

A Low-Cost Range Finder using a Visually Located, Structured Light Source

R. B. Fisher, A. P. Ashbrook, C. Robertson, N. Werghi
Division of Informatics, Edinburgh University
5 Forrest Hill, Edinburgh EH1 2QL

Abstract

In this paper we show how the cost of a structured light, range finding system can be substantially reduced by visually tracking the structured light source. To be able to recover range measurements using a structured light range finder the relative positions of the structured light source and light sensor must be known. This is typically achieved by carefully controlling the position of at least one of these components using expensive mechanical actuators. Instead, we propose that little or no control is placed on the positioning of these components and that the position of the structured light source is determined using visual feedback. A low-cost prototype system employing this principle is presented.

1 Introduction

Structured-light triangulation-based range finders are now relatively common commercial products which are used in a wide range of applications. Unfortunately, they can also be quite expensive, typically costing between \$5000 and \$100,000, leaving them inaccessible to small businesses and domestic users.

A significant proportion of the cost of conventional range finders arises from the need to accurately control the relative positions of a structured light source, a light sensor and the surface being investigated. Typically the light sensor and camera will be fixed and the surface being captured will be systematically moved or alternatively, the camera and surface will be fixed and the light source moved. This is generally achieved using expensive mechanical actuators that allow very accurate movements. In this paper we propose that the camera and surface remain fixed in position and that the position of the structured light source is controlled by a human operator. The position of the structured light source is determined by observation from a calibrated camera, which is the same sensor used to observe the projected light.

To demonstrate the principle of using visual feedback

to determine the position of the structured light source for range finding, a system employing this approach has been implemented. To keep the cost of the system as low as possible the structured light is formed by interrupting a point light source with an opaque rod, resulting in two shadow planes. The opaque rod, or wand as we refer to it, is the source of the structured light. This source of structured light has been used previously to produce a low-cost range finding system [2] but in that work the position of the shadow source is not determined explicitly. Instead the poses of the two shadow planes are determined by observing their intersection with a calibrated plane lying under the object being captured. The need for this plane limits the flexibility of the approach, as for example, objects cannot easily be captured in situ or large models constructed by stitching together a number of smaller sections. Our solution avoids the need for the calibrated plane. We estimate that the equipment cost for this range finder could be reduced to \$100 or less. All that is required is a chip camera with fixed lens, halogen lamp, PC interface card, power supply, rigid frame and wand. Even with marketing and profit considerations, this price is clearly low enough to make 3D scanning available to a consumer market.

Other systems have also employed a visible light source to generate the structured light by projecting the light through a patterned mask [1, 3]. Typically this pattern will be a simple slit, so that a plane of light is produced, although a variety of other configurations have been used. As with other conventional range finders the position of the structured light must be accurately controlled. The novelty of our approach is not in the source of the structured light but in the way the position of this source is determined.

Previously, we developed another hand-held scanner [4] that used a laser stripe source and a global position sensor and was thus much more expensive. That sensor did have the advantage that it could scan all sides of an object; however, its orientation accu-

racy allowed spatial accuracy only to about 5 mm, as compared to 1 mm here. Moreover, that sensor was sensitive to metal objects and magnetic fields in the scanned environment, whereas those problems are not encountered with the new sensor.

2 Structured Light Range Finding

Traditional structured light range finders [6] project a set of light rays onto the surface being captured and the scene is viewed with a calibrated camera. The projection of the light rays onto the surface provides visible features that are easily detected in the camera image.

The structured light source can be represented by a set of light rays which in turn are represented by a point on the ray \mathbf{l}_i and a unit direction vector \mathbf{r}_i . Any point \mathbf{x}_i along the i^{th} ray is therefore given by the expression:

$$\mathbf{x}_i = \mathbf{l}_i + \gamma_i \mathbf{r}_i \quad (1)$$

where γ_i is a scalar representing the distance of \mathbf{x}_i along the ray relative to \mathbf{l}_i . The structured light source R is simply the set of these rays.

$$R = \{\mathbf{l}_i, \mathbf{r}_i\} \quad (2)$$

Without loss of generality we assume that the position of the camera and the surface being captured are fixed and that the structured light source moves. If the structured light source is moved into position from its canonical frame by a transformation \mathbf{T} , then any point along the i^{th} ray is given by the expression:

$$\mathbf{x}_i = \mathbf{T}_1 \mathbf{l}_i + \gamma_i \mathbf{T}_2 \mathbf{r}_i \quad (3)$$

This is for a general light source. In the implemented case described below, the single light source is in a fixed position. Hence, the structured light rays are described by:

$$\mathbf{x}_i = \mathbf{l} + \gamma_i \mathbf{T} \mathbf{r}_i \quad (4)$$

Given a detected point of light, s_j , in the camera image, the camera calibration provides a line of sight vector, $\mathbf{v}(s_j)$, that passes through the detected surface point. The position of the surface point, \mathbf{x}_j , must lie somewhere along this line of sight:

$$\mathbf{x}_j = \mathbf{c} + \lambda_j \mathbf{v}(s_j) \quad (5)$$

where \mathbf{c} is the position of the camera.

The camera model used here is based on an empirical calibration of the sensor that computes a 3D line of sight for each pixel. This overcomes the need for an accurate parametric camera model when there is much

projective and lens distortion (see Fig. 5b). (Subpixel rays are computed by bilinear interpolation.) Empirical calibration is achieved by observing a calibrated regular grid at two depths and relating the known 3D points seen by each pixel in the two images. The camera center \mathbf{c} is estimated by finding the point that minimizes the distance to all of the pixel lines of sight.

When a correspondence has been established between the i^{th} structured light ray and the j^{th} detected point of light, the 3-dimensional position of the surface point is determined by equating Equations (3) and (5). To be able to calculate this intersection the position of this structured light source, represented by the transformation \mathbf{T} , must be known. Conventionally this is done by initially measuring the position of the light source, using some calibration procedure, and then controlling its movement very accurately. To substantially reduce the cost of the system we suggest that the position of the structured light source is estimated using visual feedback.

To further simplify the search for correspondences and the calculation of intersection points it is common to use a plane of light instead of individual rays. In this case the structured light source is essentially an infinite number of rays all of which lie in a plane. This plane of light projects a line onto the surface being captured. In this case there is no need to establish correspondences for detected pixels because there is only one plane, and the calculation of the position of the surface point is a simple line-plane intersection calculation.

3 A low-cost range sensor

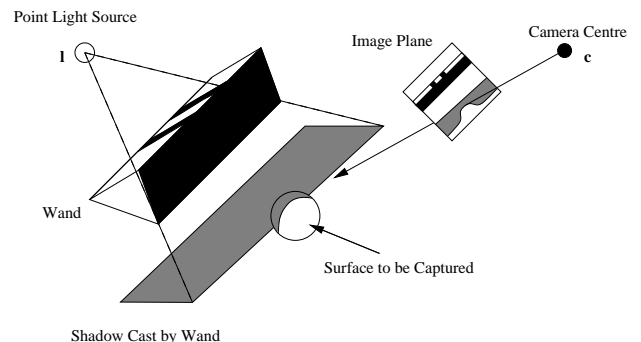


Figure 1: The design of the low-cost range finding system.

Figure 1 presents a schematic of the complete range finder. The structured light source has been realised using a point source and a wand. The wand is positioned by hand so that it interrupts the light source,

casting a shadow onto the surface being captured. To be able to recover the surface shape, the position of the wand and the position of the shadow on the surface need to be known. This is achieved by viewing both the wand and the surface using a calibrated single video camera. During the capture process the human operator simply moves the wand so that the edges of the shadow move over the surface while keeping the wand in the field of view of the camera.

The wand has been designed so that it can be easily detected in an image and its 3-dimensional pose determined from a single perspective view. The design used in this implementation is a rod with a triangular section as presented in Figure 2. To simplify the detection of the wand, the face that points towards the light source is painted in white and the opposite visible face is painted in black. The pose of the wand can be determined from the appearance of the two faces without any additional markings, although this was found to be unstable in the direction of the wand rotating away from the camera. To improve the pose estimation, two bands of known thickness and separation are painted onto the wand's white face. The length of the wand is not used in the pose estimation so the ends need not be visible from the camera.

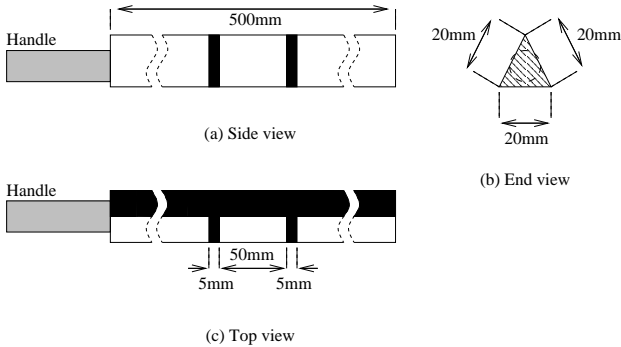


Figure 2: The design of the wand used in the implemented system.

3.1 Implementation

In this section the details of the implemented range finder are described. The complete system performs the following actions for each frame of the video sequence.

1. The pixels along the three wand edges and the four wand band edges are detected.
2. The pose of the wand is determined from the detected edge features.
3. The pixels along the shadow edges are detected.

4. The depth of pixels moving into or out of shadow are determined from their position in the image, the pose of the wand and the position of the light source.

3.1.1 Detection of the Pixels along the Wand Edges

The detection of the pixels along the observed edges of the wand requires four images, representing image statistics, constructed from two short image sequences taken prior to the main capture phase. In the first sequence the field of view only contains the object or surface to be scanned while in the second the wand is moved in such a way that by the end of the sequence it has covered the entire frame. The light source is on for both of these sequences, as it is during range data capture. The mean and standard deviation of each pixel is determined for the first sequence and these are represented by the mean and noise images I_{ij}^{mean} and I_{ij}^{σ} at each pixel (i, j) . The maximum and minimum values of each pixel in the second sequence are represented by the maximum and minimum images I_{ij}^{max} , I_{ij}^{min} . The importance of these images is that they provide an estimate of the intensities of the bright and dark faces respectively of the wand at each point in the image. This is a good assumption if the orientation of the wand axis is approximately perpendicular to the direction of motion and if the direction of motion is parallel to an imaginary line joining the light source and the camera.

For each image row the following search is conducted to find the pixels along the three edges of the wand. Initially the *first* pixel on the bright face of the wand is detected by finding a change in intensity from the mean level to the maximum level. This is expressed by the condition:

$$I_{ij} > \frac{1}{2}(I_{ij}^{max} + I_{ij}^{mean}) \quad (6)$$

Once this condition has been satisfied the position of either the current pixel, (i, j) , or the previous pixel, $(i, j - 1)$, is associated with the left edge of the wand. The choice is based on the pixel whose intensity is closest to the threshold $\frac{1}{2}(I_{ij}^{max} + I_{ij}^{mean})$.

During this first part of the search the difference between the mean and maximum levels is checked to be sufficiently large so that false detection is improbable. The difference between the mean and maximum intensities is considered to be sufficiently large if:

$$I_{ij}^{max} - I_{ij}^{mean} > 4I_{ij}^{\sigma} \quad (7)$$

If this condition is not satisfied it is uncertain whether or not the edge of the wand has been detected and the search along the row is abandoned.

In the second part of the search the central edge of the wand is detected by finding a change in intensity from the maximum value to the minimum value. This is expressed by the condition:

$$I_{ij} < \frac{1}{2}(I_{ij}^{max} + I_{ij}^{min}) \quad (8)$$

As before, once this condition is satisfied the position of either the current or previous pixel is associated with the wand edge, based on the pixel whose intensity is closer to the threshold.

Because the intensities of the two visible faces of the wand have been designed to be as different as possible there is no need to test for false detections during this part of the search.

Finally, the edge of the dark face of the wand is detected by finding a change in intensity from the minimum level back to the mean level. This is expressed by the condition:

$$I_{ij} > \frac{1}{2}(I_{ij}^{mean} + I_{ij}^{min}) \quad (9)$$

Again, once this condition is satisfied the position of either the current or previous pixel is associated with the wand edge, based on the pixel whose intensity is closer to the threshold.

During this part of the search the difference between the mean and minimum levels is checked to be sufficiently large so that false detection is improbable. The difference between the mean and minimum intensities is considered to be sufficiently large if:

$$I_{ij}^{mean} - I_{ij}^{min} > 4I_{ij}^{\sigma} \quad (10)$$

3.1.2 Detection of the Pixels along the Wand Band Edges

Once the left and middle edges of the wand have been detected a search is conducted over the left face of the wand for the observed band edges. Because the band edges are orthogonal to the wand edges this search is conducted along lines of pixels parallel to the detected left edge pixels of the wand. The algorithm is similar to that presented in the previous section.

3.1.3 Determining the Wand Position

The position of the wand in the world is determined by finding the pose that is consistent with the detected visual features. These features, namely the wand edges

and the edges of the band markings, place constraints on the position of the wand and in combination allow the pose to be completely determined. The complete procedure can be summarized as follows:

1. The *planes of sight* that contain the detected wand edges and wand band edges are determined.
2. The wand axis is determined from the plane of sight normals.
3. The planes of sight for the wand edges are re-determined using the direction of the axis as an additional constraint.
4. The *planes of sight* are used to determine 5 *lines of sight* passing through features on the wand of known separation.
5. A search is conducted for the positions of these features along the lines of sight that are consistent with the known feature separations. This gives the wand position.

Each of the straight edges of the wand and the straight edges of the band markings lie in planes which pass through the optical center of the camera. Using the positions of the detected edge pixels and the camera calibration information these *planes of sight* can be determined and represented by their normal vectors. The planes of sight passing through the left, middle and right edges of the wand are parameterized by normal vectors \mathbf{n}_l , \mathbf{n}_m and \mathbf{n}_r respectively. Similarly, the planes of sight passing through the band edges are parameterized by normal vectors \mathbf{n}_1 , \mathbf{n}_2 , \mathbf{n}_3 and \mathbf{n}_4 . This notation is presented in Figure 3. For each pixel \mathbf{w}_i that lies on one of these linear image features the camera calibration table provides a direction vector that lies in the appropriate plane of sight. The plane of sight normal is then determined by finding the vector that is orthogonal to all of these direction vectors by a least-squares method. The details of this are presented in Appendix A.

Given that the edges of the wand run parallel to the wand axis, the plane of sight normal vectors for the wand edges will be orthogonal to the axis. In principle then, the axis direction can be estimated as the cross product of any two of these normals and a better estimate made using all three (again refer to Appendix A). In practice the estimate of the axis direction formed this way is sensitive to small errors in the estimation of the planes of sight, largely because the planes of sight are close to being parallel. The axis estimation is improved by introducing an additional

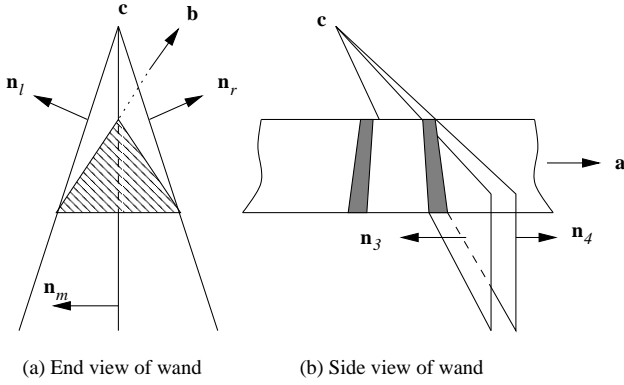


Figure 3: (a) The three planes of sight, \mathbf{n}_l , \mathbf{n}_m and \mathbf{n}_r , passing through the camera \mathbf{c} containing the left, middle and right edges of the wand. The vector \mathbf{b} is parallel to the wand side band edges. (b) The four planes of sight \mathbf{n}_1 , \mathbf{n}_2 , \mathbf{n}_3 and \mathbf{n}_4 contain the edges of the wand bands. The vector \mathbf{a} is the wand axis.

vector, \mathbf{b} , that is also orthogonal to the wand axis but is not nearly parallel to the plane of sight normals. This is the vector that is parallel to the wand band edges and may be determined from the wand band plane of sight normals. The vector \mathbf{b} is the vector that is orthogonal to all of the band plane normals (see Appendix A).

Although the plane of sight normals for the wand edges should all be orthogonal to the axis direction, this constraint is not used during the initial plane of sight estimation. This is because a closed form solution to the problem of simultaneously estimating the plane normals and the axis direction has not been found although it is desirable reasons of efficiency. Having estimated the axis direction, \mathbf{a} , from the initial planes of sight the planes of sight, the wand edges are re-fitted with the constraint that their plane normals must be orthogonal to the axis. This problem does have a simple, closed form solution and the re-fitting improves later stages of the wand position estimation. Appendix B describes how the planes of sight can be fitted when the axis constraint is imposed.

Five lines of sight are then determined which pass through detected features on the wand. These are depicted in Figure 4 for clarity. The first three of these vectors, \mathbf{v}_l , \mathbf{v}_m and \mathbf{v}_r , pass through the left, middle and right edges of the wand such that all three vectors lie in a plane orthogonal to the wand axis. In other words, these three vectors describe an orthogonal cross section through the wand. Each vector is simply determined from the appropriate wand planes of sight and the wand axis.

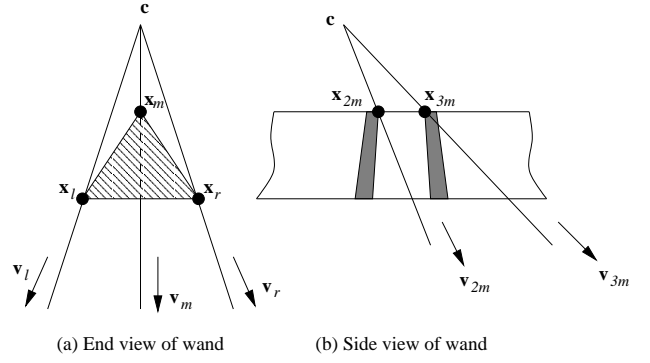


Figure 4: The five lines of sight \mathbf{v}_l , \mathbf{v}_m , \mathbf{v}_r , \mathbf{v}_{2m} and \mathbf{v}_{3m} that contain five features \mathbf{x}_l , \mathbf{x}_m , \mathbf{x}_r , \mathbf{x}_{2m} and \mathbf{x}_{3m} , with known separation.

$$\mathbf{v}_l = \mathbf{n}_l \times \mathbf{a} \quad \mathbf{v}_m = \mathbf{n}_m \times \mathbf{a} \quad \mathbf{v}_r = \mathbf{n}_r \times \mathbf{a}$$

The other two line of sight vectors, \mathbf{v}_{2m} and \mathbf{v}_{3m} , describe the lines passing from the camera center through two of the wand band corners. These vectors are easily determined from the plane of sight normal for the middle wand edge and the normals for two of the band edges.

$$\mathbf{v}_{2m} = \mathbf{n}_m \times \mathbf{n}_2 \quad \mathbf{v}_{3m} = \mathbf{n}_m \times \mathbf{n}_3$$

Finally, the wand position is determined by searching for the position of the point \mathbf{x}_m on the middle edge of the wand that best satisfies the expected distances between the features along the other lines of sight. The position of this point is given by the scalar λ_m and the line of sight vector \mathbf{v}_m , such that:

$$\mathbf{x}_m = \lambda_m \mathbf{v}_m + \mathbf{c} \quad (11)$$

Similarly, the positions of the other features depend upon scalars λ_l , λ_r , λ_{2m} and λ_{3m} . Given a particular position, \mathbf{x}_m , for the point on the middle wand edge the position of the other features relative to this point can be determined such that the appropriate distances between \mathbf{x}_m and the other features are satisfied. These distances are fixed by the physical dimensions of the constructed wand. The details of how these scalars are determined are given in Appendix C.

To find the correct value of λ_m an error function $E(\lambda_m)$ is defined that reflects the self-consistency in the distance between features. This is based on the two following constraints:

$$|\mathbf{x}_l - \mathbf{x}_r| = D_1 \quad \text{and} \quad |\mathbf{x}_{2m} - \mathbf{x}_{3m}| = D_2 \quad (12)$$

The error function is defined to be zero when these constraints are satisfied.

$$E(\lambda_m) = |\mathbf{x}_l - \mathbf{x}_r| - D_1 + |\mathbf{x}_{2m} - \mathbf{x}_{3m}| - D_2 \quad (13)$$

Note that it is not necessary to square the terms $|\mathbf{x}_l - \mathbf{x}_r| - D_1$ and $|\mathbf{x}_{2m} - \mathbf{x}_{3m}| - D_2$ as these vary monotonically with λ_m , only summing to zero at the correct solution.

The best wand pose is found by minimizing this function using a bisection search [7] over the range $\lambda_{min} \leq \lambda_m \leq \lambda_{max}$. The derivation for λ_{min} and λ_{max} are also given in Appendix C. At this point, we now know the 3D position of the wand, and can use this to estimate the shadow plane 3D positions.

3.1.4 Detection of the Shadow Edge Pixels

As the wand moves over the surface being measured, parts of the surface move into shadow and parts of the surface move out of shadow. The purpose of the shadow detection is to find the leading and trailing pixels of the moving shadow edge. These edges correspond to the intersection of the shadow planes and the surface being measured at the instant the frame is taken.

For each row, i , a search is conducted for the left shadow edge pixel, s_{li} , and the right shadow edge pixel, s_{ri} . The search begins from the detected right edge of the wand and proceeds to the right along each row. The left shadow edge is detected by finding a change in intensity from the mean level to the minimum level. This is expressed by the condition:

$$I_{ij} < \frac{1}{2}(I_{ij}^{mean} + I_{ij}^{min}) \quad (14)$$

Once this condition is satisfied the sub-pixel position of the shadow edge, s_{li} , is estimated by linear interpolation of the intensity of the current pixel and the intensity of the previous one. The difference between the mean and minimum intensities is checked to be sufficiently large so that false detection is improbable, as was done previously when detecting the wand edges. The condition which must be satisfied is given by the expression:

$$I_{ij}^{mean} - I_{ij}^{min} > 4I_{ij}^{\sigma} \quad (15)$$

The search for the shadow edges along this row is abandoned if this condition is not satisfied. After the left shadow edge has been detected the search continues along the row for the right shadow edge. This is expressed by the opposite condition:

$$I_{ij} > \frac{1}{2}(I_{ij}^{mean} + I_{ij}^{min}) \quad (16)$$

Again the sub-pixel edge location, s_{ri} , is determined by linear interpolation of the intensity of the current pixel and the intensity of the previous one. The difference between the minimum and mean intensities of each pixel is also checked to be sufficiently large as before.

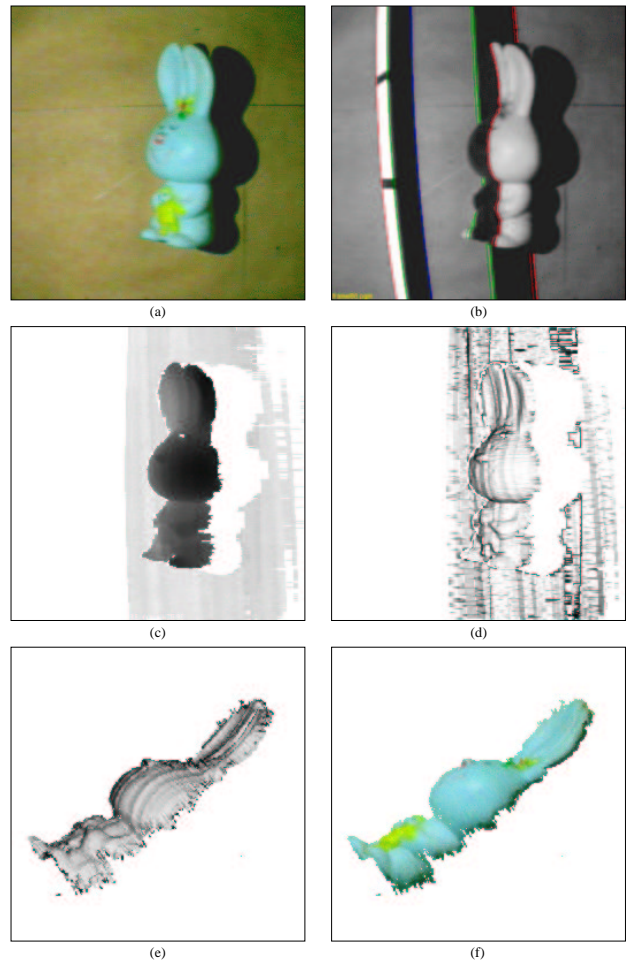


Figure 5: (a) A single color (here monochrome) image of the object. (b) A frame during the scanning process. (c) The resulting range data. (d) The shaded range data. (e) Shaded range data from a new viewpoint. (f) Colored (here monochrome) range data from a new viewpoint.

The stripe detection technique used here has some similarity to the spatio-temporal method [8, 2]. However, because we do not acquire dense shadow edge images (*e.g.* one image every 5 pixels instead of several images per pixel), that type of analysis is not easily

applicable. So, instead, our technique uses only the mean, maximum and minimum values at each pixel for determining the edge position.

3.1.5 Calculating Range Values

In each frame, a 3-dimensional point is calculated for each pixel which has either moved into or out of shadow since the previous frame. For pixels that lie along either of the current shadow edges the positions are determined by intersecting the lines of sight for those pixels with the appropriate shadow planes cast by the wand.

The left and right shadow planes P_l and P_r are determined from the known light position \mathbf{l} and the pose of the wand. Here, each plane is parameterized by a unit normal $\hat{\mathbf{m}}$ and a perpendicular distance to the origin d such that $P_l = \{\hat{\mathbf{m}}_l, d_l\}$ and $P_r = \{\hat{\mathbf{m}}_r, d_r\}$. Points \mathbf{p}_l on the left shadow plane and points \mathbf{p}_r on the right shadow plane satisfy the following constraints.

$$\mathbf{p}_l^T \hat{\mathbf{m}}_l = d_l \quad \text{and} \quad \mathbf{p}_r^T \hat{\mathbf{m}}_r = d_r \quad (17)$$

A line passing from the light source to a point on a wand edge will lie in the shadow plane as will the edge of the wand itself. The shadow plane normals can therefore be expressed as:

$$\mathbf{m}_l = (\mathbf{l} - \mathbf{x}_l) \times \mathbf{a} \quad \text{and} \quad \mathbf{m}_r = (\mathbf{l} - \mathbf{x}_r) \times \mathbf{a} \quad (18)$$

where \mathbf{a} is the axis direction, \mathbf{x}_l is a point on the left wand edge and \mathbf{x}_r is a point on the right wand edge, as discussed in Section 3.1.3. Now d_l and d_r are determined by substituting \mathbf{x}_l and \mathbf{x}_r into Equation (17).

$$d_l = \mathbf{x}_l^T \hat{\mathbf{m}}_l \quad \text{and} \quad d_r = \mathbf{x}_r^T \hat{\mathbf{m}}_r \quad (19)$$

Given an observed pixel on, for example, the left shadow edge s_{li} and the corresponding line of sight vector from the camera \mathbf{v}_{li} from the camera calibration table, the 3-dimensional position of the surface point \mathbf{x}_{li} will lie somewhere along the line of sight.

$$\mathbf{x}_{li} = \lambda_{li} \mathbf{v}_{li} + \mathbf{c} \quad (20)$$

Note that the line of sight vector, \mathbf{v}_{li} , is determined by linear interpolation of the camera calibration data given the sub-pixel position s_{li} . The actual 3-dimensional position is determined by intersecting the line of sight with the left shadow plane. This is solved by substituting Equation (20) into Equation (17). This gives the following expression for λ_{li} .

$$\lambda_{li} = \frac{d_l}{\mathbf{v}_{li}^T \mathbf{m}_l} \quad (21)$$

Substituting λ_{li} into Equation (20) gives the position \mathbf{x}_{li} .

$$\mathbf{x}_{li} = \left(\frac{d_l}{\mathbf{v}_{li}^T \mathbf{m}_l} \right) \mathbf{v}_{li} + \mathbf{c} \quad (22)$$

Similarly pixels, s_{ri} , observed on the right shadow edge provide a 3-dimensional point, \mathbf{x}_{ri} .

To determine the 3-dimensional position of each point over which the shadow edge has passed since the previous captured image frame it is necessary to estimate the position of the wand at the instance the shadow edge was at each pixel. By making the assumption that the wand is moving at a constant velocity and that the surface is approximately planar over the small distance the wand moves between image frames then the time at which the wand passes each pixel is linear. In this instance it is sufficient to linearly interpolate the depth of each pixel from the current shadow edge position back to the previous shadow edge position. This allows calculation of a range value at each pixel, not just at pixels where a shadow edge is seen. This allows much faster range image acquisition than *e.g.* [2] where multiple image frames are needed at each pixel. Moreover, as two shadow edges are observed, then each pixel gets 2 estimated range values which are averaged.

3.2 Experiments

We show three experiments have been conducted to demonstrate the flexibility and accuracy of the implemented system. In each case the captured sequence is approximately 30 frames long and each frame has a resolution of 512 by 512 pixels. The light source is a standard halogen reading light bulb separated by 140 mm from the camera, which is a standard PAL color camera with a 14 mm focal length lens. The results for the first two experiments are presented in Figures 5 and 6. Because we use a color TV camera, the texture mapping is colored; here the results are shown in monochrome, but full colored images can be seen at: http://www.dai.ed.ac.uk/daiddb/staff/personal_pages/anthonya/Striper/stripier.html. Six images are presented:

- (a) A single frame from the camera before the wand is introduced into the scene.
- (b) A frame during the capture sequence.
- (c) The captured range data where the intensity of each pixel is proportional to the depth of the surface.
- (d) Cosine shaded range data to emphasize the captured surface structure.

- (e) The shaded range data from a different viewpoint. This gives a much better impression of the structure of the captured surface.
- (f) The captured range data with the color of each pixel texture mapped onto the surface.

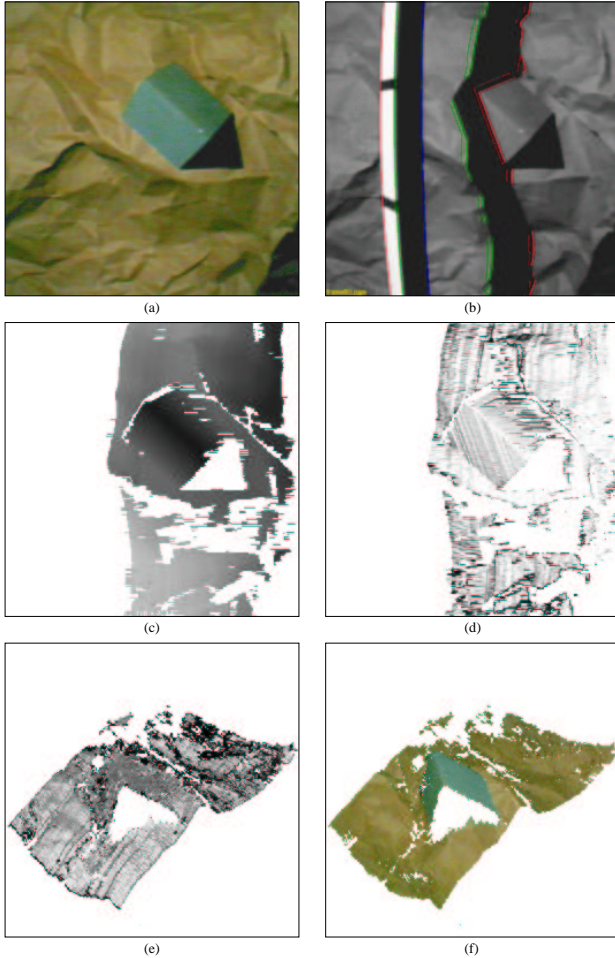


Figure 6: (a) A single color image of the object. (b) A frame during the scanning process. (c) The resulting range data. (d) The shaded range data. (e) Shaded range data from a new viewpoint. (f) Coloured range data from a new viewpoint.

In the first experiment the scene being scanned contains a single object placed on a planar surface. The object has been chosen because it is both colorful and has quite complex, free-form surfaces. It is evident that the captured surface has the expected shape and demonstrates the visual quality of the colored range data and the ability of the system to capture relatively small details. Observe the rabbit's nose which is only a few millimeters in size.

One of the advantages of the approach we have taken is that there is no need for the object being captured to be placed on some calibrated surface. This provides the flexibility to capture objects *in situ* or to capture parts of an object, for example, part of a building. This is demonstrated in the second experiment, presented in Figure 6, by placing a simple object on a rough surface of unknown geometry. A simple object has been chosen in this case to demonstrate that the surface has been captured correctly despite the lack of a calibrated surface. The holes in the surface appear because of occlusions.

These experiments demonstrate the general quality of the results but provide no quantitative measure of the accuracy of the implementation. The last experiment addresses this by fitting a plane to data captured from a planar surface. The results of the captured surface are presented in Figure 7. After fitting a plane to the data the RMS error was found to be 0.924mm.

While we have not particularly optimized the range calculation code, nor used any equipment other than a SUN Ultra 1 (200 Mhz), we can currently process 6 video frames a second. We expect that this could be easily increased to 10+ frames a second, allowing the user to scan the stick at 5 cm/sec, giving range samples every 5mm (approximately) across the scanned object.

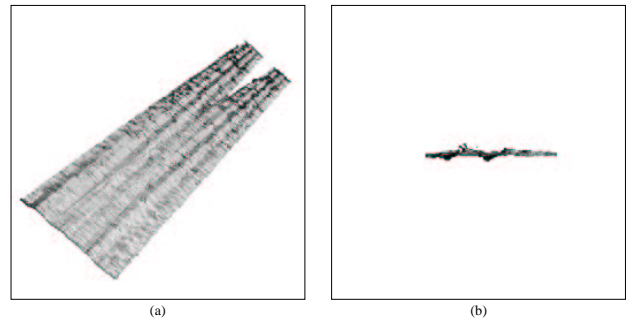


Figure 7: (a) Shaded range data of the captured plane. (b) A side on view of the plane to emphasize the captured structure.

4 Conclusions

The motivation behind the work presented in this paper is the requirement for a low-cost system capable of capturing colored 3-dimensional shape data from real objects or environments. Such a system will provide an invaluable tool for the development of detailed, realistic looking models and environments in a wide range of applications.

We have shown that we can use visual feedback to track the structured light source. This enables the

structured light source position to be controlled using much cheaper actuators or a human operator.

The full system requires only a desktop computer, a single color video camera, a point light source and a wand with some markings. Initial results of this prototype system look very promising. A brief quantitative analysis suggest that the RMS measurement error is less than a millimeter.

A Fitting a Normal to a Set of Coplanar Vectors

Given a set of vectors \mathbf{v}_i which lie in some plane we would like to determine the plane normal \mathbf{n} . The relationship between the vectors and the plane normal can be expressed as follows.

$$\mathbf{v}_i^T \mathbf{n} = 0 \quad (23)$$

In practice this condition is not precisely satisfied because of errors on each of the given vectors. Instead, the plane normal is estimated using a *least squares* approach in which the error to be minimized, E^2 , is defined as follows:

$$\begin{aligned} E^2 &= \sum_i (\mathbf{v}_i^T \mathbf{n})^2 \\ &= (\mathbf{V}^T \mathbf{n})^T (\mathbf{V}^T \mathbf{n}) \\ &= \mathbf{n}^T (\mathbf{V} \mathbf{V}^T) \mathbf{n} \\ &= \mathbf{n}^T \mathbf{H} \mathbf{n} \end{aligned} \quad (24)$$

where \mathbf{V} is a matrix formed from all of the line of sight vectors.

$$\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_N] \quad (25)$$

The solution to this that minimizes the error function is the eigenvector that corresponds to the smallest eigenvalue of the matrix \mathbf{H} .

B Fitting a Normal to a Set of Coplanar Vectors with an Orthogonal Constraint

Given a set of coplanar vectors \mathbf{v}_i we would like to determine the plane normal \mathbf{n} . In this instance, however, we would also like to impose the constraint that the plane normal is also orthogonal to a specified vector \mathbf{a} . This can be expressed as:

$$\mathbf{n}^T \mathbf{a} = 0 \quad (26)$$

Again a least squares approach is taken except this time the above constraint is introduced into the error function E^2 , such that:

$$E^2 = \mathbf{n}^T \mathbf{H} \mathbf{n} + \lambda \mathbf{n}^T \mathbf{A} \mathbf{n} \quad (27)$$

where $\mathbf{A} = \mathbf{a} \mathbf{a}^T$. This is solved by a generalized eigenvalue method.

C Line of Sight Feature Constraints

Given the intersection, $\mathbf{x}_m = \lambda_m \mathbf{v}_m + \mathbf{c}$, of the middle wand edge and the middle line of sight \mathbf{v}_m , the position of the other features can be determined such that their distances from \mathbf{x}_m are satisfied.

The distance between the intersection points \mathbf{x}_l and \mathbf{x}_r and the given point \mathbf{x}_m is equal to the width of the wand faces, D_1 , as shown in Figure 4 (a). This can be expressed as the following two constraints.

$$(\mathbf{x}_m - \mathbf{x}_l)^T (\mathbf{x}_m - \mathbf{x}_l) = (\lambda_m \mathbf{v}_m - \lambda_l \mathbf{v}_l)^T (\lambda_m \mathbf{v}_m - \lambda_l \mathbf{v}_l) = D_1^2 \quad (28)$$

And similarly.

$$(\mathbf{x}_m - \mathbf{x}_r)^T (\mathbf{x}_m - \mathbf{x}_r) = (\lambda_m \mathbf{v}_m - \lambda_r \mathbf{v}_r)^T (\lambda_m \mathbf{v}_m - \lambda_r \mathbf{v}_r) = D_1^2 \quad (29)$$

These constraints are easily rearranged to give quadratic expressions for the unknowns λ_l and λ_r in terms of the line of sight vectors and the given value of λ_m .

$$\lambda_l^2 - \lambda_l (2\lambda_m \mathbf{v}_l^T \mathbf{v}_m) + (\lambda_m^2 - D_1^2) = 0 \quad (30)$$

And similarly.

$$\lambda_r^2 - \lambda_r (2\lambda_m \mathbf{v}_r^T \mathbf{v}_m) + (\lambda_m^2 - D_1^2) = 0 \quad (31)$$

This gives two possible solutions for the points on the left and right edges but provided that the distance to the top, middle edge of the wand is less than the distance to the other two edges then the larger solution is selected in each case.

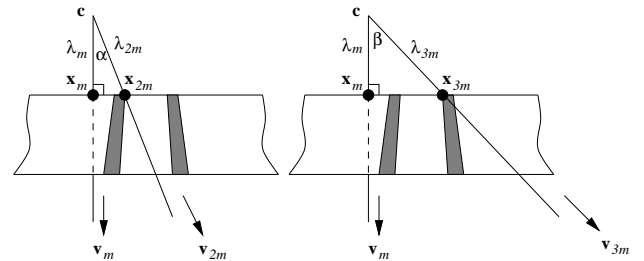


Figure 8: The relationship between λ_m , λ_{2m} and λ_{3m} .

Figure 8 presents the relationship between the given scalar distance λ_m and the distances λ_{2m} and λ_{3m} to

the two wand band features. Recall that the line of sight vector \mathbf{v}_m is orthogonal to the wand axis and is therefore also orthogonal to the middle edge of the wand. These relationships can be expressed as follows:

$$\lambda_{2m} = \frac{\lambda_m}{\cos \alpha} = \frac{\lambda_m}{\mathbf{v}_m^T \mathbf{v}_{2m}} \quad (32)$$

And similarly.

$$\lambda_{3m} = \frac{\lambda_m}{\cos \beta} = \frac{\lambda_m}{\mathbf{v}_m^T \mathbf{v}_{3m}} \quad (33)$$

As discussed in the main text, the pose of the wand is determined by searching for a value of \mathbf{x}_m that is consistent with the detected image features. This is implemented using a bisections search over the range of λ_m and so minimum and maximum values for this variable need to be determined. Clearly the wand will only be visible when it is in front of the camera and so λ_m cannot be negative. As such, the minimum possible value for λ_m is:

$$\lambda_{min} = 0 \quad (34)$$

An upper bound for λ_m , λ_{max} , derivable from simple geometry is:

$$\lambda_{max} = \frac{D_1}{\sin(\cos^{-1}(\mathbf{v}_l^T \mathbf{v}_r))} \quad (35)$$

Tighter upper and lower bounds can be derived.

Acknowledgements

This work was funded by UK EPSRC grants GR/L25110 and GR/K80938.

References

- [1] R. A. Jarvis, "A Perspective on Range Finding Techniques for Computer Vision", IEEE PAMI, Vol. 5, pp 122-139, 1983.
- [2] J. Y. Bouguet and P. Perona, "3D Photography on Your Desk", Proc. ICCV98, Bombay, India, pp. 43-50, 1998.
- [3] K. L. Boyer and A. C. Kak, "Color-Encoded Structured Light for Rapid Active Ranging", IEEE PAMI, Vol. 9, pp 14-28, 1987.
- [4] R. B. Fisher, A. Fitzgibbon, A. Gionis, M. Wright, D. Eggert, "A Hand-held Optical Surface Scanner for Environmental Modeling and Virtual Reality", Proc. Virtual Reality World '96, February 13-15, 1996, Stuttgart, Germany.
- [5] R. B. Fisher, A. P. Ashbrook, C. Robertson, N. Werghi, "A Low-Cost Range Finder using a Visually Located, Structured Light Source", patent pending (9822410.8).

- [6] D. Nitzén, A. E. Brain and R. O. Duda, "The Measurement and Use of Registered Reflectance and Range Data in Scene Analysis", Proc. IEEE, Vol. 65, pp 206-220, 1977.
- [7] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, "Numerical Recipes in C", Cambridge University Press, 1992.
- [8] B. Curless and M. Levoy "Better optical triangulation through spacetime analysis", Proc. Fifth Int. Conf. on Computer Vision, Cambridge, Massachusetts, pp 987-994, 1995.