Design and calibration of a hybrid computer vision and structured light 3D imaging system

Tom Botterill
Department of Computer Science,
University of Canterbury,
Christchurch, NZ.
Email: tom.botterill@grcnz.com.

Steven Mills
Department of Computer Science,
University of Otago,
Dunedin, NZ.

Richard Green
Department of Computer Science,
University of Canterbury,
Christchurch, NZ.

***Abstract*—This paper describes a structured light system for generating a dense 3D reconstruction of crops viewed by a robot, as the robot moves. A line laser is used to project a plane of light into the image, which is viewed by a colour camera. The laser line is localised with subpixel accuracy despite variations in colour and texture of the surface being imaged, and challenging lighting conditions. A dense depth map with errors from laser line localisation of just 0.15mm is constructed.**

A simple calibration procedure is described, where a large number of constraints are automatically extracted from images showing where the laser line crosses a checkerboard pattern. Errors in calibration are estimated at 2.3mm.

I. Introduction

This paper describes the design and calibration of a hybrid computer vision and structured light sensor, designed to generate a detailed 3D model of crops viewed by an agricultural robot. Images are captured from multiple cameras, and each camera captures a colour image of the crops, onto which a structured light pattern is projected by a set of high powered line laser modules. This paper describes how the system is calibrated, how laser lines visible in the images are detected, and how a dense depth map is constructed as the robot moves. In addition, the errors in the dense map are quantified.

Structured light (SL) systems have long been used for 3D reconstruction [1], [2], [3], [4]. These systems project a pattern of light onto an object, which is imaged with a digital camera. The position of the light pattern in the image is used to compute the 3D structure of the object.

One common SL pattern is a laser line (also known as a laser stripe): a laser diode is lensed so that a plane of light is projected into the scene [1]. The camera and laser move with respect to the object being scanned, and the positions where the line is visible are used to reconstruct the profile of the object being scanned, one slice at a time. Line-laser SL patterns can give very accurate depth measurements, however a sequence of many images are needed to build a dense depth map [3], [4].

Many more complex light patterns can also be used, for example the pattern of dots used by the Microsoft Kinect [5], multiple laser lines, or rainbow-coloured bands [6]. These patterns are often designed to provide dense reconstructions from a small number of images, however a limitation of these more complex patterns is that the localisation of the light pattern in the image can be ambiguous [2], [6]–there may be multiple possible 3D objects compatible with a particular pattern (particularly in the presence of many occlusions, for example it may be ambiguous which of two laser lines is visible if one is occluded).

In the system we are building, the SL will be used to augment a 3D reconstruction from multiple images. A limitation of multi-view reconstruction is that similar-looking features may be incorrectly matched, leading to an incorrect reconstruction, so a SL system which does not suffer from similar ambiguities is desirable in order to provide complementary measurements which can be used to resolve these ambiguities. The crops imaged by the robot are illuminated by both natural and artificial light, so a further requirement is that the light projector must be powerful enough that the light pattern is clearly visible in the images. An additional requirement is that a reasonably dense reconstruction can be made despite the high levels of occlusion present in the scene—other SL schemes sometimes assume objects are locally convex in order to fill-in gaps in a dense reconstruction, however this assumption is not valid for the tangles of branches and wires which are being imaged.

To meet these requirements, we use a line laser module. A high-power (e.g. 100mW) line laser module is used to project a single plane of laser light into the scene. By positioning the laser plane perpendicular to the camera's scanlines, the line is visible at most once in each scanline, so if line laser pixels are identified correctly, the reconstruction obtained is unambiguous. As the robot moves, the planes of 3D points are combined with an odometry estimate in order to provide a dense reconstruction.

This paper is organised as follows: Section II outlines how the location of a laser line is identified in an image; Section III outlines how the positions of detected laser lines are converted into a 3D map; Section IV describes how the system is calibrated; Section V analyses the accuracy of the reconstructions obtained; and Section VI describes our conclusions and further work.

II. Detecting a laser line in a colour image

This section describes how the position of the laser line is extracted from a colour image. Many SL systems assume

Fig. 1. The position of a laser line viewed in an image is used to reconstruct the 3D structure along its length.

that the scene is illuminated only by the SL source [7], or that other colours of light can be removed with a filter [5], however this is not the case for our system, where the scene is illuminated by both natural and artificial lighting, and we are imaging both the scene and the laser line in the same images (Figure 1). The crops which are imaged vary in colour on a small scale, and the illumination levels (even following an illumination correction) vary significantly across the image. A 1.3 megapixel colour camera is used, however a laser line is only visible for the pixels of the same colour (i.e. the red pixels for a 635nm red line laser), so only these pixels in the Bayer pattern are used for line localisation (so only a 640x480 subset of the 1280x960 pixels available are used).

To detect the line, the crops being imaged are assumed to be approximately grey, so that all colours of light are reflected approximately equally from each pixel (a reasonable assumption for the woody crops and galvanised wires being imaged). On the laser plane, there is more red light than other colours, so the crops will reflect the red light more strongly, in proportion to the reflectance of the crops at that point. In addition, the perceived colours do vary a small amount (partly due to sensor noise, and from variation in reflectivity on the same scale as the Bayer pattern). The green channel G (computed as an average of the 4 pixels surrounding every red pixel in the Bayer pattern) is used as an estimate of reflectivity. Under these assumptions, the increase in red light over other colours at a point (x, y) is then proportional to $L(x, y) = (R(x, y) - G(x, y))/G(x, y)$. In addition, the constraint $R(x, y) > G(x, y) + N$ for a noise threshold N (e.g. $N = 5$ greylevels) is used to reject increases in red due to noisy measurements. Along each scanline y, the x^* coordinate with $L(x^*, y)$ significantly higher than $L(x, y)$ for any other x on the scanline is assumed to be on the laser plane.

Naidu and Fisher [7] evaluate several methods for subpixel localisation of laser line positions in images where the only significant light source is the line laser. When the response is always greater than zero, and the sensor is never saturated, the most accurate subpixel method involves fitting a Gaussian to the pixel with the greatest illumination level. This approach does not work for these noisy images however, as L is significantly affected by noise and reflectance variation (the constraint $L(x) > 0$ is required). Instead, we use the 'Parabolic Estimator', also described by Naidu and Fisher [7], in which the laser line profile is assumed locally quadratic. A quadratic is fitted to the 3 values $L(x^* - 2, y), L(x^*, y), L(x^* + 2, y)$ (with ± 2 as pixels of the same colour in the Bayer pattern are 2 pixels apart). The x-coordinate of the maximum of this quadratic is given by $x^* + \delta$, where:

$$\delta = \frac{L(x^* - 2, y) - L(x^* + 2, y)}{L(x^* - 2, y) - 2L(x^*, y) + L(x^* + 2, y)} \tag{1}$$

This is equivalent to finding the 0-crossing point of the derivative $\frac{\partial}{\partial x} L(x, y)$ by linear interpolation between the derivatives computed by finite differences.

III. 3D RECONSTRUCTION FROM LASER LINE LOCATIONS

The camera is first calibrated using standard techniques (OpenCV's [8] implementation of the camera calibration method by Zhang [9]). The intrinsic parameters obtained define a mapping between pixel coordinates and normalised image coordinates (a normalised homogeneous image coordinate gives the direction of a ray passing through that pixel). Lens distortion (i.e. radial distortion) is corrected by this mapping.

The laser projects a plane of light into the scene (errors in the laser's optics are assumed to be negligible). The plane is a perpendicular distance K from the origin, and has unit normal $\hat{\mathbf{n}}$. Points $\mathbf{X}$ on the plane satisfy:

$$\mathbf{X}.\hat{\mathbf{n}} = K, \tag{2}$$

or equivalently

$$\mathbf{X}.\mathbf{n} = 1 \text{ where } \mathbf{n} = \hat{\mathbf{n}}/K. \tag{3}$$

To find the 3D position of a laser line point in the image, $(x, y, 1)$ in homogeneous coordinates, we need to find the scalar λ where $\lambda(x, y, 1)$ lies on the plane. This λ is given by:

$$\lambda = \frac{1}{(x, y, 1).\mathbf{n}} \tag{4}$$

and the function π projecting calibrated image points to the plane is therefore given by:

$$\pi((x, y, 1)) = \frac{1}{(x, y, 1).\mathbf{n}}(x, y, 1) \tag{5}$$

This equation maps any point in the image on the laser plane to its corresponding 3D point in the world. The full procedure for 3D reconstruction is outlined in Figure 2.

As the robot moves, a sequence of frames is captured. The 3D points reconstructed from each image are combined

For each Bayer image:
  For each scanline (y):
    For pixels of one colour (red) on scanline:
      Find two largest extrema in $L(x, y)$, x^* and x_2
      If $L(x^*, y) >> L(x_2, y)$ and noise threshold exceeded:
        Find subpixel location of extrema $(x^* + \delta)$
        Reconstruct 3D point $\pi((x^* + \delta, y, 1))$

Fig. 2. Overview of reconstruction procedure

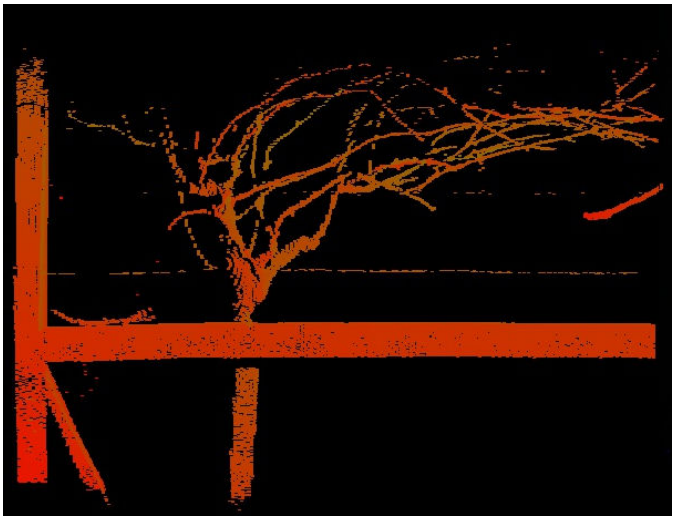

Fig. 3. As the camera and laser move, a 3D depth map is built (coloured by depth).

with the robot's odometry information to generate a dense depth map (Figure 3). Even without any additional outlier removal, very few outlier points are present. In future, the depth measurements from the SL system will be integrated with other measurements from the images to build a 3D model of the crops.

IV. Calibration using a checkerboard pattern

Calibrating the laser line SL system involves estimating the vector $\mathbf{n}$ defining the laser plane. SL calibration schemes involve measuring an object with known 3D dimensions. Early calibration schemes such as [2], [4] use an object with known position with respect to the camera, however in general only the 3D structure of the object is needed.

One suitable object for calibrating a line laser system is a checkerboard pattern. Code to detect the location of a checkerboard pattern in an image with high (subpixel) accuracy is provided in the OpenCV computer vision library [8]. The laser line is clearly visible where it crosses the checkerboard, and using a checkerboard it is easy to automatically capture a large number of images for calibration, across a high range and depth of field. Conveniently, the same checkerboard is also used for calibrating the intrinsic and extrinsic camera parameters.

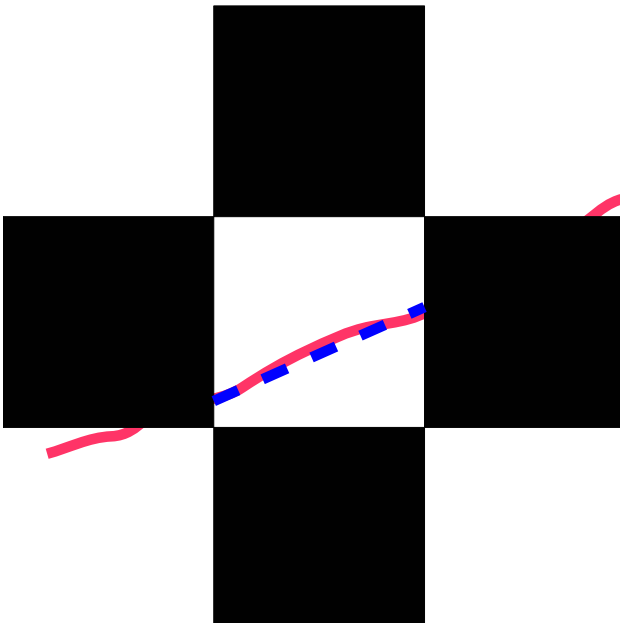

Fig. 5. Segments of the laser line crossing squares in the checkerboard are detected automatically.

The procedure for calibrating the SL system from a checkerboard pattern is outlined in Figure 4. First the checkerboard corner locations are detected with subpixel accuracy, and the locations of laser line pixels in the image are identified (Section II). All coordinates are mapped to normalised homogeneous image coordinates (all lens distortion is corrected here; in addition the laser line is assumed straight). Sections of the laser line which cross squares on the checkerboard are now identified, by iterating over the laser line pixels (sorted by y coordinate) and finding sequences with similar x coordinates (Figure 5). In general the laser line is visible on black but not white squares if lighting levels are high, and white but not black squares if lighting is lower. When a sequence is found, the four checkerboard corners surrounding it are identified, and sections not lying within the checkerboard, or crossing multiple squares, are discarded. A straight line is fitted to the sequence of laser line locations, and the intersection points $\mathbf{p}_1, \mathbf{p}_2$ where the straight line intersects the square edges are found. The size of the squares is known, so the real distance between $\mathbf{p}$ and $\mathbf{q}$,

For each frame:

1) Detect checkerboard corner locations with subpixel accuracy
2) Detect laser line locations with subpixel accuracy
3) Calibrate points
4) Find continuous sequences of pixels on the laser line
5) For each of these line sections:

    Find 2 closest checkerboard corners to each section endpoint
    If the corners found form a square:

        Fit a straight line to segment
        Find points $\mathbf{p}, \mathbf{q}$ where straight line intersects square edges
        Compute length d of line crossing square
        Add constraint: $\|\pi(\mathbf{p}) - \pi(\mathbf{q})\| - d = 0$

Levenberg-Marquardt optimisation to find $\mathbf{n}$ minimising squared errors in constraints

Fig. 4. Overview of calibration procedure

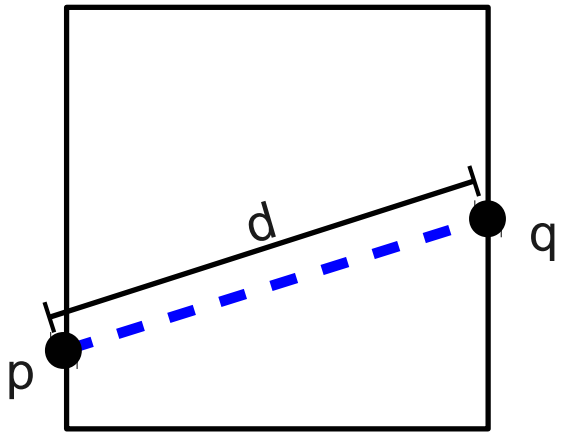

Fig. 6. The distance between points where the laser line crosses the checkerboard square boundaries is computed. Each distance measurement provides a constraint on the laser plane position; the SL system is calibrated by minimising the error in these constraints.

d, can be computed easily by Pythagoras (Figure 6). Each measurement introduces a constraint $\|\pi(\mathbf{p}) - \pi(\mathbf{q})\| - d = 0$; once many constraints, $(\mathbf{p}_i, \mathbf{q}_i, d_i), i = 1...N$, have been extracted, Levenberg-Marquardt optimisation is used to find $\mathbf{n}$ minimising the residual $\sum_i^N (\|\pi(\mathbf{p}_i) - \pi(\mathbf{q}_i)\| - d_i)^2$.

This procedure doesn't make optimal use of the information available, but is relatively simple to implement, and is robust to incorrect measurements and noise–hundreds of constraints without any errors or outliers can be extracted automatically. Examples of the measurements made are shown in Figure 7

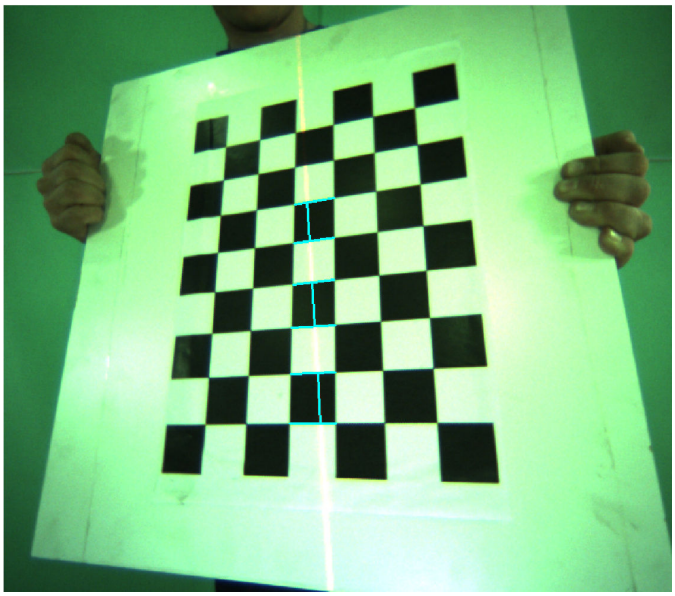

Fig. 7. 3 constraints are added where the laser line crosses a checkerboard pattern.

V. Calibration and Reconstruction Accuracy

This section analyses the errors in reconstructed depths from errors in the calibration, and from errors from locating the laser line in the image. The camera's intrinsic parameters are assumed to be accurate.

A. Quantification of errors from calibration

The SL system was calibrated using 317 constraints extracted from around 200 images (1200 images in total were captured, but many do not show the laser line and the entire checkerboard). The RMS error in the measurements made is 0.69mm, which corresponds to an error of less than 1 pixel in each measurement (the checkerboard squares have 41mm sides).

To estimate the error in the calibration from errors in extracting the constraints from these images, the system is

also calibrated on two disjoint random subsets of 50% of the constraints, and the errors in depth is estimated from the difference between the depths of a set of points (chosen throughout the image), each reconstructed with each of the different calibrations. The RMS error in depth computed using this method is 2.3mm, which corresponds to a relative error in depths of 0.6%. The errors decrease steadily as more measurements are made, and in future more calibration data will be used to further reduce errors.

B. Quantification of errors from line localisation

The second source of errors in reconstruction is from errors in locating the laser line in the image. These errors are challenging to quantify, as they vary with the range, texture, colour, and local lighting levels of the object being measured. Previous experiments have quantified errors by measuring ranges to planar Lambertian objects of constant colour [7], however these results are not necessarily applicable to real-world conditions.

To estimate a lower bound on real-world localisation errors, the line segments detected on the checkerboard pattern are considered. These lines should actually be straight, as the checkerboard pattern is planar. The checkerboard also has approximately constant colour, and no significant texture. The most likely variance about the true value of measurements is computed by fitting a straight line to segments of N points, $\mathbb{P} = \{(x_i, y_i), i = 1...N\}$, then measuring the deviation of each point from the line. If the least-squares line through $\mathbb{P}$ has equation $y = \alpha + \beta x$, then the measurement variance, σ^2, can be estimated by:

$$\sigma^2(\mathbb{P}) = \frac{\sum_1^N y_i - (\alpha + \beta x_i)}{N - 2}, \tag{6}$$

with divisor $N - 2$ as 2 degrees of freedom are lost when the line is fitted.

This estimator is influenced by random variations between point sets (particularly when point sets are small, e.g. 3 points), however it is unbiased, so by the central limit theorem, the average measurement variance over many line segments, $\{\mathbb{P}_j, j = 1..M\}$, gives a suitable estimate of point localisation error:

$$\sigma^2(\{\mathbb{P}_j\}) = \frac{\sum_1^M \sigma^2(\mathbb{P}_j)}{M}. \tag{7}$$

This estimate of measurement variance is verified to be correct by simulation (to 3 significant figures).

For straight line segments on the checkerboard, measurements are estimated to be distributed about their true values with standard deviation 0.23 pixels, which is considerably more accurate than the $1/\sqrt{3} = 0.58$ pixels which is the best which could possibly be obtained without subpixel refinement (considering only the red pixels in the Bayer pattern). For our system setup, a line localisation error of 1 pixel corresponds to a depth error of 1.0mm; hence depth measurements of the checkerboard are distributed about their true values with standard deviation of approximately 0.23mm.

TABLE I
LINE LOCALISATION ERRORS; MEASUREMENT STANDARD DEVIATIONS IN PIXELS FOR DIFFERENT SURFACES, CONSIDERING ONLY RED PIXELS IN BAYER PATTERN.

Surface	Measurement s.d. (pixels)
Wooden frame	0.19
Crops	< 0.15
Checkerboard	0.23
Best possible pixel-accurate location	0.58

For real data, the surfaces being imaged can be assumed to be approximately flat on a small scale, and the same formula can be applied for many sets of 3 approximately collinear laser line points. The estimate of variance can be considered an upper bound, as surface texture will on average increase the errors estimated using this formula. Remarkably, the results show that line localisation is more accurate on the textured crops than on the checkerboard pattern, or even on planar wooden items present in the scene. The reason for this is that the laser line is close to saturating the camera's sensor on the light-coloured wood, and white squares on the checkerboard pattern, and on the dark checkerboard squares it is less visible (so more influenced by sensor noise), and subject to some specular reflection off the black ink. The crops are more matt, and the laser line is more clearly defined here. Estimates of line localisation errors on these different surfaces are summarised in Table I.

Naidu and Fisher [7] perform a similar experiment on a similar scale where geometric shapes are imaged using a line laser system. The points reconstructed using the Parabolic Estimator (the same method used here) have errors with standard deviations of approximately 0.06mm (versus around 0.05mm for the Gaussian Estimator, which is not practical for our application; Section II). While more accurate than our system, their imaging conditions are more controlled, with little ambient light other than from the line laser, and a uniformly coloured, untextured surface.

VI. CONCLUSION

This paper described a hybrid computer vision and structured light sensor for building a dense 3D reconstruction of crops imaged by a robot. In real data, measured depths have errors from line localisation with standard deviation approximately 0.15mm.

A simple scheme to calibrate the laser line system which uses the image of the laser line on a checkerboard pattern is also described. The system is designed so that large numbers of measurements for calibration can be collected easily using a robust automated process. Errors in the calibration from 317 measurements have standard deviation of approximately 2.3mm, and accuracy will improve further with additional measurements.

Future work will focus on more robust identification of the laser line in images under variable lighting conditions, and on the integration of the measured depths into a system to build a complete 3D reconstruction.

REFERENCES

[1] Y. Shirai, "Recognition of polyhedrons with a range finder," *Pattern Recognition*, vol. 4, no. 3, pp. 243–244, 1972.

[2] D. Huynh, R. Owens, and P. Hartmann, "Calibrating a structured light stripe system: a novel approach," *International Journal of computer vision*, vol. 33, no. 1, pp. 73–86, 1999.

[3] P. Besl, "Active, optical range imaging sensors," *Machine vision and applications*, vol. 1, no. 2, pp. 127–152, 1988.

[4] F. DePiero and M. Trivedi, "3-d computer vision using structured light: Design, calibration, and implementation issues," *Advances in Computers*, vol. 43, pp. 243–278, 1996.

[5] Kinect; Wikipedia, September 2011. [Online]. Available: http://en.wikipedia.org/wiki/Kinect

[6] C. Rocchini, P. Cignoni, C. Montani, P. Pingi, and R. Scopigno, "A low cost 3d scanner based on structured light," in *Computer Graphics Forum*, vol. 20, no. 3. Wiley Online Library, 2001, pp. 299–308.

[7] D. Naidu and R. Fisher, "A comparative analysis of algorithms for determining the peak position of a stripe to sub-pixel accuracy," in *Proc. British Machine Vision Conf*, 1991, pp. 217–225.

[8] OpenCV Computer Vision Library, n.d. [Online]. Available: http://opencv.willowgarage.com/

[9] Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proceedings of the International Conference on Computer Vision*, 1999.