Accurate 3D measurement using a Structured Light System

R.J. Valkenburg, A.M. M^{c}Ivor

Imaging and Sensing, Industrial Research Ltd
P.O Box 2225, Auckland, New Zealand
Email: b.valkenburg@irl.cri.nz, a.mcivor@irl.cri.nz

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.

Keywords: structured light, triangulation, spatial intersection, calibration

1. INTRODUCTION

This paper describes a method for obtaining accurate 3D measurements using a Structured Light system (SLS).

Structured light is one of a variety of approaches for acquiring 3D information about the world. It is a general concept and there are a number of ways of exploiting it to obtain 3D data.[16,13,5,17] As such, a brief description of the specific SLS considered in this paper follows. Refer to Figure 1. The SLS consists of a projector together with a camera. The projector projects a coded sequence of 9 banded light patterns (actually 9 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). The camera captures the corresponding sequence of images. The first pattern is with the LCS fully open (full illumination). Each of the remaining patterns represents a bit in an 8 bit binary code which enables 256 stripes to be differentiated in the scene.[a] Hence for each visible point in the world there is a corresponding 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 is 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. An important property of the stripe 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 value by one. It should be

[a] This is sometimes referred to as temporal encoding.

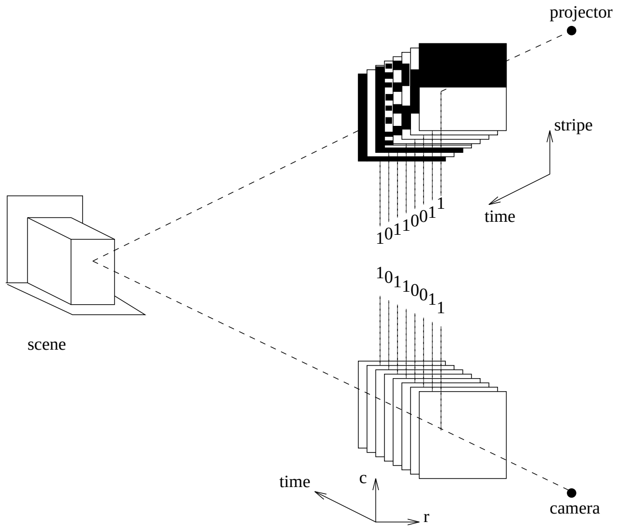

Figure 1: Idealised structured light system.

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 the its nominal location. Furthermore, the discrete nature of the CCD and LCS gives rise to uncertainty in the true pixel coordinates and stripe value. Addressing these problems is important in order to obtain accurate 3D data 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 whereas the projector has a 1D emitter. 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 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.

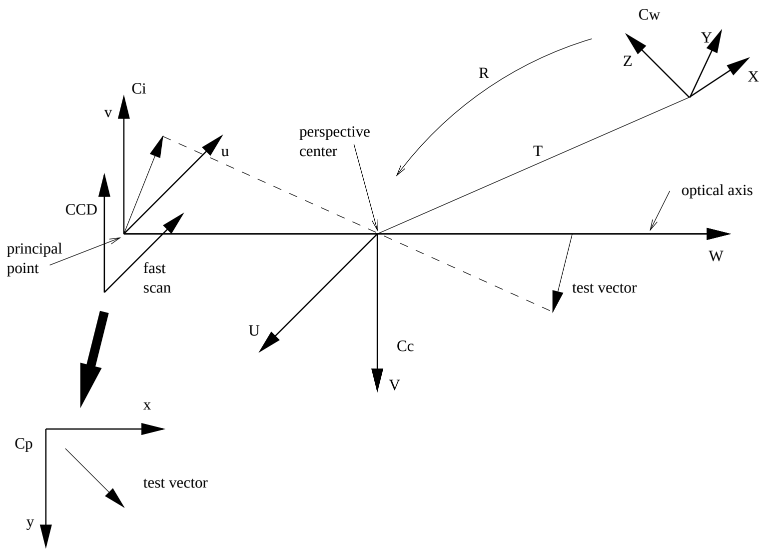

Figure 2: Camera model.

2.1. Camera model

Refer to Figure 2. Let P be a point in the world, $\mathbf{C}_w$ a *world coordinate system* and $(X\ Y\ Z)^T$ be the coordinates of P in $\mathbf{C}_w$. Define the *camera coordinate system*, $\mathbf{C}_c$, to have its z-axis parallel with the *optical axis* of the camera lens, its x-axis parallel with the x-axis of $\mathbf{C}_i$ (defined later) and origin located at the *perspective centre*. Let $(U\ V\ W)^T$ be the coordinates of P in $\mathbf{C}_c$. The coordinates $(U\ V\ W)^T$ are related to $(X\ Y\ Z)^T$ by a rigid body coordinate transformation (1) where R is a 3×3 rotation matrix and T 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 $(\bar{u}\ \bar{v})^T$ be the coordinates of p in $\mathbf{C}_i$. Then $(\bar{u}\ \bar{v})^T$ are given by (2) where f 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 $(u\ v)^T$ be actual observed image point after being subject to lens distortion. Then $(u\ v)^T$ is related to $(\bar{u}\ \bar{v})^T$ by (3) where K 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 $\mathbf{C}_p$ be the *pixel coordinate system* associated with the digital image. The pixel coordinates are related to the image coordinates by (4) where s_x, s_y are scale factors (pixel/mm), c_x, c_y are the pixel coordinates of the principal point and k is a shear coefficient (pixel/mm).

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

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

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

$$\begin{pmatrix} x_p \\ y_p \end{pmatrix} = \begin{pmatrix} s_x & k \\ 0 & s_y \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} + \begin{pmatrix} c_x \\ c_y \end{pmatrix} \tag{4}$$

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

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

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

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

$$x_s = s_x u + c_x \tag{8}$$

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 $K = 0$ and $k = 0$ 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 $K = 0$.

The shear, k, is used to model instability in the PLL line-synchronisation when the camera uses composite video output.[2] The scale factors, s_x, s_y, are governed by the inter-pixel spacing. In addition, when the sampling process is not pixel synchronous, s_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 $\{s_x, s_y, f\}$. This problem can be removed by arbitrarily fixing one of them. For the calibrations considered here f is set to unity. While this effects the interpretation of some of the parameters in the model it does not effect the range of the model. Another approach is to set the value of s_y from the inter-pixel spacing given in the camera data sheet.[2] Similar considerations are made for the projector.

Expressions to include lens distortion are usually more complex than (3) and (7) and often include more radial distortion coefficients and some decentering distortion coefficients.[14,3,2] More importantly, the lens distortion is usually regarded as a function from the measured image coordinates $(u\ v)^T$ 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. There seems little no point in making one model far more complex than the other 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 $(u\ v)^T$ as it is more difficult to include in the usual way for the projector. This is because distortion is a 2D phenomenon whereas the projector is essentially only a 1D device. However, the ranges of models using both formulations are very similar for typical system configurations[19] and so 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 projector distortion is included irrespective of how the distortion is incorporated into the model. Because only one coordinate is measured in the projector, adding additional parameters into the projector model, such decentering distortion, results in over-parameterisation problems.

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_2 in the model) will not be estimateable from a given set of data. Introducing distortion allows T_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*.[3,7]

The orientation of the axes of $\mathbf{C}_p$, $\mathbf{C}_i$, $\mathbf{C}_c$ as depicted in Figure 2 are natural. They are chosen to be consistent with the CCD scan directions so that s_x, s_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

$$\begin{aligned} \Theta_c &= (s_x\ s_y\ c_x\ c_y\ k\ K\ \omega_1\ \omega_2\ \omega_3\ T_1\ T_2\ T_3)^T \\ \Theta_p &= (s_x\ c_x\ K\ \omega_1\ \omega_2\ \omega_3\ T_1\ T_2\ T_3)^T \end{aligned} \tag{9}$$

where $\omega_k, k = 1 \ldots 3$ parameterise a rotation matrix (e.g. Euler angles). It is understood that the parameters Θ_p are different from the corresponding elements of Θ_c even though this is not made explicit in the notation. The first 6 parameters of the camera model and the first 3 parameters of projector model are referred to as *intrinsic parameters* because they are independent of the world coordinate frame. Using Θ_c and Θ_p, the transformation from world coordinates, p_w, to pixel coordinates, p_p, resulting from the composition of (1-4), and the transformation from world coordinates to the stripe value, p_s, resulting from the composition of (5-8), can be expressed succinctly as,

$$p_p = F_c(p_w; \Theta_c) \qquad \text{and} \qquad p_s = F_p(p_w; \Theta_p) \tag{10}$$

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[18] is used to estimate the centroids of the fiducial marks in the image. Centroid based techniques are attractive because they are,

- robust to noise (based on integration rather than differentiation, as are many edge detectors).
- robust to small amounts of aliasing (theoretically exact if the sampling frequency is greater than the Nyquist frequency).[1]
- computationally inexpensive.

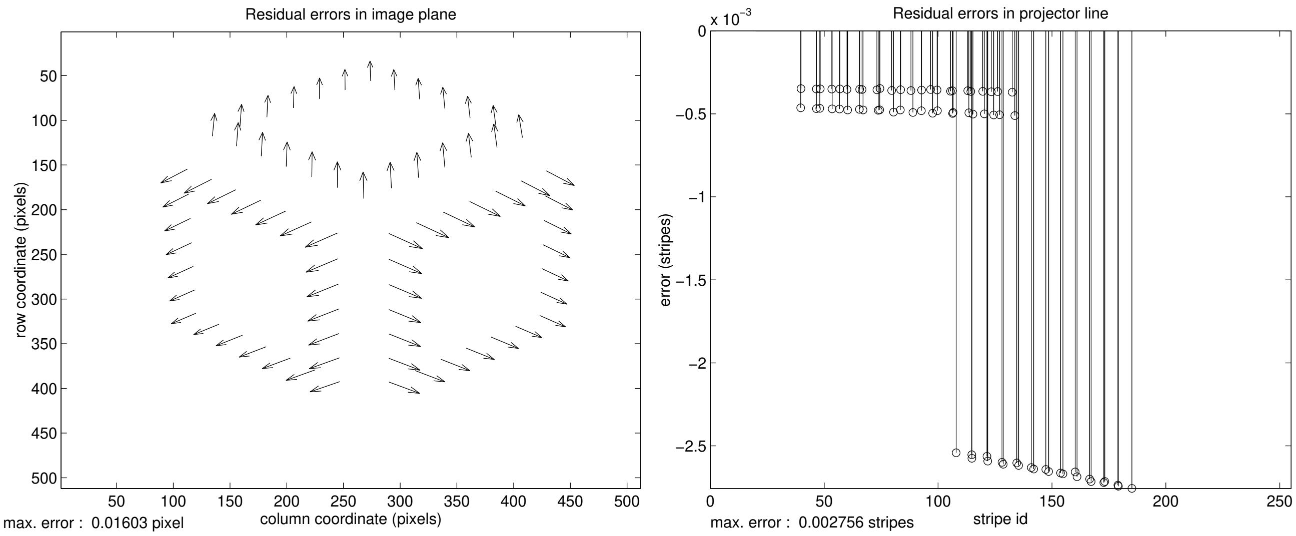

Figure 3: Centroid error (left) and stripe error (right) due to perspective distortion.

Unfortunately, in general the centroid of the image of the fiducial mark does not 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 Ω of radius r. Let n be the normal to Ω and t be the location of the centroid of Ω, both in camera coordinates. If c denotes the image of the centroid and $\hat{c}$ denotes the centroid of the image of Ω then,[19]

$$c = \frac{f}{t_3}\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \tag{11}$$

$$\hat{c} = \frac{f}{t_3^2 + r^2(n_3^2 - 1)}\begin{pmatrix} t_1 t_3 + r^2 n_1 n_3 \\ t_2 t_3 + r^2 n_2 n_3 \end{pmatrix} \tag{12}$$

As required, $\hat{c}$ reduces to c when $n = e_3$ (Ω 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. Figure 3 shows a plot of these errors (in pixel coordinates) for a typical calibration configuration[b]. 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, 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 feature e.g. edges, ridges, blobs.[12,8,18]

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 informa-

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

tion 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 surface under consideration is flat, such as the faces of the calibration reference. In this situation a planar facet is fitted to the stripe image in Ω. Assuming no lens distortion in the camera and projector then 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 Transformations Matrices (PTMs) 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,

$$\alpha \begin{pmatrix} x_s \\ 1 \end{pmatrix} = C_p T \left(C_c T \right)^{-1} \begin{pmatrix} x_p \\ y_p \\ 1 \end{pmatrix} \tag{13}$$

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

$$x_s = \frac{n_1 x_p + n_2 y_p + n_3}{d_1 x_p + d_2 y_p + d_3} \tag{14}$$

Let x_s be the true substripe value at subpixel location (x_p, y_p) calculated using (13). 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,

$$c = (A^T A)^{-1} A^T X_s \tag{15}$$

where $A = [X_p \; Y_p \; 1_n]$ is the $n \times 3$ design matrix associated with the plane. The estimate of the x_s is given by

$$\hat{x}_s = c_1 x_p + c_2 y_p + c_3 \tag{16}$$

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

Figure 3 shows a plot of these errors calculated at the centroids of the fiducial marks for a typical calibration configuration. The errors are insignificant compared with the effects which result from the quantisation of the stripe image 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 roi, Ω, which must cover a sufficient number of stripes. In practice, a roi of size 17×17 pixels and 2D third order polynomials are used.

4. 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 j^{th} 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 the following assumptions are made,

- P_w is measured without error.
- $P_p \sim \text{Normal}(F_c(P_w;\Theta_c), \sigma_p^2 I_{2n})$
- $P_s \sim \text{Normal}(F_p(P_w;\Theta_p), \sigma_s^2 I_n)$

then Maximum Likelihood estimation of $\{\Theta_c, \Theta_p\}$ gives rise to the following non-linear least squares (NLLS) problems,[11]

$$\min_{\Theta_c} \|P_p - F_c(P_w;\Theta_c)\|^2 \qquad \text{and} \qquad \min_{\Theta_p} \|P_s - F_p(P_w;\Theta_p)\|^2 \tag{17}$$

These problems can be solved using any general technique for NLLS problems such as Gauss-Newton or Levenberg-Marquardt.[10]

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[4] and projector.[9]

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 $p_p = (x_p \ y_p)^T$, 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,

$$\begin{pmatrix} p_p \\ p_s \end{pmatrix} - \begin{pmatrix} F_c(p_w;\Theta_c) \\ F_p(p_w;\Theta_p) \end{pmatrix} = 0 \tag{18}$$

This problem can be solved using any general technique such as a quasi-Newton strategy.[10]

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

$$\beta_p \begin{pmatrix} p_p \\ 1 \end{pmatrix} = C_c \begin{pmatrix} p_w \\ 1 \end{pmatrix} \qquad \beta_s \begin{pmatrix} p_s \\ 1 \end{pmatrix} = C_p \begin{pmatrix} p_w \\ 1 \end{pmatrix} \tag{19}$$

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

$$\begin{pmatrix} C_c^1 - x_p C_c^3 \\ C_c^2 - y_p C_c^3 \\ C_p^1 - x_s C_p^2 \end{pmatrix} \begin{pmatrix} \alpha p_w \\ \alpha \end{pmatrix} = 0 \tag{20}$$

A solution to (20) can be obtained (e.g. using SVD) and the estimate of p_w obtained from this homogeneous vector.

A computationally efficient solution for a large data set can be obtained as follows. A generalisation of the vector cross product for $n-1$ vectors in an n-dimension linear space can be defined implicitly by, $\langle w, v_1 \times \ldots \times v_{n-1} \rangle = \det(v_1, \ldots, v_{n-1}, w)$.[15] It is clear that $v_1 \times \ldots \times v_{n-1}$ can be written $\sum_k \det(v_1, \ldots, v_{n-1}, e_k) e_k$ and is orthogonal to v_k. Therefore a solution to (20) is given by the (generalised) cross product of the rows of the 3×4 matrix,

$$z = \sum_{k=1}^{4} \det(C_c^1 - x_p C_c^3,\ C_c^2 - y_p C_c^3,\ C_p^1 - x_s C_p^2,\ e_k) e_k \tag{21}$$

Using the linearity and antisymmetry of the determinant tensor, the k^{th} component of z is given by,

$$z^k = C_{1,2,1}^k - x_p C_{3,2,1}^k - y_p C_{1,3,1}^k - x_s C_{1,2,2}^k + x_s x_p C_{3,2,2}^k + x_s y_p C_{1,3,2}^k \tag{22}$$

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 512x512 images were captured. The calibration reference is a 150mm cube with 24 fiducial marks (of 5mm radius) on each face as shown in Figure 4. The 3D location of each fiducial mark centroid has been measured to an accuracy of 0.1mm. The nominal setup has the projector above the camera with a 12^o angle between their optic axes to give a working volume of 250mm diameter, at 1600mm in front of the camera.

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* consists 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. Figures 5 and 6 show the pixel residuals, stripe residuals and spatial intersection errors respectively, for trial 4.

Table 1 shows the average magnitude of the spatial intersection errors for all of the calibration trials. For comparison the average errors when the camera and projector models include no radial distortion coefficients and the average errors when no substripe estimation is used are also provided. Excluding either radial distortion or substripe estimation increases the error in all trials. The results show 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

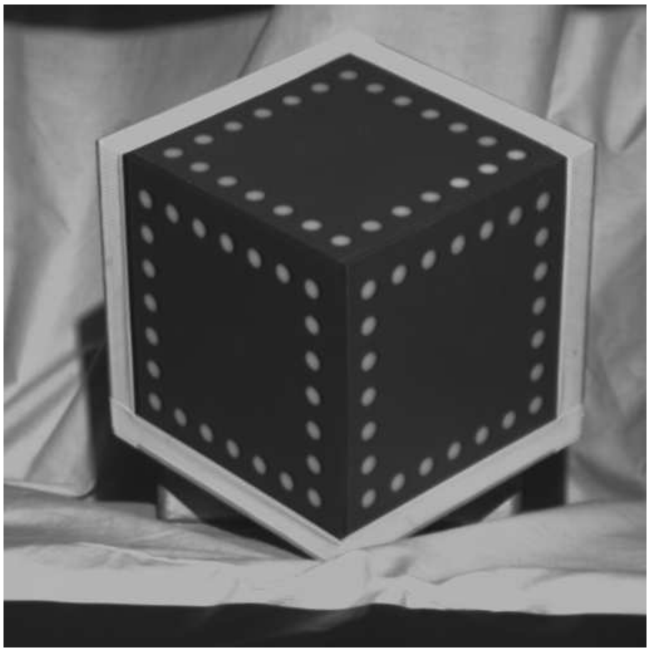

Figure 4: Calibration reference.

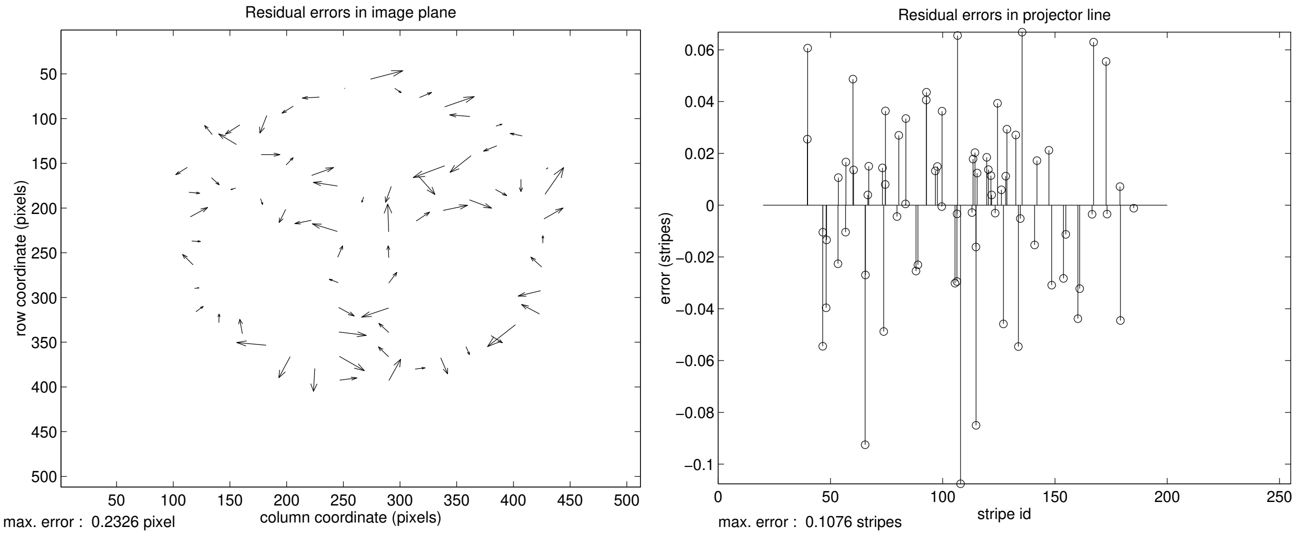

Figure 5: Pixel residual errors (left) and stripe residual errors (right).

trial	1	2	3	4	5	6	7
A	0.154	0.245	0.296	0.187	0.280	0.246	0.233
B	0.226	0.283	0.311	0.223	0.292	0.261	0.264
C	1.348	1.312	1.371	1.521	1.233	1.170	1.444

Table 1: Average magnitude of the spatial intersection error (mm) for (A) calibration trials, (B) trials with no radial distortion, (C) trials with no substripe estimation

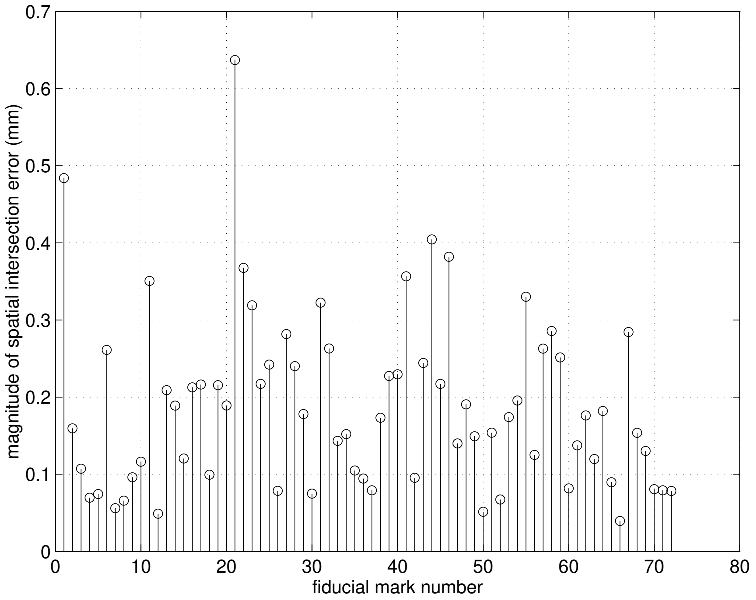

Figure 6: Spatial intersection errors.

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

Table 2: Average magnitude of the model fitting error (mm) for each calibration trial (horizontal) and each measured data set (vertical).

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.[6] The *model fitting error* is the difference between the estimated world coordinates and the transformed reference coordinates. Table 2 shows the average magnitude of the model fitting error for all of the calibration trials and observed data sets. The diagonal entries of Table 2 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 furtherest from the calibration data set.

Table 3 shows the intrinsic parameters for the SLS obtained from each trial. The results offer clear support to the manufacturers claim that the framegrabber aspect ratio is unity. The camera scale factors agree with the nominal value of 2200 pixels/mm and the camera principal point is consistent with the nominal value of $(255, 255)$ pixels. The shear coefficient is small with respect to s_x (0.14%) suggesting a very small contribution. The scale factor for the projector is close to the nominal value of 1020 stripes/mm. There are significant variations in the other parameters.

trial	1	2	3	4	5	6	7
s_x	2411.062	2399.667	2425.255	2418.480	2403.901	2443.054	2408.316
s_y	2412.349	2402.982	2425.084	2421.023	2407.875	2445.911	2410.448
c_x	289.389	283.201	301.069	274.053	271.295	314.980	272.129
c_y	271.802	303.288	258.478	275.559	309.616	261.080	259.293
k	4.214	2.316	4.308	3.285	1.938	4.006	3.136
K	-0.164	-0.176	-0.149	-0.139	-0.200	-0.219	0.035
s_x	899.157	1020.565	1151.871	1015.006	1061.155	1162.614	1059.352
c_x	159.906	240.831	130.128	183.441	233.480	132.600	181.811
K	-0.085	-0.123	0.317	-0.116	-0.138	0.384	-0.109

Table 3: The 6 intrinsic parameters for the camera followed by the 3 intrinsic parameters for the projector for each calibration trial.

7. CONCLUSION

This paper discussed a method for obtaining accurate 3D measurements using a temporally 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.

While some of the intrinsic parameters are reliably estimated, there is significant variation in the shear and radial distortion for the camera and the offset and radial distortion for the projector. However, the spatial intersection procedure is insensitive to these variations. Moreover, it is necessary to include radial distortion into the projector model in order to obtain good spatial data. This suggests that there is some redundancy in the parameterisation although it is not over-parameterised. Poor recovery of the projector parameters is also due to there being only weak information from which to estimate them (only 1 coordinate). In addition, the projector parameter T_2 is only included to allow distortion to be incorporated and is correlated with K and s_x.

The situation would 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 “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 substripe estimator are also possible by recognising the step-like structure of a stripe image. Other avenues for improving the performance include using a digital output pixel synchronous large format camera and a projector with more stripes.

8. REFERENCES

[1] B.F. Alexander and Kim Chew Ng. Elimination of systematic error in subpixel accuracy centroid estimation. *Optical Engineering*, 30(9), September 1991.

[2] H. A. Beyer. An introduction to photogrammetric camera calibration. In *Invited paper, Seminaire Orasis*, St. Malo, France, September 23-27 1991.

[3] Duane C. Brown. Close range camera calibration. *Photogrammetric Engineering*, 37(8):855–866, August 1971.

[4] O. Faugeras. *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, 1993.

[5] Ray Jarvis. Range sensing for computer vision. In A. K. Jain and P. J. Flynn, editors, *Three-Dimensional Object Recognition Systems*, pages 17–56. Elsevier Science Publishers, 1993.

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

[7] H. M. Karara, editor. *Non-Topographic Photogrametry*. American Society for Photogrammetry and Remote Sensing (ASPRS), second edition, 1989.

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

[9] Alan M. McIvor and Robert J. Valkenburg. Calibrating a Structured Light System. Report 362, Industrial Research Limited, Feburary 1995.

[10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C: the art of scientific computing*. Cambridge University Press, second edition, 1992.

[11] G. A. F. Seber and C. J. Wild. *Nonlinear Regression*. John Wiley and Sons, 1989.

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

[13] Yoshiaki Shirai. 3D computer vision and applications. In *11th International Conference on Pattern Recognition*, pages 236–245. International Association for Pattern Recognition, IEEE Computer Society Press, 1992.

[14] C. C. Slama, editor. *Manual of Photogrammetry*. American Society of Photogrammetry, 1980.

[15] Michael Spivak. *Calculus on manifolds: A modern approach to classical theorems of advanced calculus*. W. A. Benjamin, Inc., 1965.

[16] Thomas Stahs and Friedrich Wahl. Fast and versatile range data acquisition in a robot work cell. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1169–1174, July 7–10 1992.

[17] E. Trucco, R. B. Fisher, and A. W. Fitzgibbon. Direct calibration and data consistency in 3-D laser scanning. In Edwin Hancock, editor, *BMVC94: Proceedings of the 5th British Machine Vision Conference*, pages 489–498. BMVA Press, September 1994.

[18] R. J. Valkenburg, A. M. McIvor, and P. W. Power. An evaluation of subpixel feature localisation methods for precision measurement. In S. F. El-Halim, editor, *Videometrics III*, volume 2350, pages 229–238. SPIE, 1994.

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