Efficiently Combining Positions and Normals for Precise 3D Geometry

Diego Nehab[1] Szymon Rusinkiewicz[1] James Davis[2] Ravi Ramamoorthi[3]
[1]Princeton University [2]University of California at Santa Cruz [3]Columbia University

(a) *(b)* *(c)* *(d)*

Figure 1: *Rendering comparisons. (a) Rendering of 3D scanned range image, (b) same scanned geometry, augmented with a measured normal-map (from photometric stereo), (c) our hybrid surface reconstruction, which combines both position and normal constraints, (d) photograph. Renderings in this paper do not use color information in order to focus on geometric aspects. Note how our method eliminates noise from the range image while introducing real detail. The surface normals are of the same quality or better than those from photometric stereo, while most of the low-frequency bias has been eliminated.*

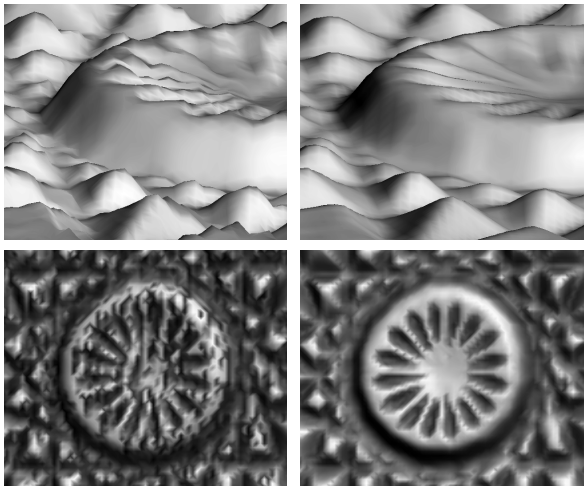
Abstract

Range scanning, manual 3D editing, and other modeling approaches can provide information about the geometry of surfaces in the form of either 3D positions (e.g., triangle meshes or range images) or orientations (normal maps or bump maps). We present an algorithm that combines these two kinds of estimates to produce a new surface that approximates both. Our formulation is linear, allowing it to operate efficiently on complex meshes commonly used in graphics. It also treats high- and low-frequency components separately, allowing it to optimally combine outputs from data sources such as stereo triangulation and photometric stereo, which have different error-vs.-frequency characteristics. We demonstrate the ability of our technique to both recover high-frequency details and avoid low-frequency bias, producing surfaces that are more widely applicable than position or orientation data alone.

1 Introduction

Increasingly, scanned 3D models of real-world objects are being used in rendering, visualization, and analysis applications. Although the absolute accuracy of such scanned data can be high, even a small amount of noise in measured positions can cause large errors when surface normals are computed. Lit renderings of such scanned models may produce low quality images (Figure 1a).

An alternative is to measure a normal field independently of the depth estimates, instead of computing it from measured positions. For example, there exist technologies based on *shape from shading* or *photometric stereo* [Woodham 1980] that directly measure surface orientations. Using these independently-measured fields as normal maps allows for high-quality renderings from certain viewpoints (Figure 1b), even when the actual mesh has low quality or low resolution. One drawback to simply pasting a "good" normal field onto "bad" geometry, however, is incorrect parallax and occlusion at grazing views (Figure 2, top left). In addition, some rendering and mesh processing effects such as shadowing or accessibility

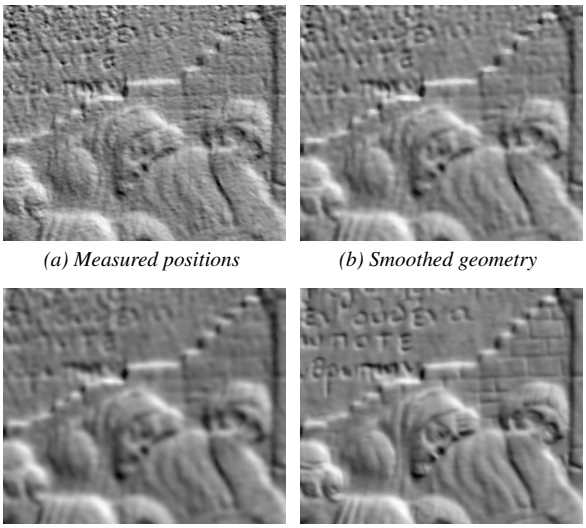


3D scanned range image with measured normal map *Hybrid reconstruction using our algorithm*

Figure 2: *Rendering at grazing angles (top) and accessibility shading (bottom). High-quality normal maps are not appropriate in these cases. The examples require the precise geometry our method can deliver.*

shading [Miller 1994] inherently operate on only the surface positions, not the normals. Thus, the poor performance of such techniques on noisy geometry (e.g., as shown in Figure 2, bottom left) cannot be directly ameliorated by the availability of high-quality normals. Although it is sometimes possible to directly integrate the high-resolution normal field to produce a surface without using additional geometric information, the lack of constraints between multiple disconnected patches, as well as the frequent presence of low-frequency distortion (as shown later in the paper), can lead to bias in the reconstruction.

We present a hybrid algorithm that produces a surface that optimally conforms to given surface positions and normals, taking advantage of the information available from both sources (Figures 1c and 2, right). Although we focus in this paper on combining the depth information from a triangulation scanner with the normal information from photometric stereo, the technique is applicable to positions and normals (or bump maps) acquired through scanning, manual editing, or signal processing algorithms. Our method is efficient on dense, real-world datasets, since we formulate our optimization in a way that requires solving only a sparse linear system.



(a) Measured positions *(b) Smoothed geometry*

(c) Further smoothing *(d) Our optimization*

Figure 3: *Range scanners frequently have noise in position measurements that is on the order of sample spacing, leading to noise in estimated normals (a). To use this scanned geometry for graphics, a popular approach is to downsample or smooth the raw measurements, which either leaves remaining noise (b) or blurs out relevant surface detail (c). In contrast, our method combines measured positions with measured normals, leading to noise elimination while adding real detail (d).*

In addition, only the most reliable frequency components of each source are considered, resulting in a reconstruction that both preserves high-frequency details and avoids low-frequency bias.

Our method is motivated by an analysis of the common error characteristics of measured normals and positions (section 3). It proceeds in two stages: first, we correct for low-frequency bias in the measured normal field with the help of measured surface positions (section 4.1). Then, we optimize for the final surface positions (sections 4.2 and 4.3) using linear constraints and an efficient sparse solver. We analyze the performance and quality of the algorithm on a variety of datasets (section 5), demonstrating its ability to reconstruct accurate and precise geometry suitable for rendering and mesh processing.

2 Relation to previous work

Positional measurement: Of the many range scanning technologies available (see [Besl 1989; Poussart and Laurendeau 1989; Curless 1997] for surveys), methods based on triangulation have become popular, since they can be flexible, inexpensive, and accurate. However, a fundamental limitation of triangulation-based methods is that their depth accuracy is some fixed multiple of their (horizontal) sample spacing, which typically results in noisy estimated normals. Although downsampling or blurring the raw measurements is sometimes acceptable, it frequently leads to oversmoothing of detail (Figure 3). The only effective means of reducing noise is to combine multiple measurements. Curless and Levoy [1996] have investigated a method for combining multiple range scans, showing that their VRIP algorithm is essentially a least-squares estimator that can average away noise while keeping detail. In contrast, we investigate combining the fundamentally different data types of measured positions and measured normals.

Orientation measurement: The photometric stereo method [Woodham 1980], which obtains surface orientation from shading information, is part of a larger set of methods known as *shape-from-shading* [Horn 1970]. These include methods whose outputs are surface normals, or sometimes simply single components of surface normals. Although it is possible to integrate the normals to find the shape of the surface, this approach is fragile when accurate surface reconstruction is desired, largely because integrating normals is prone to introducing low-frequency biases (recall that the action of the integration operator in frequency space is to scale energy at some frequency ω by an amount proportional to $^1/_\omega$, hence exaggerating any low-frequency noise). Furthermore, when the surface consists of multiple disconnected components, integration has no way of determining the relative position of those components. Therefore, most of the use of measured surface orientations in graphics has been in rendering, by directly using the measured normals as normal maps [Bernardini et al. 2002], which produces effective results in some but not all applications (Figure 2).

Combining positions and orientations: Much of the existing work on combining measured positions and orientations is intended for rough surface reconstruction in computer vision [Terzopoulos 1988; Banerjee et al. 1992; Fua and Leclerc 1994; Lange 1999], rather than for accurate reconstruction of surfaces for rendering or mesh manipulation. Accordingly, orientation measurements are taken into account mostly because dense estimates are available, and can help fill holes left by stereo correspondences based on sparse features.

Of methods that are general, one class integrates normals to yield a surface, then merges the resulting mesh with the measured positions as a final step. As mentioned above, this typically has the side effect of introducing bias and robustness problems. Some of these methods [Cryer et al. 1995; Mostafa et al. 1999] address the problem of low-frequency deformation by performing frequency-dependent processing. Although our method has a similar effect, it is inherently more stable in the presence of bias and disconnected components, as are other techniques that *avoid* explicit integration of normals.

A final class of methods combines positions and normals by formulating the reconstruction as a nonlinear optimization of constraints provided by the different measurements [Ikeuchi 1983, 1987; Chen et al. 2003]. However, for the typical scanned meshes used in computer graphics, which contain 10^5 polygons and above, nonlinear optimization methods can be prohibitively expensive. Our formulation of the optimization using a sparse, typically diagonally-dominant, linear system results in efficient optimization for large meshes, such as the ones shown throughout this paper and accompanying demo.

3 Motivation and Quality Assessment

In this section we motivate our method by showing some of the typical error characteristics of acquired positions from depth scanners and acquired normals from photometric stereo techniques. It should be noted that, although we focus on this application, our method can be applied to a wider range of scanning, modeling or mesh processing techniques, being independent of the specific methodology or experimental setup. In the remainder of this section, we first describe the specific scanner we used in this work, followed by an assessment of the quality of recovered positions and normals.

3.1 Experimental Setup and Hybrid Scanner Design

Consider the setup shown in Figure 4. To acquire positions, we use a temporal stereo triangulation scanner [Zhang et al. 2003; Davis et al. 2005], which combines the accuracy of active triangulation with the ease of calibration of stereo. It is composed of two Sony DFW-X700 firewire cameras and a Toshiba TLP511 projector. The cameras are calibrated using the toolbox by Bouguet [2004] and synchronized by an external trigger. The projector flashes a series of stripe patterns onto the object while the cameras simultaneously capture images from two different viewpoints.

Because each surface point on a given epipolar plane receives a unique lighting profile through time, it is possible to establish

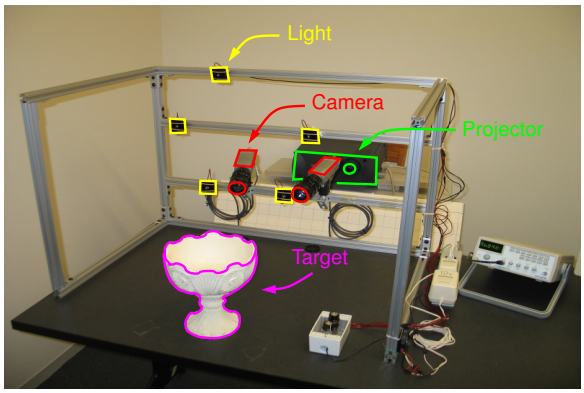


Figure 4: *Scanner setup. For position measurement, two cameras capture synchronized images while a projector flashes random stripes at the object. For normal acquisition, one of the cameras takes pictures under illumination by 5 different, calibrated light sources. Part of this object is pictured in Figures 2 and 5.*

unique correspondences by simply correlating the intensity variations in both images. There is no need to calibrate or even synchronize the projector with the cameras. Given the correspondences, the 3D positions of each point as seen by either camera can be determined by triangulation.

To acquire normals, we augment our scanner with a number of fixed light sources and use the photometric stereo technique [Woodham 1980]. Although there are more sophisticated formulations for the method [Tagare and deFigueiredo 1991; Georghiades 2003], we rely, as others [Epstein et al. 1996], on the redundancy provided by extra light sources in order to avoid regions that deviate from the Lambertian assumption. We use 5 white Luxeon® LED emitters arranged in a pentagon around the reference camera.

In our setup, the triangulation scanner and the photometric stereo scanner share the same reference camera. Therefore, normals and positions are automatically registered. Other hybrid scanner designs use independent cameras for depth and normal capture, usually acquiring normals at a higher resolution [Bernardini et al. 2002]. In these cases, after registration, depth data can be upsampled to match the measured normal field, resulting in an output similar to that of our scanner.

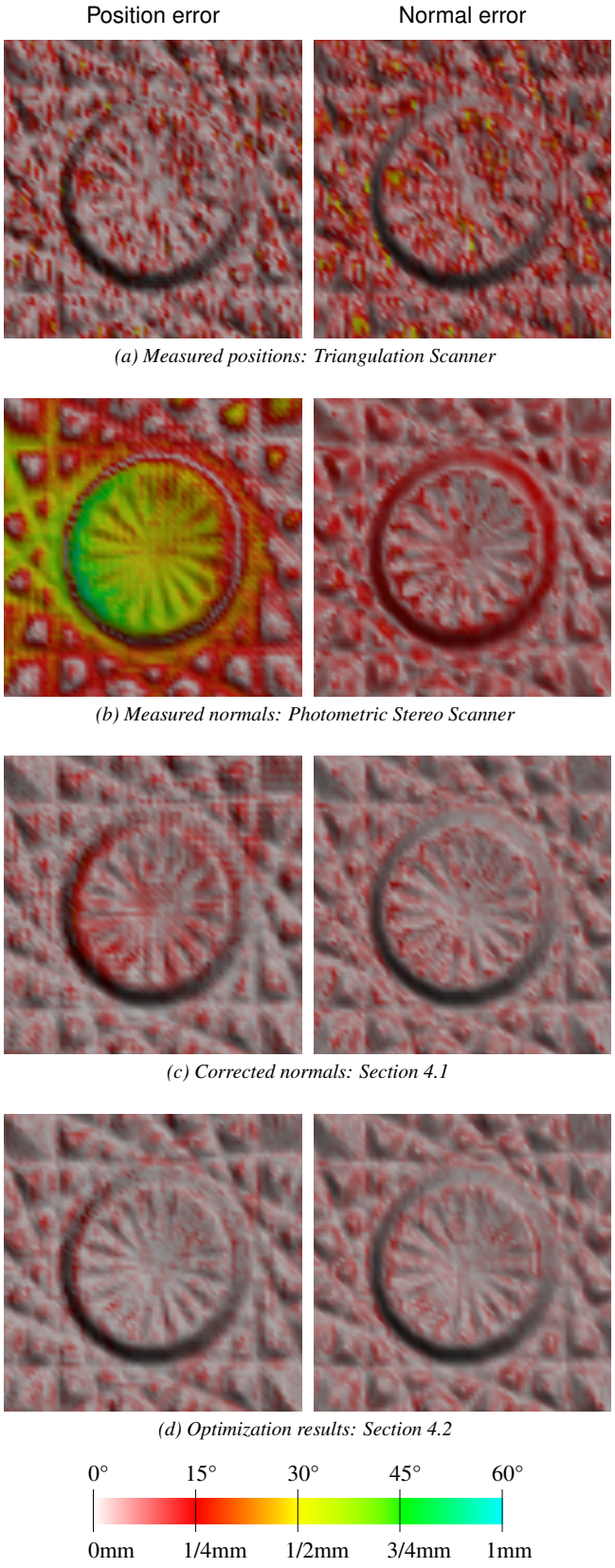
3.2 Quality assessment

We need a way to assess the quality of measured normals and positions, and the accuracy of our results. Our triangulation scanner has been tested by scanning simple objects with known geometry, by comparing results with that of other scanners, and by analyzing how well different scans from the same object align together. We are confident that position errors are below 0.2mm.

One way to obtain ground truth for further comparisons is to produce a high-resolution range image with our triangulation scanner and use it as ground-truth when comparing to data obtained from much lower resolution input images. In practice, we compare full resolution scans with quarter resolution scans.

Following this idea, Figures 5a and 5b present comparisons between the output of our scanners and ground truth for a closeup of the target object seen in Figure 4. Position errors are defined as the distance to ground truth samples along lines of sight of the camera. Normal field errors give the angular deviation with respect to ground truth normals. To facilitate visualization, error values are mapped to colors.

Figure 5a shows the results for directly measured positions from our triangulation scanner. We can also compute normals from this geometry. Position errors in this example were below 0.5mm and can result from noise in the captured images, from speckles or from



(a) Measured positions: Triangulation Scanner

(b) Measured normals: Photometric Stereo Scanner

(c) Corrected normals: Section 4.1

(d) Optimization results: Section 4.2

Figure 5: *Quality assessment. (a-b) Notice how measured data contains considerable error. This takes the form of high-frequency noise for measured positions from a geometric scanner, and low-frequency bias for measured normals from a photometric scanner. (c-d) The combination of the two sources of data corrects most of the error present in the original data.*

imperfections in the subpixel estimation of correspondences. The errors essentially take the form of random high-frequency noise throughout the object. Although relatively small, such errors can produce deviations in excess of 30° in the estimated normal field. When rendered with shading, these deviations are responsible for the distracting bumps seen in Figure 1a.

Figure 5b shows the results for normals measured by our photometric stereo scanner. Errors in measured normals are low-frequency in nature and often manifest themselves in terms of a systematic bias that can reach 30°. These can be the result of interreflections, shadows, or of the oversimplification of the lighting model [Rushmeier and Bernardini 1999; Lensch et al. 2003]. When these normals are used to integrate for positions, large errors appear in the reconstructed surface.

4 Hybrid reconstruction algorithm

On the basis of the preceding error analysis, we conclude that the prevailing errors in measured normals, such as from photometric stereo, are low-frequency in nature, whereas measured positions, such as from a geometric depth scanner, contain mostly high-frequency noise. Therefore, our algorithm for combining these data sources considers frequency components independently. While it is possible to do this in a single step, we find that it is more efficient to proceed in two stages: first *correcting* the bias in the normals, then *optimizing* the geometry to conform to both the corrected normals and the measured positions.

Efficiency is one of our main design goals. Our normal correction step is very efficient and simply involves low-pass filtering and rotations. Combining the corrected normals and measured positions into an improved surface is more challenging. We formulate this as an optimization problem, with careful development of the objective function to enable solution as a sparse linear system. This allows our algorithm to operate efficiently on the large meshes typically found in computer graphics.

We start with the normal correction, then we describe an algorithm specialized to work on single range images, and finally we describe a *full model* method that operates on entire, arbitrarily tessellated triangle meshes.

4.1 Using positions to improve normals

A method to eliminate the bias in the measured normals that takes advantage of the underlying measured positions was presented by Rushmeier and Bernardini [1999]. This method is specific to the photometric stereo setting and therefore we developed a technique that can be applied to a wider range of input data.

As we have seen, the bias present in measured normals is low-frequency. On the other hand, the noise in normals computed from measured positions is high-frequency (compare Figures 5a and 5b). By combining the appropriate frequency bands, we obtain higher quality normal estimates.

Let N^p and N^m be the normal field indirectly obtained from measured positions and the directly measured normal field, respectively. Conceptually, we wish to replace the low-frequency component of N^m with data from N^p. We start by smoothing both fields by the same amount, which should be enough to eliminate the high-frequency noise present in N^p and the high-frequency detail present in N^m. The smoothing can be performed by individually convolving the coordinate functions of the normals with a Gaussian and then renormalizing (3D distances can be used instead of geodesic distances). The resulting smoothed fields, $\mathrm{S}(N^p)$ and $\mathrm{S}(N^m)$, correspond to the low-frequency components of the original fields.

We then compute a rotation field R representing the rotations which move each normal in $\mathrm{S}(N^m)$ to the corresponding normal in N^m. Finally, we compute the corrected normal field $N^c = R\,\mathrm{S}(N^p)$ by applying the rotation field to the smoothed normal field obtained from measured positions. The rotation field captures the high-frequency detail in N^m, but is free of low-frequency information. The detail is transferred uniformly from one normal field to the other, regardless of the angular distance between the corresponding smoothed fields. Notice that since the normal field N^p is only used after severe smoothing, it can be obtained by virtually any method that produces normals from positions.

Figure 5c shows the resulting corrected normals and the improved positions obtained from their integration. We have eliminated the high-frequency noise in Figure 5a, and significantly reduced the bias in Figure 5b. There are still some errors, especially in terms of surface positions due to the inherent problems of integrating the surface normals without depth information. These will be addressed in the next subsection.

4.2 Using normals to improve positions

Measured positions usually come organized in a range-image. The pixel coordinates on the reference camera induce a natural parametrization of the corresponding surface. Accordingly, under perspective projection, the coordinates of a surface point can be written in terms of a depth function $Z(x,y)$. In other words, given the pixel coordinates, the position of the corresponding surface point $P(x,y)$ has only one degree of freedom, $Z(x,y)$:

$$P(x,y) = \left[\begin{array}{ccc} -\frac{x}{f_x}Z(x,y) & -\frac{y}{f_y}Z(x,y) & Z(x,y) \end{array} \right]^T \qquad (1)$$

where f_x and f_y are the camera focal lengths in pixels. Our problem is to find a depth function that conforms to the estimates we have for the position and normal of each point. To do so, we choose the depth function that minimizes the sum of two error terms: the *position error* E^p and the *normal error* E^n.

The position error is defined as the sum of squared distances between the optimized positions and the measured positions:

$$E^p = \sum_i ||P_i - P_i^m||^2 \qquad (2)$$

where P_i is the i^{th} optimized position, and P_i^m is the corresponding measured point. To evaluate the position error, we transform depth values to distances from the center of projection, along lines of sight:

$$||P_i - P_i^m||^2 = \mu_i^2 (Z_i - Z_i^m)^2, \quad \text{where} \qquad (3)$$

$$\mu_i^2 = \left(\frac{x_i}{f_x}\right)^2 + \left(\frac{y_i}{f_y}\right)^2 + 1 \qquad (4)$$

The normal error could be defined in a number of different ways, including the sum of angular errors between corresponding normals in the optimized surface and the corrected normal field, or the sum of squared distances between each normalized or un-normalized pair. However, most formulations lead to a *non-linear* optimization problem, which is impractical for the size of our datasets. Our solution is to consider the tangents to the optimized surface instead. The corrected normals (which are constant) and the tangents to the optimized surface should be perpendicular. Recall that the surface tangents T_x and T_y at a given pixel can be written as linear functions of the depth values and their partial derivatives:

$$T_x = \frac{\partial P}{\partial x} = \left[\begin{array}{ccc} -\frac{1}{f_x}\left(x\frac{\partial Z}{\partial x} + Z\right) & -\frac{1}{f_y}y\frac{\partial Z}{\partial x} & \frac{\partial Z}{\partial x} \end{array} \right]^T \qquad (5)$$

$$T_y = \frac{\partial P}{\partial y} = \left[\begin{array}{ccc} -\frac{1}{f_x}x\frac{\partial Z}{\partial y} & -\frac{1}{f_y}\left(y\frac{\partial Z}{\partial y} + Z\right) & \frac{\partial Z}{\partial y} \end{array} \right]^T \qquad (6)$$

We now define

$$E^n = \sum_i [T_x(P_i) \cdot N_i^c]^2 + [T_y(P_i) \cdot N_i^c]^2 \qquad (7)$$

where N_i^c is the corrected normal corresponding to P_i. E^n is the sum of squared projections of the tangents to the optimized surface into the corrected normal field. It is minimized when all tangents

are perpendicular to the corrected normals. To evaluate $T_x(P_i)$ and $T_y(P_i)$, we compute the partial derivatives of the depth function and substitute in equations 5 and 6. Since we are dealing with a uniform discrete sampling of the depth function, we can approximate the partial derivatives by considering 3×3 neighborhoods and the following convolution kernels, assuming all neighbors are available:

$$\frac{\partial Z}{\partial x} = Z * \frac{1}{12}\begin{array}{|c|c|c|}\hline -1 & 0 & 1 \\ \hline -4 & 0 & 4 \\ \hline -1 & 0 & 1 \\ \hline \end{array}, \quad \frac{\partial Z}{\partial y} = Z * \frac{1}{12}\begin{array}{|c|c|c|}\hline 1 & 4 & 1 \\ \hline 0 & 0 & 0 \\ \hline -1 & -4 & -1 \\ \hline \end{array} \tag{8}$$

Occasionally, around boundaries and depth discontinuities, some neighbors will not be present. We can detect these cases by analyzing the measured positions and use the best possible discrete derivative, down to simple one-sided derivatives. When there are no neighbors, the point can be removed from the minimization.

The optimal surface is then given by

$$\arg\min_{Z} \lambda E^p + (1-\lambda)E^n \tag{9}$$

where the parameter $\lambda \in [0,1]$ controls how much influence the positions and normals have in the optimization. The two error terms are measured in units of squared distance and therefore λ is dimensionless. When λ is 0, the algorithm considers normals exclusively, with help from measured positions only in boundary conditions, much like shape-from-shading (in fact, this is the method we use to *integrate* normals for comparison purposes in this paper). When λ is 1, the algorithm simply returns the original positions. For intermediate values, the method finds the optimal weighted combination of normals and positions.

Each pixel generates at most 3 equations: one for the position error, and one for the normal error in each of the x and y directions. Before squaring, the equations for the error terms are linear in the depth values we are solving for. Therefore, the entire minimization can be formulated as a large over-constrained linear system to be solved by least squares:

$$\begin{bmatrix} \lambda\mu I \\ \cdots \\ N \cdot T \end{bmatrix} \begin{bmatrix} Z \end{bmatrix} = \begin{bmatrix} \lambda\mu Z^m \\ \cdots \\ 0 \end{bmatrix} \tag{10}$$

Here, $N \cdot T$ represents a matrix that, when multiplied by the unknown vector Z, produces a vector with two rows per point, evaluating the normal constraints $(1-\lambda)N^c \cdot T_x$ and $(1-\lambda)N^c \cdot T_y$. Note that T_x and T_y are linear in Z and N^c is independent of Z. Additionally, the method can be extended to use individual confidence estimates for each normal and position constraint by simply pre-multiplying the system by a diagonal weight matrix (i.e., using weighted least squares).

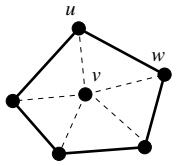
With 3 equations per pixel, a 1024×768 scan can generate hundreds of thousands of equations. Fortunately, the matrix is very sparse. In fact, the number of non-zero entries is linear in the number of pixels because there are at most 7 non-zero entries per row (one coefficient for the depth of the reference pixel and at most six others for the neighbors used to find the partial derivatives). The Paige and Saunders [1982] implementation of their Conjugated Gradient method for solving sparse least squares problems can easily handle this type of system. The resulting range image is accurate in both positions and normals, as the example in Figure 5d shows.

4.3 Full model optimization

In some cases, we might be given a full 3D model with a normal map, and no access to the original range maps. This might be because the range maps never existed (e.g., a depth acquisition method that does not yield regularly-spaced range images was used), or because the normals were obtained from several different sources (as in Figure 10). A generalization of the 2.5D method presented in the previous section to arbitrarily tessellated 3D meshes can be useful in these cases.

The number of variables in the optimization and the number of equations will usually be larger than for the 2.5D case. Not only are there more vertices, but we are optimizing for 3 independent coordinates instead of one depth value per vertex. It is, therefore, important that our method remains efficient and this entails finding a linear formulation for the objective function.

In defining the new normal error $\hat{E}^n$, we no longer have a trivial way to compute the partial derivatives of the optimized surface at each vertex (needed to compute tangents), because there is no obvious parametrization. A solution that proved adequate is to consider the polygon formed by the neighbors of each vertex as an approximation to its tangent space. For each edge in each polygon, we add a term to the normal error that favors edges that are perpendicular to the measured normal at the central vertex. This approximation has the advantage of being linear.



$$\hat{E}^n = \sum_v \sum_{u,w} [N_v^m \cdot (P_u - P_w)]^2 \tag{11}$$

The new position error $\hat{E}^p$ also requires additional thought. If we allow too much freedom in moving the vertices, the optimization might result in a self-intersecting model. In practice, this does not happen when the vertices move relatively little compared to the edge lengths. This is the case if we start from a reasonable approximation to the optimal positions. If, in addition, we restrict vertex motion to the normal direction, self-intersections are extremely unlikely. Fortunately, this can be achieved without breaking the linearity of the objective function.

$$\hat{E}^p = \sum_v ||M_v(P_v - P_v^m)||^2 \tag{12}$$

Here, $M_v = \alpha N_v^m N_v^{mT} + \beta(I - N_v^m N_v^{mT})$ is a 3×3 matrix that has N_v^m and its tangent space as eigenvectors with α and β respective eigenvalues. Choosing a relatively small value for α allows P_v to move away from P_v^m in the N_v^m direction without adding much to the position error. On the other hand, deviations in the tangent direction are magnified by β and are therefore not favored. Thus, M_v helps prevent self-intersections in the optimized model, which is given by

$$\arg\min_{P} \lambda\hat{E}^p + (1-\lambda)\hat{E}^n \tag{13}$$

Since all our constraints are linear, the optimization can proceed by least squares just like the 2.5D version of the algorithm, with λ playing the same role. The resulting full model has less high-frequency noise and more high-frequency detail than the simple merging of range scans. Examples of full model optimization are shown in Figures 10 and 9, and discussed in the results section.

5 Results

In this section, we first evaluate the algorithm, discussing the setting of parameter values, and the accuracy and efficiency. We then discuss some applications, showing examples of some of the reconstructions we have produced using our method.

Accuracy and robustness: The precision of our method is one of its key features. The color-coded error renderings of Figures 5c and 5d provide an indication of the quality of the results. Another way to assess the accuracy of our method is to analyze depth profiles for a reference object, as in Figure 6. The plots show that measured positions are noisy and integration of corrected normals creates extraneous detail, whereas the optimized surface eliminates noise while closely following the ground truth.

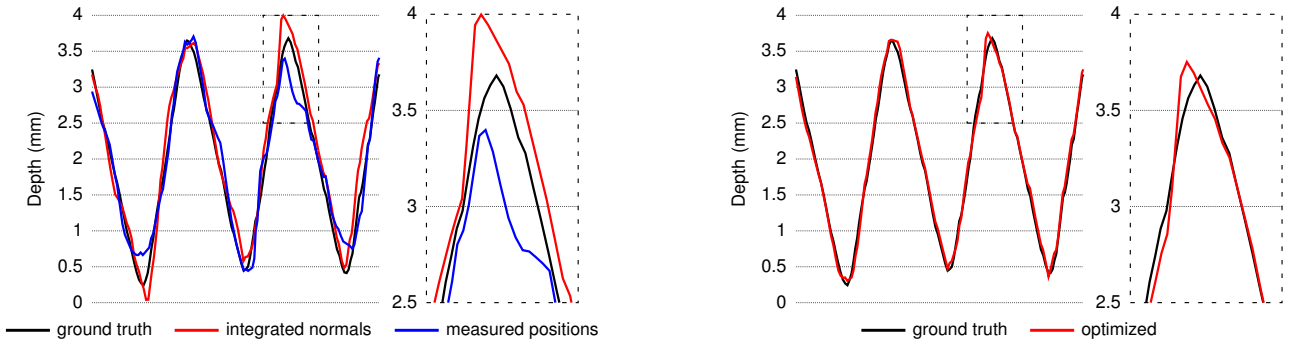


Figure 6: *Depth profiles (in 0.5mm) for a reference object. (Left) Measured positions and integration from corrected normals, (Right) optimized surface. Note how the optimized profile follows the ground truth profile closer than the others.*

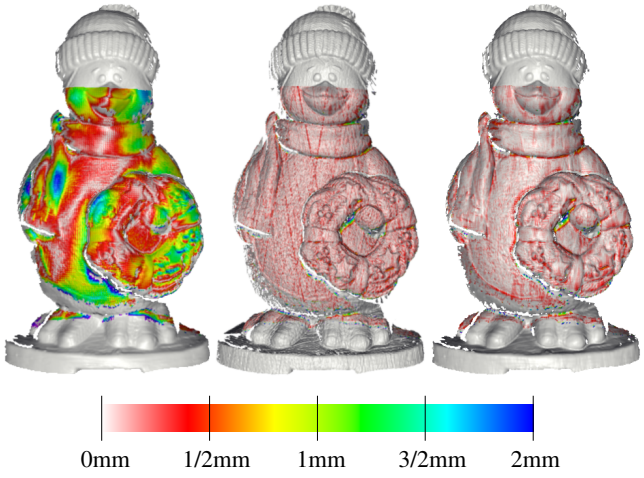


Figure 7: *Distance between two aligned scans of a 30cm plaster penguin. (Left) Integration of normals, (Center) directly measured positions, (Right) optimized surfaces ($\lambda = 0.1$). Our method produces unwarped results that align very well.*

Another important measure of surface precision is how well partial scans align with respect to each other. In a standard scanning pipeline, partial scans are first manually brought close to their aligned positions. Then, ICP [Besl and McKay 1992] is run to precisely align scans pairwise. A global registration pass [Pulli 1999] is used to spread registration error evenly across pairwise alignments. Finally, all scans are merged into a single model using the VRIP volumetric method [Curless and Levoy 1996].

Good alignments are vital if details are to be preserved. In addition, since warped scans do not align properly, it is possible to use alignment quality to verify that scans are free of warp. Figure 7 compares our results against the alignment obtained by directly measured positions and by integration of photometric stereo normals. Low-frequency warps are clearly visible in the results of integration of normals. Conversely, high-frequency noise can be observed in the directly measured positions. By contrast, the results of our method are free of warp and align very well.

In general, our method is robust with regard to outliers both in positions and normal estimates. Position outliers can be easily detected and eliminated with the analysis of shapes and sizes of the triangles produced from the measured range images. Although harder to detect, the influence of normal outliers is limited to a small neighborhood.

Relative importance of positions and normals—setting λ: The λ parameter provides simple and effective control over the behavior of the algorithm, which consistently produces good results. Recall that λ is a dimensionless parameter in the range $[0,1]$ and controls the relative contribution of normals and positions in the optimization. For instance, when the measured positions are noisy, a smaller λ, giving less weight to geometric data, will smooth the noise away. In general, we find that values in the range of $[0.10, 0.30]$ are most suitable for λ, and the method is not very sensitive to parameters in that range, with satisfactory results obtained for all the examples we tested. Values outside this range can be appropriate when the quality of positions and normals is less balanced.

Efficiency: One important characteristic of our method is its efficiency. By formulating only linear constraints, we are able to solve the minimization by linear least squares. Since each constraint refers to a constant number of variables, and the number of constraints is linear in the number of variables, the memory requirements are also linear. As the weight λ assigned to measured positions increases, the identity part of the linear system dominates (see equation 10) and convergence is greatly improved. Hence, optimizing a range image with $\lambda = 0$ (corresponding to integrating the surface normals directly) may require several minutes. However, using a value of λ within the suitable range [0.10, 0.30] allows us to obtain high quality results within seconds on models having hundreds of thousands of vertices.

Reconstruction and Rendering of Complex Objects: Our method can produce very high-quality reconstructions of complex objects with hundreds of thousands of triangles, including sharp and high-frequency spatial features. Our results can be used for rendering, as well as other visualization and signal-processing tasks. In the course of this research, we have used our algorithm to create reconstructions of several objects, some of which are shown in Figures 1 and 8 (at the end of paper). The top row in Figure 8 shows a closeup of 16 aligned scans of the plaster penguin used in Figure 7. The middle row shows a closeup of illustrations on a porcelain gravy boat. The last row shows a model obtained from a sea shell. As discussed in the introduction, errors in the normals estimated from the scanned geometry produce unpleasant bumps (a). Using the normals acquired by photometric stereo (b) eliminates the bumps, but introduces bias. The optimized geometry (c), on the other hand, is almost free of noise or bias. Furthermore, we obtain accurate renderings at grazing angles, and with mesh operations like accessibility shading (Figure 2). Indeed, many rendering tasks like ray-tracing reflections, silhouette computations, suggestive contours, and lighting will benefit from the noise free, unbiased normals of our method.

Full Model Optimization: Two examples are presented for the full model version of our algorithm (subsection 4.3). The model shown in the left part of Figure 9 is the result of the alignment and merging of 15 raw range scans into a full model. Normals for the

same object were measured by another 15 photometric stereo scans. These normals were then corrected for bias (as in section 4.1) and mapped to the vertices of the merged model. Finally, the vertex positions were then optimized to conform to a number of overlapping normal constraints. The result, shown in the right of Figure 9, clearly has less high-frequency noise and more detail.

As an example of application of our method to unconventional sources of surface normal and geometry information, consider the quarter in Figure 10. The model was produced by the mesh optimization of flat geometry with a bump map. The flat geometry was generated procedurally, from the specification of the coin dimensions. The bump maps for the two sides of the coin were produced with a flat-bed scanner. The bump map for the edge was produced procedurally from the specification of the number of ridges. The figure shows a model lit by normals computed from the optimized geometry.

6 Conclusion

3D geometry is commonplace in computer graphics. Typically, this information comes in the form of depth and vertex positions, or normals and surface orientations. However, neither depth information alone, normal information alone, nor a simple combination such as a bump map, provide a complete solution for computer graphics modeling and rendering. In this paper, we have shown that when we have access to positions and normals for a 3D model, it is possible to combine these two sources of information into an optimal surface. If only the most reliable component of each source of information is considered, the resulting surface will be more precise than that obtainable by each source independently. We present an analysis of the common error characteristics of standard position and normal acquisition systems, and design our algorithm to account for these types of errors. By formulating the problem in a particular way, we reduce it to solving a sparse linear system, enabling very efficient computation of optimal surfaces for large meshes. Our algorithm represents the first practical technique in computer graphics of combining position and normal information for precise surface reconstruction.

Acknowledgements

We thank the many people, especially the Tiggraph reviewers, who have offered their suggestions on this paper. The coin normal maps were acquired by Lucas Pereira for Stanford's CS348B, Spring 1996. We thank the National Science Foundation for funding that partially supports this project (Szymon Rusinkiewicz: CCF-0347427, "Practical 3D Model Acquisition"; Ravi Ramamoorthi: CCF-0446916, "Mathematical and Computational Fundamentals of Visual Appearance for Computer Graphics").

References

BANERJEE, S., SASTRY, P., and VENKATESH, Y. 1992. Surface reconstruction from disparate shading: An integration of shape-from-shading and stereopsis. In *IAPR*, volume 1, pages 141–144.

BERNARDINI, F., MARTIN, I., MITTLEMAN, J., RUSHMEIER, H., and TAUBIN, G. 2002. Building a digital model of michelangelo's florentine pietà. *IEEE Computer Graphics and Applications*, 22(1):59–67.

BESL, P. J. 1989. *Active Optical Range Imaging Sensors*, pages 1–63.

BESL, P. J. and MCKAY, N. D. 1992. A method for registration of 3-d shapes. *PAMI*, 14(2):239–256.

BOUGUET, J.-Y. 2004. *Camera Calibration Toolbox for Matlab*. URL `http://www.vision.caltech.edu/bouguetj/calib_doc`.

CHEN, C.-Y., KLETTE, R., and CHEN, C.-F. 2003. Shape from photometric-stereo and contours. In N. Petkov and M. Westenberg, editors, *CAIP*, Springer-Verlag, pages 377–384.

CRYER, J. E., TSAI, P.-S., and SHAH, M. 1995. Integration of shape from shading and stereo. *Pattern Recognition*, 28(7): 1033–1043.

CURLESS, B. 1997. *New Methods For Surface Reconstruction From Range Images*. PhD thesis, Stanford University.

CURLESS, B. and LEVOY, M. 1996. A volumetric method for building complex models from range images. In *SIGGRAPH*, ACM Press, pages 303–312.

DAVIS, J., NEHAB, D., RAMAMOORTHI, R., and RUSINKIEWICZ, S. 2005. Spacetime stereo: A unifying framework for depth from triangulation. *PAMI*, 27(2):296–302.

EPSTEIN, R., YUILLE, A. L., and BELHUMEUR, P. N. 1996. Learning object representations from lighting variations. In *ECCV*, pages 179–199.

FUA, P. and LECLERC, Y. G. 1994. Using 3-dimensional meshes to combine image-based and geometry-based constraints. In J.-O. Eklundh, editor, *ECCV*, volume 2, Springer-Verlag, pages 281–291.

GEORGHIADES, A. 2003. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *ICCV*, volume 2, pages 816–823.

HORN, B. K. P. 1970. *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View*. PhD thesis, Massachusetts Institute of Technology.

IKEUCHI, K. 1983. Constructing a depth map from images. A.I Memo AIM-744, Artificial Intelligence Laboratory, MIT.

IKEUCHI, K. 1987. Determining a depth map using a dual photometric stereo. *IJRR*, 6(1):15–31.

LANGE, H. 1999. Advances in the cooperation of shape from shading and stereo vision. In *3DIM*, pages 46–58.

LENSCH, H. P. A., KAUTZ, J., GOESELE, M., HEIDRICH, W., and SEIDEL, H.-P. 2003. Image-based reconstruction of spatial appearance and geometric detail. *ACM Trans. Graph.*, 22(2): 234–257.

MILLER, G. 1994. Efficient algorithms for local and global accessibility shading. In *SIGGRAPH*, pages 319–326.

MOSTAFA, M., YAMANY, S., and FARAG, A. 1999. Integrating shape from shading and range data using neural networks. In *CVPR*, volume 2, IEEE Computer Society, pages 15–20.

PAIGE, C. C. and SAUNDERS, M. A. 1982. LSQR: An algorithm for sparse linear equations and sparse least squares. *TOMS*, 8(1): 43–71.

POUSSART, D. and LAURENDEAU, D. 1989. *3-D Sensing for Industrial Computer Vision*, pages 122–159.

PULLI, K. 1999. Multiview registration for large datasets. In *3DIM*, pages 160–168.

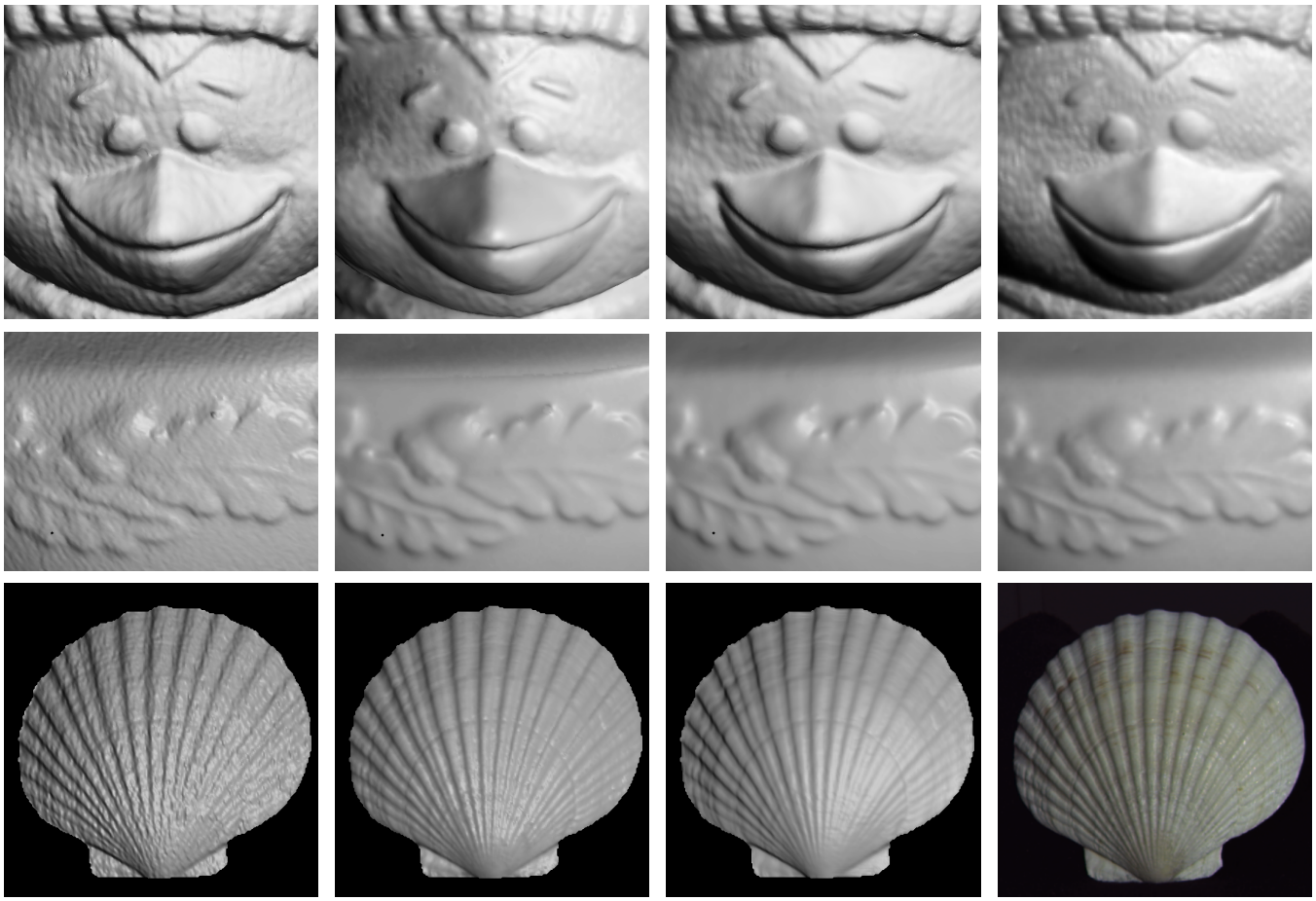
RUSHMEIER, H. and BERNARDINI, F. 1999. Computing consistent normals and colors from photometric data. In *3DIM*, pages 99–108.

TAGARE, H. D. and DEFIGUEIREDO, R. J. P. 1991. A theory of photometric stereo for a class of diffuse non-lambertian surfaces. *PAMI*, 13(2):133–152.

TERZOPOULOS, D. 1988. The computation of visible-surface representations. *PAMI*, 10(4):417–438.

WOODHAM, R. J. 1980. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19 (1):139–144.

ZHANG, L., CURLESS, B., and SEITZ, S. M. 2003. Spacetime stereo: Shape recovery for dynamic scenes. In *CVPR*, pages 367—374.



(a) (b) (c) (d)

Figure 8: *Rendering comparisons. (a) Rendering of 3D scanned range image, (b) same scanned geometry, augmented with a measured normal-map (from photometric stereo), (c) our hybrid surface reconstruction, (d) photograph. The top row is the result of aligning several range images for the penguin shown in Figure 7. The middle row shows a closeup of a porcelain gravy boat. The bottom row shows scans of a sea shell.*

Figure 9: *Full model optimization. (Left) Several range scans were aligned and merged. (Right) The result was later optimized with mapped normals coming from several independent photometric stereo scans.*

Figure 10: *Full model optimization, with geometry and normals from a variety of sources. A model of a quarter was produced from initially flat geometry and a bump map. The figure shows optimized model with normals recomputed from geometry.*