

3D Slit Scanning with Planar Constraints[★]

Matthew J. Leotta^{1†}, Austin Vandergon^{2‡} and Gabriel Taubin¹

¹Brown University Division of Engineering, Providence, RI, USA

²Microsoft, USA

Abstract

We present a planarity constraint and a novel three-dimensional (3D) point reconstruction algorithm for a multiview laser range slit scanner. The constraint is based on the fact that all observed points on a projected laser line lie on the same plane of laser light in 3D. The parameters of the plane of laser light linearly parametrize a homography between a pair of images of the laser points. This homography can be recovered from point correspondences derived from epipolar geometry. The use of the planar constraint reduces outliers in the reconstruction and allows for the reconstruction of points seen in only one view. We derive an optimal reconstruction of points subject to the planar constraint and compare the accuracy to the suboptimal approach in prior work. We also construct a catadioptric stereo rig with high quality optical components to remove error due to camera synchronization and non-uniform laser projection. The reconstruction results are compared to prior work that uses inexpensive optics and two cameras.

Keywords: catadioptric stereo, laser, planar curves, 3D scanning, slit scanner

ACM CCS: Categories and subject descriptors: I.4.1 [IMAGE PROCESSING AND COMPUTER VISION]: Digitization and Image Capture – Scanning; I.4.8 [IMAGE PROCESSING AND COMPUTER VISION]: Scene Analysis – Shape, Stereo; I.3.5 [COMPUTER GRAPHICS]: Computational Geometry and Object Modeling

1. Introduction

Laser range scanners (such as those offered by Laser Design [LsD], NextEngine [NxE] and HandyScan [HnS]) provide an excellent way to recover shape data. Some sort of tracking, such as the ‘articulated arm’ in the Perceptron ScanWorks package [Per] or FASTRACK in the Polhemus FastScan [Pol] system, is also often necessary. Our system, by contrast, was conceived with simple scanning in mind. The setup includes a synchronized stereo pair of video streams and a hand-

held straight line projector, as well as a display where visual feedback in the form of incremental three-dimensional (3D) reconstruction is provided. Figure 1 illustrates the configuration of the system. This rig uses a catadioptric configuration, employing four mirrors and a lens, to create a stereoscopic view in the frame of a single video camera. The virtual cameras from the left and right halves of the image are fully calibrated with respect to the world coordinate system, and the fields of view overlap. The intersection of these two fields of view is the working volume where 3D data can be captured. The line projector generates an arbitrary and unknown plane of light, which intersects the working volume and generates curves which should be visible in both images.

The same physical setup is suggested by Davis and Chen [DC01] as a greatly improved laser range scanner. It maintains many of the desirable traits of other laser range scanners (accuracy, robustness) while eliminating actuated components, thereby reducing calibration complexity and concomitantly increasing maintainability and scan

[★]Based on ‘Interactive 3D Scanning Without Tracking’, by M. J. Leotta, A. Vandergon and G. Taubin which appeared in *Proceedings of the XX Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)*

[†]This research was conducted in part while the author was supported by a fellowship from the Rhode Island Space Grant Program, which is part of the National Space Grant College and Fellowship Program.

[‡]This research was conducted while the author was at Brown University.

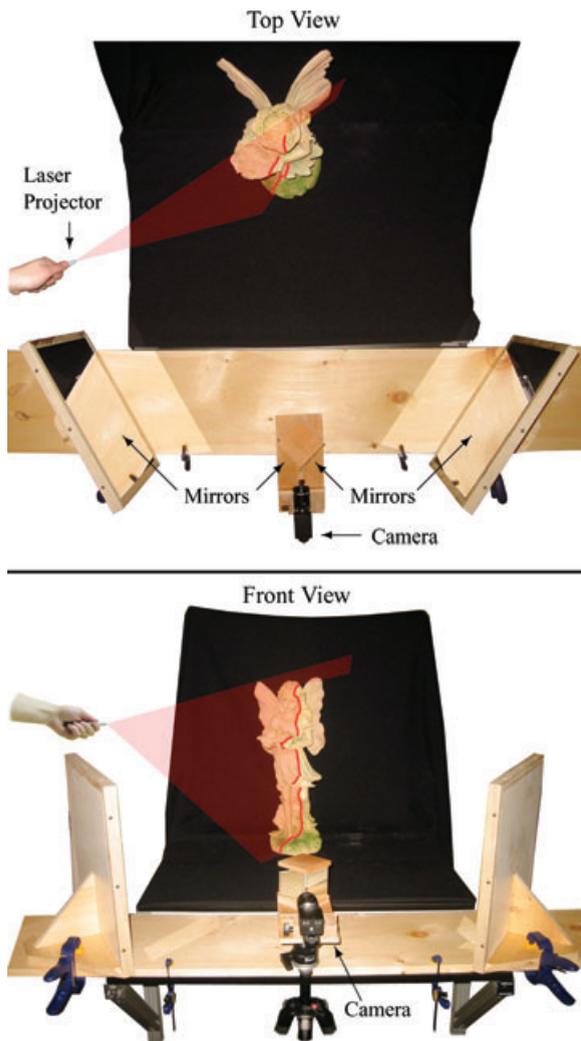


Figure 1: The catadioptric scanning rig.

repeatability. While standard triangulation used by Davis and Chen can only recover points in the working volume, our approach can recover additional points visible from only one view. Our extended working volume is the union of the fields of view as illustrated in Figure 2. This improvement is possible because a planar constraint is enforced.

In a recent paper [LVT07], we introduced a planarity constraint for the reconstruction of 3D points in a laser slit scanner with two cameras. This paper extends the previous work with a new reconstruction algorithm that computes the optimal reconstruction of points constrained to the plane. The scanning rig shown in Figure 1 also differs from the pair of stereo cameras used in the previous work. The rig is combined with a high quality laser and line generating lens. We show that this results in a better reconstruction, even for the old suboptimal methods.

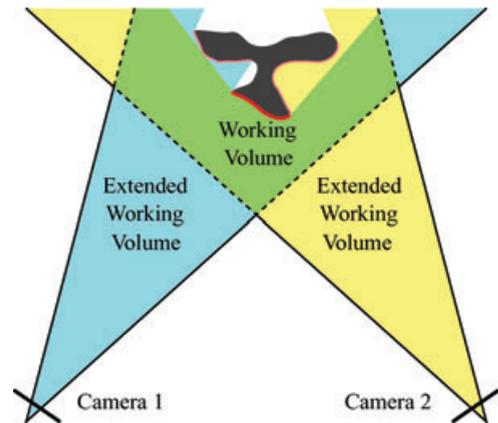


Figure 2: A top view of the working volume and scannable surfaces of a T-shaped object.

The novel reconstruction method presented in this paper has two parts: plane estimation and planar reconstruction. In the first part, the detected image point correspondences are used to estimate the parameters of the plane of projected laser light. Next, the corresponding points are triangulated with an additional constraint forcing them to lie on the plane. The remaining image points that have no correspondence are back projected onto the plane. The former reconstruction method results in improved accuracy while the latter results in improved coverage in an extended working volume.

The rest of this paper is organized as follows: Section 2 discusses related work. Section 3 describes catadioptric stereo. Section 4 derives the equations for estimation of the plane, and Section 5 derives the equations for the novel reconstruction method. In Section 6, this method is applied in the implementation of an interactive scanning system. Section 7 provides error analysis and scanning results, and Section 8 concludes with a summary and future work.

2. Related Work

Slit scanners have been continuously improving since their development early in the history of range scanners. Section 2.2 of F. Blais' survey on the subject [Bla04] provides a good overview of slit scanners' associated advantages and disadvantages. Work on this particular system started as an extension of a real-time implementation of the shadow scanner created by Bouget and Perona [BP98]. This shadow scanner works using two calibrated planes on which the shadow cast from a stick illuminated by a point light source can be observed. By estimating the location of the front edge of the shadow plane in space, a reconstruction of an object can be obtained using a single camera via ray-plane intersection. Notably, in Section 5, they suggest extending the shadow scanner to integrate multiple views. This suggestion is made, however, with the intention of integrating

multiple scans easily to obtain a full 3D model of an object. Bouget and Perona also expanded upon their previous work in [BWP99], using an arbitrary (rather than predictably oriented) set of shadow planes to eliminate the necessity of a background plane when doing reconstructions. Merging scans was addressed in [BP99], but the simultaneous use of multiple views and arbitrary shadows is not explored in their literature.

Subsequent research revealed that the same system setup we planned to test had been used by Davis and Chen [DC01]. In this paper, reconstructions were done by recovering point correspondences using epipolar lines, then doing triangulation. This paper provided both confirmation that a stereo pair with an uncalibrated light source was indeed capable of accurate reconstruction using triangulation alone, as well as a good reference concerning important implementation details such as laser line detection.

The specific setup used by Davis and Chen is constructed using a catadioptric stereo system, where mirrors are used to create two virtual cameras, and only one physical camera is needed for capture. With a global shutter, this means all captured image pairs are synchronized by design. The use of catadioptric systems has been explored in many contexts. In 1987, a system named ‘Sphereo’ [Nay88] constructed by S. K. Nayar was used to determine depth using two specular spheres and a single camera. Real-time systems were constructed in [GN98] and [FHM*93]. A system designed to rectify images before capture was constructed in [GN00].

Another system involving a plane of laser light was recently constructed by Zagorchev and Goshtasby [ZG06], in which depth calculation was done via ray-plane intersection. The reconstructions were done from single views, however, and integrated using a reference frame which was also used to recover the laser plane. A third system, in principle basically the same as Bouget and Perona’s shadow scanner with an added fast surface registration technique, was presented by Winkelbach *et al.* in [WMW06].

Also closely related is the work done by Trucco *et al.* [TFFN98], which discusses many constraints on reconstructed points in a system with two cameras and a plane of laser light. Like our work, their system reconstructs points from either one or two views depending upon visibility. Unlike our work, their system relies on actuated components – specifically, a sliding table – and uses a ‘direct calibration’ method to skirt around the use of camera models and plane parameterization. As their system is fully calibrated, the use of a second camera is not strictly required, but helps to constrain the reconstruction and reduce outliers. Direct calibration simply involves using an object of known geometry to calibrate the entire system, rather than calibrating the cameras and the plane individually. Therefore, it retains many of the disadvantages (such as constantly requiring recalibration) of complex laser range scanners.

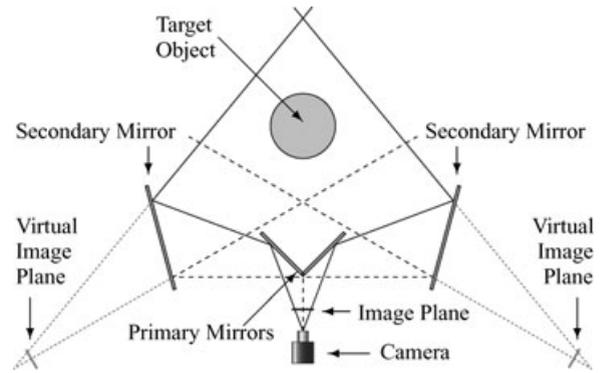


Figure 3: Catadioptric stereo diagram.

A final group of papers [BSA98; BZ99; IAW98; ST98] describe methods for 3D reconstruction which either assume or take advantage of planarity in the observed scene. These are typically planar objects, however, such as tabletops or faces of buildings; none are planar illumination systems such as a laser line projector. In a paper by Szeliski *et al.* [ST98], a planarity constraint is used for reconstructing points that are determined to lie on a detected plane in the scene. Notably, insignificant additions to accuracy are reported due to their planarity constraint.

3. Catadioptric Stereo

When constructing a stereoscopic vision system, one typically uses a pair of cameras to observe the scene from different viewpoints. An alternative approach is to use a single camera and mirrors. The mirrors divide the image into regions, each appearing to view the scene from a different viewpoint. Figure 3 is a diagram showing the path of the light rays in our catadioptric system from an overhead vantage point. Images of the actual rig are shown in Figure 1.

Tracing backward from the camera, the rays first encounter a pair of primary mirrors forming a ‘V’ shape. The rays from the left half of the image are reflected to the left, and those from the right half are reflected to the right. Next, the rays on each side encounter a secondary mirror that reflect them back towards the centre and forward. The viewing volumes of the left and right sides of the image intersect in a location where the target object to be scanned is placed. Each half of the resulting image may be treated as a separate camera for image processing. The standard camera calibration techniques for determining camera position and orientation still apply to each half. The location of the virtual cameras determined by calibration is also shown in Figure 3.

There are advantages to using catadioptric stereo as opposed to multiple camera stereo. One advantage is cost. Mirrors, even high quality first surface mirrors, tend to be less expensive than purchasing a second camera. Another



Figure 4: A sample image taken with the catadioptric rig under ambient illumination.

advantage is synchronization. Assuming the imaging sensor uses a global (not rolling) shutter, the stereo views of the catadioptric system will be in perfect synchronization automatically. When using multiple cameras, special circuitry is usually needed to trigger the cameras simultaneously. Lack of synchronization translates into reconstruction error in this laser scanning application. Synchronization is the principal reason for choosing catadioptric stereo in this work.

Catadioptric stereo is not without disadvantages. With respect to cost, for equivalent performance the single camera in the catadioptric system needs to have at least twice the resolution of the cameras in a traditional stereo pair. Using multiple cameras also provides much more freedom on the choice of each viewpoint. The use of a single lens induces the constraint that the optical paths to the object of the left and right views must be approximately equal to keep both views in focus. Furthermore, the finite size of the lens aperture results in some blending of the left and right views at their dividing line rather than a sharp boundary. As a result, there may be a portion of the image that needs to be ignored. Figure 4 shows a sample image taken with the catadioptric rig. The vertical grey band is caused by a small strip of non-mirrored surface where the primary mirrors meet. Even with this gap, there is still a small amount of blending between the left and right images.

The algorithm described in the next sections applies to either a catadioptric stereo rig or a pair of synchronized cameras as in [LVT07]. In the catadioptric case, the camera parameters describe the virtual camera positions and orientations. The details of the components used in our catadioptric stereo rig are withheld until Section 6.

4. Plane Estimation

Using the pinhole model for image formation, the equation of projection of a 3D point p onto image point u in homogeneous

coordinates is

$$\lambda u = K(Rp + T),$$

where λ is a non-zero scalar value, K is an upper triangular 3×3 matrix, R is a 3×3 rotation matrix and T is 3D translation vector. K , R and T are all parameters of the camera. For the remainder of the paper, it is assumed that all cameras are calibrated. That is, K , R and T are known for each camera.

As K is known, all points in pixel coordinates u can be converted to normalized image coordinates u' and

$$\lambda u' = \lambda K^{-1}u = Rp + T.$$

To simplify notation, all image measurements will refer to the normalized coordinates. Hence, for a pair of cameras and a common point, the image formation equations become:

$$\begin{cases} \lambda_1 u_1 = R_1 p + T_1 \\ \lambda_2 u_2 = R_2 p + T_2. \end{cases}$$

For more information on image formation and camera calibration the reader is referred to [HZ00].

The unknown plane of light is written as follows:

$$\Pi = \{p : n_1 p_1 + n_2 p_2 + n_3 p_3 + n_4 = 0\},$$

where the coefficient vector $[n_1 \ n_2 \ n_3 \ n_4]^t$ is non-zero. Because this coefficient vector is determined up to a multiplicative scale factor, the family of 3D planes has three degrees of freedom. An alternative vector notation is also used in this paper:

$$\Pi = \{p : n^t p - d = 0\},$$

where $n = [n_1 \ n_2 \ n_3]^t$ is a unit vector and $d = -n_4$ is the distance from the plane to the origin of the coordinate system.

4.1. Planar curves and homographies

If an object is placed inside the working volume, the set of points on the object illuminated by the line projector form a 3D curve, C (see Figure 5). As a result of depth discontinuities, the curve may be composed of various disconnected segments. However, the entire curve is planar and lies in the plane Π . As a result, the two image curves, c_1 and c_2 , captured by the pair of cameras are related by a homography, H . This homography is the composition of the two perspective homographies; one from the first image plane to the plane of light, H_1^{-1} , followed by a second one from the plane of light to the second image plane, H_2 . As this homography is parametrized by the plane of light, the family of homographies produced by this process has only 3 degrees of freedom instead of the 8 degrees of freedom of a general unconstrained homography. Further information on plane induced homography constraints is available in Chapter 12.1 of [HZ00].

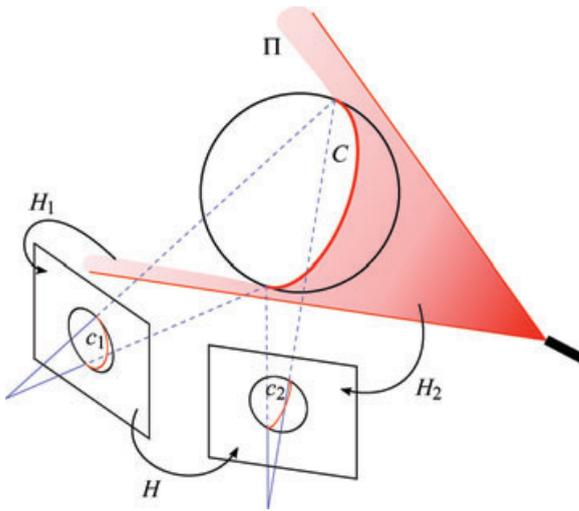


Figure 5: Homographies between the laser plane and image planes.

4.2. Planarity constraint

As the stereo pair is calibrated, corresponding points in the two image curves can be determined using epipolar geometry. By traditional triangulation, each pair of corresponding image curve points determines a point on the 3D curve. This is the point which minimizes the sum of the square distances to the two rays back projected through the image points. Due to measurement noise, these estimated points do not satisfy the co-planarity constraint. However, the constraint imposed by the reduced numbers of degrees of freedom in the parameterization of the homography allows for the estimation of the plane of light. Subsequently, the location of points on the 3D curve are estimated under the constraint of belonging to the estimated plane. This method produces more accurate results as demonstrated in Section 7.

4.3. Homography parameterization

The first step is to derive an expression for the homography $\xi u_2 = H u_1$ which transforms points on the first image plane onto points on the second image plane. Geometrically, this homography is defined by the following steps: (1) given a first image point u_1 , compute the intersection point p of the ray corresponding to the image point u_1 with the plane Π ; and (2) compute the image point u_2 as the projection of the point p onto the second image plane.

Algebraically, a point along the ray corresponding to u_1 can be written as

$$p = \lambda_1 v_1 + q_1, \quad (1)$$

in world coordinates, for a direction vector $v_1 = R_1^t u_1$, a centre of projection $q_1 = -R_1^t T_1$, and some value of $\lambda_1 \neq 0$.

For this point to be on the plane Π , the following condition must be satisfied as well:

$$0 = n^t p - d = \lambda_1 (n^t v_1) + (n^t q_1 - d), \quad (2)$$

or equivalently

$$\lambda_1 = \frac{d - n^t q_1}{n^t v_1}. \quad (3)$$

Note that the denominator is zero if and only if the ray defined by the first image point u_1 is parallel to the plane of light, which should never be the case here. Replacing the value of λ_1 just computed in the expression (1) for p we obtain

$$p = \frac{d - n^t q_1}{n^t v_1} v_1 + q_1. \quad (4)$$

The projection of this point onto the second image is an image point u_2 which satisfies the projection equation

$$\lambda_2 v_2 = p - q_2 = \frac{d - n^t q_1}{n^t v_1} v_1 + q_1 - q_2 \quad (5)$$

for some scalar value $\lambda_2 \neq 0$. Simple algebraic operations transform this expression into the following equivalent one:

$$\xi v_2 = [(d - n^t q_1) I + (q_1 - q_2) n^t] v_1 \quad (6)$$

for some scalar value $\xi \neq 0$ ($\xi = \lambda_2 (n^t v_1)$ in the previous expression), and where I is the 3×3 identity matrix. If A denotes the matrix inside the brackets, then the homography $\xi u_2 = H u_1$ which transforms points on the first image plane onto points on the second image plane is defined by the 3×3 matrix $H = R_2 A R_1^t$. Now, note that the matrix A can be written as a linear homogeneous combination of 3×3 matrices which only depends on the calibration parameters

$$A = n_1 A_1 + n_2 A_2 + n_3 A_3 + n_4 A_4 \quad (7)$$

with the coefficients of the plane of light as linear combination coefficients. As a result, so does H :

$$H = n_1 H_1 + n_2 H_2 + n_3 H_3 + n_4 H_4. \quad (8)$$

Explicitly,

$$\begin{cases} A_1 &= (r_{11}^t T_1) I + (R_2^t T_2 - R_1^t T_1) e_1^t \\ A_2 &= (r_{12}^t T_1) I + (R_2^t T_2 - R_1^t T_1) e_2^t \\ A_3 &= (r_{13}^t T_1) I + (R_2^t T_2 - R_1^t T_1) e_3^t \\ A_4 &= -I, \end{cases} \quad (9)$$

where r_{11}, r_{12}, r_{13} are the three columns of the rotation matrix $R_1 = [r_{11} \ r_{12} \ r_{13}]$ and e_1, e_2, e_3 are unit basis vectors (e.g. $e_1 = [1 \ 0 \ 0]^t$). Finally $H_j = R_2 A_j R_1^t$ for $j = 1, 2, 3, 4$.

4.4. Homography estimation

Pairs of image points (u_1^j, u_2^j) , $j = 1, \dots, N$, corresponding to the same point on the 3D curve are determined using epipolar geometry. For now, assume that each epipolar line

intersects the imaged curve at exactly one point. Thus, the corresponding image points are uniquely determined. The general case of epipolar matching is discussed in Section 6. Each of the image point pairs satisfy the homography equation $\xi_j u_2^j = H u_1^j$ for a different scale factor ξ_j . The scale factor is eliminated using a cross product, yielding two equations in H for each point pair:

$$\widehat{u}_2^j H u_1^j = 0 \tag{10}$$

where, if

$$u_2^j = \begin{bmatrix} u_{21}^j \\ u_{22}^j \\ 1 \end{bmatrix} \quad \text{then} \quad \widehat{u}_2^j = \begin{bmatrix} 1 & 0 & -u_{21}^j \\ 0 & 1 & -u_{22}^j \end{bmatrix}.$$

Equations (8) and (10) are combined to obtain the following matrix equation:

$$\left[\widehat{u}_2^j H_1 u_1^j \mid \widehat{u}_2^j H_2 u_1^j \mid \widehat{u}_2^j H_3 u_1^j \mid \widehat{u}_2^j H_4 u_1^j \right] n = 0. \tag{11}$$

Denote the 2×4 matrix within the brackets as L_j , and the $2N \times 4$ matrix resulting from vertically concatenating L_1, \dots, L_N as L . In the absence of measurement noise the linear equation $Ln = 0$ should be satisfied, which implies that the matrix L should be rank-deficient, i.e. $\text{rank}(L) < 4$. The solution is unique if $\text{rank}(L) = 3$, which is the typical case. In practice, there is measurement noise, and the solution is computed using the Singular Value Decomposition of the matrix L as the right singular vector corresponding to the minimum singular value. The second smallest singular value should be significantly larger than the minimum one.

If the points illuminated by the plane of light happen to be colinear, then a plane containing them is not uniquely determined. This singular case corresponds to $\text{rank}(L) = 2$. The handling of this degenerate case is discussed in Section 5.3. Note, however, that the location of the 3D points can still be estimated from triangulation.

5. 3D Point Reconstruction

Once, the plane of light has been estimated, there are several ways to reconstruct the 3D location of the points. First consider the non-singular case when a unique plane of light Π can be determined. If a point p is visible from only one camera (due to occlusion or indeterminate correspondence), it can still be reconstructed by ray-plane intersection. Equations (1) and (3) can be used to compute p as the intersection of the ray defined by the image point u_1 (or u_2) with the plane Π .

For points visible in both views, it is beneficial to use the data from both views. One approach is to triangulate the points. This is the approach taken by Davis and Chen [DC01]. While both views are used, the planarity constraint induced by the light plane is ignored.

Another approach is to reconstruct using the light plane and each view independently (as described above) and then average the resulting 3D points. This is the approach taken by Trucco *et al.* [TFFN98]. It works well when the angles between the plane and ray for each view are similar. With a hand-held light projector and fixed cameras this condition is not generally met. The magnitude of the error in the reconstruction from each view depends on the angle between the ray and plane. As a result, one reconstruction may have much larger error than the other. Averaging these points with equal weight does not account for this imbalance and results in larger error than triangulation.

A third approach is to triangulate the points and then project them to the closest point on the plane. This is the approach used in our prior work [LVT07]. It combines the stability of triangulation with the constraint of planarity. However, there is no guarantee that the projected point is the optimal point in the plane.

In this paper, we introduce a fourth approach that computes the optimal triangulation constrained to the plane. While our experiments use only two views, the derivations in the rest of this section are general and apply to any number of views greater than or equal to two. Below we review triangulation and then derive a planar constrained version of triangulation that computes the optimal point in the plane in the sense of least squares distance to the rays.

5.1. Triangulation

The goal of triangulation is to find a point in 3D space that is as close as possible to rays back projected from two or more images. A ray in 3D can be represented by the intersection of two planes. For reasons that will become apparent, each ray will be represented by the planes connecting the horizontal and vertical lines through a point in the image plane with the camera centre. Let π_x and π_y denote the homogeneous vectors defining these planes. Each plane has the form $\pi = [n_1 \ n_2 \ n_3 \ -d]^t$ with normal vector $n = [n_1 \ n_2 \ n_3]^t$ and distance d from the origin. If the planes are normalized such that $\|\pi\| = 1$, then the distance of point $p = [X \ Y \ Z]^t$ to the plane is simply $n^t p - d$ (or $\pi [p^t \ 1]^t$ in homogeneous coordinates). As the planes are chosen to be orthogonal, the square of the distance from p to the ray is $(n_x^t p - d_x)^2 + (n_y^t p - d_y)^2$. This is easy to express for k rays from multiple views using matrix notation. Define the $2k \times 3$ matrix B and the $2k$ vector c as

$$B = \begin{bmatrix} n_{x1}^t \\ n_{y1}^t \\ \vdots \\ n_{xk}^t \\ n_{yk}^t \end{bmatrix} \quad c = \begin{bmatrix} d_{x1} \\ d_{y1} \\ \vdots \\ d_{xk} \\ d_{yk} \end{bmatrix},$$

then $\|Bp - c\|^2$ is the sum of squared distances to the k rays.

The planes π_x and π_y are computed easily given the image coordinates of a point (x, y) and the 3×4 camera projection matrix $P = K[R | T]$. Express P in terms of its row vectors

$$P = \begin{bmatrix} P'_1 \\ P'_2 \\ P'_3 \end{bmatrix},$$

then any point p on the horizontal plane π_x projects to a point on the horizontal image line at the vertical coordinate y . That is

$$\lambda y = P'_2 \begin{bmatrix} p \\ 1 \end{bmatrix} \quad \text{and} \quad \lambda = P'_3 \begin{bmatrix} p \\ 1 \end{bmatrix}. \quad (12)$$

Equating terms gives $(yP'_3 - P'_2) [p^t \ 1]^t = 0$ which implies $\pi_x = yP'_3 - P'_2$. Similarly, $\pi_y = xP'_3 - P'_1$. These vectors should be normalized by dividing by the magnitude of the plane normals vectors.

Solving for the triangulated point p now requires constructing B and c and solving for the p that minimizes $\|Bp - c\|^2$. This equation expands to

$$p^t B^t B p - 2c^t B p + c^t c. \quad (13)$$

Differentiating with respect to p and setting the derivative equal to zero gives an equation for finding the minimum of (13). The solution is

$$p = (B^t B)^{-1} B^t c, \quad (14)$$

which is simply the Moore–Penrose pseudoinverse of B multiplying c . This result is the well known least squares solution to the minimization problem.

Alternatively, the triangulation problem can be posed in a homogeneous form. In the homogeneous case, the goal is to find a p as a four-dimensional (4D) vector with an additional constraint on its magnitude. The advantage of the homogeneous solution is the ability to accurately reconstruct p even when the solution is at or near infinity. However, points at infinity are not of interest in laser range finding, and the inhomogeneous form presented above is more suitable to adding in the additional planar constraint described in the next section.

5.2. Triangulation constrained to a plane

In this work, there is an additional constraint on the reconstructed points. All points must lie in the plane of laser light Π . In previous work [LVT07], this was accomplished by a two stage process. First a point was triangulated (as described above), then the point was projected orthogonally onto Π . A triangulated point is optimally reconstructed in the sense of L^2 distance to the rays. However, there is no guarantee that, after projecting onto Π , the projected point is the optimal point in the plane.

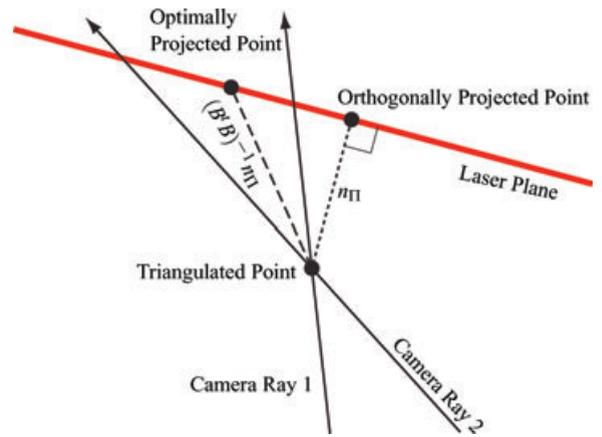


Figure 6: A 2D view of triangulation and both orthogonal and optimal projection onto the laser plane.

To achieve the optimal point in the plane, the planar constraint must be applied directly in the triangulation problem. Let n_Π be the normal vector of Π and let d_Π be the distance between the origin and the closest point on Π . That is, $\Pi = [n_\Pi^t \ -d_\Pi]^t$. Equation (13) can be modified with a Lagrange multiplier λ to include the constraint $n_\Pi^t p - d_\Pi = 0$:

$$p^t B^t B p - 2c^t B p + c^t c - 2\lambda(n_\Pi^t p - d_\Pi). \quad (15)$$

The -2 factor in front of λ is added to simplify the form of the solution. Setting the derivative of this function to zero and solving results in

$$p = (B^t B)^{-1} B^t c + \lambda(B^t B)^{-1} n_\Pi, \quad (16)$$

where λ is still an unknown multiplier which must be chosen to make p satisfy the constraint $n_\Pi^t p - d_\Pi = 0$. It is interesting to note that optimal solution for p consists of two terms. The first term is the original triangulated point, and the second term is a projection of that point onto the plane Π in the direction $(B^t B)^{-1} n_\Pi$. In fact, the only difference between the optimal reconstruction method and the orthogonal projection method of the prior work is the direction along which the projection is made. Figure 6 shows the difference between these projections in the analogous two-dimensional (2D) problem.

What remains is to determine the value of λ . Note that Equation (16) defines p as a point along a ray exactly as in Equation (1). Hence, using ray-plane intersection, the value of λ is given by Equation (3). By equating terms in Equations (1) and (16) and substituting into Equation (3), the value of λ is

$$\lambda = \frac{d_\Pi - n_\Pi(B^t B)^{-1} B^t c}{n_\Pi^t (B^t B)^{-1} n_\Pi}. \quad (17)$$

5.3. Singular reconstruction

Now consider the singular case when the numerical rank of the matrix L is 2. The two right singular vectors associated with the two smallest singular values of L define two planes whose intersection is the line supporting all the illuminated points. Hence, the points lie on a line in three dimensions. There is no longer a unique plane that fits the points; any plane in the null space of L is a good fit.

In practice, the measurement noise prevents the system from ever becoming completely singular. However, if the vast majority of illuminated points lie nearly on a line in 3D, the singular condition is approached and the plane estimate becomes unstable. To measure the degree of instability, we define the plane condition number κ_{Π} as the ratio of the second smallest to the largest singular value of L . When κ_{Π} is small, the rank approaches 2 and the plane estimate becomes more unstable. Interestingly, points visible in both views can still be reconstructed as before, even as the plane becomes degenerate. The estimated plane will still fit the data well – even if the solution is not unique. A problem arises only with reconstructions from one view using ray-plane intersection. These reconstructions rely on an accurate estimate of the true light plane. A small pivot in the plane orientation around a nearly linear set of 3D points has little effect on those linear points but generates large error in points away from this line. To avoid this problem, single view reconstructions are not computed when κ_{Π} exceeds a threshold. In the results shown in Section 7, only points visible in both views are reconstructed when $\kappa_{\Pi} < 0.01$.

When the singular case occurs, it may seem that a stronger constraint can be applied. It is possible to estimate the best 3D line that fits the data and then constrain the reconstruction to this line. While this might reduce the reconstruction error when the points are actually collinear, this is not the case in general when the rank of L is nearly 2. It is possible – and common – for the cross section of an object to have little variation in one direction within the light plane. Yet, this variation may contain important shape information. Without prior information about the scene, it is difficult to distinguish between this situation and truly linear data in the presence of measurement noise and outliers. For this reason, we do not apply a hard rank 2 constraint.

One strategy to reduce the occurrence of the singular condition is to place the object in front of a light coloured background. If the background reflects the laser light then it provides additional points to constrain the estimate of the laser plane. The separation between object and background ensures that the added points will not lie on the same line as points on the object. The reconstructed background can be cropped out of the resulting point cloud. This approach is quite similar to background planes used in [BWP99]. The key difference is that the background in this method need not have a simple or known shape and no additional calibration is

required. For example, a wrinkled white sheet of cloth could be draped behind the object. The reconstructions in this paper do not use a reflective background for support, but the results in the previous work [LVT07] did.

6. Scanner Implementation

Using the reconstruction technique in Sections 4 and 5, we implemented an interactive scanning system. The system uses the catadioptric configuration discussed in Section 3. The hardware components used in the system are described in the first subsection below. The image processing steps are detailed in the second subsection. Image processing is required to produce the point correspondences used in reconstruction. In the last subsection, a protocol is defined for how the light projector should be moved during scanning.

6.1. System hardware

The scanning system consists of three key components: the pair of cameras (or equivalently, the catadioptric rig), the laser plane projector and computer with a display. Each is discussed in detail.

The experiments in this paper use the catadioptric rig shown in Figure 1 and diagramed in Figure 3. In prior work [LVT07], we used a pair of 1394b cameras at a resolution of 1024×768 . The catadioptric rig uses a single 1394b camera with a resolution of 1600×1200 . The image is split horizontally allocating 800×1200 pixels for each view. This configuration provides 22% more pixels per view. Notably, the aspect ratio is also changed making the views taller than they are wide. Both the primary and secondary mirrors are first surface. That is, the reflective surface is on the front rather than the back of the glass. Ordinary back surface mirrors generate an unwanted secondary reflection. The lens on the camera has a 12.5 mm focal length with negligible radial distortion. Intrinsic calibration of the camera found camera parameters very close to the factory specified values, so the factory specified intrinsic parameters are used in the experiments.

The laser plane projector used in these experiments has been upgraded from the inexpensive laser pointer and glass rod configuration used in [DC01] and [LVT07]. The system uses an industrial quality 5 mW, 630 nm, focusable laser module. The cylindrical lens (glass rod) is now replaced with a 45° line generating Powell lens. A Powell lens is a lens designed to generate a laser line with approximately uniform intensity distribution. In comparison, the intensity of the laser line produced by a cylindrical lens is higher in the middle and falls off at the ends.

The computer used in the experiments is a standard workstation PC. The CPU operates at 3.2 GHz and has 1 MB of L3 cache. The system has 2GB of RAM. The display is an

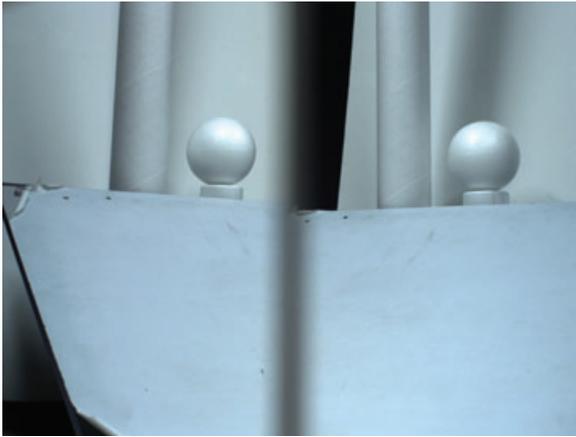


Figure 7: Stereo image of simple objects used for error analysis.

LCD monitor. This computer system is capable of processing the images and updating the reconstruction on the display in real time. Immediate, iteratively updated reconstruction results allow the user to direct the laser to regions of sparse or missing data. Currently the code is run serially in a single-threaded application with graphics hardware assisting only in the display of the results. We have been able to achieve frame rates as high as 15 fps by skipping the image smoothing step described below. However, smoothing is required for best results. Actual frame rates with smoothing are closer to 6 fps. The slowest steps are well suited to parallel computing. It should be easy to achieve full processing at 15 fps or more using a multithreaded or graphics hardware assisted implementation.

6.2. Line detection and matching

The laser line is detected independently in each view. A mean image is computed and subtracted from the current image. This removes the appearance from ambient lighting and leaves only the laser line. Gaussian smoothing is applied to the image to reduce image noise and laser speckle. Intensity peaks are located in horizontal scan lines, and parabolic interpolation is used to compute the sub-pixel peak locations. The peaks are linked by scanning vertically and connecting neighbouring peaks into piecewise linear curves. An isotropic ridge detection algorithm would probably generate better curves. However, the anisotropic method can be implemented very efficiently and works well on lines that are close to vertical. This agrees with the horizontally spaced camera configuration, which is better suited to reconstruct points on a vertical plane.

Correspondences are computed by rectifying the detected curves using homographies that map the epipoles to infinity, align the epipolar lines and roughly preserve scale (refer

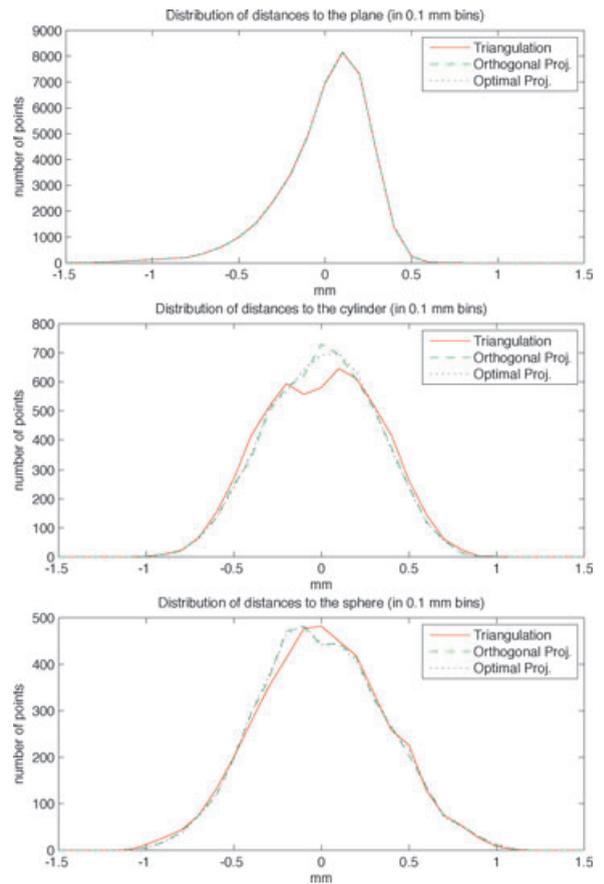


Figure 8: Histograms showing the distribution of distances from the fitted shape for each reconstruction method.

Table 1: Standard deviation of errors measured in mm.

Shape	Reconstruction method		
	Triangulation	Orthogonal proj.	Optimal proj.
Plane	0.2589	0.2587	0.2583
Cylinder	0.3267	0.3106	0.3097
Sphere	0.3610	0.3578	0.3586

to Chapter 10.12 of [HZ00]). The curves are resampled at equal intervals in the rectified space resulting in a set of correspondences along epipolar lines. In practice, some points may have multiple correspondences. Davis and Chen [DC01] address this problem by discarding all ambiguous correspondences. The approach in this paper allows these ambiguities to be resolved by using the estimated homography, H , resulting in a more detailed reconstruction. However, the homography parameters must first be estimated as in Section 5 using unique correspondences.

Table 2: Rated accuracies at specified ranges of other devices.

Product	Accuracy (mm)	Range (mm)
Laser Design [LsD]	0.1016	Unspecified
NextEngine [NxE]	0.1270	127–254
HandyScan [HnS]	0.0400	300
ScanWorks [Per]	0.0340	144
Polhemus FastScan [Pol]	1.0000	200

Note that the listed accuracies are not necessarily directly comparable, as different methods are often used to determine their values. Please see the referenced products' documentation for more details.

Estimating the homography from unique correspondences is still problematic. An ambiguous correspondence combined with an occlusion can produce a unique, but incorrect, correspondence. These incorrect correspondences are outliers and make up a small portion of the data in practice. Yet a single outlier can introduce a significant amount of error in a linear estimate. To compensate for this, RANSAC [FB81] is used. RANSAC is applied to the set of all correspondences to get a robust estimate of the homography and discard outliers. RANSAC uses subsets of three correspondences to find many non-degenerate solutions to the estimation problem resulting from Equation (11). Given a randomly selected subset, the parameters of plane Π are computed and H is recovered as in Equation (8). For each correspondence (u_1^j, u_2^j) , the symmetric transfer error is computed:

$$E^j = \sqrt{\left(H u_1^j - u_2^j\right)^2 + \left(H^{-1} u_2^j - u_1^j\right)^2}, \quad (18)$$

where distances in Equation (18) are computed using inhomogeneous coordinates. E^j is small for correspondences agreeing with H (inliers) and large for others (outliers). Applying a threshold to E^j classifies correspondence j as an inlier or outlier. In practice, RANSAC is not very sensitive to the value of this threshold. The experiments in Section 7 use a threshold value of 2 pixels. After RANSAC finds the solution with the most inliers, a final estimate of Π is computed using all the inliers.

6.3. Scanning protocol

When operating the scanner, the user has the freedom to move the light projector freely in space. We have not observed any configuration of the light plane that causes the presented algorithm to fail. Nevertheless, there are some configurations that are clearly superior and produce a greater volume of data points or more accurate reconstructions. Below we discuss configurations to favour and those to avoid. Then, we outline the scanning protocol used in our experiments.

The laser plane should be kept more or less vertical. The reason for this is twofold. First, the cameras are distributed

horizontally, which in turn means the epipolar lines run more horizontal than vertical in the images. The correspondences along the detected laser lines are most accurate when the laser lines are perpendicular to the epipolar lines. Second, the fast anisotropic ridge detector described above favours vertical lines. In practice, we tend to keep the light plane within about 45 degrees from vertical. With larger rotations the number of detected points drops off quickly.

The light plane should not be positioned such that it lies near a virtual camera centre. When this happens, the homography between images becomes singular. The estimation of the plane parameters [Equation (11)] is still valid, but the outlier test [Equation (18)] is less reliable. In effect, the in-plane camera is used only to determine the parameters of the plane, and the system is reduced to ray-plane intersection with the other camera. When this occurs the accuracy might be reduced, but the results are not disastrous.

The light plane projector should be both translated and rotated in space. The surface can be swept out by rotation alone, but with a fixed position there are points that will be self-occluded and not receive light. Moving the centre of projection allows more of these points to be illuminated.

In the experiments in Section 7, we apply the above guidelines using the following scanning protocol.

1. Position the laser projector to the left of both camera centres, between both centres, and to the right of both centres.
2. At each position, sweep the plane horizontally across the object back and forth.
3. While sweeping, vary the angle of the plane and move the projector up and down. Adjust the sweeping action according to the real-time feedback to fill in areas with sparse coverage.

7. Evaluation and Results

To evaluate the accuracy of the algorithm, three simple objects of known dimension – a plane, a cylinder, and a sphere – were scanned. The diameter of the cylinder is 3.125 inches (79.375 mm) and the diameter of the sphere is 4 inches (101.6 mm). For comparison, points visible in both views were reconstructed by triangulation, triangulation with orthogonal planar projection, and triangulation with optimal planar projection. The same points were considered for all three methods and only points deemed to be inliers (as described in Section 6.2) were used. This restriction reduces the bias from outliers in this experiment. Figure 7 shows an image of the scene containing the simple objects for error analysis.

Plane, cylinder, and sphere models were fitted to each of the reconstructions using Gauss-Newton iteration to find a

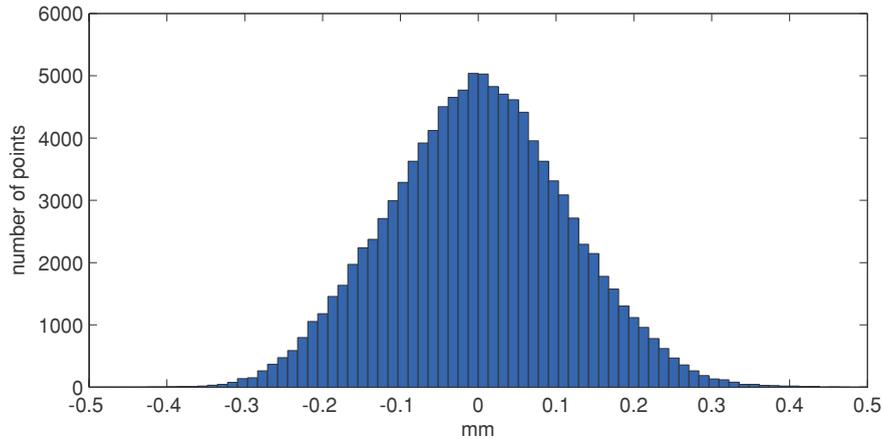


Figure 9: Histogram of distances from triangulated points to the estimated laser planes.

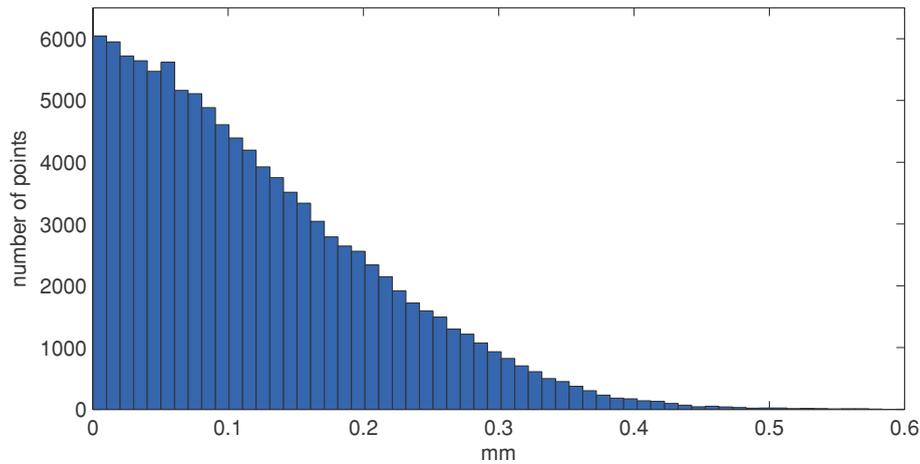


Figure 10: Histogram of distances from the planar optimal points to the camera rays.

least-squares solution. Points were manually cropped from the full reconstructions using the neighbourhood of each shape to select the set of points belonging to it. Figure 8 shows histograms of the distances of the points to their respective estimated surfaces. By construction, these distributions all have zero mean. The standard deviations of the errors are given in Table 1.

There are some observations to be made about these results. First, the error distribution for plane is nearly identical across reconstruction methods. This is a result of the singularity condition. In many of the frames, most or all of the laser light was visible only on the plane. In these cases, the light plane was poorly estimated and did not provide an accurate constraint. It is important to note that even when the singular case occurs frequently, the error is not worse than unconstrained triangulation on average.

A second observation is that the optimal method has slightly higher standard deviation of error in the case of the sphere. These experiments measure error in the direction normal to the object surface. The planar projection methods reduce error only in the direction normal to (or nearly normal to) the light plane. During much of the scanning these two directions are almost orthogonal. As a result only a fraction of the observable error reduction is actually measured in these experiments. Furthermore, there is no guarantee that the surface normal error is decreased by using the optimal method. As demonstrated below, however, the distance to the rays is always decreased.

Overall – when considering only inliers – there is only a slight improvement gained by enforcing the planarity constraint. The difference between the orthogonal and optimal projection methods is insignificant. Compared to our earlier

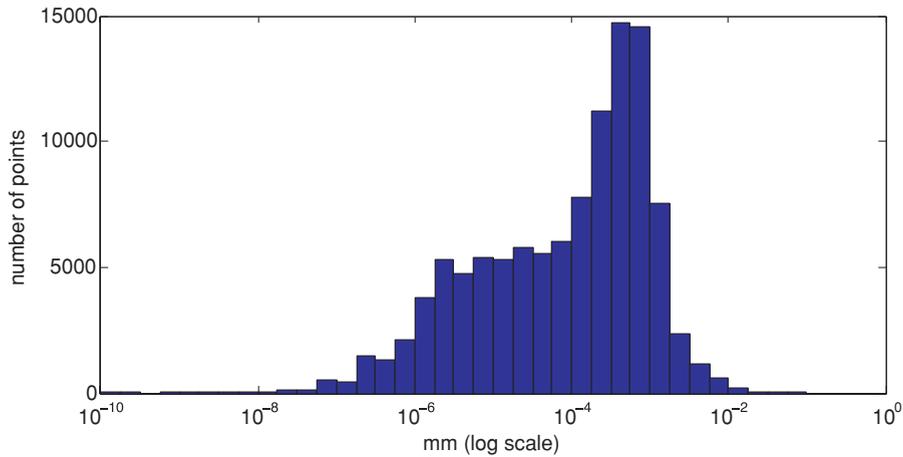


Figure 11: Histogram (in logarithmic scale) of differences in distances to the camera rays between optimal and orthogonal projections



Figure 12: Catadioptric camera view and reconstruction results using triangulation and optimal planar projection. Points in (c) are reconstructed from both views and (d) shows the addition of points seen only from one view.

work [LVT07], the accuracy of these reconstructions have improved by a factor of 2 for the cylinder and a factor of 5 for the sphere. This improvement is the result of improved synchronization and improved laser optics. The accuracies of some of the previously mentioned commercial scanners are listed in Table 2. The range of the objects for our system is 1200–1600 mm. Given the order of magnitude increase in range over the commercial systems, our system is very competitive on accuracy.

The orthogonal signed distances of the points to the estimated laser plane were computed for all points in the simple objects scene. A histogram of these distances is shown in Figure 9. The distribution of distances from the plane looks Gaussian and has a mean of -6.543×10^{-5} mm (nearly zero) and a standard deviation of 0.1152 mm. This represents the error that the plane projection methods are designed to eliminate. Unfortunately, the previous experiments do not

fully demonstrate the error reduction ability of these methods for reasons discussed above.

After projection onto the laser plane using the optimal method, the distances to the camera rays are measured. Figure 10 shows the distribution of these distances. The values shown here are the magnitude of the vector of distances to each of the rays. This is the square root of the value minimized in Equation (15) and also equal to $\sqrt{2}$ times the RMS distance to the rays. Because correspondences are chosen along epipolar lines, this error measure is zero for triangulated points.

The distances to the rays for the orthogonal projection are nearly the same as those in the optimal method. Figure 11 shows the distributions of differences between the orthogonal projection distance and the optimal projection distance. The majority of differences are small, but all are greater than or

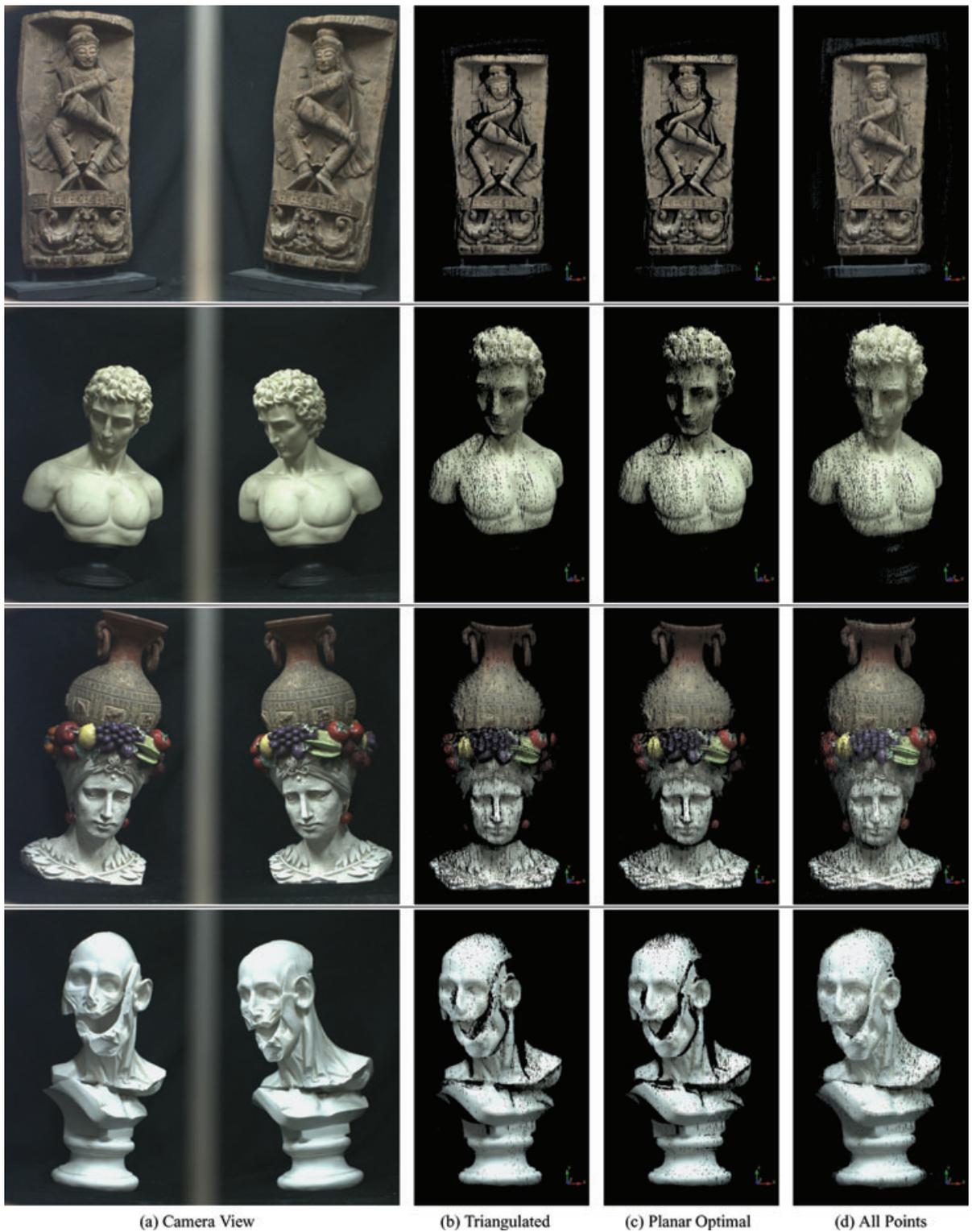


Figure 13: Catadioptric camera view and reconstruction results using triangulation and optimal planar projection. Points in (c) are reconstructed from both views and (d) shows the addition of points seen only from one view.

equal to zero. This result helps to confirm the optimality of the new method. The distance to the rays using the optimal method is always less than or equal to the distance using the orthogonal method, but in this case, the improvement is small.

Figure 12 shows the reconstructions of the fairy statue that was used as an example above. Compared are the results of triangulation, optimal planar projection and optimal planar projection plus ray-plane intersection from single views. The 3D points have been coloured by projecting back into the mean ambient light image. The black speckled noise arises from aliasing in the rendering of the points clouds. As the light plane is kept roughly vertical during scanning, the points tend to have closer neighbours in the vertical direction (in plane) than in the horizontal direction (between planes). As a result, the noise appears vertically biased. Between plane distances can be reduced by increasing frame rate, slowing the motion of the light plane or making multiple passes.

In Figure 12, note the reduction of outliers from (b) to (c) and the increased number of points from (b) to (d). The fairy model results in 360 102 points from triangulation. The optimal planar projection method reconstructs 333 091 points from both views because many triangulated points are deemed to be outliers. These outliers are especially apparent near the wing on the left side of the image. An additional 130 242 and 143 192 points were reconstructed from only the right and left views, respectively. In total, our method results in 606 525 points, almost twice that of standard triangulation. Reconstructions of various other objects are shown in Figure 13.

8. Conclusions and Future Work

This paper has presented a planar constraint for the 3D reconstruction of points in a stereo laser scanning system. The plane is estimated directly from the image data with no need for tracking the laser projector. The estimated plane is used to detect and remove outlier points, constrain the points to lie within the plane, and reconstruct additional points observed only by one camera. A new reconstruction method was derived which produces the optimal reconstruction constrained to the plane. A new catadioptric scanning rig with high quality optical components resulted in reduction of error over the previous work [LVT07] when applying the same methods. The optimal reconstruction method also provides an improvement in accuracy, but the improvement is small.

There are several areas that could be improved in future work. One is the image processing responsible for detecting the laser lines in the images. The current approach should be compared to an isotropic ridge detection method. This step is probably the largest source of error in the current system configuration. For improved usability, it is also desirable to improve the processing speed. Using graphics hardware or multiple CPU cores, it should be possible to achieve frame

rates of 15 or even 30 fps while maintaining the accuracy of the results.

Acknowledgments

The authors would like to thank the reviewers of this paper for pointing out additional related references and guiding us to clarify several points. The authors would also like to thank Jennifer Leotta for proofreading and offering constructive criticism.

References

- [Bla04] BLAIS F.: Review of 20 years of range sensor development. *Journal of Electronic Imaging* 13, 1 (2004), 231–243.
- [BP98] BOUGUET J.-Y., PERONA P.: 3D photography on your desk. In *Proceedings of the Sixth International Conference on Computer Vision (ICCV)* (1998), p. 43.
- [BP99] BOUGUET J.-Y., PERONA P.: 3D photography using shadows in dual-space geometry. *International Journal of Computer Vision* 35, 2 (1999), 129–149.
- [BSA98] BAKER S., SZELISKI R., ANANDAN P.: A layered approach to stereo reconstruction. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)* (1998), pp. 434–441.
- [BWP99] BOUGUET J.-Y., WEBER M., PERONA P.: What do planar shadows tell about scene geometry? In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)* (Fort Collins, Colorado, USA, 1999), Vol. 1, p. 1514.
- [BZ99] BAILLARD C., ZISSERMAN A.: Automatic reconstruction of piecewise planar models from multiple views. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)* (Fort Collins, Colorado, USA, 1999), Vol. 2, pp. 559–565.
- [DC01] DAVIS J., CHEN X.: A laser range scanner designed for minimum calibration complexity. In *Proceedings of the International Conference on 3-D Digital Imaging and Modeling (3DIM)* (2001), Vol. 00, p. 91.
- [FB81] FISCHLER M. A., BOLLES R. C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 24, 6 (1981), 381–395.
- [FHM*93] FAUGERAS O., HOTZ B., MATHIEU H., VIÉVILLE T., ZHANG Z., FUA P., THÉRON E., MOLL L., BERRY G., VUILLEMIN J., BERTIN P., PROY C.: *Real Time Correlation Based Stereo: Algorithm Implementations and Applications*. Tech. Rep. RR-2013, INRIA, 1993.

- [GN98] GLUCKMAN J., NAYAR S.: A real-time catadioptric stereo system using planar mirrors. In *Proceedings of the DARPA Image Understanding Workshop (IUW)* (November 1998), pp. 309–313.
- [GN00] GLUCKMAN J., NAYAR S.: Rectified catadioptric stereo sensors. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2000), Vol. 2, pp. 380–387.
- [HnS] <http://www.creaform3d.com/en/handyscan3d/default.aspx>, 12 August 2008.
- [HZ00] HARTLEY R., ZISSERMAN A.: *Multiple View Geometry in Computer Vision*, First ed. Cambridge University Press, 2000.
- [IAW98] IRANI M., ANANDAN P., WEINSHALL D.: From reference frames to reference planes: Multi-view parallax geometry and applications. *Lecture Notes in Computer Science 1407* (1998), 829–845.
- [LsD] <http://www.laserdesign.com/quick-attachments/hardware/low-res/dt-series.pdf>, 12 August 2008.
- [LVT07] LEOTTA M. J., VANDERGERON A., TAUBIN G.: Interactive 3D scanning without tracking. In *Proceedings of the XX Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI)* (Belo Horizonte, MG, Brazil, October 2007), pp. 205–212.
- [Nay88] NAYAR S.: Sphero: Determining depth using two specular spheres and a single camera. In *Proceedings of the SPIE Conference on Optics, Illumination, and Image Sensing for Machine Vision III* (November 1988), pp. 245–254.
- [NxE] <https://www.nextengine.com/indexSecure.htm>, 12 August 2008.
- [Per] <http://www.perceptron.com/ports.html>, 12 August 2008.
- [Pol] http://www.polhemus.com/?page=Scanning_Fastscan, 12 August 2008.
- [ST98] SZELISKI R., TORR P. H. S.: Geometrically constrained structure from motion: Points on planes. In *Proceedings of the European Workshop on 3D Structure from Multiple Images of Large-Scale Environments (SMILE)* (1998), pp. 171–186.
- [TFFN98] TRUCCO E., FISHER R., FITZGIBBON A., NAIDU D.: Calibration, data consistency and model acquisition with a 3-D laser striper. *International Journal of Computer Integrated Manufacturing 11*, 4 (1998), 292–310.
- [WMW06] WINKELBACH S., MOLKENSTRUCK S., WAHL F.: Low-cost laser range scanner and fast surface registration approach. In *Proceedings of the 28th Annual Symposium of the German Association for Pattern Recognition (DAGM)* (2006), pp. 718–728.
- [ZG06] ZAGORCHEV L., GOSHTASBY A.: A paintbrush laser range scanner. *Computer Vision and Image Understanding 101*, 2 (2006), 65–86.