G. Taylor and L. Kleeman, “Stereoscopic Light Stripe Scanning: Interference Rejection, Error Minimization and Calibration”, International Journal Robotics Research, Vol 23 No 12, Dec 2004, pp 1141-1156.

Stereoscopic Light Stripe Scanning: Interference Rejection, Error Minimization and Calibration

Geoffrey Taylor and Lindsay Kleeman
Department of. Electrical. and Computer Systems Engineering
Monash University, Clayton 3800
Victoria, Australia
{Geoffrey.Taylor;Lindsay.Kleeman}@eng.monash.edu.au

Abstract

This paper addresses the problem of rejecting interference due to secondary specular reflections, cross-talk and other mechanisms in an active light stripe scanner for robotic applications. Conventional scanning methods control the environment to ensure the brightness of the stripe exceeds that of all other features. However, for a robot operating in an uncontrolled environment, valid measurements are likely to be hidden amongst noise. Robust scanning methods already exist, but suffer from problems including assumed scene structure, acquisition delay, lack of error recovery, and incorrect modelling of measurement noise. We propose a robust technique that overcomes these problems, using two cameras and knowledge of the light plane orientation to disambiguate the primary reflection from spurious measurements. Unlike other robust techniques, our validation and reconstruction algorithms are optimal with respect to sensor noise. Furthermore, we propose an image-based procedure to calibrate the system using measurements of an arbitrary non-planar target, providing robust validation independently of ranging accuracy. Finally, our robust techniques allow the sensor to operate in ambient indoor light, allowing colour and range to be implicitly registered. An experimental scanner demonstrates the effectiveness of the proposed techniques.

1 Introduction

Light stripe ranging is an active, triangulation-based technique for non-contact surface measurement that has been studied for several decades [22, 1]. By projecting a known feature onto the measured surface, active scanners provide a more robust solution to the measurement problem than passive ranging techniques. A review of light stripe scanning and related range sensing techniques can be found in [13, 4, 5]. Range sensing is an important component of many robotic applications, and light stripe ranging has been applied to a variety of tasks including navigation [19, 2], obstacle detection [12], object recognition for grasping [3, 21], and visual servoing [7].

The drawback of conventional single-camera light stripe ranging techniques is that favourable lighting conditions and surface reflectance properties are required, which allow the stripe to be identified as the brightest feature in the captured image. However, the range sensor presented in this paper is intended for use on a humanoid robot operating in an uncontrolled domestic environment [24, 25]. Under these conditions, various noise mechanisms interfere with the sensor to defeat conventional stripe detection techniques: smooth surfaces cause secondary reflections, edges and textures may have a stripe-like appearance, and cross-talk can arise when multiple robots scan the same environment. The motivation for this work was to develop a robust light stripe sensor suitable for operation in these noisy conditions.

A number of techniques for improving the robustness of light stripe scanners have been proposed in other work, using both stereo and single-camera configurations. Magee *et al.* [16] develop a scanner for industrial inspection using stereo measurements of a single stripe. Spurious reflections are eliminated by combining stereo fields via a minimum intensity operation. This technique depends heavily on user intervention and *a priori* knowledge of the scanned target. Trucco *et al.* [27]

also use stereo cameras to measure a laser stripe, and treat the system as two independent single-camera sensors. Robustness is achieved by imposing a number of consistency checks to validate the range data, the most significant of which requires independent single-camera reconstructions to agree within a threshold distance. Another constraint requires valid scanlines to contain only a single stripe candidate, but a method for error recovery in the case of multiple candidates is not proposed. Thus, secondary reflections cause both the true and noisy measurements to be rejected.

Nakano *et al.* [18] develop a similar method to reject false data by requiring consensus between independent scanners, but using two laser stripes and only a single camera. In addition to robust ranging, this configuration provides direct measurement of the surface normal. The disadvantage of this approach is that each image only recovers a single range point at the intersection of the two stripes, resulting in a significant acquisition delay for the complete image.

Other robust scanning techniques have been proposed using single-camera, single-stripe configurations. Nygards and Wernersson [20] identify specular reflections by moving the scanner relative to the scene and analyzing the motion of reconstructed range data. In [12], periodic intensity modulation distinguishes the stripe from random noise. Both of these methods require data to be associated between multiple image, which is prone to error. Furthermore, intensity modulation does not disambiguate secondary reflections, which vary in unison with the true stripe. Alternatively, Clark *et al.* [6] use linearly polarized light to reject secondary reflections from metallic surfaces, based on the observation that polarized light changes phase with each specular reflection. However, the complicated acquisition process requires multiple measurements through different polarizing filters.

Unlike the above robust techniques, the method proposed in this paper uniformly rejects interference due to secondary reflections, cross-talk, background features and other noise mechanisms. The proposed algorithm solves the association problem to allow the true stripe to be disambiguated from a set of noisy candidates. Thus, our method offers a significant improvement over previous techniques for error detection, by providing both noise rejection and *recovery* of valid measurements. The depth data reconstructed by our stereo scanner is *optimal* with respect to sensor noise (unlike the stereo techniques in [18, 27]), and fuses stereo measurements with the light plane parameters. Thus, our sensor provides greater precision than a single-camera configuration, as demonstrated experimentally. Furthermore, we develop an image-based calibration procedure using a scan of a non-planar but otherwise arbitrary surface. Our calibration procedure provides robust validation independently of ranging accuracy. Finally, the ability of our sensor to operate in ambient indoor light allows surface colour and range to be captured in the same cameras. Since the intended application is to support object modelling and tracking [24], this implicit registration of colour and range is highly desirable. The system described in this work was first proposed in [26], and is extended here by providing a more rigorous treatment with respect to measurement noise.

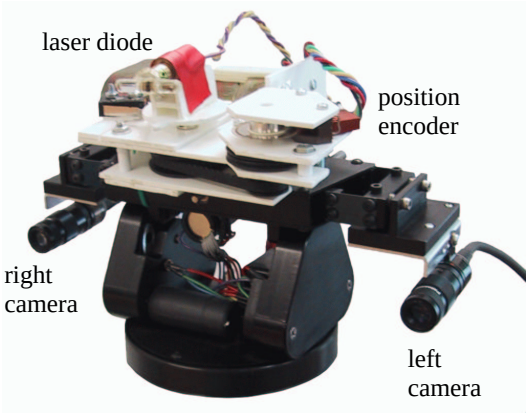


Figure 1: Experimental stereoscopic light stripe scanner

The following Section briefly describes the hardware configuration of our stereo stripe scanner. Section 3 develops the underlying framework for optimal noise rejection and reconstruction. Image-based calibration of the system is described in Section 4. Implementation details are discussed in Section 5, and Section 6 presents experimental results to compare the performance of our system with other stripe scanning techniques.

2 Overview

2.1 Basic Operation

Figure 1 shows the components of the experimental stereoscopic light stripe scanner, which is mounted on a pan/tilt active head. A vertical light plane is generated by a laser diode module with a cylindrical lens, and is

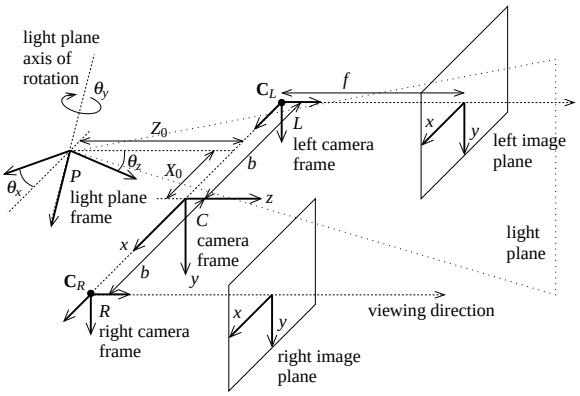


Figure 2: Light stripe camera system model.

scanned across a scene by rotating the laser about a vertical axis. The angle of rotation is measured by an optical encoder connected to the output shaft via a toothed belt. PAL colour cameras capture stereo images of the stripe at 384×288 pixel (half-PAL) resolution.

Captured image are processed to extract all possible locations of the light stripe on each scanline. Scanlines typically contain multiple candidates, with the true stripe (primary reflection) hidden amongst spurious measurements. The true stripe is identified on each scanline by searching for the pair of stereo measurements most likely corresponding to the projection of a 3D point on the light plane, given the current encoder measurement. The validation process is detailed in Section 3. Finally, the validated measurements are used to reconstruct the 3D profile of the illuminated surface.

As the stripe is scanned across the scene, the laser profiles are assembled into an array of 3D points, which is referred to as a *range map*. A colour image is captured and registered with the range map at the completion of a scan. Captured frames are processed at PAL frame-rate (25 Hz) on the 2.2 GHz dual Xeon host PC. Motor control and optical encoder measurements are implemented on a PIC microcontroller, which communicates with the host PC via an RS-232 serial link.

2.2 Coordinate Frames and Notation

The following sections adopt the convention of representing 3D points in upper-case, such as $\mathbf{X}$, and 3D points on the image plane using lower-case, such as $\mathbf{x}$, with all vectors in *homogeneous* coordinates. Coordinate frames are specified in super-script, such as ${}^{A}\mathbf{X}$, and the homogeneous transformation matrix ${}^{B}\mathrm{H}_{A}$ transforms points from frame A to frame B as ${}^{B}\mathbf{X} = {}^{B}\mathrm{H}_{A}\,{}^{A}\mathbf{X}$. Figure 2 shows the relevant coordinate frames for the experimental scanner: L and R denote the left and right camera frames, P is rigidly attached to the light plane and W is the world frame.

The cameras are modelled by the 3×4 projection matrices ${}^{L}\mathrm{P}$ and ${}^{R}\mathrm{P}$, which project points from the world frame onto the image planes according to the homogeneous transformation ${}^{L,R}\mathbf{x} = {}^{L,R}\mathrm{P}\,{}^{W}\mathbf{X}$. The cameras are arranged in a rectilinear stereo configuration (parallel axes and coplanar image planes), with optical centres at $\mathbf{C}_{L,R} = (\mp b, 0, 0)^{\top}$ (taking the negative sign for L and positive for R) in the world frame. In practice, the cameras can verge about the y-axis (violating the conditions of rectilinear stereo) but rectilinear measurements are recovered by applying *projective rectification* to captured frames (see [10]). Thus, without loss of generality, the analytical models consider only the rectilinear arrangement shown in Figure 2.

2.3 System Model

A theoretical model of the experimental scanner is now developed to identify points on the light plane. Since the light plane is fixed in P, points ${}^{P}\mathbf{X}$ on the plane may be represented by the plane equation ${}^{P}\boldsymbol{\Omega}^{\top}\,{}^{P}\mathbf{X} = 0$, where ${}^{P}\boldsymbol{\Omega}$ represents the parameters of the light plane. Now, P is defined such that the light plane is approximately vertical and parallel to the z-axis, which allows ${}^{P}\boldsymbol{\Omega}$ to be expressed as

$$ {}^{P}\boldsymbol{\Omega} = (1,\ B_0,\ 0,\ D_0)^{\top} \tag{1} $$

where B_0 is related to the small angle between the plane and the y-axis ($B_0 \ll 1$ for an approximately vertical plane), and D_0 is the distance of the plane from the origin. During a scan, frame P rotates about its y-axis with angle θ_y, where $\theta_y \triangleq 0$ when the light plane is parallel to the optical axes of the cameras. The rotation axis intersects the xz-plane of the world frame at $(X_0, 0, Z_0)^{\top}$, and the orientation of the rotation axis relative to the y-axis of the world frame is defined by the small fixed angles θ_x and θ_z. The scan angle (θ_y in Figure 2) is linearly related to the measured optical encoder value e via two additional parameters m and k:

$$ \theta_y = me + k \tag{2} $$

Now let ${}^{W}\mathrm{H}_{P}$ represent the homogeneous coordinate transformation from P to W. If $\mathrm{R}_x(\theta_x)$, $\mathrm{R}_y(\theta_y)$

and $\mathrm{R}_z(\theta_z)$ represent the homogeneous transformations for rotations about the x, y and z-axes, and $\mathrm{T}(X_0, 0, Z_0)$ represents the transformation for translation by $(X_0, 0, Z_0)^\top$, ${}^W\mathrm{H}_P$ can be expressed as

$$ {}^W\mathrm{H}_P = \mathrm{T}(X_0, 0, Z_0) \cdot \mathrm{R}_z(\theta_z) \cdot \mathrm{R}_x(\theta_x) \cdot \mathrm{R}_y(\theta_y) \qquad (3) $$

It is straightforward to show that if ${}^W\mathrm{H}_P$ is the coordinate transformation from P to W, the plane parameters transform from P to W as

$$ {}^W\mathbf{\Omega} = ({}^W\mathrm{H}_P)^{-\top} \cdot {}^P\mathbf{\Omega} \qquad (4) $$

Finally, combining equations (3) and (4) and making the simplifying assumptions $B_0, \theta_x, \theta_z \ll 1$ to eliminate insignificant terms, the laser plane parameters can be expressed in the world frame by the approximate model:

$$ {}^W\mathbf{\Omega} \approx \begin{pmatrix} \cos\theta_y \\ \theta_x \sin\theta_y + \theta_z \cos\theta_y + B_0 \\ -\sin\theta_y \\ -X_0 \cos\theta_y + Z_0 \sin\theta_y + D_0 \end{pmatrix} \qquad (5) $$

where $c_i = \cos\theta_i$ and $s_i = \sin\theta_i$, for $i = x, y, z$. Now, using equations (2) and (5), points in the world frame can be identified as coincident with the light plane (given the system parameters and current encoder measurement) using the plane equation:

$$ {}^W\mathbf{\Omega}^\top (B_0, D_0, X_0, Z_0, \theta_x, \theta_y, \theta_z)\, {}^W\mathbf{X} = 0 \qquad (6) $$

3 Robust Stripe Scanning

3.1 Problem Statement

As discussed earlier, the problems associated with light stripe scanners result from ambiguity in identifying the primary reflection. The following sections describe an optimal strategy to resolve this ambiguity and robustly identify the stripe in the presence of secondary reflections, cross-talk and other sources of noise by exploiting redundancy.

Figure 3 shows a simplified plan view of the scanner to demonstrate the issues involved in robust stripe detection. The surface at $\mathbf{X}$ causes a primary reflection that is measured (using a noisy process) at ${}^L\mathbf{x}$ and ${}^R\mathbf{x}$ on the stereo image planes. However, a secondary specular reflection causes another stripe to appear at $\mathbf{X}'$, which is measured on the right image plane at ${}^R\mathbf{x}'$ but obscured from the left camera (in practice, such noisy measurements are produced by a variety of mechanisms other

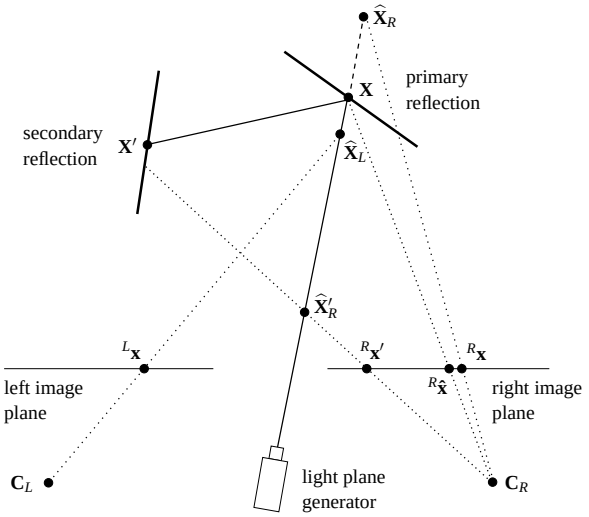


Figure 3: Validation/reconstruction problem.

than secondary reflections). The 3D reconstruction for each measurement, labelled $\widehat{\mathbf{X}}_L$, $\widehat{\mathbf{X}}_R$ and $\widehat{\mathbf{X}}'_R$ in Figure 3, is recovered as the intersection of the light plane and a ray back-projected through the image plane measurements. These points will be referred to as the *single-camera reconstructions*. As a result of noise on the CCD (exaggerated in this example), the back-projected rays do not intersect the physical reflections at $\mathbf{X}$ and $\mathbf{X}'$.

The robust scanning problem may now be stated as follows: given the laser plane position and the measurements ${}^L\mathbf{x}$, ${}^R\mathbf{x}$ and ${}^R\mathbf{x}'$, one of the left/right candidate pairs, $({}^L\mathbf{x}, {}^R\mathbf{x})$ or $({}^L\mathbf{x}, {}^R\mathbf{x}')$, must be chosen as representing stereo measurements of the primary reflection. Alternatively, all candidates may be rejected. This task is referred to as the *validation problem*, and a successful solution in this example should identify $({}^L\mathbf{x}, {}^R\mathbf{x})$ as the valid measurements. The measurements should then be combined to estimate the position of the ideal projection ${}^R\hat{\mathbf{x}}$ (arbitrarily chosen to be on the right image plane) of the actual point $\mathbf{X}$.

Formulation of optimal validation/reconstruction algorithms should take account of measurement noise, which is not correctly modelled in previous related work. In [27] and [18], laser stripe measurements are validated by applying a *fixed* threshold to the difference between corresponding single-camera reconstructions ($\widehat{\mathbf{X}}_L$, $\widehat{\mathbf{X}}_R$ and $\widehat{\mathbf{X}}'_R$ in Figure 3). Such a comparison requires a uniform reconstruction error over all depths,

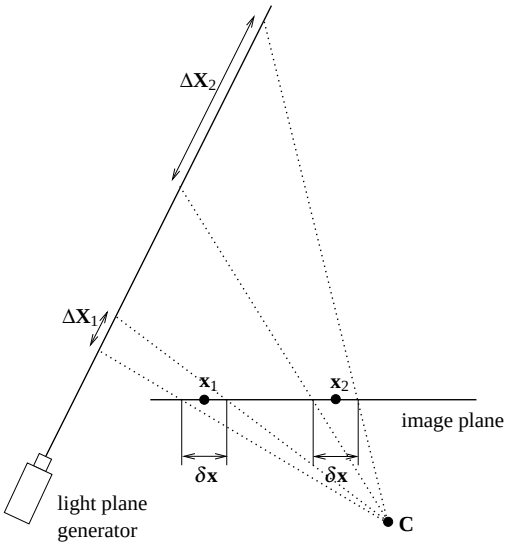


Figure 4: Variation of single-camera reconstruction error with depth.

which Figure 4 illustrates is clearly not the case. Two independent measurements at $\mathbf{x}_1$ and $\mathbf{x}_2$ generally exhibit a constant error variance on the image plane, as indicated by the interval $\delta\mathbf{x}$. However, projecting $\delta\mathbf{x}$ onto the laser plane reveals that the reconstruction error increases with depth, since $\Delta\mathbf{X}_1 < \Delta\mathbf{X}_2$. Thus, the validation threshold on depth difference should increase with depth to account for measurement noise, otherwise validation is more lenient for closer reconstructions. Similarly, taking either $\widehat{\mathbf{X}}_L$, $\widehat{\mathbf{X}}_R$ or the arithmetic average $\frac{1}{2}(\widehat{\mathbf{X}}_L + \widehat{\mathbf{X}}_R)$ as the final reconstruction is generally suboptimal for noisy measurements.

The following sections present optimal solutions to the validation/reconstruction problem, based on an error model with the following features (the assumptions of the error model are corroborated with experimental results in Section 6.2):

1. Light stripe measurement errors are independent and Gaussian distributed with uniform variance over the entire image plane.

2. The dominant error in the laser plane position is the angular error about the axis of rotation due to the shaft encoder (e in equation (2)), which couples all the image plane measurements in a given frame.

3. All other parameters described in Section 2.3 (apart from the visual measurements and shaft encoder position) are assumed to be known with sufficient accuracy that any uncertainty can be ignored for the purpose of validation and reconstruction.

3.2 General Solution

We now present a general analytical solution to the validation/reconstruction problem. Let ${}^L\mathbf{x}$ and ${}^R\mathbf{x}$ represent noisy *candidate* measurements of the stripe on corresponding epipolar lines, and let $\mathbf{\Omega} = (A,B,C,D)^\top$ represent the parameters of the light plane. The *cost* of associating the measurements with the primary reflection of the stripe can be formulated as an error distance on the image plane (similar to the reconstruction method proposed in [11]). Let $\widehat{\mathbf{X}}$ represent the minimum mean squared error 3D reconstruction, which is constrained to coincide with the laser plane, but not necessarily at the intersection of the back-projected rays from the noisy image plane measurements. Errors in the light plane parameters $\mathbf{\Omega}$ are considered in the next section, but for now the light plane parameters are assumed to be known exactly. The optimal reconstruction $\widehat{\mathbf{X}}$ corresponds to the *ideal* projections, ${}^L\hat{\mathbf{x}}$ and ${}^R\hat{\mathbf{x}}$, on the left and right image planes according to ${}^{L,R}\hat{\mathbf{x}} = {}^{L,R}\mathrm{P}\widehat{\mathbf{X}}$. Now, the sum of squared errors between the ideal and measured points can be used to determine whether the candidate measurement pair ($\mathbf{x}_L$,$\mathbf{x}_R$) corresponds to a point $\widehat{\mathbf{X}}$ on the light plane:

$$E = d^2({}^L\mathbf{x}, {}^L\mathrm{P}\widehat{\mathbf{X}}) + d^2({}^R\mathbf{x}, {}^R\mathrm{P}\widehat{\mathbf{X}}) \tag{7}$$

where $d(\mathbf{x}_1,\mathbf{x}_2)$ is the Euclidean distance between $\mathbf{x}_1$ and $\mathbf{x}_2$. For a given candidate pair, the optimal reconstruction $\widehat{\mathbf{X}}$ with respect to image plane error is found by a constrained minimization of E with respect to the condition that $\widehat{\mathbf{X}}$ is on the laser plane:

$$\mathbf{\Omega}^\top \widehat{\mathbf{X}} = 0 \tag{8}$$

When multiple ambiguous correspondences exist, equation (7) is optimized with respect to the constraint in (8) for all possible candidate pairs, and the pair with minimum error is chosen as the most likely correspondence. Finally, the result is validated by imposing a threshold on the maximum allowed squared image plane error E.

Performing the constrained optimization of equations (7)-(8) is analytically cumbersome. Fortunately, the problem may be reduced to an unconstrained optimization by formulating the ideal left projection ${}^L\hat{\mathbf{x}}$ as a function of the ideal right projection ${}^R\hat{\mathbf{x}}$ for points on the

light plane. To determine this relationship, the 3D reconstruction is first expressed as a function of ${}^R\hat{\mathbf{x}}$ by taking the intersection of the laser plane and the ray back-projected from ${}^R\hat{\mathbf{x}}$. Plücker matrices provide a concise notation for the intersection of planes and lines (see [9]). If $\mathbf{A}$ and $\mathbf{B}$ represent the homogeneous vectors of two points on a line, the Plücker matrix L describing the line is

$$\mathrm{L} = \mathbf{A}\mathbf{B}^\top - \mathbf{B}\mathbf{A}^\top \tag{9}$$

Then, the intersection $\mathbf{X}$ of a plane $\mathbf{\Omega}$ and the line described by L is simply

$$\mathbf{X} = \mathrm{L}\mathbf{\Omega} = (\mathbf{A}\mathbf{B}^\top - \mathbf{B}\mathbf{A}^\top)\mathbf{\Omega} \tag{10}$$

The Plücker matrix for the back-projected ray from ${}^R\hat{\mathbf{x}}$ can now be constructed from two known points on the ray: the camera centre $\mathbf{C}_R$, and ${}^R\mathrm{P}^{+R}\hat{\mathbf{x}}$, where ${}^R\mathrm{P}^+$ is the pseudo-inverse of the camera projection matrix ${}^R\mathrm{P}$ such that ${}^R\mathrm{P}({}^R\mathrm{P}^{+R}\hat{\mathbf{x}}) = {}^R\hat{\mathbf{x}}$, given by

$${}^R\mathrm{P}^+ = {}^R\mathrm{P}^\top({}^R\mathrm{P}{}^R\mathrm{P}^\top)^{-1} \tag{11}$$

Applying these to equation (9), the Plücker matrix for the back-projection of ${}^R\hat{\mathbf{x}}$ is

$$\mathrm{L} = \mathbf{C}_R({}^R\mathrm{P}^{+R}\hat{\mathbf{x}})^\top - ({}^R\mathrm{P}^{+R}\hat{\mathbf{x}})\mathbf{C}_R^\top \tag{12}$$

The intersection of L with the laser plane $\mathbf{\Omega}$, can now be expressed using equation (10) as:

$$\widehat{\mathbf{X}} = \mathrm{L}\mathbf{\Omega} = [\mathbf{C}_R({}^R\mathrm{P}^{+R}\hat{\mathbf{x}})^\top - ({}^R\mathrm{P}^{+R}\hat{\mathbf{x}})\mathbf{C}_R^\top]\mathbf{\Omega} \tag{13}$$

Finally, the ideal projection ${}^L\hat{\mathbf{x}}$ corresponding to ${}^R\hat{\mathbf{x}}$ is obtained by projecting $\widehat{\mathbf{X}}$ onto the left image plane:

$$\begin{aligned} {}^L\hat{\mathbf{x}} &= {}^L\mathrm{P}\widehat{\mathbf{X}} \\ &= {}^L\mathrm{P}[\mathbf{C}_R({}^R\mathrm{P}^{+R}\hat{\mathbf{x}})^\top - ({}^R\mathrm{P}^{+R}\hat{\mathbf{x}})\mathbf{C}_R^\top]\mathbf{\Omega} \\ &= {}^L\mathrm{P}\mathbf{C}_R({}^R\mathrm{P}^{+R}\hat{\mathbf{x}})^\top\mathbf{\Omega} - {}^L\mathrm{P}({}^R\mathrm{P}^{+R}\hat{\mathbf{x}})(\mathbf{C}_R^\top\mathbf{\Omega}) \end{aligned}$$

Using the identity $({}^R\mathrm{P}^{+R}\hat{\mathbf{x}})^\top\mathbf{\Omega} = \mathbf{\Omega}^\top({}^R\mathrm{P}^{+R}\hat{\mathbf{x}})$, and noting that $(\mathbf{C}_R^\top\mathbf{\Omega})$ is scalar, the common factors and collected to simplify the above expression to

$$\begin{aligned} {}^L\hat{\mathbf{x}} &= {}^L\mathrm{P}(\mathbf{C}_R\mathbf{\Omega}^\top)({}^R\mathrm{P}^{+R}\hat{\mathbf{x}}) - {}^L\mathrm{P}(\mathbf{C}_R^\top\mathbf{\Omega})({}^R\mathrm{P}^{+R}\hat{\mathbf{x}}) \\ &= \left({}^L\mathrm{P}[\mathbf{C}_R\mathbf{\Omega}^\top - (\mathbf{C}_R^\top\mathbf{\Omega})\mathrm{I}]{}^R\mathrm{P}^+\right){}^R\hat{\mathbf{x}} \end{aligned} \tag{14}$$

Equation (14) is of the form ${}^L\hat{\mathbf{x}} = \mathrm{H}{}^R\hat{\mathbf{x}}$ and simply states that points on the laser plane induce a homography between coordinates on the left and right image planes, which is consistent with known results [9]. Finally, the error function becomes

$$E = d^2({}^L\mathbf{x}, \mathrm{H}{}^R\hat{\mathbf{x}}) + d^2({}^R\mathbf{x}, {}^R\hat{\mathbf{x}}) \tag{15}$$

where $\mathrm{H} = {}^L\mathrm{P}[\mathbf{C}_R\mathbf{\Omega}^\top - (\mathbf{C}_R^\top\mathbf{\Omega})\mathrm{I}]{}^R\mathrm{P}^+$. The reconstruction problem can now be formulated as an unconstrained optimization of equation (15) with respect to ${}^R\hat{\mathbf{x}}$. The validation problem is solved by finding the pair of candidates on each scanline with the minimum squared error E, and the optimal reconstruction $\widehat{\mathbf{X}}$ is recovered from equation (13) and the ideal projection ${}^R\hat{\mathbf{x}}$.

3.3 Special Case: Rectilinear Stereo and Pin-Hole Cameras

The results of the previous section apply to general camera models and stereo geometry. However, the special case of rectilinear stereo and pin-hole cameras is important as it reduces equation (15) to a single degree of freedom. Furthermore, rectilinear stereo applies without loss of generality (after projective rectification), and the pin-hole model is a good approximation for CCD cameras (after correcting for radial lens distortion). The stereo cameras used in this work are assumed to have unit aspect ratio and no skew, and the pin-hole models are parameterized by identical focal length f. For rectilinear stereo cameras with optical centres at $C_{L,R} = (\mp b, 0, 0, 1)^\top$, the projection matrices ${}^{L,R}\mathrm{P}$ are given by

$${}^{L,R}\mathrm{P} = \begin{pmatrix} f & 0 & 0 & \pm fb \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{16}$$

where the positive sign is taken for L and the negative for R. Now, substituting equation (16) and $C_{L,R} = (\mp b, 0, 0, 1)^\top$ into equation (14), the homography between the projections of a point on the laser plane can be expressed as:

$${}^L\hat{\mathbf{x}} = \begin{pmatrix} Ab - D & 2Bb & 2Cbf \\ 0 & -(Ab + D) & 0 \\ 0 & 0 & -(Ab + D) \end{pmatrix} {}^R\hat{\mathbf{x}} \tag{17}$$

With ${}^R\hat{\mathbf{x}} = ({}^R\hat{x}, {}^R\hat{y}, 1)^\mathrm{T}$ and ${}^L\hat{\mathbf{x}} = ({}^L\hat{x}, {}^L\hat{y}, w)^\mathrm{T}$, the above transformation can be evaluated as

$$\begin{pmatrix} {}^L\hat{x} \\ {}^L\hat{y} \\ w \end{pmatrix} = \begin{pmatrix} (Ab - D){}^R\hat{x} + 2Bb{}^R\hat{y} + 2Cbf \\ -(Ab + D){}^R\hat{y} \\ -(Ab + D) \end{pmatrix} \tag{18}$$

Expressed in inhomogeneous coordinates, the relationship between ${}^L\hat{\mathbf{x}}$ and ${}^R\hat{\mathbf{x}}$ is

$$ {}^L\hat{x} = -\frac{(Ab-D){}^R\hat{x}+2Bb{}^R\hat{y}+2Cbf}{Ab+D} \quad (19) $$

$$ {}^L\hat{y} = {}^R\hat{y} \quad (20) $$

Since the axes of L and R are parallel (rectilinear stereo), the notation $\hat{y} \triangleq {}^L\hat{y} = {}^R\hat{y}$ replaces equation (20). Rectilinear stereo gives rise to epipolar lines that are parallel to the x-axis, so the validation algorithm need only consider possible correspondences on matching scanlines in the stereo images. Any measurement error in the stripe detection process (see Section 5.1) is assumed to be in the x-direction only, while the y-coordinate is fixed by the height of the scanline. Thus, the y-coordinate of the optimal projections are also fixed by the scanline, ie. $\hat{y} = y$, where y is the y-coordinate of the candidate measurements ${}^L\mathbf{x}$ and ${}^R\mathbf{x}$.

Finally, substituting equations (19)-(20) with $\hat{y} = y$ into (15), the image plane error E can be expressed as a function of a *single* unknown, ${}^R\hat{x}$:

$$ E = ({}^Lx+\alpha{}^R\hat{x}+\beta y+\gamma f)^2+({}^Rx-{}^R\hat{x})^2 \quad (21) $$

where the following change of variables in the plane parameters is introduced (noting that $\mathbf{\Omega}$ has only three degrees of freedom due to the unconstrained scale):

$$ \alpha = (Ab-D)/(Ab+D) \quad (22) $$

$$ \beta = 2Bb/(Ab+D) \quad (23) $$

$$ \gamma = 2Cb/(Ab+D) \quad (24) $$

For the experimental scanner, with $\mathbf{\Omega}$ given by equation (5), α, β and γ can be written as:

$$ \alpha = -\frac{k_1\cos\theta_y+k_2\sin\theta_y+k_3}{\cos\theta_y+k_2\sin\theta_y+k_3} \quad (25) $$

$$ \beta = \frac{(1-k_1)(\theta_x\sin\theta_y+\theta_z\cos\theta_y+B_0)}{\cos\theta_y+k_2\sin\theta_y+k_3} \quad (26) $$

$$ \gamma = \frac{(k_1-1)\sin\theta_y}{\cos\theta_y+k_2\sin\theta_y+k_3} \quad (27) $$

where $\theta_y = me+c$ and the following change of variables is made in the system parameters:

$$ k_1 = -(b+X_0)/(b-X_0) \quad (28) $$

$$ k_2 = Z_0/(b-X_0) \quad (29) $$

$$ k_3 = D_0/(b-X_0) \quad (30) $$

Optimization of equation (21) now proceeds using standard techniques, setting $\frac{dE}{d{}^R\hat{x}} = 0$ and solving for ${}^R\hat{x}$. Let ${}^R\hat{x}^*$ represent the optimal projection resulting in the minimum squared error, E^*. It is straightforward to show that the optimal projection is given by

$$ {}^R\hat{x}^* = [{}^Rx-\alpha({}^Lx+\beta y+\gamma f)]/(\alpha^2+1) \quad (31) $$

and the corresponding minimum squared error E^* is:

$$ E^* = ({}^Lx+\alpha{}^Rx+\beta y+\gamma f)^2/(\alpha^2+1) \quad (32) $$

For completion, substituting equation (31) and ${}^R\hat{y}^* = y$ into (19) gives the corresponding optimal projection on the left image plane as

$$ {}^L\hat{x}^* = [\alpha^2{}^Lx-\alpha{}^Rx-(\beta y+\gamma f)]/(\alpha^2+1) \quad (33) $$

Finally, the optimal 3D reconstruction $\widehat{\mathbf{X}}^*$ is recovered from equation (13). Evaluating (13) for rectilinear, pinhole cameras with the change of variables in equations (22)-(24), and expressing the result in inhomogeneous coordinates, the relationship between $\widehat{\mathbf{X}}^*$ and the optimal projection ${}^R\hat{x}^*$ is:

$$ \begin{pmatrix} \widehat{X}^* \\ \widehat{Y}^* \\ \widehat{Z}^* \end{pmatrix} = \begin{pmatrix} \frac{b[\beta y+\gamma f+(\alpha-1){}^R\hat{x}^*]}{(\alpha+1){}^R\hat{x}^*+\beta y+\gamma f} \\ \frac{-2by}{(\alpha+1){}^R\hat{x}^*+\beta y+\gamma f} \\ \frac{-2bf}{(\alpha+1){}^R\hat{x}^*+\beta y+\gamma f} \end{pmatrix} \quad (34) $$

Finally, substituting ${}^R\hat{x}^*$ from equation (33) into the above, the optimal reconstruction from candidate measurements ${}^Lx$ and ${}^Rx$ on the scanline at height y is:

$$ \begin{pmatrix} \widehat{X}^* \\ \widehat{Y}^* \\ \widehat{Z}^* \end{pmatrix} = \begin{pmatrix} \frac{[(\alpha-1)(\alpha{}^Lx-{}^Rx)-(\alpha+1)(\beta y+\gamma f)]b}{(\alpha+1)(\alpha{}^Lx-{}^Rx)+(\alpha-1)(\beta y+\gamma f)} \\ \frac{2by(\alpha^2+1)}{(\alpha+1)(\alpha{}^Lx-{}^Rx)+(\alpha-1)(\beta y+\gamma f)} \\ \frac{2bf(\alpha^2+1)}{(\alpha+1)(\alpha{}^Lx-{}^Rx)+(\alpha-1)(\beta y+\gamma f)} \end{pmatrix} \quad (35) $$

Now, let ${}^Lx_i$, $i = 1\ldots n_L$, and ${}^Rx_j$, $j = 1\ldots n_R$ represent candidate measurements of the stripe on corresponding scanlines at height y. Furthermore, let e represent the current measured encoder value for the scanner (see Section 3.4). The validation problem, that is, determining which pair of measurements correspond to the primary reflection, can now be solved as follows:

1. Light plane parameters α, β and γ are calculated from e and the system parameters using equations (25)-(27) and (2).

2. For every possible pair of measurements $({}^{L}x_i, {}^{R}x_j)$, the optimal reconstruction error E^* is calculated from equation (32). Then, the pair $({}^{L}x_i^*, {}^{R}x_j^*)$ with the minimal reconstruction error are chosen as the most likely valid measurements, given by:

$$({}^{L}x_i^*, {}^{R}x_j^*) = \arg\min_{({}^{L}x_i, {}^{R}x_j)} E^*({}^{L}x_i, {}^{R}x_j, \alpha, \beta, \gamma) \quad (36)$$

3. A final test, $E^*({}^{L}x_i^*, {}^{R}x_j^*, \alpha, \beta, \gamma) < E_{th}$, ensures that the optimal candidates are valid, where E_{th} is calculated off-line from the expected measurement error. If this test is violated, no reconstruction is recovered in current frame.

4. For valid measurements, the optimal reconstruction $\widehat{\mathbf{X}}^*({}^{L}x_i^*, {}^{R}x_j^*)$ is finally calculated from equation (35). The reconstructions from all scanlines in the current frame are added to the range map.

3.4 Laser Plane Error

The above solution is optimal with respect to the error of image plane measurements, and assumes that the parameters of the laser plane are known exactly. In practice, the encoder measurements are likely to suffer from both random and systematic error due to acquisition delay and quantization. Unlike the image plane error, the encoder error is constant for all stripe measurements in a given frame and thus cannot be minimized independently for candidate measurements on each scanline.

Let ${}^{L}\mathbf{x}_i$ and ${}^{R}\mathbf{x}_i$, $i = 1 \ldots n$ represent valid corresponding measurements of the laser stripe on the n scanlines in a frame. The reconstruction error $E_i^*(e)$ for each pair can be treated as a function of the encoder count via the system model in equations (25)-(27). The total error $E_{\text{tot}}^*(e)$ over all scanlines for a given encoder count e is calculated as:

$$E_{\text{tot}}^*(e) = \sum_{i=1}^{n} E_i^*(e) \quad (37)$$

Finally, an optimal estimate of the encoder count e^* is calculated from the minimization

$$e^* = \arg\min_{e} [E_{\text{tot}}^*(e)] \quad (38)$$

Since $E_{\text{tot}}^*(e)$ is a non-linear function, equation (38) is implemented using the LM minimization from MINPACK [17], with the measured value of e as the initial estimate.

As noted above, valid corresponding measurements must be identified before calculating $E_{\text{tot}}^*(e)$. However, since the correspondences are determined by minimizing E^* over all candidate pairs given the plane parameters, the correspondences are also a function of the encoder count. Thus, the refined estimate e^* may relate to a different set of optimal correspondences than those from which it was calculated. To resolve this issue, the optimal correspondences and encoder count are calculated recursively. In the first iteration, correspondences are calculated using the measured encoder value e to yield the initial estimate e_0^* via equations (37)-(38). A new set of correspondences are then extracted from the raw measurements using the refined encoder value e_0^*. If the new correspondences differ from the previous iteration, an updated estimate of the encoder value e_1^* is calculated (using e_0^* as the initial guess). The process is repeated until a stable set of correspondences is found.

The above process is applied to each captured frame, and the optimal encoder count e^* and valid corresponding measurements are substituted into equation (35) to finally recover the optimal 3D profile of the laser.

3.5 Additional Constraints

As already described, robust stripe detection is based on minimization of the image plane error in equation (32). However, the minimum image plane error is a necessary but insufficient condition for identifying valid stereo measurements. In the practical implementation, two additional constraints are employed to improve the robustness of stripe detection.

The first constraint simply requires stripe candidates to be moving features; a valid measurement must not appear at the same position in previous frames. This is implemented by processing only those pixels with sufficiently large intensity difference between successive frames. While this constraint successfully rejects static stripe-like edges or textures in most scenes, it has little effect on cross-talk or reflections, since these also appear as moving features.

The second constraint is based on the fact that valid measurements only occur within a sub-region of the image plane, depending on the angle of the light plane. From equation (34), the inhomogeneous z-coordinate of a optimal reconstruction $\widehat{\mathbf{X}}$ can be expressed as a function of the image plane projection ${}^{R}\hat{\mathbf{x}} = ({}^{R}\hat{x}, y)^\top$ as

$$\widehat{Z} = \frac{-2bf}{(\alpha+1){}^{R}\hat{x} + \beta y + \gamma f} \quad (39)$$

Rearranging the above, the projected x-coordinate of a point on the light plane may be expressed as a function of depth $\widehat{Z}$ and the height y of the scanline:

$$ {}^{R}\hat{x} = -\frac{\beta y + \gamma f}{\alpha + 1} - \frac{2bf}{\widehat{Z}(\alpha + 1)} \tag{40} $$

The extreme boundaries for valid measurements can now be found by taking the limit of equation (40) for points on the light plane near and far from the camera. Taking the limit for distant reflections gives:

$$ \lim_{\widehat{Z}\to\infty} {}^{R}\hat{x} = -\frac{\beta y + \gamma f}{\alpha + 1} \tag{41} $$

Taking the limit $\widehat{Z} \to 0$ for close reflections gives the other boundary at ${}^{R}\hat{x} \to -\infty$. Now, if w is the width of the captured image, valid measurements on the right image plane at height y must be constrained to the x-coordinate ranges

$$ -\frac{w}{2} \;<\; {}^{R}x \;<\; -\frac{\beta v + \gamma f}{\alpha + 1} \tag{42} $$

Following a similar development, it is straightforward to show that the limits of valid measurements on the left image plane are:

$$ -\frac{\beta v + \gamma f}{\alpha + 1} \;<\; {}^{L}x \;<\; +\frac{w}{2} \tag{43} $$

Stripe extraction is only applied to pixels within the boundaries defined in (43)-(42); pixels outside these ranges are immediately classified as invalid. In addition to improving robustness, sub-region processing also reduces computational expense by halving the quantity of raw image data and decreasing the number of stripe candidates tested for correspondence.

4 Active Calibration

In this section, determination of the unknown parameters in the model of the light stripe scanner are considered. Let the unknown parameters be represented by the vector

$$ \mathbf{p} = (k_1, k_2, k_3, \theta_x, \theta_z, B_0, m, c) $$

where k_1, k_2 and k_3 were introduced in equations (28)-(30). Since most of the parameters relate to mechanical properties, the straightforward approach to calibration is manual measurement. However, such an approach would be both difficult and increasingly inaccurate as parameters vary through mechanical wear. To overcome this problem, a strategy is now proposed to optimally estimate $\mathbf{p}$ using only image-based measurements of a non-planar but otherwise arbitrary surface with favourable reflectance properties (the requirement of non-planarity is discussed below). This allows calibration to be performed cheaply and during normal operation.

The calibration procedure begins by scanning the stripe across the target and recording the encoder and image plane measurements for each captured frame. Since the system parameters are initially unknown, the validation problem is approximated by recording only the brightest pair of features per scanline. Let ${}^{L}\mathbf{x}_{ij}$ and ${}^{R}\mathbf{x}_{ij}$, $i = 1 \ldots n_j$, $j = 1 \ldots t$ represent the centroids of the brightest corresponding features on n_j scanlines of t captured frames, and let e_j represent the measured encoder value for each frame. As described earlier, image plane measurements have independent errors, while the encoder error couples all measurements in a given frame. Thus, optimal system parameters are determined from iterative minimization of the stripe measurement and encoder errors, based on the algorithm first described in Section 3.4. First, the total image plane error E^*_{tot} is summed over all frames:

$$ E^*_{\text{tot}} = \sum_{j=1}^{t} \sum_{i=1}^{n_j} E^*({}^{L}\mathbf{x}_{ij}, {}^{R}\mathbf{x}_{ij}, e_j, \mathbf{p}) \tag{44} $$

where E^* is defined in equation (32). The requirement of a non-planar calibration target can now be justified. For a planar target, the stripe appears as a straight line and the image plane measurements obey a linear relationship of the form $x_{ij} = a_j y_{ij} + b_j$. Then, the total error E^*_{tot} reduces to the form

$$ E^*_{\text{tot}} = \sum_{j=1}^{t} \sum_{i=1}^{n_j} (A_j y_{ij} + B_j)^2 \tag{45} $$

Clearly, the sign of A_j and B_j cannot be determined from equation (45), since the total error remains unchanged after substituting $-A_j$ and $-B_j$. The geometrical interpretation of this result is illustrated in Figure 5, which shows the 2D analogue of a planar target scan. For any set of encoder values e_j and collinear points $\mathbf{X}_j$ measured over t captured frames, there exist two symmetrically opposed laser plane generators capable of producing identical results. This ambiguity can be overcome by constraining the calibration target to be non-planar. It may also be possible for certain non-planar targets to produce ambiguous results, but the cur-

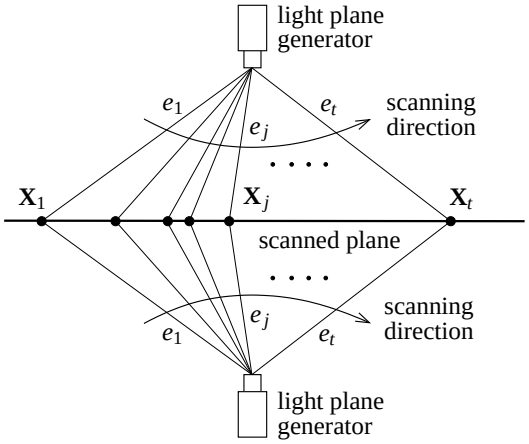


Figure 5: Possible positions of the light plane from the scan of a planar calibration target. The position of the light plane generator is ambiguous.

rent implementation assumes that such an object will rarely be encountered.

An initial estimate $\mathbf{p}_0^*$ of the parameter vector is given by the minimization

$$\mathbf{p}_0^* = \arg\min_{\mathbf{p}}[E_{\text{tot}}^*(\mathbf{p})] \tag{46}$$

using the measured encoder values e_j and stereo correspondences ${}^L\mathbf{x}_{ij}$ and ${}^R\mathbf{x}_{ij}$. Again, equation (46) is implemented numerically using LM minimization in MINPACK. The stripe measurements ${}^L\mathbf{x}_{ij}$ and ${}^R\mathbf{x}_{ij}$ are likely to contain gross errors resulting from the initial coarse validation constraint (in the absence of known system parameters). Thus, the next calibration step refines the measurements by applying out-lier rejection. Using e_j and the initial estimate $\mathbf{p}_0^*$, the image plane error $E^*({}^L\mathbf{x}_{ij}, {}^R\mathbf{x}_{ij}, e_j, \mathbf{p}_0^*)$ in equation (32) is calculated for each stereo pair. The measurements are then sorted in order of increasing error, and the top 20% are discarded.

The system parameters and encoder values are then sequentially refined in an iterative process. The initial estimate $\mathbf{p}_0^*$ is only optimal with respect to image plane error, assuming exact encoder values e_j. To account for encoder error, the encoder value is refined for each frame using the method described in Section 3.4 with the initial estimate $\mathbf{p}_0^*$ of the system model. The resulting encoder estimates $e_{j,0}^*$ are optimal with respect to $\mathbf{p}_0^*$. A refined system model $\mathbf{p}_1^*$ is then obtained from equation (46) using the latest encoder estimates $e_{j,0}^*$ and in-lier image plane measurements. At the k^{th} iteration, the model is considered to have converged when the fractional change in total error E_{tot}^* is less then a threshold δ:

$$\frac{E_{\text{tot},k-1}^* - E_{\text{tot},k}^*}{E_{\text{tot},k-1}^*} < \delta \tag{47}$$

The final parameter vector $\mathbf{p}_k^*$ is stored as the near-optimal system model for processing regular scans using the methods described in Section 3. A final check for global optimality is performed by comparing the minimum total error $E_{\text{tot},k}^*$ to a fixed threshold, based on an estimate of the image plane error. The rare case of non-convergence (less than 10% of trials) is typically due to excessive out-liers introduced by the sub-optimal maximum intensity validation constraint applied to the initial measurements. Non-convergence is resolved by repeating the calibration process with a new set of data.

The calibration technique presented here is practical, fast and accurate, requiring only a single scan of any suitable non-planar scene. Furthermore, the method does not rely on measurement or estimation of any metric quantities, and so does not require accurate knowledge of camera parameters b and f. Thus, image-based calibration allows the validation and correspondence problems to be solved robustly and independently of reconstruction accuracy.

5 Implementation

This section describes the signal processing used to implement stereoscopic light stripe scanning on the experimental system introduced in Section 2. The output of the scanner is a 384×288 element range map, with each element recording the 3D position of the surface as viewed from the right camera. During normal operation, the laser is scanned across a scene and the shaft encoder and stereo images are recorded at regular 40 ms intervals (25 Hz PAL frame-rate). The laser is mechanically geared to scan at about one pixels of horizontal motion per captured frame, so a complete scan requires approximately 384 processed frames (15 seconds).

5.1 Light Stripe Measurement

Laser stripe extraction is performed using intensity data only, which is calculated by taking the average of the colour channels. As noted in Section 3.5, the motion of the stripe distinguishes it from the static background, which is eliminated by subtracting the intensity values

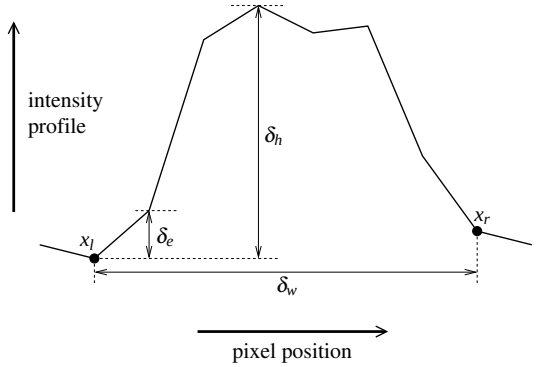


Figure 6: Thresholds for robust extraction of multimodal pulses.

in consecutive frames and applying a minimum difference threshold. The resulting *difference image* is morphologically eroded and dilated to reduce noise and improve the likelihood of stripe detection. In Section 3.5 it was also shown that valid measurements occur in a predictable sub-region of the image. This is calculated from equations (43)-(42) and the measured encoder value, and pixels outside this region are set to zero in the difference image. Further processing is only applied to pixels with non-zero difference.

The intensity profile on each scanline is then examined to locate candidate stripe measurements. If the stripe appeared as a simple unimodal pulse, the local maxima would be sufficient to detect candidates. However, mechanisms including sensor noise, surface texture and saturation of the CCD interfere and perturb the intensity profile. These issues are overcome by extracting pulses using a more sophisticated strategy of intensity edge extraction and matching. On each scanline, left and right edges are identified as an increase or decrease in the intensity profile according to the thresholds defined in Figure 6. Processing pixels from left to right, the location of a left edge x_l is detected when the intensity difference between successive pixels exceeds a threshold δ_e, and the closest local intensity maxima to the right of x_l exceeds the intensity at x_l by a larger threshold δ_h. Right edges x_r are extracted by processing the scanline in reverse. Finally, the edges are examined to identify left/right pairs without intervening edges. When x_l and x_r are closer than a threshold distance δ_w, the pair are assumed to delimit a candidate pulse. The pulse centroid is calculated to sub-pixel accuracy as the mean pixel position weighted by the intensity profile within these bounds.

The result of the above process is a set of candidate stripe locations on each scanline of the stereo images. Along with the measured encoder value, these candidates are analyzed using the techniques described in Section 3 to refine the laser plane angle, identify valid corresponding measurements and reconstruct an optimal 3D profile. The reconstruction on each scanline is stored in the range map at the location of the corresponding measurement in the right image.

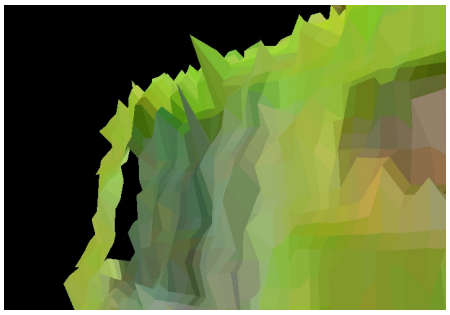
5.2 Range Data Postprocessing

Post-processing is applied *after* each complete scan to further refine the measured data. Despite robust scanning, the raw range map may still contain outliers as the stripe validation conditions are occasionally satisfied by spurious noise. Fortunately, the sparseness of the outliers make them easy to detect and remove using a simple thresholding operation: the minimum distance between each 3D point and its eight neighbours is calculated, and when this exceeds a threshold (10 mm in the current implementation), the associated point is removed from the range map.

Holes (pixels for which range data could not be recovered) may occur in the range map due to specular reflections, poor surface reflectivity, random noise and outlier removal. A further post-processing step fills these gaps with interpolated depth data. Each empty pixel is checked to determine whether it is bracketed by valid data within a vertical or horizontal distance of two pixels. To avoid interpolating across depth discontinuities, the distance between the bracketing points must be less than a fixed threshold (30 mm in the current implementation). Empty pixels satisfying these constraints are assigned a depth linearly interpolated between the valid bracketing points. The effect of both outlier rejection and interpolation on a raw scan is demonstrated in Figure 7.

Finally, a colour image is registered with the range map. Since robust scanning allows the sensor to operate in normal light, the cameras used for stripe detection also capture colour information. However, depth and colour cannot be sampled simultaneously for any given pixel, since the laser masks the colour of the surface. Instead, a complete range map is captured before registering a colour image from the right camera (assuming the cameras have not moved). Each pixel in this final image yields the colour of the point measured in the corresponding element of the range map.

(a) Before post-processing



(b) After post-processing

Figure 7: Removal of impulse noise and holes (eliminated features are circled on the left).

6 Experimental Results

6.1 Robustness

To evaluate the performance of the proposed robust active sensing method, the scanner is implemented along with two other common techniques on the same experimental platform. The first method is a simple single-camera scanner without any optimal or robust properties. A single-camera reconstruction is calculated from equation (13), using image plane measurements from the right camera only. Since no validation is possible in this configuration, the stripe is simply detected as the brightest feature on each scanline. The second alternative implementation will be referred to as a *double-camera* scanner. This approach is based on the robust techniques proposed in [18, 27], which exploit the requirement of consensus between two independent single-camera reconstructions. The single-camera reconstructions $\widehat{\mathbf{X}}_L$ and $\widehat{\mathbf{X}}_R$ are calculated independently from measurements on the left and right image planes, and discarded when $|\widehat{\mathbf{X}}_L - \widehat{\mathbf{X}}_R|$ exceeds a fixed distance threshold. For valid measurements, the final reconstruction is calculated as $\frac{1}{2}(\widehat{\mathbf{X}}_L + \widehat{\mathbf{X}}_R)$.

The performance of the three methods in the presence of a phantom stripe (secondary reflection) was assessed using the test scene shown in Figure 8. A mirror at the rear of the scene creates a reflection of the objects and scanning laser, simulating the effect of cross-talk and secondary reflections. To facilitate a fair comparison, the three methods operated simultaneously on the same raw measurements captured during a single scan of the scene.

Figure 9(a) shows the colour/range data captured by the single-camera scanner. As a result of erroneous associations between the phantom stripe and laser plane, numerous phantom surfaces appear in the scan without any physical counterpart. Figure 9(b) shows the output of the double-camera scanner, which successfully removes the spurious surfaces. However, portions of real surfaces have also been rejected, since the algorithm is unable to disambiguate the phantom stripe from the primary reflection when both appear in the scene. Finally, Figure 9(c) shows the result using the techniques presented in this paper (see also Extension 1). The portions of the scene missing from Figure 9(b) are successfully detected using the proposed robust scanner, while the phantom stripe has been completely rejected. Also noteworthy is the implicitly accurate registration of colour and range.

The single-camera result highlights the need for robust methods when using light stripe scanners on a domestic robot. While the double-camera scanner successfully rejects reflections and cross-talk, the high rejection rate for genuine measurements may cause problems for segmentation or other subsequent processing. In contrast, segmentation and object classification have been successfully applied to the colour/range data from the proposed robust scanner to facilitate high-level domestic tasks [25]. Extensions 2 and 3 provide additional colour/range maps of typical domestic objects to demonstrate the robustness of our sensor in this application.

6.2 Error Analysis

The results in this section experimentally validate of the system and noise models used to derive the theoreti-

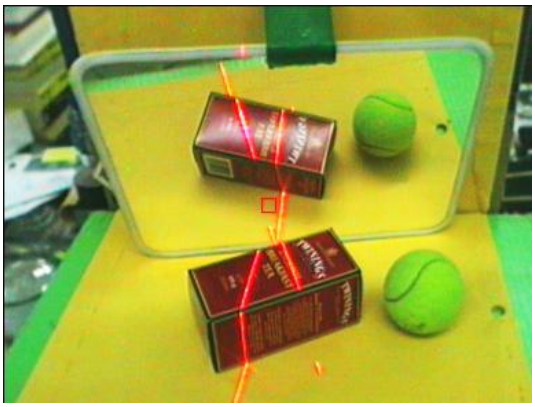
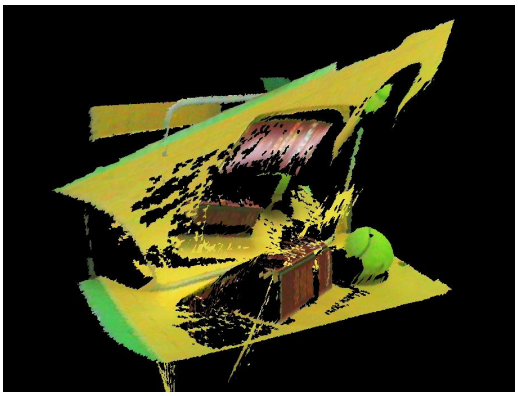
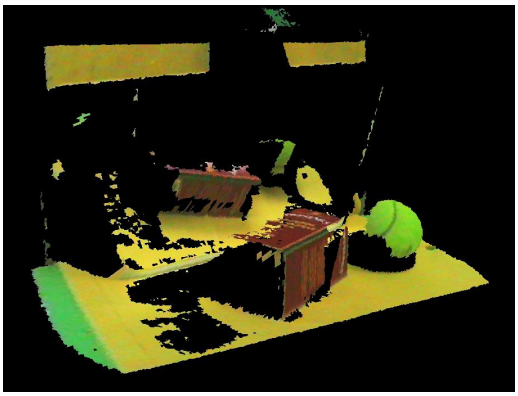


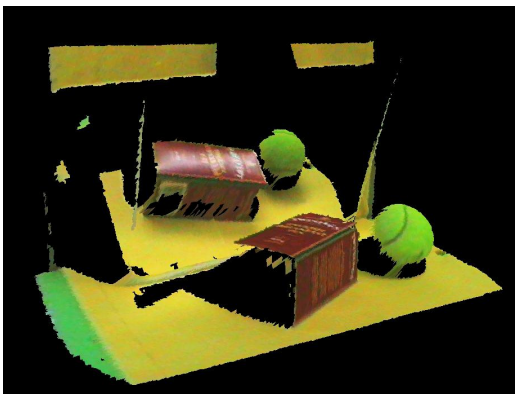
Figure 8: Robust scanning experiment.



(a) Single-camera



(b) Double-camera



(c) Robust scanner

Figure 9: Comparison of range scanner results in the presence of secondary reflections.

cal results. In particular, the encoder angle estimation and calibration techniques described in Sections 3.4 and 4 are shown to be sufficiently accurate that any uncertainty in the system parameters and encoder values can be ignored for the purposes of optimal validation and reconstruction.

First, the calibration procedure described in Section 4 was performed using the corner formed by two boxes as the calibration target, and the valid image plane measurements and encoder values for each frame were recorded. Using the estimated system parameters, the optimal projections ${}^{L,R}\hat{x}^*$ and residuals $({}^{L,R}x - {}^{L,R}\hat{x}^*)$ were calculated from equations (31) and (33) for all measurements. Figure 10 shows the histogram of residuals for measurements on the right image plane, and a Gaussian distribution with the same parameters for comparison. The residuals are approximately Gaussian distributed as expected, and assuming the light stripe measurement errors are similarly distributed, the error model proposed in Section 3.1 is found to be valid. The variance of the image plane measurements are shown in the first two rows of Table 1.

The variance of the system parameters and encoder values were determined using statistical bootstrapping. In this process, the residuals were randomly and uniformly sampled (with replacement) from the initial data, and added to the optimal projections ${}^{L,R}\hat{x}^*$ to generate a new set of pseudo-measurements. The system parameters were then estimated for each new set of pseudo-measurements using the calibration process described in Section 4. A total of 5000 re-sampling experiments were performed, and the resulting variance in the estimated system parameters are shown in the third

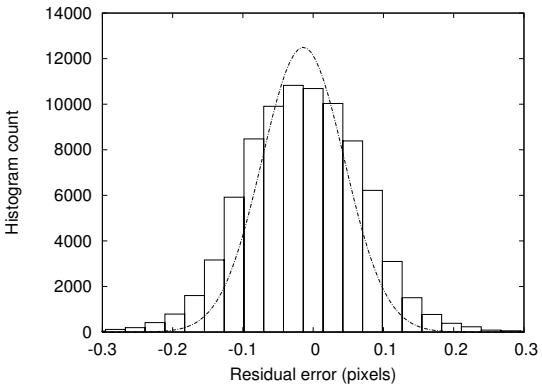


Figure 10: Distribution of residual reconstruction errors on the image plane.

Table 1: System parameter errors and contribution to reconstruction error.

p_i	$\bar{p_i}$	$\mathrm{var}(p_i)$	$\mathrm{var}({}^R\hat{x}^*)_i$
x_L	59.2	8×10^{-3}	2×10^{-3}
x_R	-47.4	7×10^{-3}	2×10^{-3}
y	6.0	0.0	0.0
e	451.5	5×10^{-5}	2×10^{-5}
k_1	-1.1033	2×10^{-8}	1×10^{-5}
k_2	0.1696	4×10^{-8}	2×10^{-8}
k_3	0.0717	1×10^{-8}	2×10^{-5}
θ_x	-0.0287	3×10^{-10}	2×10^{-12}
θ_z	0.0094	5×10^{-8}	2×10^{-6}
B_0	-0.0004	5×10^{-8}	2×10^{-6}
m	0.001105	7×10^{-16}	3×10^{-5}
c	-0.4849	1×10^{-10}	4×10^{-5}

column of Table 1.

Finally, the optimality of the reconstruction is assessed by calculating the contribution from each parameter to the variance of the optimal projection ${}^R\hat{x}^*$ in equation (31). Representing the components of the parameter vector as $\mathbf{p} = \{p_i\}$, and assuming the parameters have independent noise, the contribution of parameter p_i (with all other parameters fixed) to the variance of ${}^R\hat{x}^*$, represented as $\mathrm{var}({}^R\hat{x}^*)_i$, is calculated as

$$\mathrm{var}({}^R\hat{x}^*)_i = \left(\left. \frac{\partial {}^R\hat{x}^*}{\partial p_i} \right|_{\mathbf{p}} \right)^2 \cdot \mathrm{var}(p_i) \qquad (48)$$

The independence of the parameters is readily verified from the covariance matrix of $\mathbf{p}$.

Choosing a test point near the centre of the image plane, the contribution of each parameter to the total variance was calculated from equation (48) and the results are shown in the far right column of Table 1. Importantly, the errors due to ${}^Rx$ and ${}^Lx$ are two orders of magnitude greater than the contribution from the system parameters and encoder value. For comparison, the variance in ${}^R\hat{x}^*$ measured from the bootstrapping process was 0.0035 pixels^2, which agrees well with the sum of contributions from ${}^Lx$ and ${}^Rx$. Finally, it should be noted that the variance of ${}^R\hat{x}^*$ is about half the variance of ${}^Rx$, indicating that the optimal reconstruction has a higher precision than a single-camera reconstruction.

These results demonstrate the reliability of the image-based techniques presented in Sections 3.4 and 4 for estimating the encoder value and calibrating the system parameters in the presence of noisy measurements. Furthermore, the main assumptions in deriving equations (31) and (32) are now justified: any uncertainty in the system parameters and encoder value can be reasonably ignored for the purpose of validation and reconstruction.

7 Discussion

In addition to providing a mechanism for validation, the error distance E^* in equation (32) could be used to measure the random error of range points. As discussed in Section 3.1, the error variance of a 3D reconstruction increases with depth as the reconstruction problem becomes ill-conditioned. This systematic uncertainty can be calculated directly from the reconstruction in equation (35). In contrast, E^* measures the random uncertainty due to sensor noise. A suitable function of these systematic and random components could be formulated to provide a unique confidence interval for each 3D point, which would be useful in subsequent processing. For example, parametric surface fitting could be optimized with respect to measurement error by weighting each point with the confidence value.

One of the main limitations of light stripe scanning (compared to methods such as passive stereo) is the acquisition rate. In the current implementation, the PAL frame-rate of 25 Hz results in a 15 second measurement cycle to capture a complete half-PAL resolution range map of 384 stripe profiles. Clearly, such a long acquisition time renders the sensor unsuitable for dynamic

scenes. However, a more subtle issue is that the robot must remain stationary during a scan to ensure accurate registration of the measured profiles. Obviously, the acquisition rate could be improved using high-speed cameras and dedicated image processing hardware; high-speed CMOS cameras are now capable of frame-rates exceeding 1000 Hz. Assuming the image processing could be accelerated to match this speed, the sensor could be capable of acquiring 2-3 range maps per second. An example of a high-speed monocular light stripe sensor using a "smart" retina is described in [8].

To minimize complexity and cost, the experimental prototype uses a red laser diode to generate the light plane. Consequently, the scanner only senses surfaces which contain a high component of red. Black, blue and green surfaces reflect insufficient red laser light and are effectively invisible to the sensor. Since the light plane is not required to be coherent or monochromatic, the laser diode could be replaced by a white light source such as a collimated incandescent bulb. However, laser diodes have particular design advantages including physical compactness, low power consumption and heat generation, and are thus more desirable than other light sources. To solve the colour deficiency problem while retaining these advantages, the light plane could be generated using a triplet of red, green and blue laser diodes. Currently, the main obstacle to this approach is the high cost of green and blue laser diodes.

As with colour, surfaces with high specular and low Lambertian reflection may appear invisible, since insufficient light is reflected back to the sensor. This limitation is common to all active light sensors and can also defeat passive stereopsis, since the surface appears differently to each viewpoint. To illustrate this effect, Figure 11 shows the raw image and resulting scan of a highly polished object. The only visible regions appearing in the range map are the high curvature edges that provide a specular reflection directly back to the sensor. The best that can be achieved is to ensure that secondary reflections do not interfere with range data acquisition, as demonstrated in this result.

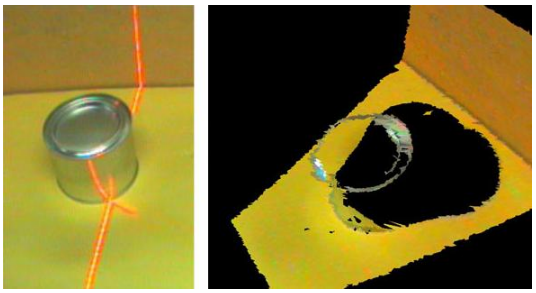


Figure 11: Scan of a highly polished object. The light stripe (with specular reflections) is shown on the left, and the resulting range map is shown on the right.

8 Summary and Conclusions

We have presented the theoretical framework and implementation of a robust light stripe scanner for a domestic robot, capable of measuring arbitrary scenes in ambient indoor light. The scanner uses the light plane orientation and stereo camera measurements to robustly identify the light stripe in the presence of secondary reflections, cross-talk and other sources of interference. The validation and reconstruction framework is based on minimization of error distances measured on the image plane. Unlike previous stereo scanners, this formulation is optimal with respect to measurement error. An image-based procedure for calibrating the light plane parameters from the scan of an arbitrary non-planar target is also demonstrated.

Results from the experimental scanner demonstrate that our robust method is more effective at recovering range data in the presence of reflections and cross-talk than comparable light stripe methods. Experimental results also confirm the assumptions of our noise model, and show that image-based calibration produces reliable results in the presence of noisy image plane measurements. Finally, the optimal reconstructions from our proposed scanner are shown to be more precise than the reconstructions from a single-camera scanner.

The ability to solve the stripe association problem may provide interesting future research in the development of a multi-stripe scanner. Multi-stripe scanners have the potential to solve a number of issues associated with single-stripe scanners: illuminating a target with two stripes could double the acquisition rate, and projecting the stripes from different positions reveals points that would otherwise be hidden in shadow. Single-camera multi-stripe systems rely on colour [23], sequences of illumination [15] or epipolar constraints [14] to disambiguate the stripes. However, the method proposed in this paper could allow the stripes to be uniquely identified using the same principles that provide validation for a single stripe.

Finally, the over-arching goal in this development is

to allow a robot to model and manipulate arbitrary objects in a domestic environment. Our other results in this area [25] already demonstrate that the scanner provides sufficiently robust measurements to achieve this goal.

Acknowledgment

This project was funded by the *Strategic Monash University Research Fund for the Humanoid Robotics: Perception, Intelligence and Control* project at IRRC.

Appendix: Index to Multi-Media Extensions

The mult-imedia extensions to this article can be found online by following the hyperlinks from www.ijrr.org. These results may also be found at www.irrc.monash.edu.au/laserscans.

Extension	Media type	Description
1	Data	VRML 97 model (half scanner resolution) of mirror scene, (requires VRML plug-in).
2	Data	VRML 97 model (half scanner resolution) of typical domestic objects.
3	Data	VRML 97 model (half scanner resolution) of typical domestic objects.

References

[1] G. J. Agin and T. O. Binford. Computer description of curved objects. In *Proc. 3rd Int. Joint Conf. on Artifical Intelligence*, pages 629–640, 1973.

[2] M. J. Aldon and L. Le Bris. Mobile robot localization using a light stripe sensor. In *Proc. of the Intelligent Vehicles '94 Symposium*, pages 255–259, 1994.

[3] H. M. M. Alshawish and C. R. Allen. 3d object recognition using coded light projection for robot assembly applications. In *Proc. 21st International Conference on Industrial Electronics, Control and Instrumentation*, volume 2, pages 1420–1247, 1995.

[4] C. M. Bastuscheck. Techniques for real-time generation of range images. In *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, pages 262–268, 1989.

[5] P. J. Besl. Active, optical range image sensors. *Machine Vision Applications*, 1(2):127–152, 1988.

[6] J. Clark, E. Trucco, and H-F. Cheung. Improving laser triangulation sensors using polarization. In *Proc. 5th Int. Conf. on Computer Vision*, pages 981–986, 1995.

[7] D. Khadraoui et al. Visual servoing in robotics scheme using a camera/laser-stripe sensor. *IEEE Transactions on Robotics and Automation*, 12(5):743–750, 1996.

[8] A. Gruss, L. R. Carley, and T. Kanade. Integrated sensor and range-finding analog signal processor. *IEEE Journal of Solid-State Circuits*, 26(3):184–191, Mar 1991.

[9] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[10] R. I. Hartley. Theory and practice of projective rectification. *Int. Journal of Computer Vision*, 35(2):115–127, 1999.

[11] R. I. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.

[12] J. Haverinen and J. Röning. An obstacle detection system using a light stripe identification based method. In *Proc. IEEE International Joint Symposium on Intelligence and Systems*, pages 232–236, 1998.

[13] M. Hebert. Active and passive range sensing for robotics. In *Proc. IEEE International Conference on Robotics and Automation*, pages 102–110, 2000.

[14] G. Hu and G. Stockman. 3-D surface solution using structured light and constraint propagation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(4):390–402, apr 1989.

[15] D. Q. Huynh, R. A. Owens, and P. E. Hartmann. Calibrating a structured light stripe system: A novel approach. *International Journal of Computer Vision*, 33(1):73–86, 1999.

[16] M. Magee, R. Weniger, and E. A. Franke. Location of features of known height in the presence of reflective and refractive noise using a stereoscopic light-striping approach. *Optical Engineering*, 33(4):1092–1098, April 1994.

[17] J. J. Moré, B. S. Garbow, and K. E. Hillstrom. Users' guide for MINPACK-1. Technical Report ANL-80-74, Applied Math. Div., Argonne National Laboratory, 1980.

[18] Kouichi Nakano, Yasuo Watanabe, and Sukeyasu Kanno. Extraction and recognition of 3-dimensional information by projecting a pair of slit-ray beams. In *Proceedings of the 9th International Conference on Pattern Recognition*, pages 736–738, 1988.

[19] J. Nygards, T. Högström, and Å. Wernersson. Docking to pallets with feedback from a sheet-of-light range camera. In *Proc. 2000 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1853–1859, 2000.

[20] J. Nygards and Å. Wernersson. Specular objects in range cameras: Reducing ambiguities by motion. In *Proc. of the IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems*, pages 320–328, 1994.

[21] K. Rao, G. Medioni, H. Liu, and G. A. Bekey. Shape description and grasping for robot hand-eye coordination. *IEEE Control Systems Magazine*, pages 22–29, February 1989.

[22] Y. Shirai and M. Suwa. Recognition of polyhedrons with a range finder. In *Proc. 2nd Int. Joint Conf. on Artifical Intelligence*, pages 80–87, 1971.

[23] J. Tajima and M. Iwakawa. 3-d data acquisition by rainbow range finder. In *Proc. 10th Int. Conf. on Pattern Recognition*, volume 1, pages 309–313, 1990.

[24] G. Taylor and L. Kleeman. Flexible self-calibrated visual servoing for a humanoid robot. In *Proc. 2001 Australian Conference on Robotics and Automation*, pages 79–84, 2001.

[25] G. Taylor and L. Kleeman. Grasping unknown objects with a humanoid robot. In *Proc. 2002 Australiasian Conference on Robotics and Automation*, pages 191–196, 2002.

[26] G. Taylor, L. Kleeman, and Å. Wernersson. Robust colour and range sensing for robotics applications using a stereoscopic light stripe scanner. In *Proc. 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 178–183, 2002.

[27] E. Trucco and R. B. Fisher. Acquisition of consistent range data using local calibration. In *Proc. of the 1994 IEEE Int. Conf. on Robotics and Automation*, volume 4, pages 3410–3415, 1994.