A Self-Calibrating Method for Photogeometric Acquisition of 3D Objects

Daniel G. Aliaga Yi Xu

Abstract— We present a self-calibrating photogeometric method using only off-the-shelf hardware that enables quickly and robustly obtaining multi-million point-sampled and colored models of real-world objects. Some previous efforts use a priori calibrated systems to separately acquire geometric and photometric information. Our key enabling observation is that a digital projector can be simultaneously used as either an active light source or as a virtual camera (as opposed to a digital camera which cannot be used for both). We present our self-calibrating and multi-viewpoint 3D acquisition method, based on structured-light, which simultaneously obtains mutually registered surface position and surface normal information and produces a single high-quality model. Acquisition processing freely alternates between using a geometric setup and using a photometric setup with the same hardware configuration. Further, our approach generates reconstructions at the resolution of the camera and not only the projector. We show the results of capturing several high-quality models of real-world objects.

Index Terms— Three Digitization and Image Capture, Scene Analysis, Geometric Modeling.

1 INTRODUCTION

WE present a new self-calibrating method for acquiring highly-detailed models for 3D objects. Our method uses off-the-shelf uncalibrated digital projectors and cameras and enables a computational trade-off from coarse and fast acquisitions to highly-detailed and optimized models. Our most-detailed models are multi-million point-sampled and colored representations of real-world objects with a sampling resolution of 0.1mm (using only consumer hardware). Our method simultaneously uses a photometric-based and a geometric-based approach in order to produce a *photogeometric* modeling system. This combination enables self-calibration, capitalizes on the high-visual details of photometric methods, and supports the precision of a geometric technique.

Our key enabling observation is that a digital projector can be simultaneously used as either an active light source or as a virtual camera. This fact permits acquiring per-pixel photometric and geometric observations of objects and from one or more viewpoints. Our photogeometric method differs from previous approaches in several ways. While Helmholtz reciprocity [28] and Dual Photography [21] methods have identified a projector as the dual to a camera, they do not explicitly obtain 3D models. Although the term photogeometric was first coined by [14], previous efforts have captured photometric and geometric observations using separate a priori calibrated systems (e.g., [16][19]). Our photometric and geometric systems are one and the same thus acquisition is free to change from one setup to the other.

Traditional geometry-only acquisition techniques are either passive and depend less robust feature correspondence or active. While our method builds upon structured light, typically such techniques obtain samples that are limited by projector resolution and are pre-calibrated. A typical projector has a resolution of one mega-pixel (MP) but digital cameras are easily available with an order of magnitude higher-resolution (e.g., 10 MP). Photometric-based approaches (e.g., shape-from-shading and photometric stereo) can use images of multiple MPs to obtain per-pixel surface normals, hence their effectiveness in high-quality image-based relighting. However, the recovered normals suffer from low-order distortions which make it difficult to recover the global shape (e.g., General Bas-Relief transformation [4]).

In contrast, our photogeometric method obtains both the high-resolution detail of photometric techniques and the metric accuracy of geometric approaches. Our algorithm enables smoothly changing from a photometric-only solution, to a photogeometric-solution, and then to a geometric-only solution. On the one hand, a photometric solution has the advantage of fast processing and high-resolution but bad global shape recovery. On the other hand, a geometric solution provides metric accuracy but often needs more compute time and is at lower-resolution than a photometric solution. Our photogeometric solution is effectively able to trade nonlinear geometric computations for faster linear computations and is able to incorporate the additional resolution and details of a photometric method. Furthermore, the combination supports performing self-calibration of the intrinsic and extrinsic parameters of the used hardware. Also, our methodology scales well with model size enabling quick processing of large multi-million point sampled models.

Our approach uses one of several configurations of digital projectors and cameras to obtain a large collection of multi-viewpoint samples of the object each with geometric and photometric observations. Our method uses as little as one projector and one camera. We demonstrate (1) using the projectors as light sources and then as virtual cameras to produce a fast geometric reconstruction whereby an uncalibrated photometric-based initialization obtains estimates for a self-calibration of the geometric setup, and (2) simultaneously using the projectors as both light sources and virtual cameras to calculate a photogeometric reconstruction using a linear optimization and producing very high-quality multi-viewpoint models. Our method produces highly-detailed models of 3D objects (ranging from 250k to 6M triangles) without any a priori calibration, yields both surface position and surface normal data, and uses only off-the-shelf hardware. The main contributions of our paper are:

- *D.G. Aliaga is with the Department of Computer Science at Purdue University, West Lafayette, IN 47907. E-mail: aliaga@cs.purdue.edu.*
- *Y. Xu is with the Department of Computer Science at Purdue University, West Lafayette, IN 47907. E-mail: xu43@cs.purdue.edu.*

Figure 1. Photogeometric Acquisition. a) Our method captures photometric (top) and geometric (bottom) observations of an object. b) A novel viewpoint rendering of a 3D model is produced using our photogeometric method. c) A close-up of the model rendered with texture mapping. d) An even closer view of the model now rendered using wireframe and synthetic shading showing details beyond those possible using only geometric observations. Our approach uses the same hardware as standard structured-light but is fully self-calibrating and able to capture models at the resolution of the camera (in this example camera resolution is 10x greater than projector resolution). The average triangle edge length is 0.09035 or about 0.1 mm.

- a photogeometric method for fast, robust, and detailed reconstruction of 3D objects,
- a method to use projectors as light sources and then as virtual cameras to obtain a fast and self-calibrating geometric reconstruction, and
- an algorithm for photogeometric optimization producing multi-viewpoint models best complying to both photometric and geometric measurements, without having to align various reconstructions.

Advantages and Limitations: Our method has several advantages over standard structured light systems. First, our method is self-calibrated and thus easier to deploy and use. Second, our method supports reconstructions at the resolution of the camera which often is greater than that of the projector (Figure 1). Third, photogeometric optimization improves the quality of the acquired models and provides surface normal data in addition to surface position data. One limitation of our method is that uncalibrated Lambertian photometric stereo is used to initialize the self-calibration. However, we do not rely on the surface to be Lambertian. As shown later, the computed lighting directions only need to be approximate estimates that will be snapped to the correct location during geometric processing.

2 Related Work

Geometric-based Methods: Geometric-based acquisition produces detailed models but the process is often time consuming and does not necessarily produce smooth and accurate normals. On the one hand, passive geometric methods are unobtrusive, but rely on natural features (e.g., [18]). On the other hand, active geometric methods explicitly generate correspondences, but laser-based systems acquire geometry (i.e., no color), can be expensive, and require lengthy cleanup and post-processing to obtain smooth, accurate surface normals and geometry. Further, most laser-based systems assume Lambertian objects, though [17] is an exception. Typical structured-light methods use a priori calibration and reconstruct mostly-Lambertian objects (e.g., [5][20][27]). Structured-light methods for dynamic scenes often obtain lower-resolution geometry and/or need custom hardware and calibration (e.g., [12][24]). Though some self-calibrating structured-light systems exist [7], most require pre-calibration. Self-calibration methods typically rely on features and on either assumed scene or geometry constraints to estimate parameters [9] and generally convergence is difficult [22].

Photometric-based Methods: As opposed to photometric stereo and shape-from-shading, our approach actively generates robust features and is able to overcome the typical low-frequency deformations. In particular, we surmount the ambiguity of the generalized bas-relief transformation [4] and, thus, obtain the 3D model up to a global scale factor. To improve upon the linear ambiguity, some previous methods rely on feature tracking and on a structure-from-motion refinement (e.g., [13][26]). While this may reduce global deformations, it relies on feature correspondence and cannot produce densely-sampled geometric data points. Another self-calibrating photometric option builds closed models of an object placed on a turntable and sequentially illuminated by three distant lights [10]. Many images of the object at constant rotational steps are captured. Silhouettes are extracted and used to form a 3D visual hull, which is perturbed to find matching photometric normals and mesh normals. This approach depends on silhouette detection, initial estimates for object orientation, and distant light sources (i.e., lights 3-4 meters away for a 15 cm object).

Photogeometric Methods: While some previous approaches have acquired photometric and geometric data, each from a different viewpoint and using two pre-calibrated acquisition setups (e.g., [16][19]), we improve such systems in several ways. Our method is both self-calibrating and multi-viewpoint. This makes the setup significantly more flexible, practical, and able to obtain more complete 3D models. We use the exact same equipment for both geometric and photometric acquisition. The single capture system also has the benefit of removing the rotational alignment (as in [16]) or positional-and-rotational alignment (as in [19]) needed between photometric and geometric samples and does not need co-location of the hardware via beam splitters.

Camera-Light Duality: Duality between cameras and light sources has been exploited for novel image generation and for surface reconstruction. For instance, a stereo method can use a projector as a virtual camera in the matching error metric formulation (e.g., [5]). Dual Photography [21] demonstrates how the view from a light source (e.g., projector) can be obtained by transposing the light transport matrix from projector to camera; however, no object geometry is obtained. Helmholtz stereopsis [28] 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 reflec-

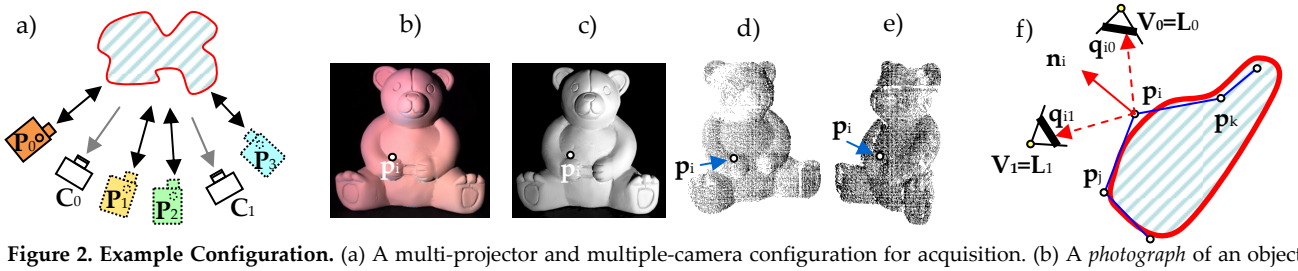

Figure 2. Example Configuration. (a) A multi-projector and multiple-camera configuration for acquisition. (b) A *photograph* of an object to be captured. (c) Photometric reconstruction using 3 images illuminated from 3 directions. (d-e) Two corresponded virtual views between two projectors and the camera picture in b. (f) Photogeometric setup surrounding point p_i.

tance distribution function (BRDF). While the initial Helmholtz stereopsis method requires pre-calibration, it has been extended to an uncalibrated approach requiring known epipolar geometry [29] and to an uncalibrated approach using reciprocal image features [30]. 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.

The work presented in this article is an extension of our conference publication [1]. However, we generalize the approach to a photogeometric method supporting several camera-projector configurations, introduce new equations to force a sample spacing similar to the original, and present additional results and analysis about the time-quality tradeoff performed with our photogeometric approach and comparisons to standard structured light.

3 Photogeometric Acquisition

For photogeometric acquisition, we use one of three fundamental configurations with $C \geq 1$ cameras and $R \geq 1$ projectors to obtain object point samples $S = \{S_i\}$ where $i \in [1..N]$ and N is desired to be large. The configurations are a generalization of structured-light systems and support multi-viewpoint reconstructions. They all provide self-calibration and increased resolution to almost any structured-light system. An object point sample is defined as $S_i = (p_i, n_i)$, where p_i is the position and n_i is the normal of the sample. Each sample is computed from a set $G_i = \{q_{ij} \mid i \in [1..N], j \in [1..R]\}$ of at least two geometric observations needed for geometric/classical stereo and a set $P_i = \{c_{ij} \mid i \in [1..N], j \in [1..R]\}$ of at least three photometric observations needed for photometric stereo, where $q_{ij} = (u_{ij}, v_{ij})$ is the projection of sample i onto the image plane of projector j, and $c_{ij} = (c_{ij}^r, c_{ij}^g, c_{ij}^b)$ is the RGB color of point i under the illumination of projector j.

3.1 Configurations

Our configurations hinge on the notion that a digital projector can be either a virtual camera or a digitally controlled light -- a digital camera, however, cannot be both. This leads to the following three possible configurations.

Single-camera Configuration (SC): SC has one static camera and one projector moved to several locations. A static camera-projector configuration provides the two viewpoints needed for a geometric reconstruction, but does not provide the multiple light sources for a photometric reconstruction. Moving a single projector to several distinct locations increases both the number of viewpoints and the number of light sources. This configuration is the simplest for obtaining the minimum needed geometric and photometric observations per S_i. One disadvantage of this configuration is that as the number of projector locations increases, the number of object points visible from the camera and all projectors decreases.

Multi-camera Configuration (*MC*): MC has several static cameras and one projector moved to several locations. In this configuration, object points have correspondences in multiple cameras, and collectively sample a larger portion of the object as compared to single camera configuration.

Multi-projector Configuration (*MP*): MP has several identical projectors and one or more static cameras. This approach assumes the same focal length for all projectors and avoids moving the projector. It is suitable for real-time processing but requires more hardware. Figure 2(a) shows a multiple camera configuration. It can also be realized using one single projector.

3.2 Image Capture

For each projector, a sequence of $A + B$ patterns are projected where $A \geq 1$ patterns are used for photometric observations P_i and $B \geq 1$ patterns are used for geometric observations G_i. Changing the number of patterns enables different time-quality trade-offs. In the limit, using $B = 0$ yields a purely photometric-based capture while using $A = 0$ produces a purely geometric-based capture.

The image data for the photometric observations P_i consists of the color intensity of the samples S_i as lit by the projectors (Figure 3a). For the first phase of uncalibrated photometric stereo, we assume a Lambertian object illuminated by each projector acting as a diffuse light source. For the second phase, after geometric self-calibration, the light sources are known and it facilitates one of several improved photometric reconstructions.

The image data for geometric observations G_i is captured using Gray code patterns [11], which are a set of coarse to fine multi-level binary patterns (Figures 3d-e). Our method projects Q pairs of (horizontal and vertical) stripe patterns from each of the projectors and captures images for all cameras. To prevent determining surface albedos, we project the patterns and their inverses; resulting in $4Q$ patterns in total per projector. These patterns permit robustly corresponding about $(2^Q - 1)^2$ surface points between a camera and a projector. For a projector with 1400x1050 pixels, the maximum Q is 10; resulting in $\sim 10^6$ samples. To yield fewer points and faster processing, we can use small Q values (e.g., 5). Other structured light patterns can also be used for point correspondence (e.g. grid pattern, color coded pattern, etc.). We choose to use temporally-coded Gray code patterns due to its robustness against different surface albedos and its ability to achieve high resolution. In addition, changing the levels of Gray code pattern can be used to control the density of

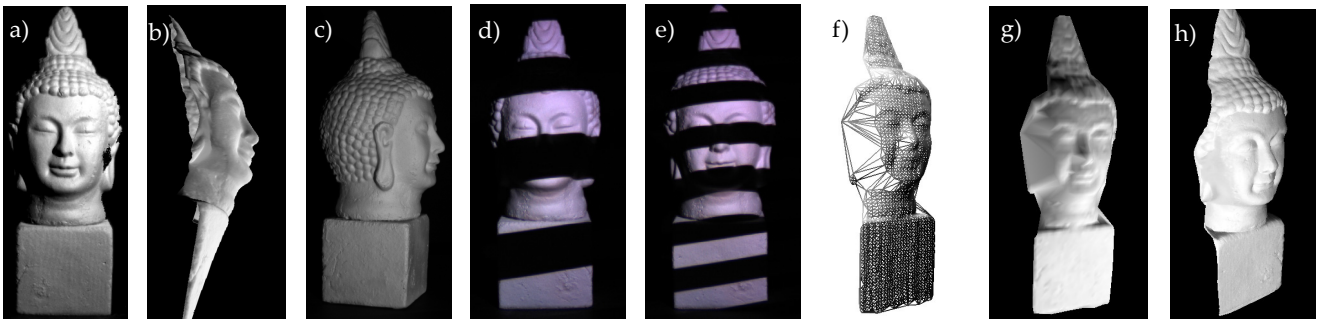

Figure 3. Processing. a) Illuminated by projector. b) Photometric reconstruction. c) Side-view *photograph* (compare to b). d-e) Structured light input images. (f-g) Wireframe and filled view of coarse self-calibrated model. (f) Final model after up-sampling and photogeometric optimization (rendered using synthetic lighting).

geometric measurements easily. When multiple cameras are available, space-time stereo is also an alternative to obtain higher resolution correspondence with a small number of images [5][27]. Please refer to [3] for a detailed discussion on different coding strategies.

3.3 Object Point Sampling

In order to convert the projectors into virtual cameras, we need to correspond pixels amongst as many projectors (and cameras) as possible. Examining the Gray code sequence seen by the cameras and emitted by the projectors produces the camera-to-projector correspondence. We then re-sample the camera-to-projector correspondences to produce dense projector-to-projector correspondences.

After capture, each camera has observed all projector patterns and has a set of point correspondences with the projectors. A camera and a projector each creates a 2D triangulation of the points they have in common with each other. The camera then computes the barycentric coordinates for a regular grid of 2D points on the camera's image plane and uses the barycentric coordinates to compute corresponding 2D points on the projector's image plane. A newly created point in the projector's view is then corresponded with points on all the other cameras' regular grid provided the projector triangle is visible in the other camera. 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. Points 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).

The projector-projector correspondences become geometric observations G_i and the projector-camera correspondence are used to extract photometric observations P_i.

4 Photometric Modeling

Our method uses photometric modeling to first perform an uncalibrated photometric stereo reconstruction of the object and then again later during the process of multi-view photogeometric optimization. In the first stage, we use a Lambertian photometric stereo formulation and its results of lighting directions and surface estimates are fed to the next stage of geometric modeling.

4.1 Uncalibrated Photometric Stereo

Uncalibrated photometric stereo recovers surface normals and lighting directions. The well-known model for Lambertian objects and distant directional lights is

$$NL^T = C \quad (1)$$

where N is a k x 3 matrix of k outward-facing surface normals, L is a l x 3 matrix of l light directions pointing towards the light, and C is a k x l matrix of the observed pixel intensities. Based on [2] and [25], we look closely at the case of one pixel/normal and three lights, namely k=1 and l=3. Matrix N becomes a row vector normal n^T and the matrix C becomes a row vector c^T. Equation (1) can be written as $n^TL^T=c^T$. Next, we define a new matrix $D=(L^T)^{-1}$ and rewrite equation (1) as $n^T=c^TD$. Since we desire unit surface normals, we seek n^Tn=1 which can also be written as $(c^TD)(D^Tc)=1$, or as

$$c^TEc = 1 \quad (2)$$

where $E=DD^T$ is symmetric and positive definite consisting of six unknowns. When expanded, it takes the form

$$e_{11}c_1^2 + e_{22}c_2^2 + e_{33}c_3^2 + 2e_{12}c_1c_2 + 2e_{13}c_1c_3 + 2e_{23}c_2c_3 = 1. \quad (3)$$

The equation writes the six unknowns of E as a linear combination of the known components c_i of the vector c. Given at least six pixels and three intensities per pixel, the e_{ij} variables are over-constrained and solved using linear least squares. Given E, observe that

$$E^{-1} = (DD^T)^{-1} = ((L^T)^{-1})(((L^T)^{-1})^T)^{-1} = LL^T. \quad (4)$$

From equation (4) we recover the magnitude of and angles between light directions, up to an unknown global rotation. Then, using L^T we compute the matrix N.

A surface height field can be calculated from N up to an unknown global rotation R and to a generalized bas-relief (GBR) transform G [4]. The GBR transform (λ, μ, ν) of a surface is the 3x3 matrix $G = [1\ 0\ 0; 0\ 1\ 0; \mu\ \nu\ \lambda]$. Since typically the same projector is used at multiple locations, we assume equal light source intensity and simplify the GBR transform to $(\lambda, 0, 0)$ [4]. Using one arbitrary object, we manually estimate a λ, which transforms the photometric surface to an shape similar to the real object. Registration does not need to be very precise and λ=0.3 works well with all our objects. We also determine a 3D rotation to bring the lighting setup into registration with the xy image plane using a simple user interface.

4.2 Surface Estimation

To compute a photometrically-estimated surface $z(x,y)$, we integrate the surface normals. The geometric normals of $z(x,y)$ are given by (z_x, z_y, -1) where z_x and z_y denote partial derivatives. On a grid, the partial derivatives are

$$z_x = z(x+1, y) - z(x, y) = -n_x/n_z$$
$$z_y = z(x, y+1) - z(x, y) = -n_y/n_z \quad (5)$$

where (n_x, n_y, n_z) is the photometrically-estimated normal at pixel (x, y). Then, we integrate and construct a surface height field using the method of [6]. Although depth discontinuities cause problems when integrating normals, our method does not assume the integrability of the surface. Our geometric self-calibration algorithm is able to handle certain amount of self-occlusions using a rough photometric estimate as shown later.

5 Geometric Modeling

Our geometric modeling uses the photometrically-estimated surface, approximate lighting directions, and tailored reprojection equations to obtain a self-calibrated reconstruction of a subset $S' = \{S_i'\}$ of the object point samples S_i. The size of S' affects the time and quality of the resulting reconstruction. Our approach estimates both the focal length and pose of the projectors, acting as virtual cameras. We do not correct for radial distortion and thus assume long focal lengths and/or high-quality lenses. For all configurations, the projectors will be necessarily fully calibrated but the physical cameras are only optionally calibrated. Geometric modeling seeks to minimize reprojection error expressed by the well-known nonlinear system of equations

$$\sum_j \sum_i \left(\frac{1}{h_{ij_z}} \begin{bmatrix} h_{ij_x} \\ h_{ij_y} \end{bmatrix} - \begin{bmatrix} u_{ij} \\ v_{ij} \end{bmatrix} \right)^2 \text{ where } h_{ij} = F_j(R_j p_i + T_j) \quad (6)$$

and R_j, T_j, and F_j are the unknown 3x3 rotation matrix, 3D translation vector, and 3x3 perspective projection matrix.

5.1 Geometric Initialization

During initialization, a sparse and uniformly-distributed subset of object point samples of S' are used to estimate the distance from each projector to the object's center as well as the global projector focal length f. Initial values for p_i come from the photometrically-estimated surface fragment. The calculated lighting (or virtual viewing) directions l_j and an assumed up-vector of w=[0 1 0]^T define an initial oriented orthogonal coordinate system for each projector, $M_j = [l_j \times w \quad (l_j \times w) \times l_j \quad -l_j]$. The free parameters are the focal length f and distances z_j from the origin to each projector j along l_j. To bring the reprojection of the object points into rough alignment with the observed projections (u_{ij}, v_{ij}), we optimize the following simplified nonlinear system of equations of only R+1 unknowns (f and z_j for $j \in [1, R]$) and where $\hat{p}_{ij} = M_j p_i$:

$$\sum_j \sum_i \left(\frac{\hat{p}_{ij_x} f}{\hat{p}_{ij_z} + z_j} - u_{ij} \right)^2 + \left(\frac{\hat{p}_{ij_y} f}{\hat{p}_{ij_z} + z_j} - v_{ij} \right)^2. \quad (7)$$

5.2 Geometric Reconstruction

Next, our method optimizes for a linear correction to each projector location and performs a global bundle adjustment. In particular, in equation (6) F_j is replaced by a perspective projection matrix parameterized by f, T_j is replaced by [0 0 z_j]^T and R_j is replaced by $Q_j M_j$. Each matrix Q_j is computed using the following linear system of equations in the 8 unknowns of the matrix (i.e., q_{33}=1)

$$\sum_i \begin{bmatrix} q_{ij} - u_{ij}(q_{ij} + z_j)/f \\ q_{ij} - v_{ij}(q_{ij} + z_j)/f \end{bmatrix} \text{ where } \hat{q}_{ij} = Q_j M_j p_i \quad (8)$$

Using an iterative process we then include all the remaining object point samples of S' and optimize the projector poses, object points, and remove outliers. First, we fix projector pose parameters and use a sparse bundle-adjustment optimization of all object points in equation (7). Second, we use sparse nonlinear bundle adjustment to refine both projector pose and all object points. Image-space and world-space point culling criteria are also applied. The culling criteria are related to the inter-sample distance in the image plane and in world space. Thus, the same criterion is applied to similar configurations (e.g. similar object size, camera/projector to object distances, and resolutions). The optimization and culling repeats until convergence; M_j is updated to the final pose matrix.

6 Photogeometric Processing

Finally, we combine photometric and multi-view geometric data in a single linear optimization. Our approach enables a time-quality tradeoff whereby a variable amount of geometric modeling is performed and an approximation of the missing details is obtained from the faster photometric processing. Effectively, the traditional nonlinear modeling of multi-million point samples is converted into a fast and specialized nonlinear optimization of a small set of points followed by a linear upsampling and linear multi-view optimization of all points.

6.1 Upsampling

First, we increase the sampling density of the object point samples to that of the camera. While the relative pose of the photometrically-reconstructed model with respect to the geometrically-reconstructed model is unknown, the known image-space correspondence defines a piecewise linear mapping; both observation types are from the same viewpoint (e.g., a projector). Thus points from the photometric-surface can be warped to the geometric-surface.

To perform the up-sampling, our method computes 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, the barycentric coordinates and the corresponding vectors formed by pairing the vertices of the photometric-surface triangle with the corresponding vertices of the geometric-surface triangle are used to compute a displacement. A new geometric-surface point p_{G_i} corresponding to photometric-surface point p_{P_i}, is computed by

$$p_{G_i} = p_{P_i} + \alpha_i(a_{G_i} - a_{P_i}) + \beta_i(b_{G_i} - b_{P_i}) + \gamma_i(c_{G_i} - c_{P_i}) \quad (9)$$

where $(a_{P_i}, b_{P_i}, c_{P_i})$ is the triangle of photometrically-reconstructed points surrounding p_{P_i}, and $(a_{G_i}, b_{G_i}, c_{G_i})$ are the corresponding geometrically-calibrated points. The pixel observation of the up-sampled point is determined by interpolating the triangle's pixel observations.

6.2 Multi-View Optimization

Our optimization alters object points so as to best match both photometric and geometric measurements. We search for a displacement that follows the properties of

(i) minimizing reprojection error onto the projectors,
(ii) keeping a similar relative position of points, and
(iii) reducing the difference between photometrically- and geometrically-computed normals.

An important aspect of our multi-view optimization method is to prevent undesired "flipping" and "self-intersection" of the mesh of object points. One option is to

keep each object point restricted to lie on the corresponding projector ray emanating from the center-of-projection and passing through its observation on the image plane (e.g., as in [16]). However, such an approach is not suitable for multi-viewpoint processing. Only letting the point move along one or more projector rays does not support freely compensating for inaccuracies in the estimated pose (and focal length). In our method, each object point moves freely in 3D space and is able to accommodate larger corrections to the surface. The equations that attempt to ensure property (i) keep the points near the correct geometric location. The equations for property (ii) assist in yielding a distribution of points over the object's surface that is similar to the original one. Finally, property (iii) guides the upsampled points to an arrangement resembling the additionally captured photometric detail and provides a smoothly reconstructed surface.

6.2.1 Improved Photometric Method

The surface normals computed via the initial photometric processing are now updated to better represent the configuration. We upgrade to point lights which more accurately imitate the true setup because projectors are kept relatively close to the objects. Surface normal n_i is computed as

$$n_i = \frac{L_i^{-1}\left[c_{ij_1} \cdots c_{ij_R}\right]^T}{\left\|L_i^{-1}\left[c_{ij_1} \cdots c_{ij_R}\right]^T\right\|} \text{ where } L_i = \begin{bmatrix} [M_{j_1}^{-1}[0\ 0\ 0\ ]^T - p_i] \\ \cdots \\ [M_{j_R}^{-1}[0\ 0\ 0\ ]^T - p_i] \end{bmatrix}. \quad (10)$$

Other non-Lambertian photometric methods can also be used with our approach to produce a better photometric solution for non-Lambertian surfaces (e.g., [8][15]).

6.2.2 Formulation

The linear equation that satisfies the aforementioned triple of properties and that we wish to minimize is

$$e_t = (1-\alpha)(1-\beta)\kappa_g e_g + \beta\kappa_r e_r + \alpha\kappa_p e_p \rightarrow 0 \quad (11)$$

$$e_g = \sum_j \sum_i \begin{bmatrix} \hat{p}_{ij_x} - \left(\frac{u_{ij}\hat{p}_{ij_z}}{f}\right) \\ \hat{p}_{ij_y} - \left(\frac{v_{ij}\hat{p}_{ij_z}}{f}\right) \end{bmatrix} \quad (12)$$

$$e_r = \sum_i \delta_{ik}((p_i - p_k) - d_{ik}) \quad (13)$$

$$e_p = \sum_i \delta_{ik}(n_i \cdot (p_i - p_k)) \quad (14)$$

where d_{ik} is the initial distance between point p_i and p_k, δ_{ik} is 1 when p_k is considered a neighbor of p_i and 0 otherwise, and the unknowns are the 3D coordinates of each p_i. Equations (12-14) correspond to properties (i-iii) respectively. To determine the neighbors of a point p_i and to define δ_{ik}, we create a local Delaunay Triangulation using a set of neighboring object points projected onto the tangent plane. Equation (11) can be written as Ax=b and solved using over-constrained sparse linear least squares. Given K equal to the average number of neighbors per object point, $N(2R+K) \geq 3N$ ensures number of equations is larger than or equal to the number of unknowns.

To control the tradeoff between geometric error, photometric error, and relative distance error, we scale the individual error terms to the range [0, 1] (using κ_g, κ_p, and κ_r) and optimize a weighted linear combination of them. The errors are combined using α and β; e.g., a low value for α implies low photometric importance and high geometric importance; a low value for β implies lack of importance of keeping the same relative distance between points. A solution must balance the three properties.

Name	Configuration	# Ps	# C	# Pts	Error
Beethoven	MP	7	3	311564	0.978
Bear	MC	4	2	214949	0.88
Statue	MC	4	2	131756	0.9
Buddha	SC	3	1	204678	0.75
Green Pot	SC	3	1	303467	0.55
Textured Pot	SC	3	1	3037143	0.27
Ornament	SC	3	1	124749	1.2
3 Objects	SC	3	1	354021	0.95

Table 1. Datasets. We show the configuration type (see Section 3.1), number of projector views, number of camera views, number of reconstructed points, and final self-calibration error (pixels).

7 Results

We have implemented a prototype of our method in C/C++, use either 1400x1050 Optoma EP910 or 800x600 Mitsubishi Handheld PK10 projectors, and a 3888x2592 Canon Digital Rebel XTi camera or a 1024x768 Point Grey Research Flea. Table 1 presents a summary of our test datasets. Total processing time for any model is less than 15 minutes: self-calibration takes a few minutes, upsampling needs a few seconds, and photogeometric optimization requires from 1 to 10 minutes (e.g., for "Pot"). All final calibration errors are less than one pixel (except the shiny *Ornament*) indicating good self-calibration.

Figure 4 demonstrates how our method captures the *Bear* object (a) by combining a photometrically-computed surface (b) with that of a geometrically-computed surface (c) to yield a single higher quality surface (d) which captures the same details as in (a). Figure 4(e) shows the effectiveness of our method compared to a direct bundle adjustment (BA) solution. BA uses the photometrically-reconstructed surface as initial guess and performs non-linear optimization using Equation (6) directly. In Figure 4(e), we vary the number of geometrically-calibrated

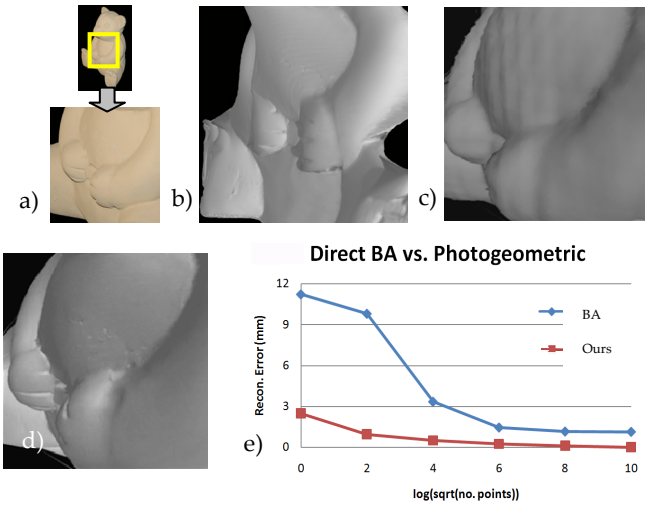

Figure 4. Photogeometric Reconstruction. a) *Photograph* and close-up of original object. b) Photometrically-computed surface (with obvious global deformations). c) Geometrically-computed surface (globally accurate but missing fine details). d) Our photogeometrically-computed surface. e) Comparison: our method vs. bundle adjustment.

points used in a photogeometric reconstruction by changing the level of Gray code patterns. The total number of reconstructed points remains the same (equal to the number of camera pixels). With the initial photometric surface having a reconstruction error of 15%, our approach reconstructs a surface with almost up to an order of magnitude less error (8.5:1 on average) as compared to BA. Error is shown as the difference, in percent of the model diagonal, of the indicated method to our best solution (i.e. all pixels used in self-calibration/photogeometric optimization).

Figure 5 shows the benefit of the photogeometric optimization using the *Textured Pot* object. The reconstruction of the object is shown in Figure 1. For graph (a), we vary the photogeometric parameter α in order to yield a best combination of geometric and photometric observations (β is fixed at 0.1). For graph (b), we compare a standard structured light scheme to our approach of an optimized combination of sparse geometric and dense photometric observations. To fairly perform this comparison, we reduce the sampling resolution to approximately yield one object point sample per projector pixel (for our hardware, this implies up to one million points). In the graph, we vary the number of projector pixels that are reconstructed geometrically, upsample the remaining projector pixels using photometric data, and perform the photogeometric optimization; a value of k in the horizontal axis of the graph impies $2^k x 2^k$ geometric points. In (c), we show a color-coded difference image between the standard structured-light reconstruction and our reconstruction at $k = 7$, or 64 photometric points for every geometric point. Our photogeometric processing always improves the upsampled surface to within 0.006mm, on average, of the full-resolution structured-light surface. For our hardware and this setup, a projector pixel has a size of 0.5mm. Thus, our solution of coarse geometry and photometrically-upsampled and optimized details is at least as accurate as a full-resolution structured-light solution. We extrapolate similar quality results occur when full-resolution structured-light is used with full-resolution photometric observations and photogeometric optimization. Figure 5(d) shows the reconstructed object with significant displacement of the vertices but still a small reconstruction error ($\beta = 0.01$). $\beta = 0.1$ produces an equally accurate solution but vertices positioned similar to original locations.

Figures 6 demonstrates the performance of our algorithm on non-Lambertian objects: a plastic *Green Pot* and a shiny porcelain *Ornament*. Although Lambertian photometric stereo can only produce an approximate solution (especially for *Ornament*) in initialization, our geometric self-calibration still snaps the light directions to the correct direction. In these cases, our system is similar to standard structured-light without the benefit of photometric data. Thus, strong highlights cause artifacts and the reconstructed model is noisy due to the lack of the optimization step as shown in Figure 6(c). Figure 7 contains additional results and close-ups of models. Figure 7(k-l) shows a 3 *object*s scene captured using our system. Depth discontinuity causes problems for integrating surface using photometric data, but not in our geometric self-calibration.

8 Conclusions and Future Work

We have presented our self-calibrating and multi-viewpoint 3D photogeometric acquisition method. Our technique uses the same hardware setup to obtain photometric and geometric observations and to perform a single acquisition and modeling effort. Our approach successfully extracts from the photometric data an increased amount of surface detail and extracts the metric accuracy from the geometric data in order to compensate for the global deformations typical of pure photometric methods. Our work can be applied to one of several configurations, including standard structured light, but provides additional quality and avoids the need for a priori calibration.

Regarding limitations, our current formulation for photometric reconstruction is restricted to capturing at most a hemisphere of the object. Also, the number of object points that are reconstructed is less than those of a typical stereo system. This is because the points must be visible from the viewpoint of the camera and at least three projector viewpoints. Thus, a multi-viewpoint configuration is needed more often than expected.

For future work, we would like to incorporate more sophisticated photometric methods (e.g. subpixel precision methods [23] and non-Lambertian photometric stereo [8][15]), explore adaptive geometric and photometric schemes, and extend our method to dynamic objects.

Acknowledgments

This work supported in part by NSF CCF 0434398.

References

[1] Aliaga D., Xu Y., Photogeometric Structured Light: A Self-Calibrating and Multi-viewpoint Framework for Accurate 3D Modeling, *IEEE CVPR*, 2008.

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

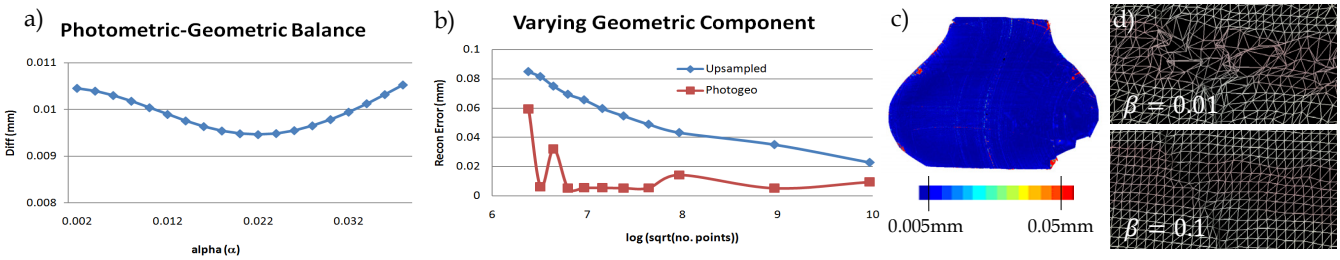

Figure 5. Photogeometric Optimization. a) Close-up of smallest reconstruction errors resulting from varying the photogeometric balance α. b) Improvement achieved by photogeometric optimization using different amounts of geometrically-computed points. c) Color-coded difference between a photogeometrically-computed surface and an ideal obtained by a standard structured light approach. d) View of the distribution of the reconstructed points of same reconstruction accuracy but using β=0.01 (top) and β=0.1 (bottom).

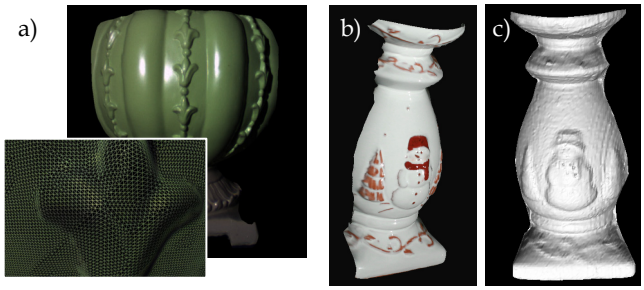

Figure 6. Non-Lambertian Objects. a) Reconstruction of a specular plastic object (with a close-up wireframe view in the inset). b) Side view of the texture mapped reconstruction of a shiny porcelain *Ornament*. c) A synthetically illuminated *Ornament*.

[3] Batlle J., Mouaddib E., and Salvi J., Recent Progress in Coded Structured Light as a Technique to Solve the Correspondence Problem: A Survey, *Pattern Recognition*, 31, 7, 963-982, 1998.

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

[5] Davis J., Nehab D., Ramamoorthi R., Rusinkiewicz S., Spacetime Stereo: A Unifying Frame-work for Depth from Triangulation, *IEEE TPAMI* 27(2), 296-302, 2005.

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

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

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

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

[10] Hernandez C., Vogiatzis G., Cipolla R., Multi-view Photometric Stereo, *IEEE PAMI*, 30(3), 2008.

[11] Inokuchi S., Sato K., and Matsuda F., Range Imaging System for 3-D Object Recognition, *ICPR*, 806-808, 1984.

[12] Koninckx T., Griesser A., van Gool L., Real-time Range Scanning of Deformable Surfaces by Adaptively Coded Structured Light, *3DIM,* 293-300, 2003.

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

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

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

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

[17] Park J., Kak A., 3D Modeling of Optically Challenging Objects, *IEEE TVCG*, 14(2), 246-262, 2008.

[18] 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.

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

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

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

[22] 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.

[23] Tan P., Lin S., Quan L., Subpixel Photometric Stereo, *IEEE PAMI,* 30(8), 2008.

[24] Weise T., Leibe B., van Gool L., Fast 3D Scanning with Automatic Motion Compensation, *IEEE CVPR,* 2007.

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

[26] 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.

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

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

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

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

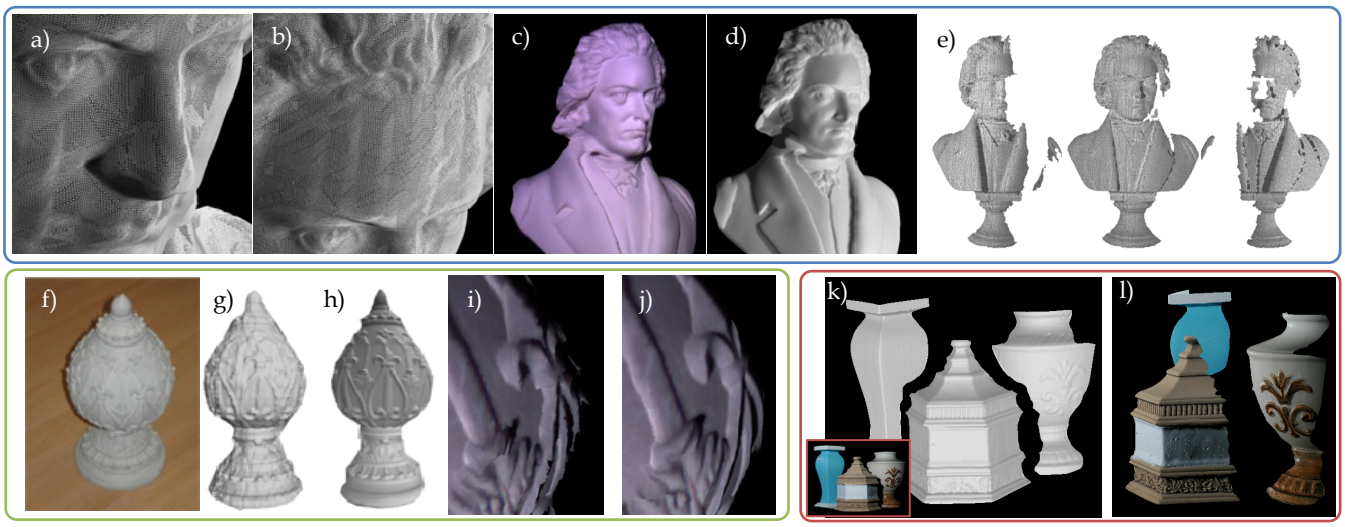

Figure 7. Examples. *Beethoven*: a-b) synthetically illuminated wireframe close-ups, c) texture-mapped rendering of full model, and d) synthetically illuminated model with lighting direction very different than the physical one. e) Visualization of the subset of points for each of three viewpoints. *Statue*: f) *photograph* of objects, g) rendering after self-calibration and upsampling, h) rendering after photogeometric optimization, i) close-up before photogeometric optimization, and j) close-up after photogeometric optimization. *3 Objects*: k) synthetically illuminated models with a photograph of the scenes from the capturing viewpoint in the inset. l) Texture mapped models from a very different novel viewpoint.