# Photogeometric Structured Light:
# A Self-Calibrating and Multi-Viewpoint Framework for Accurate 3D Modeling

Daniel G. Aliaga          Yi Xu

Department of Computer Science

Purdue University

`{aliaga|xu43}@cs.purdue.edu`

## Abstract

*Structured-light methods actively generate geometric correspondence data between projectors and cameras in order to facilitate robust 3D reconstruction. In this paper, we present Photogeometric Structured Light whereby a standard structured light method is extended to include photometric methods. Photometric processing serves the double purpose of increasing the amount of recovered surface detail and of enabling the structured-light setup to be robustly self-calibrated. Further, our framework uses a photogeometric optimization that supports the simultaneous use of multiple cameras and projectors and yields a single and accurate multi-view 3D model which best complies with photometric and geometric data.*

## 1. Introduction

Creating high-quality 3D models of detailed and complex objects is crucial to a wide range of uses, such as in virtual reality, telepresence, and interactive walkthroughs. The robust correspondences generated by structured-light based methods, and other active techniques (e.g., lasers), make them appealing. Our work provides a self-calibrating structured-light framework whereby both photometric and geometric data are used to robustly obtain highly-detailed multi-viewpoint models.

Passive correspondence approaches are based on observing naturally occurring features on scene objects and reconstructing the geometry from these observations. Classical stereopsis methods typically assume a diffuse surface reflectance model and seek features to use in a 3D reconstruction [17] or to use in a geometrically-based structure-from-motion formulation [13]. Photometric stereopsis [1][21] and Helmholtz stereopsis [24] enable the reconstruction of surfaces of a known reflectance model and of surfaces of an arbitrary unknown reflectance model, respectively. Photometrically-based methods then reconstruct the object and are able to capture fine surface details. However, they are known to suffer from global warps due to a fundamental linear ambiguity in the reconstruction [2]. To reduce the warp, methods use geometric data, such as tracked image features [8][22] or features in reciprocal image pairs [20].

Active correspondence methods benefit from the increased robustness of actively generating features and thus reliably establishing correspondence. Techniques have been developed using scanning laser-light and structured-light patterns. Laser-scanning requires highly specialized and expensive equipment. Structured-light employs standard digital projectors and cameras. While structured-light recovers the global object shape well, fine details are often averaged away or undersampled because of limited projector resolution. In contrast, cameras typically obtain photometric data at higher resolution and photometrically-based processing can be less sensitive to surfaces straying from a diffuse reflectance assumption.

In this paper, we present a self-calibrating structured-light method for robustly performing a photogeometric reconstruction that in addition obtains multi-viewpoint data and is self-calibrating (Figure 1). Photogeometric reconstruction, as defined by [9], refers to using both photometric and geometric measurements to reconstruct a surface. As opposed to pre-calibrated and single-viewpoint photogeometric systems [12][14] and to photogeometric systems requiring the collocation or physical swapping of cameras and light sources [24], our approach uses non-collocated and un-calibrated capture devices. Akin to methods using Helmholtz reciprocity (e.g., Dual Photography [18]), we exploit the duality of cameras and projectors. However, unlike Dual Photography we obtain both geometric and photometric data from each of multiple perspective viewpoints and perform a 3D reconstruction. Further, instead of the standard approach of using a projector to link multiple camera views, our approach uses the cameras to link multiple projector views. This subtle difference enables us to use the same hardware setup for both photometric and geometric measurements, to obtain data from multiple viewpoints, and to perform a global photogeometric optimization that prevents needing to align multiple reconstructions.

Moreover, our framework reduces the number of structured-light patterns and thus decreases image capture time. The reduction comes at the expense of a coarser initial reconstruction; however, our photogeometric process robustly improves the reconstruction so as to be
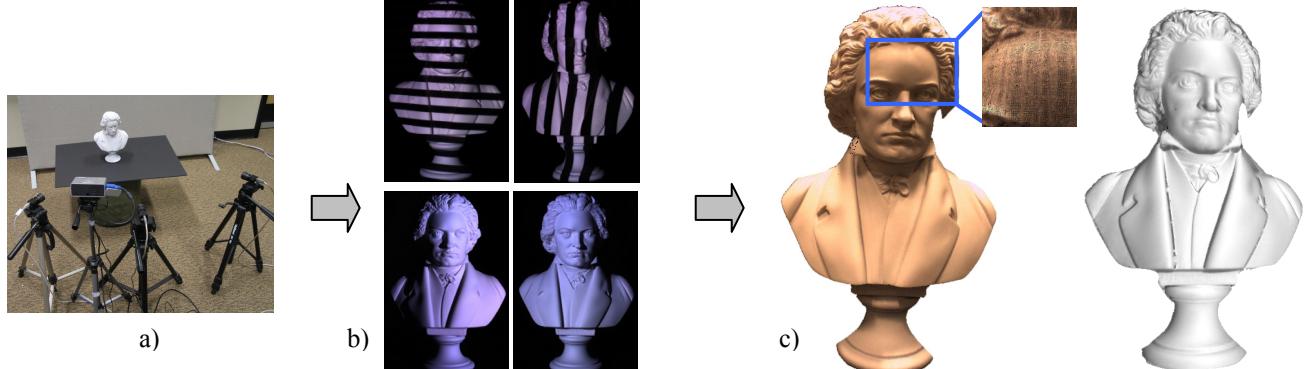
**Figure 1. Photogeometric Structured Light.** a) Setup consists of one projector placed at multiple uncalibrated positions and several uncalibrated cameras. b) Our approach uses geometric (top) and photometric (bottom) techniques from the same effective viewpoints. c) The result is self-calibrated accurate multi-view models (left=texture-mapped, right=synthetic illum., middle=wireframe close-up).

similar to or better than when all patterns are used.

The end result is accurate, robust, and flexible reconstruction of detailed objects. The overall capture and reconstruction process is simple to use, automatic, and completes in as little as 15 seconds. We report and analyze our approach on a variety of example reconstructed objects spanning from approximately 50k to 300k points. Our main contributions include:

- a self-calibrating structured-light method for robustly capturing photogeometric data per pixel using uncalibrated hardware, without requiring beam splitters or co-location,

- a photogeometric optimization for obtaining a single multi-viewpoint model complying to both photometric and geometric measurements, without having to align various reconstructions, and

- a complete and automatic framework for the acquisition of highly-detailed 3D objects obtaining accurate photometric and geometric data.

## 2. Related Work

Our approach builds upon previous efforts in structured-light acquisition, photometric stereo, and their combination. In particular, as opposed to conventional calibrated structured-light (e.g., [15][16][23]) and uncalibrated structured-light (e.g., [5]) approaches, our method also acquires and optimizes high-surface-detail photometric information which serves the double purpose of robustly initializing the self-calibration process and of significantly improving the accuracy of 3D modeling. Self-calibration methods typically rely on features and on either assumed scene or geometry constraints to estimate camera parameters [7][10]. While convergence is sometimes feasible, it is difficult and sometimes not possible [19]. It is our aim to use the initial viewing parameters estimated via a photometric method to help initialize self-calibration of a structured-light system.

In contrast with photometric stereo and shape-from-shading, our structured-light approach actively generates robust features and is able to overcome the typical low-frequency deformation of the geometry. In particular, we surmount the ambiguity of the generalized bas-relief transformation [2] and, thus, obtain the 3D model up to a global scale factor. To improve upon the linear ambiguity, some previous methods rely on tracking passive features from image-to-image and simultaneously performing a structure-from-motion refinement [8][22]. While this may help to reduce global shape deformations, it relies on fragile feature correspondence and cannot produce densely-sampled geometric data points.

Helmholtz reciprocity has been exploited for novel image generation and for surface reconstruction. For instance, Dual Photography [18] demonstrated how the view from a light source (e.g., projector) can be obtained by simply inverting the light transport matrix from projector to camera; however, no scene geometry is obtained. In [3], they use a setup similar to Dual Photography and use optical flow for subsets of the scene to perform relighting from interpolated positions but no geometry is extracted. Helmholtz stereopsis [24] physically swaps a light source and camera so as to enable their co-location and to reconstruct an object without having to make assumptions about its bidirectional reflectance distribution function (BRDF). While the initial Helmholtz stereopsis method required pre-calibration it has been extended to an uncalibrated approach requiring known epipolar geometry [25] and to an uncalibrated approach using reciprocal image features [26]. Both still require co-location of the camera and light source and the latter method also depends on the presence of either 'texture' features or 'specularity' features.

While some previous approaches have acquired photogeometric from a single viewpoint using two pre-calibrated acquisition setups [12][14], we improve such systems in several ways. Our method is both self-

calibrating and multi-viewpoint, hence making the setup significantly more flexible, practical, and able to obtain more complete 3D models. In addition, we use the exact same equipment for both geometric and photometric data acquisition. This does require us to first estimate a solution to the photometric model. Then, after self-calibration, we change the photometric model to a more truthful point light model and recompute surface normals. The single capture system and coordinate framework also has the benefit of removing the rotational alignment (as in [12]) or positional-and-rotational alignment (as in [14]) needed between photometric and geometric samples.

## 3. Photogeometric Structured Light

Our photogeometric structured light approach reconstructs an object observed and illuminated from multiple locations. First, an initialization phase calculates approximate projector poses (using a photometrically-based formulation) and a set of point correspondences between projectors views (using binary structured-light patterns). The user can choose between using a faster and more approximate initialization with coarse initial geometry information or using a slower and more accurate initialization with denser initial geometry information. Second, the projectors, now acting as virtual cameras, are used in a self-calibrating method to estimate a multi-view surface model of the object. Third, the multi-view surface geometry is optionally up-sampled by warping to it a denser surface obtained from the integration of the photometrically-computed normals. Fourth, the 3D model is optimized so as to best match both the photometric and the geometric observations. Since the location of the lights is now known, they are more accurately treated as point-lights and are used to update the photometrically-computed normals. Acquisition calibrates the projector (its focal-length and external parameters) and estimates the geometry and surface normals of the object. If desired, the cameras' internal and external parameters can be computed. In the following, we describe image capture and steps 1-3. We explain step 4 in the following section.

### 3.1 Image Capture

Our capturing method consists of sequentially placing one projector at $n$ unknown locations around the object and acquiring images of the object from $m$ different cameras at fixed (but unknown) locations. Our method projects $P$ pairs of binary patterns. To prevent determining surface albedos, we project the patterns and their inverses. Our patterns permit robustly corresponding about $(2^P-1)^2$ surface points between a camera and a projector. To yield fewer points and faster processing, we use small $P$ values (e.g., 5). To avoid deciding which cameras to use for a projector location, we capture images from all cameras.

A multi-viewpoint modeling approach observes a 3D object from various viewpoints using several cameras or one camera sequentially placed at multiple locations. However, our method requires the observer to be either a passive camera or an active light. Thus, we use a simple consequence of Helmholtz reciprocity and exploit the per-pixel correspondence data of a structured-light setup to enable the projector to be either a virtual camera or a point-light source. Figure 2a illustrates the setup. In particular, a projector is sequentially moved to locations $P_0$-$P_3$. At each location, it illuminates the object with binary stripe patterns and their inverses. Cameras $C_0$ and $C_1$ are statically positioned and capture images of the patterns projected onto the objects. The poses and focal lengths of the cameras and projector are unknown.

### 3.2 Photometry: Initialization

Using the captured images, we initialize the viewing direction of the projectors and compute a per-pixel correspondence between them. By considering the projector to be a directional light, we use uncalibrated photometric stereo to recover estimates of the lighting directions $l_j$. Then, by considering the projector to be virtual camera $V_j$ and using the structured-light patterns observed by the actual cameras, we generate a correspondence between the multiple projector-views.

#### 3.2.1  Photometric Processing

Uncalibrated Lambertian photometric processing enables recovering estimates of the surface normals $\bar{s}_i$ and the surface lighting $l_j$ directions using a linear least squares optimization. In particular, it solves $NL^T=C$ where $N$ is an unknown $p$ x 3 matrix of $p$ outward-facing surface normals, $L$ is an unknown $n$ x 3 matrix of $n$ light directions pointing towards the light, and $C$ is a $p$ x $n$ matrix of the pixel intensities observed by a camera. The solution is known up to a linear ambiguity, which includes a global rotation. Nevertheless, our approach uses each estimated lighting direction as the viewing direction of the corresponding projector. Thus, future geometry estimates will implicitly be in the same space as the normal field. It is worth noting that although we use diffuse photometric processing, our method is not limited to Lambertian surfaces. The photometric method initializes the lighting directions, which will be snapped to the correct solution by the subsequent geometric self-calibration. Therefore, accuracy of the initial estimation is not key as long as it is not too far off. Our method can handle non-Lambertian surfaces with a sufficient diffuse component.

Figures 2b-c show typical results from photometric stereo processing. Figure 2b contains a photograph of a diffuse object illuminated by a projector with the first (all-white) pattern of the structured-light sequence. Figure 2c
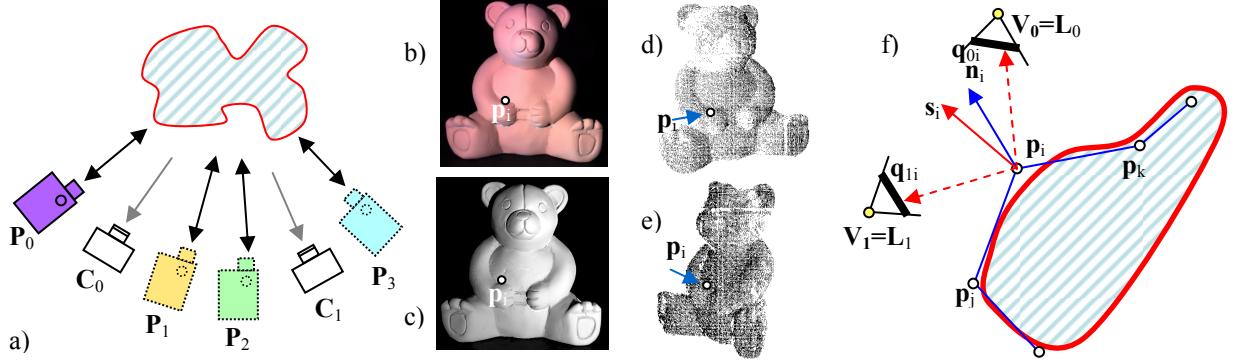
**Figure 2. Acquisition Setup.** (a) A one projector and multiple-camera setup for performing acquisition. (b) A photograph of an object to be captured. (c) Photometric reconstruction using 3 images illuminated from 3 directions. (d-e) Two corresponded virtual views between two projectors and the camera picture in b. (f) Photogeometric setup surrounding point $p_i$.

uses the estimated normals to show the pixel luminance values using a differently-oriented synthetic light. Each pixel $p_i$ has both a color and a normal associated with it.

### 3.2.2 *Virtually Exchanging Projectors and Cameras*

The structured-light patterns have generated a correspondence between the *n* projector locations and *m* cameras. Similar to Dual Photography [18], we can invert a projector-to-camera light transport matrix to obtain a view from the projector. However, in order to convert the projectors into virtual cameras, we need to correspond as many pixels as possible between all projectors (and cameras). Thus, our method is to re-sample and use the camera-to-projector correspondence data for producing dense projector-to-projector correspondence data.

Our algorithm obtains a near-regular grid of object points corresponded amongst as many projectors (and cameras) as possible. After capture, each camera has observed all projector patterns and has a set of point correspondences with the projectors. Camera $C_i$ ($i \in [1..m]$) and projector $P_j$ ($j \in [1..n]$) each create a triangulation of the points they have in common with each other. Camera $C_i$ then computes the barycentric coordinates for a regular grid of points and uses the barycentric coordinates to obtain corresponding points on projector $P_j$. The newly created points on projector $P_j$ are then projected onto all other cameras provided the projector triangle is visible in the other camera. This process is repeated for all other cameras. The result is a large set of points which at most are visible in all projectors and all cameras and at the least equals the correspondence between one camera and one projector. For example, point $p_i$ in Figures 2b and 2c is corresponded with the projector views in Figures 2d and 2e. Point too close to others in all images are eliminated and the final outcome is a near-regular distribution of points corresponded between projectors (and cameras).

## 3.3 Geometry: Surface Estimation

Our method uses the previously estimated *n* projector viewing directions and correspondence data to perform a self-calibration and geometric reconstruction. Given at least $n \geq 2$ projector views, geometric processing seeks to estimate projector parameters and the 3D location of points on the observed object so as to minimize the reprojection error expressed by the following well-known nonlinear system of equations

$$\sum_j \sum_i \left(\frac{1}{q_{ij_z}}\begin{bmatrix} q_{ij_x} \\ q_{ij_y} \end{bmatrix} - \begin{bmatrix} u_{ij} \\ v_{ij} \end{bmatrix}\right)^2 \text{ where } q_{ij} = F_j(R_j p_i + T_j) \quad (1)$$

and $(u_{ij}, v_{ij})$ are the known projections of object point $i$ onto projector $V_j$, $p_i$ are the unknown world coordinates, and $R_j$, $T_j$, and $F_j$ are the unknown projector 3x3 rotation matrices, 3D translation vectors, and 3x3 perspective projection matrices, respectively.

In an initial phase, a sparse and uniformly-distributed subset of object points corresponded across all projectors (or across a large subset of them) are used to optimize the distance from each projector to the object's center and a global projector focal length. To obtain the initial values for the object points $p_i$, our approach integrates the per-pixel normals obtained during photometric processing using the method of [4]. The photometrically-calculated virtual viewing directions $l_j$ and an assumed up-vector of $w = [0\ 1\ 0]^T$ are then used to create an oriented orthogonal coordinate system for each projector, represented by a matrix $M_j$ (i.e., its columns correspond to the unit vectors $l_j \times w$, $(l_j \times w) \times l_j$, and $-l_j$). The free parameters are the distances $z_j$ from the origin to each projector $j$ along the $l_j$ vector and a single global focal length $f$ used by all projectors. To bring the re-projection of the object points into rough alignment with the observed projections, we optimize the following simplified nonlinear system of equations of only $n+1$ unknowns ($f$ and $z_j$ for $j \in [1, n]$)

$$\sum_j \sum_i \left(\frac{\hat{p}_{ij_x} f}{\hat{p}_{ij_z} + z_j} - u_{ij}\right)^2 + \left(\frac{\hat{p}_{ij_y} f}{\hat{p}_{ij_z} + z_j} - v_{ij}\right)^2 \text{ where } \hat{p}_{ij} = M_j p_i. \ (2)$$

Having already estimated projector orientations, object-projector distances, and a global focal length, our method

optimizes for a linear correction to each projector location. In particular, in equation set (1), $F_j$ is replaced by a perspective projection matrix using $f$, $T_j$ is replaced by $[0\ 0\ z_j]^T$ and $R_j$ is replaced by $Q_j M_j$. The 3x3 matrix $Q_j$ is to be computed via the optimization. For each projector, the equations are simplified to the following linear system of equations in the 8 unknowns of the $Q_j$ matrix (i.e., $q_{33}=1$)

$$\sum_i \begin{bmatrix} q_{ij_x} - u_{ij}(q_{ij_z} + z_j)/f \\ q_{ij_y} - v_{ij}(q_{ij_z} + z_j)/f \end{bmatrix} \text{ where } \hat{q}_{ij} = Q_j M_j p_i . \quad (3)$$

The unknowns are solved for using an over-constrained sparse linear least squares formulation. The $Q_j$ matrix can be forced to only be a rotation matrix by adding the constraints of keeping its columns mutually orthogonal. However, in practice we found this not to be necessary.

In a final iterative phase, we include all corresponded object points and optimize the projector poses, object points, remove outliers, and repeat until convergence. Since so far we have only operated on a subset of the corresponded object points, we bring the rest of the corresponded object points up to speed by keeping the projector parameters fixed and using a linear least squares optimization to update the 3D coordinates of all object points. Using the estimated initial values, we use sparse nonlinear bundle adjustment to refine both projector pose parameters and object points. Outliers are removed using image- and world-space criteria. This optimization and culling process is repeated until convergence.

## 3.4 Up-sampling Geometric Data

To obtain a more densely sampled geometric surface, our approach warps denser photometrically-computed points to the coarser geometric model. Since we have per-pixel correspondences between the photometric and geometric points (i.e., they are both viewed by the same projector), we use a piecewise linear mapping to warp points from the photometric-surface to the geometric-surface. In particular, we first compute a 2D triangulation of all geometrically-calibrated points and then the barycentric coordinates ($\alpha_i$, $\beta_i$, $\gamma_i$) of all photometrically-computed points within this triangulation. To warp a photometrically-computed point to the geometric surface, we use its barycentric coordinates to compute a displacement using the corresponding displacement vectors of the surrounding triangle. The new geometric-surface point $p_{G_i}$ corresponding to photometric surface point $p_{P_i}$ is computed by

$$p_{G_i} = p_{P_i} + \alpha_i(a_{G_i} - a_{P_i}) + \beta_i(b_{G_i} - b_{P_i}) + \gamma_i(c_{G_i} - c_{P_i}) \quad (4)$$

where ($a_{G_i}$, $b_{G_i}$, $c_{G_i}$) are the vertices of the triangle of geometrically-calibrated points surrounding $p_{P_i}$, and ($a_{P_i}$, $b_{P_i}$, $c_{P_i}$) are the corresponding photometrically-calibrated points. The pixel observation of an up-sampled point in

the projector view is estimated by linearly interpolating the pixel observations of the surrounding triangle vertices.

## 4. Photogeometric Optimization

Our optimization process alters the object points so as to best match both photometric and geometric measurements. We search for a displacement of the current object point estimates that simultaneously minimizes reprojection error onto the projectors (i.e., geometric error) and reduces the difference between photometrically-computed surface normals and surface normals computed using neighboring object points (i.e., photometric error).

We seek to efficiently optimize a large number of object points. Figure 2f depicts the optimization setup. A 3D point $p_i$ is observed by two projectors $V_0$ and $V_1$. Given at least two views, its location is determined by minimizing the pixel reprojection error with the observations $q_{0i}$ and $q_{1i}$ by the projector. Figures 2d and 2e show in the projector's (virtual) image plane the location of the point $p_i$. Simultaneously, point $p_i$ is observed by the camera in Figure 2b. Given at least three differently illuminated projector views $L_0$, $L_1$, and $L_2$ of the object, point $p_i$'s desired surface normal can also be computed (Figure 2c). Since the projector $V_j$ and the point-light $L_j$ are one and the same, a system of linear equations can reduce the geometric point error. To improve photometric agreement, we ensure the tangent plane of all object points is perpendicular to the normals $\bar{s}_i$. The equations are

$$\sum_j \sum_i \left\| \begin{matrix} \hat{p}_{ij_x} - (u_{ij}\hat{p}_{ij_z}/f) \\ \hat{p}_{ij_y} - (v_{ij}\hat{p}_{ij_z}/f) \end{matrix} \right\|^2 \to 0 \quad (5)$$

$$\sum_i \sum_k \delta_{ik}(\bar{s}_i \cdot t_{ik})^2 \to 0 \quad (6)$$

where $\hat{p}_{ij} = M_j p_i$, $u_{ij}$ and $v_{ij}$ are the projections of $p_i$ onto image $j$, $t_{ik}$ are vectors in the tangent plane of point $p_i$, $\delta_{ik}$ is 1 when $p_k$ is a neighbor of $p_i$ and 0 otherwise, and the unknowns are the 3D coordinates of each $p_i$. Equation (5) represents the geometric term of the optimization and it is easy to compute. Equation (6) requires estimating tangents and that process is described in Section 4.2.

## 4.1 Changing to Point-Light Sources

Prior to optimization, the surface normals computed via the initial photometric processing (Section 3.2.1) are updated to reflect the newly improved projector positions. As opposed to directional lights, we observed improved quality by assuming point lights. A directional light effectively requires the light to be distant relative to the size of the model. However, for structured-light acquisition using a standard projector, the projectors (and

cameras) are placed as close as possible to the object in order to maximize point correspondence robustness and density and to improve sampling resolution. Thus, each surface normal $\bar{s}_i$ is recomputed using $(L_j\text{-}p_i)$ as the light vector, where $L_j = M_j^{-1}[0\ 0\ 0]^{\mathrm{T}}$ is the world light position. For simplicity, we recomputed the normals for a Lambertian surface. This can be easily replaced with another photometric method. For example, the methods in [6][11] can be used without any new capture.

## 4.2 Estimating Tangent Vectors

We estimate tangent vectors for our point cloud corresponded over several views using a local-reprojection assumption. In general, there are various ways that tangent vectors can be estimated from point-sampled data. For example, Nehab et al. [12] define two methods for estimating positions and normals. For the single view case, an efficient option is to organize points in raster order on the image plane and parameterize them by their uncertainty along the ray through the center-of-projection. Another option is to assume the points are connected via a tessellated 3D mesh and a normal map is provided externally. This information can be used to estimate tangent vectors. Since no views are available, they endeavor to prevent self-intersections (or "flipping") by restricting points to move along the normals mapped to each point. In contrast, our multi-view observations implicitly provide the control to help keep the points near their correct place. Thus, we do not need to restrict movement to be along the measured normals and are able to accommodate larger adjustments of the surface. Moreover, we use the measured normals to locally re-project the object points and can avoid assuming a tessellated 3D mesh is provided as input.

Our method estimates a set of tangent vectors using the collection of neighbors surrounding each point $p_i$. The photometrically-computed normal $s_i$ and the point $p_i$ define the desired tangent plane of the object point. We create a local Delaunay Triangulation using the set of neighboring object points (i.e., $H_i=\{p_k \mid k$ is a neighbor of $i\}$) projected onto the aforementioned tangent plane. Neighboring points outside a radius or too distant from the tangent plane are rejected. Using the points of $H_i$, we define non-normalized tangent vectors as $t_{ik}=p_i\text{-}p_k$.

## 4.3 Linear Optimization

Equations (5) and (6) can be written as Ax=b and solved as an over-constrained sparse linear least squares problem. There are $3N$ unknowns corresponding to the coordinates of the $N$ object points. The geometric equations (5) yield $2MN$ equations for $M$ images and $N$ points. The photometric equations (6) produce additional $KN$ equations and no new unknowns, where $K$ is the

| Name | # Ps | # Cs | # Pts | Calib Err | Improv. |
|---|---|---|---|---|---|
| Beethov | 7 | 3 | 311564 | 0.978 | 34:1 |
| Bear | 4 | 2 | 214949 | 0.88 | 77:1 |
| Column | 4 | 2 | 272644 | 0.67 | 63:1 |
| Statue | 4 | 2 | 131756 | 0.9 | 91:1 |
| Vase | 4 | 1 | 47478 | 0.65 | 83:1 |

**Table 1. Datasets**. We show no. projector locations, no. cameras, no. points, self-calibration error (pixels), and error ratio of before to after photogeometric optimization.

average number of neighbors per object point (e.g., typically $K{\geq}3$). Thus as long as $N(2M+K){\geq}3N$, there are at least as many equations as unknowns. Given that there is at least one image and one neighbor, the solution is always over-constrained. Further, the units of equation (5) and (6) do not match, thus we use a user-defined scale to balance the errors of equations (5) to those of equations (6).

## 5. Results

We have captured several real-world objects using our approach and prototype implementation (see supplemental video file). Our system is implemented in C/C++, 3.2 GHz CPU, and 2GB memory. The camera used for acquisition is a Point Grey Flea camera capturing 1024x768 @ 30Hz color images. To project patterns, we use a Mitsubishi PK10 projector of 800x600 pixels (thus maximum number of pattern pairs $P$=9). To initialize self-calibration, we use the manufacturer's specified focal length estimate, a pixel size to obtain a reasonable world-scale, and a guessed camera-to-scene distance (1 meter in all our examples).

Table 1 presents a summary of our datasets. The statistics are shown for captures using a full-set of structured-light patterns. Self-calibration time ranges from an average of 15 seconds for a fast and coarse set of points (i.e., number of patterns $P$=5) to an average of 27 minutes when using all points (i.e., $P$=9). The up-sampling process only takes a few seconds. Photogeometric optimization averages 32 minutes (dominated by a naïve method to estimate geometric normals -- simple caching and tuning can vastly improve this) and reduces its error term by almost 100x (the numerical value is not reported because
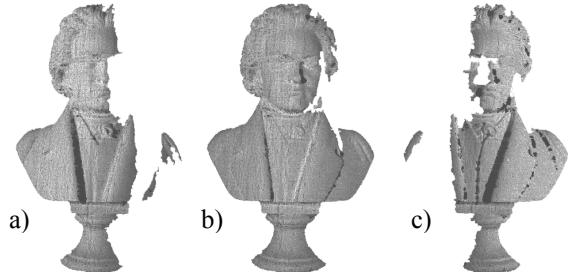


**Figure 3. Multi-view Acquisition.** We show object points visible from three exemplary (virtual) viewpoints. a) View from left; b) view from front; and c) view from right.
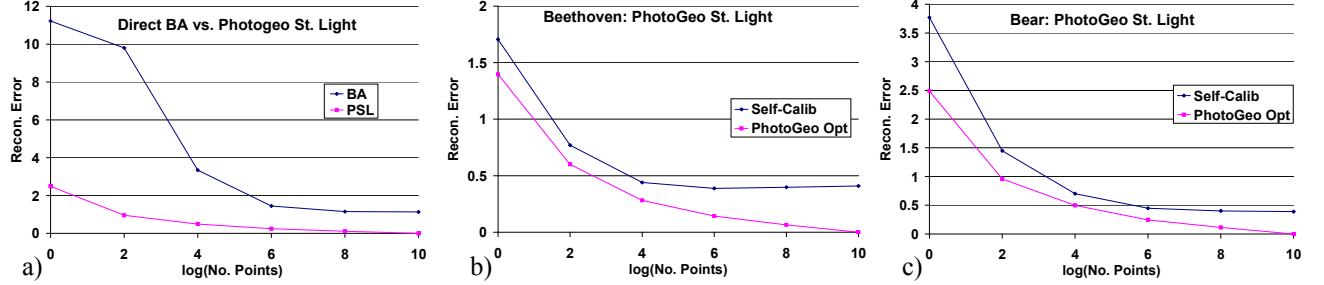
**Figure 4. Reconstruction Comparisons.** a) Comparing direct application of bundle adjustment to our full approach. b-c) Shows improvement after applying our self-calibrating method and after our optimization. Horiz. axis is approximate number of points resulting from using increasingly more structured-light patterns. Vertical axis is reconstruction error as percent of model diagonal.

it does not have an intuitive physical interpretation).

Our multi-view acquisition enables us to reconstruct object points that are at least in a user-selectable minimum number of images. Increasing the minimum-image threshold increases accuracy and robustness but usually decreases density of point samples. Figure 3 shows Beethoven and the subset of points visible from three of the (virtual) viewpoints -- the entire reconstruction is shown in Figure 1. From the initial set of all corresponded object points, at a minimum a point must be in two views so that it may be geometrically reconstructed.

Figure 4 shows the reconstruction errors at various stages of our pipeline and for various numbers of structured-light patterns. Figure 4a uses the Bear model to compare our method to the reconstruction error obtained by directly using bundle adjustment (BA) with the initial estimates provided via photometric processing. Our reconstruction is up to an order of magnitude more

accurate (8.57:1 on average), even when many correspondences are available. The original per-pixel reconstruction error of the photometrically-computed surface is 15% and 14% for Bear and Beethoven, respectively. Figures 4b-c each show the error after using our surface estimation method (Section 3.3) and after using photogeometric optimization (Section 4). The reconstruction error is always noticeably smaller than using BA directly. Moreover, even when using fewer patterns, and thus coarse geometry, photogeometric optimization can consistently further improve the result. Reconstruction error is indicated as the difference, in percent of the model diagonal, of the indicated method to our highest quality solution (i.e. every pixel is used in the self-calibration and in the photogeometric optimization).

Figure 5 contains images after photometric reconstruction, self-calibration, up-sampling, and photogeometric optimization. Noise in geometry-only
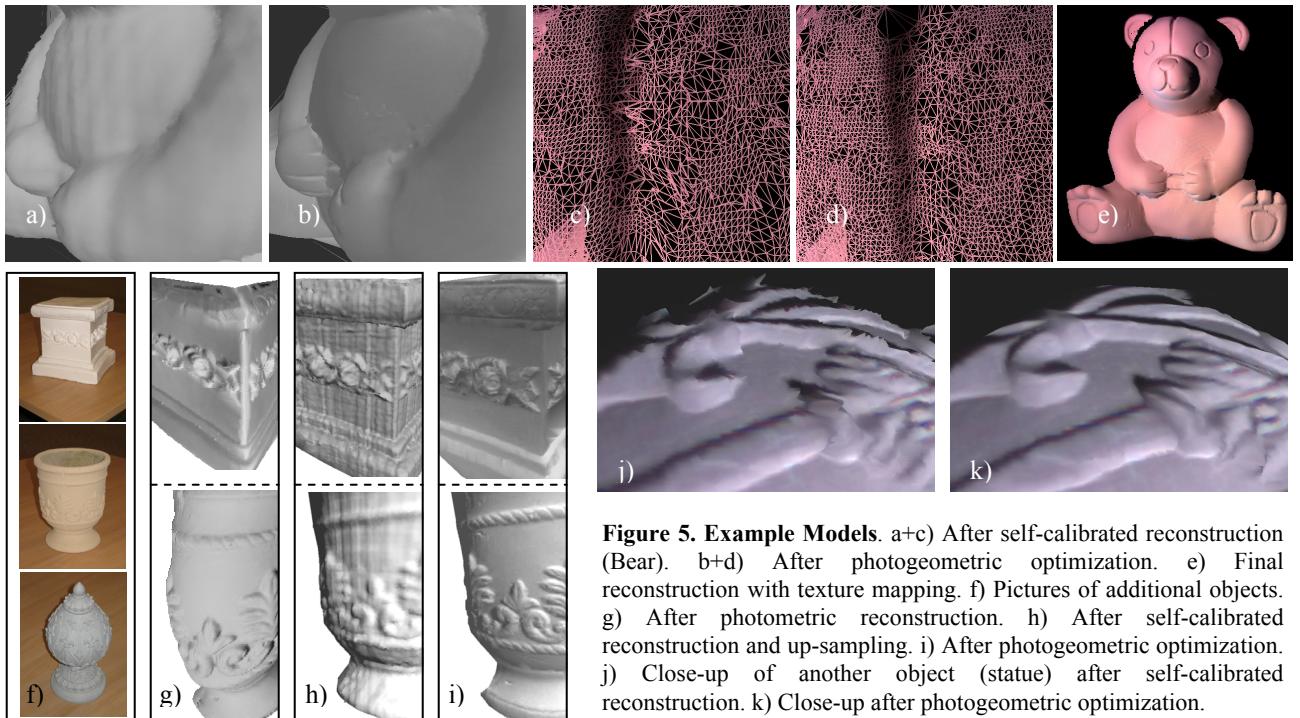




**Figure 5. Example Models**. a+c) After self-calibrated reconstruction (Bear). b+d) After photogeometric optimization. e) Final reconstruction with texture mapping. f) Pictures of additional objects. g) After photometric reconstruction. h) After self-calibrated reconstruction and up-sampling. i) After photogeometric optimization. j) Close-up of another object (statue) after self-calibrated reconstruction. k) Close-up after photogeometric optimization.

approximation is clear (Fig. 5a/c/h/j). After optimization, surfaces are better represented as observed using synthetic shading (Fig. 5b/d/i). Even when used for texture-mapping the improvement is large (Fig. 5e/k).

Regarding limitations, our current approach is restricted to capturing at most a hemisphere of the object. This is because for our current formulation for Lambertian surfaces there must be points that are visible in all views so as to produce a coherent global solution for the lights.

## 6. Conclusions and Future Work

We have presented our self-calibrating and multi-viewpoint 3D acquisition method, based on structured-light, which simultaneously obtains mutually registered surface position and normal data and produces a single high-quality model. Unlike previous methods, our automated processing is able to switch between a geometric setup and a photometric setup at will, using the same hardware and the same effective viewpoint. Self-calibration makes capturing easier since the devices do not have to be carefully and permanently located. Being multi-viewpoint enables more robustness and object coverage. We have also shown our results using several real-world objects and rendered them using either synthetic illumination or texture-mapping.

We are interested in several avenues of future work. First, we would like to incorporate more sophisticated photometric techniques that, for instance, handle specularities. Second, we would like to capture more than a hemisphere of the object by partitioning the viewing space into overlapping regions. Third, we believe significant more research will occur in photogeometric approaches. The resolution increase of digital cameras makes an approach such as ours very appealing.

## Acknowledgements

## References

[1] Basri R., Jacobs D., Kemelmacher I., Photometric Stereo with General Unknown Lighting, *IJCV*, 72(3), 239-257, 2007.

[2] Belhumeur P., Kriegman D., Yuille A., The Bas-Relief Ambiguity, *IJCV*, 35(1), 33-44, 1999.

[3] Chen B., Lensch H., Light Source Interpolation for Sparsely Sampled Reflectance Fields, *VMV*, 461-469, 2005.

[4] Frankot R, Chellappa R., A Method for Enforcing Integrability in Shape from Shading Algorithms, *IEEE Trans. on PAMI*, 10(4), 439-451, 1988.

[5] Furukawa R. and Kawasaki H., Uncalibrated Multiple Image Stereo System with Arbitrarily Movable Camera and Projector for Wide Range Scanning, *3DIM*, 302-309, 2005.

[6] Goldman D., Curless B., Hertzmann A., Seitz S., Shape and Spatially-Varying BRDFs from Photometric Stereo, *IEEE ICCV*, 341-348, 2005.

[7] Hemayed E, A Survey of Camera Self-Calibration, *IEEE Advanced Video Signal Based Surveillance*, 351-357, 2003.

[8] Lim J., Ho J., Yang M., Kriegman D., Passive Photometric Stereo from Motion, *IEEE ICCV*, 1635-1642, 2005.

[9] Lu J., Little J., Reflectance and Shape from Images using a Collinear Light Source, *IJCV*, 32(3), 213-240, 1999.

[10] Lu Y., Zhang J., Wu J., Li Z., A Survey of Motion-Parallax-Based 3-D Reconstruction Algorithms, *IEEE Trans. Systems, Man, & Cybernetics*, 34(4), 532-548, 2004.

[11] Mallick S., Zickler T., Kriegman D., Belhumeur P., Beyond Lambert: Reconstructing Specular Surfaces Using Color, *IEEE CVPR*, 619-626, 2005.

[12] Nehab D., Rusinkiewicz S., Davis J., Ramamoorthi R., Efficiently Combining Positions and Normals for Precise 3D Geometry, *ACM TOG* 24(3), 536-543, 2005.

[13] Pollefeys M., van Gool L., Vergauwen M., Verbiest F., Cornelis K., Tops J., Koch R., Visual Modeling with a Hand-held Camera, *IJCV*, 59(3), 207-232, 2004.

[14] Rushmeier H., Bernardini F., Computing Consistent Normals & Colors from Photometric Data, *3DIM*, 99-108, 1999.

[15] Salvi, J., Pages J., Batlle, J., Pattern Codification Strategies in Struct. Light Sys., *Patt. Recognition*, 37, 827-849, 2004.

[16] Scharstein, D. Szeliski, R., High-Accuracy Stereo Depth Maps Using Structured Light, *IEEE CVPR*, 195-202, 2003.

[17] Seitz S., Curless B., Diebel J., Scharstein D., Szeliski R. A Comparison and Evaluation of Multi-view Stereo Reconstruction Algorithms, *IEEE CVPR*, 519-528, 2006.

[18] Sen P., Chen B., Garg G., Marschner S., Horowitz M., Levoy M., Lensch H., Dual Photography, *ACM TOG* 24(3), 745-755, 2005.

[19] Sturm P., Critical Motion Sequences for the Self-calibration of Cameras and Stereo Systems with Variable Focal Length, *Image and Vision Computing*, 20(5-6), 415-426, 2002.

[20] Tan P., Mallick S., Quan L., Kriegman D., Zickler T., Isotropy, Reciprocity and the Generalized Bas-Relief Ambiguity, *IEEE CVPR*, 1-8, 2007.

[21] Woodham R., Iwahori Y., Barman R., Photometric Stereo: Lambertian Reflectance and Light Sources with Unknown Direction and Strength, UBC, TR-91-18, 1991.

[22] Zhang L., Curless B., Hertzmann A., Seitz S., Shape and Motion under Varying Illumination: Unifying Structure from Motion, Photometric Stereo, and Multi-view Stereo, *IEEE ICCV*, 618-625, 2003.

[23] Zhang L., Curless B., Seitz S., Spacetime Stereo: Shape Recovery for Dynamic Scenes, *IEEE CVPR*, 367-374, 2003.

[24] Zickler T., Belhumeur P., and Kriegman D., Helmholtz Stereopsis: Exploiting Reciprocity for Surface Reconstruction, *ECCV*, 869-884, 2002.

[25] Zickler T., Belhumeur P., Kriegman D., Toward a Stratification of Helmholtz Stereopsis, *IEEE CVPR*, 548-555, 2003.

[26] Zickler T., Reciprocal Image Features for Uncalibrated Helmholtz Stereopsis, *IEEE CVPR*, 1801-1808, 2006.