Accurate 3D measurement using a Structured Light System

R.J. Valkenburg and A.M. McIvor

Machine Vision Team, Industrial Research Limited
P.O. Box 2225, Auckland
New Zealand
Ph: +64 9 3034116, Fax: +64 9 3028106
E-mail: b.valkenburg@irl.cri.nz, a.mcivor@irl.cri.nz
URL: http://kauri.auck.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.

1 Introduction

Structured light is a general concept and there are a number of ways of exploiting it to obtain 3D information e.g. [1]. The structured light system (SLS) considered in this paper consists of a projector and a camera. The projector projects a sequence of band patterns which together identify one of 256 stripe values at each visible point in the world. Given the parameters of the SLS, obtained during system calibration, each stripe value defines a plane in the world and each camera pixel defines a ray in the world. The intersection of this ray and plane defines a unique 3D location in the world.

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 second problem is addressed through the use of subpixel and substripe operators.

2 Model of SLS

This section describes models for the CCD camera and projector which comprise the SLS. More details are provided in [2]. First consider the camera model. 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 $\mathbf{R}$ 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. If $(\bar{u}\ \bar{v})^T$ denote the coordinates of p in $\mathbf{C}_i$ then $(\bar{u}\ \bar{v})^T$ are given by (2) where f is the *principal distance*. Radial lens distortion is incorporated into the model in the following way. If $(u\ v)^T$ is the 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 the radial distortion coefficient. 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).

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

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

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

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

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. Consequently, (3) and (4) are replaced by,

$$u = \bar{u}\left[1 + K(\bar{u}^2 + \bar{v}^2)\right] \tag{5}$$

$$x_s = s_x u + c_x \tag{6}$$

Denoting the camera and projector parameter vectors by Θ_c and Θ_p allows the transformations 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, to be expressed more succinctly as,

$$p_p = F_c(p_w; \Theta_c) \quad \text{and} \quad p_s = F_p(p_w; \Theta_p) \tag{7}$$

3 Localisation of coordinates

During calibration, the subpixel location of the centroids of the fiducial marks on the calibration reference are required. The greyscale centroid operator is used to estimate the centroids of the fiducial marks in the image. Centroid based techniques are robust to noise (because they are based on integration), robust to small amounts of aliasing (theoretically exact if the sampling frequency is greater than the Nyquist frequency) and computationally inexpensive. Unfortunately, the centroid of the image of the fiducial mark does not in general coincide with 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.

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. Then the difference between the image of the centroid and the centroid of the image of Ω is given by [2],

$$\epsilon_{\hat{p}_i} = \frac{f}{t_3}\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} - \frac{f}{t_3^2 + r^2(n_3^2 - 1)}\begin{pmatrix} t_1 t_3 + r^2 n_1 n_3 \\ t_2 t_3 + r^2 n_2 n_3 \end{pmatrix} \tag{8}$$

This error can be transformed into pixel coordinates using the scale factors and shear coefficient. A maximum error of 0.016 pixel was obtained for a typical calibration configuration[1]. 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 dictated by a feature in the world and the other where it can be freely chosen. The need to use subpixel estimation does not arise in the second situation, e.g. during the generation of a dense range map with image locations on the pixel lattice. The type of subpixel operator used during the first situation depends on the nature of feature (e.g. edges, ridges).

During calibration and spatial intersection the substripe values of specified world points are required. Refer to [2] for more details on the problem of substripe estimation. The array formed from the 8-bit stripe values at each pixel location 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. A substripe estimate, x_p, is obtained by interpolating a least squares polynomial facet fitted to the stripe image. The facet is fitted to the stripe image in a roi Ω centred at the subpixel location, (x_p, y_p), where x_s is required.

When the surface under consideration is flat, such as the faces of the calibration reference, 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 are related by a rational function (the composition of two projective maps). Hence a planar facet does not model perspective effects precisely. The errors can be estimated as described in [2]. A maximum error of 0.003 stripes was 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.

When the surface is not flat, 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 SLS calibration

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 and 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 μ_{P_p} and μ_{P_s} are the true values of P_p and P_s respectively, then the observations are given by $P_p = \mu_{P_p} + \epsilon_p$ and $P_s = \mu_{P_s} + \epsilon_s$. If it is assumed that P_w is measured without

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

error, $P_p \sim \mathrm{N}(\mu_{P_p}, \sigma_p^2 I_{2n})$ and $P_s \sim \mathrm{N}(\mu_{P_s}, \sigma_s^2 I_n)$ then Maximum Likelihood estimation of $\{\Theta_c, \Theta_p\}$ gives rise to the following non-linear least squares (NLLS) problems [3],

$$\min_{\Theta_c} \|P_p - F_c(P_w; \Theta_c)\|^2 \quad \text{and} \quad \min_{\Theta_p} \|P_s - F_p(P_w; \Theta_p)\|^2 \tag{9}$$

These problems can be solved using any general technique for NLLS problems such as Gauss-Newton or Levenberg-Marquardt [3], with initial estimates obtained by decomposing the PTMs for the camera [4] and projector.

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. 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,

$$\begin{pmatrix} p_p \\ p_s \end{pmatrix} - \begin{pmatrix} F_c(p_w; \Theta_c) \\ F_p(p_w; \Theta_p) \end{pmatrix} = 0 \tag{10}$$

This problem can be solved using any general technique (e.g. quasi-Newton [5]).

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. $\begin{pmatrix} p_p\ 1 \end{pmatrix}^T = C_c \begin{pmatrix} p_w\ 1 \end{pmatrix}^T$ and $\begin{pmatrix} p_s\ 1 \end{pmatrix}^T = C_p \begin{pmatrix} p_w\ 1 \end{pmatrix}^T$ can be rearranged into the 3×4 linear homogeneous system,

$$\begin{pmatrix} C_c^1 - x_p C_c^3 \\ C_c^2 - y_p C_c^3 \\ C_p^1 - x_s C_p^2 \end{pmatrix} \begin{pmatrix} \alpha p_w \\ \alpha \end{pmatrix} = 0 \tag{11}$$

A solution to (11) can be obtained (e.g. using SVD) and the estimate of p_w obtained from this homogeneous vector. When there is a large number of points to process substantial computational savings can be made in obtaining this initial estimate [2].

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. 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 as described in Section 4.

The *spatial intersection errors* are the difference between the reference coordinates of the fiducial marks and their coordinates estimated by spatial intersection. 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

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 in the SLS model and (C) trials with no substripe estimation.

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).

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 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.

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 in-

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.

corporate 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. Ideas for further improvements are described in [2]. 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 overparameterised. 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.

References

1. 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.

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

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

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

5. 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.

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