridge University Press, 2nd ed., 1992.

[19] G.A.F. Seber, C.J. Wild, Nonlinear Regression, Wiley, New York, 1989.

[20] O. Faugeras, Three-dimensional Computer Vision: a Geometric Viewpoint, MIT Press, 1993.

[21] A.M. McIvor, R.J. Valkenburg, Calibrating a Structured Light System, Report 362, Industrial Research Limited, February 1995.

[22] M. Spivak, Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus, W.A. Benjamin, Inc., 1965.

[23] K. Kanatani, Geometric Computation for Machine Vision, Oxford University Press, 1993.