



Surround structured lighting: 3-D scanning with orthographic illumination

Douglas Lanman ^{*}, Daniel Crispell, Gabriel Taubin

Brown University, Division of Engineering, Providence, RI 02906, USA

ARTICLE INFO

Article history:

Received 1 May 2008

Accepted 27 March 2009

Available online 11 April 2009

Keywords:

3-D reconstruction

Structured lighting

Gray codes

Orthographic projection

Full object scanning

ABSTRACT

This paper presents a new system for rapidly acquiring complete 3-D surface models using a single orthographic structured light projector, a pair of planar mirrors, and one or more synchronized cameras. Using the mirrors, we project structured light patterns that illuminate the object from all sides (not just the side of the projector) and are able to observe the object from several vantage points simultaneously. This system requires that projected planes of light to be parallel, so we construct an orthographic projector using a Fresnel lens and a commercial DLP projector. A single Gray code sequence is used to encode a set of vertically-spaced light planes within the scanning volume, and five views of the illuminated object are obtained from a single image of the planar mirrors located behind it. From each real and virtual camera we recover a dense 3-D point cloud spanning the entire object surface using traditional structured light algorithms. A key benefit of this design is to ensure that each point on the object surface can be assigned an unambiguous Gray code sequence, despite the possibility of being illuminated from multiple directions. In addition to presenting a prototype implementation, we also develop a complete set of mechanical alignment and calibration procedures for utilizing orthographic projectors in computer vision applications. As we demonstrate, the proposed system overcomes a major hurdle to achieving full 360° reconstructions using a single structured light sequence by eliminating the need for merging multiple scans or multiplexing several projectors.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Recovering the 3-D shape of a physical object is a fundamental challenge in computer vision. Historically, research has focused on two broad strategies: active vs. passive methods. Passive methods attempt to reconstruct a model of the object by recording how ambient light interacts with the scene. Typical passive approaches include shape-from-silhouette algorithms (e.g., the visual hull [1]) and space-carving [2]. These methods more generally fall under the wider umbrella of multi-view stereo [3], wherein the shape of an object is recovered by recording a sparse set of images from multiple viewpoints. In general, passive methods require robust feature-matching algorithms in order to establish corresponding points between views—an ongoing topic of research within the field [3]. Active methods, in contrast, utilize controlled illumination in order to solve the correspondence problem directly. Typical active systems include laser-stripe scanners, structured lighting, and time-of-flight range scanners [4]. In this work we focus our attention on structured light designs, since they continue to be one of the lowest-cost active systems that can be easily built using off-the-shelf components.

Although there exists a wide variety of structured light 3-D scanning systems, previously none were able to capture a full 360° object scan using a single camera position. As shown in Fig. 1, typical structured light systems are composed of a single camera/projector pair that is used to illuminate the scene using a set of temporally-coded images. Unfortunately, such implementations can only reconstruct a small portion of the object's surface—that which is both illuminated and imaged. As a result, obtaining complete object models requires multiple scans, often achieved by placing the object on a turntable, or multiplexing multiple cameras and projectors. Afterwards, post-processing alignment of the scans must be performed using registration algorithms such as Iterated Closest Point (ICP) [5]. As a result, single camera/projector structured lighting systems cannot be used for real-time acquisition of dynamic scenes.

In this paper, we present a novel modification of a traditional single camera/projector structured light system that allows full 360° surface reconstructions, without requiring turntables or multiple scans. As shown in Fig. 2, the basic concept is to illuminate the object from all directions with a structured pattern consisting of horizontal planes of light, while imaging the object from multiple views using a single camera and mirrors. A key benefit of this design is to ensure that each point on the object surface can be assigned an unambiguous Gray code sequence, despite the possibility of being illuminated from multiple directions. We discuss our implementation of the system and how we plan to lever-

* Corresponding author.

E-mail addresses: dlanman@brown.edu, dlanman@gmail.com (D. Lanman), daniel_crispell@brown.edu (D. Crispell), taubin@brown.edu (G. Taubin).

URL: <http://mesh.brown.edu/dlanman> (D. Lanman).

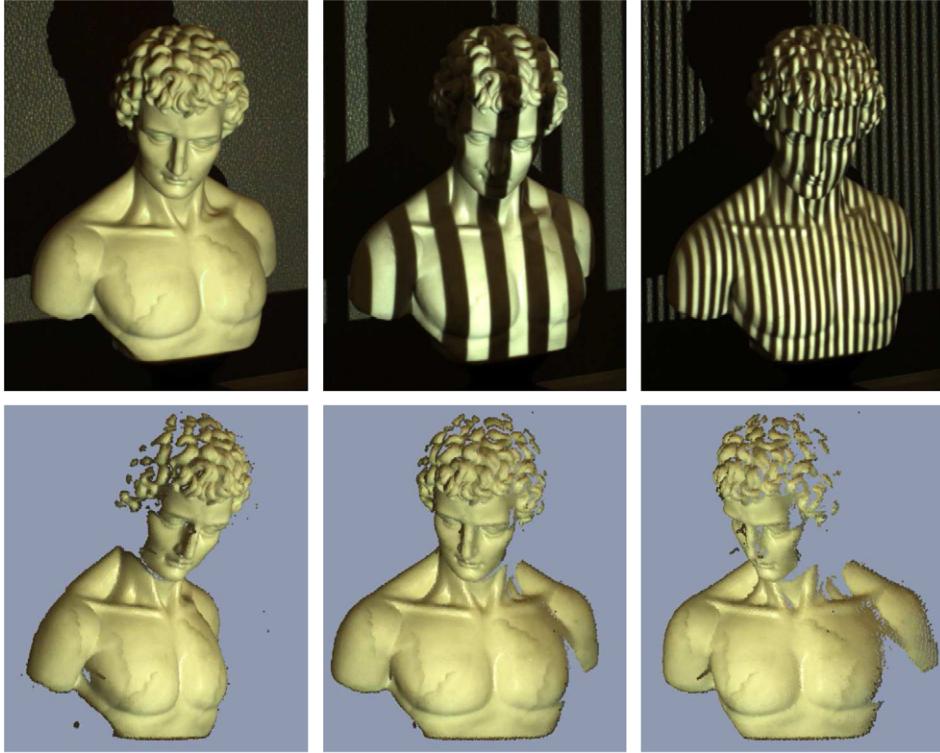


Fig. 1. Traditional structured lighting surface reconstruction using a single camera/projector pair. (Top row, from left to right) Image of a sculpture with all projector pixels illuminating the scene, the scene as illuminated by the sixth Gray code, and the scene illuminated by the eighth Gray code. (Bottom row) Three views of the reconstructed surface, displayed as point-based renderings.

age the design into a full 360° real-time scanner, ideal for motion capture applications. The paper is laid out as follows: In Section 2 we discuss related scanning systems and other works preceding our system. In Section 3 we describe the general system design and construction, and the calibration procedures in Section 4. In Section 5 we present a structured light reconstruction algorithm for our design. In Section 6 we present both qualitative reconstruction results, as well as an analysis of the reconstruction accuracy. Finally, in Sections 7 and 8 we discuss the limitations of our design and our future plans.

1.1. Contributions

We present a set of hardware modifications to existing structured light systems to allow low-cost, rapid acquisition of complete 3-D surface models. Specific technical contributions are as follows:

- i. We propose a new combination of a structured light projector and a pair of planar mirrors to create a surround structured lighting system that allows near real-time capture of the 3-D surface of general objects.
- ii. We analyze the benefits and limitations of the surround structured lighting system by constructing an initial prototype using off-the-shelf components. We present complete calibration and reconstruction procedures required for successful operation, and document both the qualitative and quantitative reconstruction performance for expected applications.
- iii. We propose a set of practical methods for constructing an orthographic illumination system using a digital projector and a Fresnel lens. These methods include a novel procedure for aligning a projector, with known intrinsic calibration parameters, to a given Fresnel lens using a printed pattern affixed to the lens surface.

- iv. We describe a simple method for estimating the position and pose of a planar mirror with respect to a camera using one or more photographs containing a direct and reflected image of a calibration pattern.

2. Related work

This paper draws on three areas of active research within the computer vision and graphics communities: (1) structured lighting for 3-D surface reconstruction, (2) multi-view systems which exploit one or more planar mirrors to obtain virtual views of an object, and (3) recent work on orthographic projectors by Nayar et al. [6]. As described in the survey article by Salvi et al. [7], coded structured light is a reliable, inexpensive method for recovering 3-D surfaces. In its simplest form, structured light reconstruction requires a single calibrated projector-camera pair. By illuminating the surface with a known sequence of coded images, the correspondence between projector and camera pixels can be uniquely identified. In this paper, we use the classic binary Gray code sequence originally proposed by Inokuchi et al. [8] (see Fig. 1), in which each pattern is composed of a sequence of black and white stripes oriented along the horizontal image axis of the projector (i.e., the projector rows). By illuminating the object with a temporally-multiplexed sequence of increasingly-fine Gray code patterns, the corresponding projector row can be identified for each camera pixel. Each pattern simply encodes one bit plane of the Gray code for each projector row index [8]—significantly accelerating the correspondence process when compared to traditional laser striping [4]. Afterwards, the appropriate ray-plane intersection can be used to recover a 3-D point on the surface.

The idea of using planar mirrors to create virtual structured light projectors was first presented by Epstein et al. [9]. In their system, one or more planar mirrors are illuminated by a projector

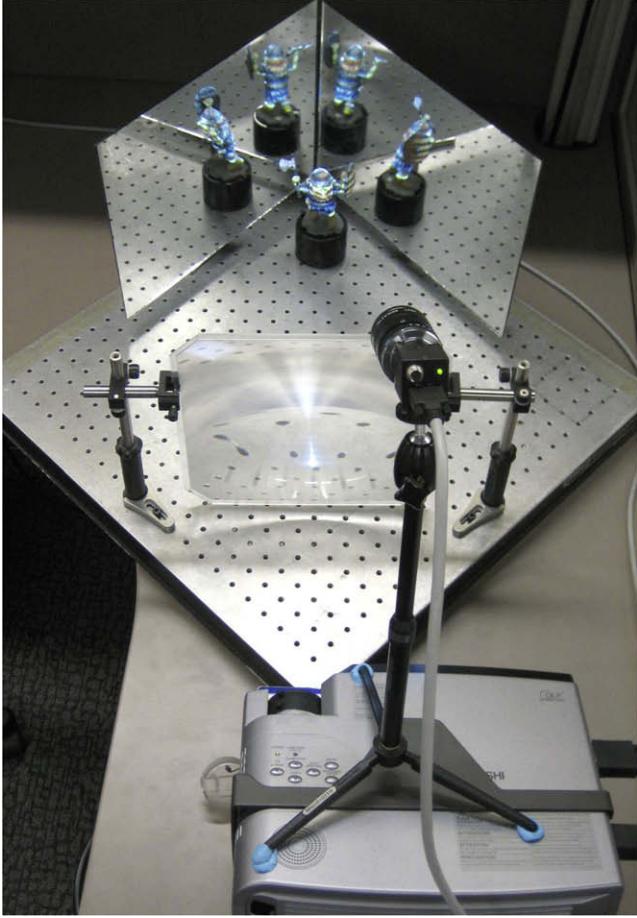


Fig. 2. Surround structured lighting system architecture: DLP projector, Fresnel lens, planar mirrors, and a digital camera.

displaying a modified Gray code sequence which is invariant to mirror reflections. By visually tracking the relative camera, projector, and mirror positions and by interactively selecting a conservative object bounding box, the authors mask the projected patterns to ensure that each surface point is illuminated from a single direction in each image. While eliminating the need for multiplexing several projectors to obtain complete object models, this system still suffers from several limitations. Foremost, it increases the number of required patterns since the directly and indirectly viewed portions of the object surface cannot be illuminated simultaneously. Using our system, however, a single pass of a conventional Gray code sequence can be used to recover the full object surface.

The concept of using planar mirrors to obtain virtual viewpoints has recently been explored by several authors [10–13]. As discussed by Gluckman and Nayar [11], the virtual cameras created by planar mirror reflections have several benefits over multiple camera systems, including automatic frame synchronization, color calibration, and identical intrinsic calibration parameters. While Gluckman and Nayar restricted their attention to stereo catadioptric systems in that work, Forbes et al. [10] have explored precisely the planar mirror configuration used in this paper: a pair of planar mirrors oriented such that an object placed between the two mirrors will produce one real and four virtual viewpoints, resulting from the first and second reflections (see Fig. 12). In their original work, the authors obtained a complete 3-D surface model by estimating the visual hull [1] defined by the five object silhouettes. Viewed in the context of these systems, our approach can be seen

as complementary. Rather than relying on conventional stereo matching or background segmentation algorithms, we employ a structured light sequence to uniquely identify the correspondence between projector rows and camera pixels.

Finally, we note that Nayar and Anand [6] previously presented a similar orthographic projection system using a DLP projector and a Fresnel lens. In their application, the orthographic projector illuminated passive optical scatterers to create a volumetric display. To our knowledge, their system is the first demonstration of orthographic projectors within the computer vision and graphics community.

3. System design and construction

Traditional structured light projectors, for example those using Gray code sequences, cannot be used to simultaneously illuminate an object from all sides due to interference. If such a configuration was used, then there is a high probability that certain points would be illuminated by multiple projectors. In such circumstances, multiple Gray codes would interfere, resulting in erroneous reconstruction due to decoding errors. In this paper we propose a novel optical design that resolves this fundamental limitation. Rather than using multiple projectors (each with a single center of projection), we propose using a single *orthographic* projector and a pair of planar mirrors. In the following section we describe how such a system can be applied to prevent interference of individual Gray codes and to obtain full 360° object reconstructions using a single projector–camera pair.

The key components of the proposed scanning system are an orthographic projector, two planar mirrors aligned such that their normal vectors are contained within the plane of light created by each projector row, and a single high-resolution digital camera. If any structured light pattern consisting of horizontal binary stripes is implemented, then the object can be fully illuminated on all sides due to direct and reflected projected light (see Fig. 3). As shown in Fig. 2, if the camera's field of view contains the object and both mirrors, then it will record five views of the illuminated object: one direct view, two first reflections, and two second reflections [10]. By carefully aligning the mirrors so that individual projector rows are always reflected back upon themselves, we can ensure that only a single Gray code sequence will be assigned to each projector row—ensuring that each vertically-space plane in the reconstruction volume receives a unique code. The full struc-

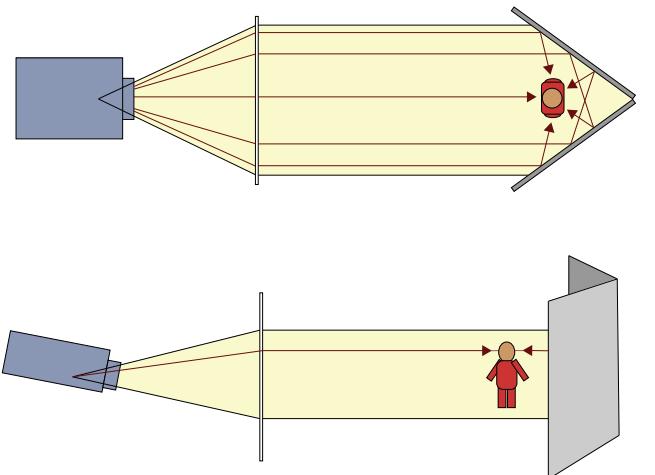


Fig. 3. Diagram showing the orthographic projection system and multiple rays from a single projector row illuminating the object from different angles. The top and bottom diagrams show top and side views of the system, respectively.

tured light pattern combined with the five views provides sufficient information for a nearly complete surface reconstruction from a single camera position.

The required orthographic projector can be implemented using a standard off-the-shelf DLP projector and a Fresnel lens, similar to that used by Nayar and Anand [6] for their volumetric display. The Fresnel lens converts light rays diverging from the focal point to parallel rays and can be manufactured in large sizes, while remaining lightweight and inexpensive. In our implementation, we use a Mitsubishi XD300U (1024×768 resolution) projector and a 10.5 in. square Fresnel lens (No. 54 from Fresnel Technologies, Inc. [14]) with 200 grooves per inch and a focal length of 24.0 in., placed such that the focal point is located at the projector's center of projection. Although we model the projector as a pinhole projector, it has a finite aperture lens, and therefore a finite depth of field in practice. This makes conversion to a perfectly-orthographic set of rays impossible with the Fresnel lens, but an acceptable approximation is still feasible, as we demonstrate in our discussion of the calibration (Section 4). We focus the projector at its maximum distance in order to provide the sharpest images in the reconstruction volume, and adjust the field of projection to its largest angle in order to maximize this volume. The quality of the Fresnel lens also limits the accuracy of the orthographic approximation, with a higher "grooves per inch" value generally indicating a lens of better quality (i.e., a closer approximation to a spherical lens). While our initial prototype used low-cost second-surface mirrors obtained from a hardware store, we found that these mirrors added refractive effects to the system due to the protective layer of glass covering the reflective surface. In order to avoid this complication, our current implementation uses 16 in. square optical grade first-surface mirrors [15]. The mirrors are positioned such that their surface normals are roughly 72° apart (as per Forbes et al. [10]) in order to provide the five equally-spaced views, and such that their surface normals are perpendicular to the surface normals of the projected light planes. The mirrors are mounted on gimbals with fine tuning knobs in order to facilitate precise positioning.

To complete our prototype implementation, we used a single 1600×1200 resolution digital camera (Model GRAS-20S4C-C from Point Grey Research, Inc. [16]) with a 12.5 mm focal length lens. The final assembled prototype is shown from the side in Fig. 4. Typical images acquired by our imaging system, while using orthographic illumination, are shown in Fig. 5.

4. Calibration

Because of the unique design of the scanning system, calibration of the multiple components is a non-trivial task. We note that experienced individuals have required up to several hours to perform all necessary steps, with the majority of the time spent refining the mechanical alignment. Our calibration procedure is divided into three stages: (1) configuration of the orthographic projector, (2) alignment of the planar mirrors, and (3) calibration of the camera/mirror system. We explain, in a similar progression, the details of each of these stages in the following three subsections.

4.1. Projector calibration and alignment

The reconstruction algorithm relies heavily on the existence of a projection system that produces parallel (or close to parallel) light rays. Here we describe how we achieve the mechanical alignment and calibration necessary for such a system in practice.

The first step requires estimating the mapping from projector pixel coordinates to projected rays in 3-D. To date, a wide variety of approaches for general projector calibration have been proposed [17–21], some of which have been specifically tailored to structured light configurations [22,21]. For our system, we follow the well-established method of utilizing a calibrated camera to subsequently determine the intrinsic and extrinsic projector calibration [18]. We first estimate the intrinsic camera parameters (i.e., focal length, principal point, and skew coefficient), as well as a fourth-order radial lens distortion model, using the *Camera Calibration Toolbox for Matlab* [23]. Typical camera calibration results are shown in Fig. 6. Afterwards, as shown in Fig. 7, we project a fixed checkerboard pattern and observe its image in a set of white planes at various orientations throughout the scene. From a set of known fiducials attached to the planes, we recover the 3-D position of each projected checkerboard corner. Given the set of correspondences between projector pixels and 3-D points, we then use a nonlinear optimization procedure to estimate the intrinsic calibration of the projector, its position within the camera coordinate system, as well as a fourth-order radial lens distortion model. (Note that the principal point of the projector lens is located far from the image center, due to the typical floor or ceiling placement of DLP projectors in home theater environments. As a result, the projector must be tilted downward, as shown in Fig. 4.)

The second step involves the correct placement of the DLP projector with respect to the Fresnel lens as described in Section 3.



Fig. 4. Surround structured lighting prototype viewed from the side. Note the key components, from left to right: the first-surface planar mirrors mounted on fine-tuning gimbals, the Fresnel lens, a high-resolution digital camera, and a DLP projector mounted on a geared tripod head.

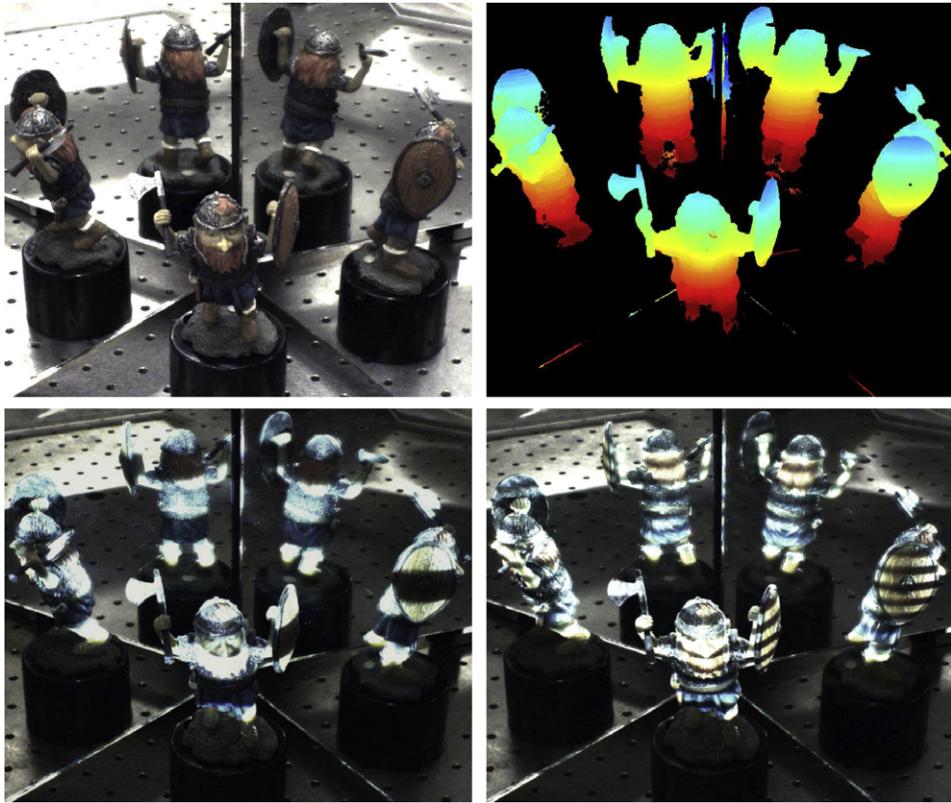


Fig. 5. Example of an orthographic Gray code pattern and recovered projector rows. (Top-left) Scene, as viewed under ambient illumination, for use in texture mapping. (Top-right) Per-pixel projector rows indices recovered by decoding the projected Gray code sequence (shaded by increasing index, from red to blue). (Bottom-left) Fourth projected Gray code. (Bottom-right) Sixth projected Gray code. (For interpretation of the references to colours in this figure legend, the reader is referred to the web version of this paper.)

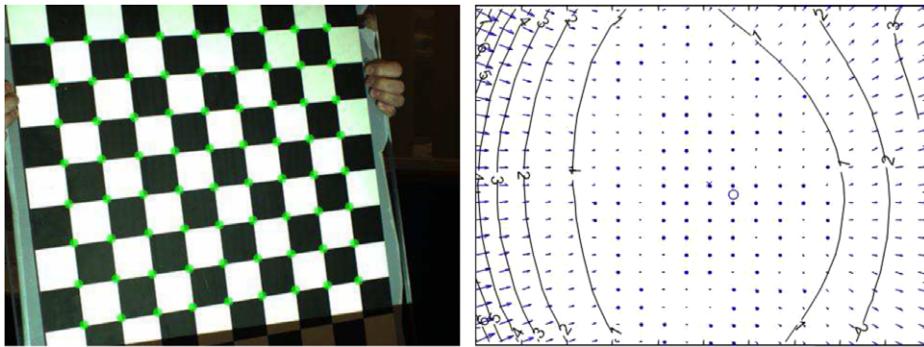


Fig. 6. Estimating the intrinsic parameters of the camera. (Left) Calibration image collected using a printed checkerboard. A least-squares procedure is used to simultaneously optimize the intrinsic and extrinsic camera parameters in order to minimize the difference between the predicted and known positions of the checkerboard corners (denoted as green circles). (Right) The resulting fourth-order radial lens distortion model for the camera, where isocontours denote the displacement (in pixels) between an ideal pinhole camera image and that collected with the actual lens. (For interpretation of the references to colours in this figure legend, the reader is referred to the web version of this paper.)

Using the projector calibration and the focal length of the Fresnel lens (provided by the manufacturer), we are able to predict the image of a projected calibration pattern as it should appear in the Fresnel lens plane, assuming that the projector is in the correct position with its center of projection located at the focus of the lens. We create a printed version of the desired projection pattern and affix it to the lens surface, aligning a marked point on the pattern to the lens center. The exact center of the lens is visibly apparent as the center of the concentric ridges on the Fresnel surface. We then project the original pattern and fine-tune the projector's position and orientation until the patterns are aligned on the lens surface (see Fig. 8). While theoretically providing a perfect alignment, in practice some difficulty arises due to the finite depth of focus of

the projector. Since the projector is generally tuned to be in focus in the scanning volume, the projected calibration pattern will typically be out of focus on the Fresnel lens surface.

The final stage of projector calibration involves mapping rows of projected images (e.g., using Gray codes) to planes in 3-D (as defined in the camera coordinate system). If the illumination system is close to orthographic, these planes should all be approximately parallel. Here we present two alternative methods for estimating the plane coefficients for each projector row.

4.1.1. Orthographic projector calibration using planar homographies

In our first method to recover an estimate of the projected orthographic light planes, we place a calibration board into the

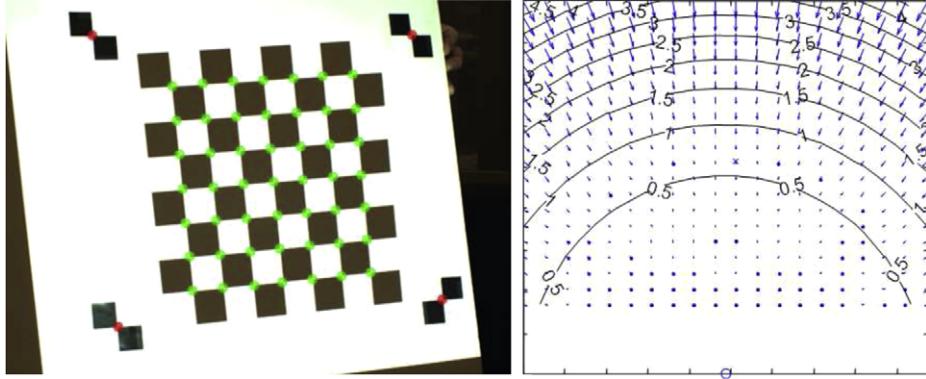


Fig. 7. Estimating the intrinsic parameters of the projector using a calibrated camera. (Left) Calibration image collected using a white plane marked with four fiducials in the corners (denoted as red circles). The position of the plane, in the camera coordinate system, is determined using the fiducials and the camera calibration; afterwards, the intrinsic parameters of the projector are recovered by projecting a checkerboard on the blank region of the calibration plane. Considering the projector as an “inverse camera”, a least-squares procedure is used to optimize the projector calibration in order to minimize the difference between the predicted and known position of the projected checkerboard corners (denoted as green circles). (Right) The resulting fourth-order lens distortion model for the projector. (For interpretation of the references to colours in this figure legend, the reader is referred to the web version of this paper.)

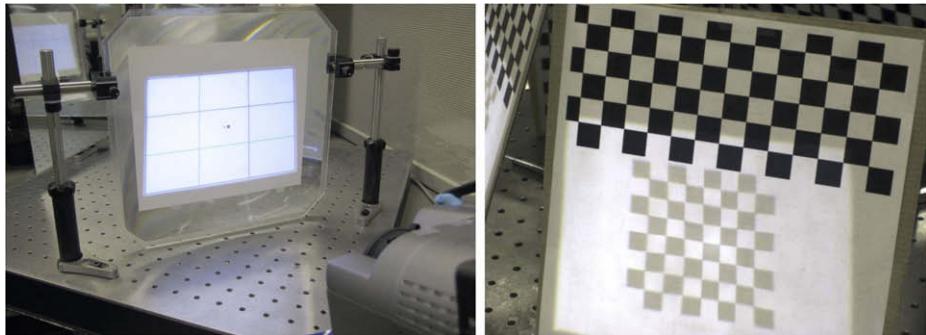


Fig. 8. Orthographic projector alignment and calibration. (Left) Aligning the projector with the Fresnel lens using a printed pattern derived from the projector’s intrinsic calibration. (Right) Estimating the plane coordinates for each orthographic projector row using printed and projected checkerboards and planar homographies.

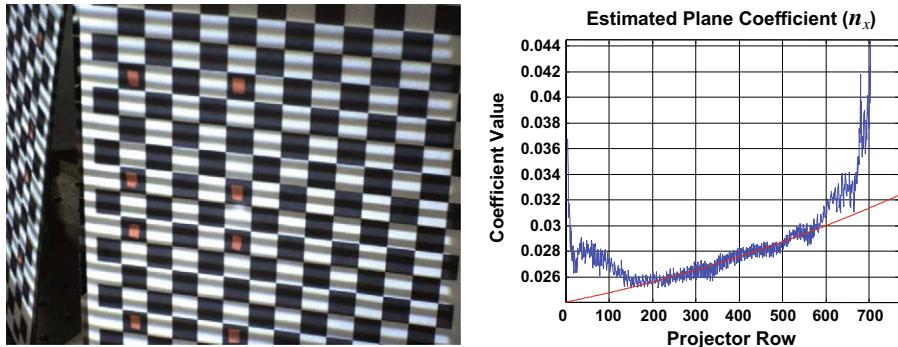


Fig. 9. Orthographic projector calibration using structured light patterns. (Left) A planar surface with a printed checkerboard pattern is placed within the reconstruction volume. The correspondence between camera pixels and projected planes is established using Gray codes, yielding a labeled 3-D point cloud. Afterwards, a plane is fit independently to each set of 3-D points corresponding to a given orthographic projector row. (Right) The plane coefficients are filtered and extrapolated using a least-squares quadratic polynomial fit to each parameter.

reconstruction volume that contains a printed checkerboard pattern on the upper half, and blank whitespace on the lower half and project a known checkerboard pattern onto the blank half (see Fig. 8). Using the fixed and projected checkerboard patterns, we can recover the planar homographies which map image coordinates to the world and projector coordinate systems, respectively. Using the estimated planar homographies and the camera’s intrinsic calibration, we then determine a projected line in 3-D for each row of the projector and each position of the calibration board.

Using two or more calibration board positions, we can then determine the plane of projected light for each row of the projector. In practice, we use five positions to provide a robust estimate. Experimentally, we find that the estimated planes are close to parallel, with the surface normals being no more than 2.4° apart in the worst case. These results demonstrate that we are able to achieve a close approximation to an orthographic projector despite the practical limitations of using a low-cost Fresnel lens with the proposed homography-based calibration procedure.

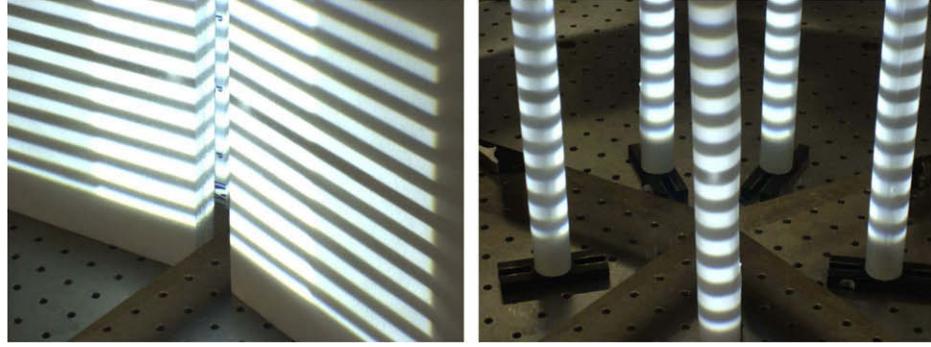


Fig. 10. Manual mirror alignment procedure. (Left) Image of mirrors collected when one mirror is covered with a blank surface and the orthographic projector is used to display the sixth Gray code. (Right) Image of mirrors collected when a cylinder is placed in the center of the reconstruction volume and the sixth Gray code is projected.

4.1.2. Orthographic projector calibration using structuring lighting

As an alternative to the method presented above, we also developed an independent procedure that utilizes Gray codes to obtain the orthographic projector calibration. As shown in Fig. 9, we again place a planar surface with a printed checkerboard pattern within the scanning volume. The correspondence between camera pixels and projector rows is then established using the familiar Gray codes (and again repeated for at least two calibration plane positions). This yields a “labeled” 3-D point cloud, where each point is reconstructed by intersecting the optical ray (for each camera pixel) with the calibration plane. The labels are generated from the Gray codes and provide the direct correspondence between a 3-D point and the projector row which generated it. Afterwards, a plane is fit independently to each set of 3-D points corresponding to a given orthographic projector row. Finally, the plane coefficients are filtered and extrapolated by using a least-squares quadratic polynomial fit to each parameter. Experimentally, we report that this calibration procedure tends to produce more accurate reconstructions than the previous method based on planar homographies.

4.2. Planar mirror alignment

At this point, we have described how to both calibrate and align a digital projector to create an orthographic illumination system using a Fresnel lens. In order to create the necessary surround illumination, as shown in Fig. 3, we must precisely align the planar mirrors such that the plane spanned by their surface normals is parallel to the orthographic illumination rays. Once again, we propose a simple solution based on Gray codes. As shown in Fig. 10, we begin by covering one mirror with a blank flat surface and projecting a given Gray code. Afterwards, we adjust the orientation of the opposite mirror until the reflected and projected stripes coincide on the blank surface. We then repeat this procedure for the remaining mirror. At this point, the orientation of the planar mirrors is close to that required for surround structured lighting. In order to fine tune our initial alignment, we conclude by placing a cylindrical object within the scanning volume and slightly adjust the mirrors until the direct and reflected Gray code stripes coincide around the entire surface. Conveniently, the Gray code stripes allow the mirrors to be adjusted in a hierarchical fashion, where at each stage the mirrors are aligned to a given Gray code pattern before progressing to a finer scale.

4.3. Camera and planar mirror calibration

The final calibration stage requires estimating the position and pose of the planar mirrors relative to the fixed, calibrated camera. As mentioned in Section 2, several authors have developed meth-

ods for calibrating and exploiting multi-view systems containing planar mirrors [10–13]. Briefly, these methods include: obtaining mirror calibration from object silhouettes extracted from the real and reflected images [10], exploiting the epipolar geometry of a real and reflected image pair [11], and using a bundle adjustment procedure to align actual and predicted images [13]. Because our system requires precise mirror alignment and calibration, we utilize an accurate bundle adjustment procedure inspired by Lin et al. [13].

To obtain an initial estimate of the left (M_1) and right (M_2) mirror calibration, we begin by recording a pair of images of a planar checkerboard (approximately 3 mm in thickness) held against each mirror surface. Afterwards, we recover the initial rotation $\{\mathbf{R}_{M_1}, \mathbf{R}_{M_2}\}$ and translation $\{\mathbf{T}_{M_1}, \mathbf{T}_{M_2}\}$ using the set of known 3-D checkerboard corners and corresponding image coordinates. Given these estimates, a point \mathbf{x}_{c0} in the camera coordinate system will map to points $\{\mathbf{x}_{M_1}, \mathbf{x}_{M_2}\}$ in the mirror coordinate systems as follows.

$$\mathbf{x}_{c0} = \mathbf{R}_{M_1} \mathbf{x}_{M_1} + \mathbf{T}_{M_1} \quad (1)$$

$$\mathbf{x}_{c0} = \mathbf{R}_{M_2} \mathbf{x}_{M_2} + \mathbf{T}_{M_2}. \quad (2)$$

Next, we collect a series of images for each mirror containing a planar checkerboard pattern and its reflection at various orientations throughout the scene (see Fig. 11). For each image, we manually select a set of corresponding points in the real and reflected images. We observe that the correct mirror parameters should allow the prediction of the projected real checkerboard corners (shown in green in Fig. 11). That is, given the intrinsic camera calibration, we can trace an optical ray from the camera center towards a reflected checkerboard corner (shown in blue in Fig. 11). The reflection of this ray by the corresponding mirror will intersect the known calibration checkerboard plane. The projection of this location (shown in red in Fig. 11) should be as close as possible to the projection of the real checkerboard corner. The following bundle adjustment procedure explains how we utilize this constraint to refine the initial mirror parameters.

Note that the reflection \mathbf{x}'_{c0} about the left mirror of a point \mathbf{x}_{c0} in the camera coordinate system is given by

$$\mathbf{x}'_{c0} = \mathbf{Q}_{M_1} \mathbf{x}_{M_1} + (\mathbf{I} - \mathbf{Q}_{M_1}) \mathbf{T}_{M_1}, \quad (3)$$

where

$$\mathbf{Q}_{M_1} = \mathbf{R}_{M_1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \mathbf{R}_{M_1}^T. \quad (4)$$

(A similar equation can be used to reflect points about the right mirror.) The reflections $\{\mathbf{v}_{M_1}, \mathbf{v}_{M_2}\}$ about each mirror of an optical ray \mathbf{v}_{c0} , defined in the camera coordinate system by the line passing

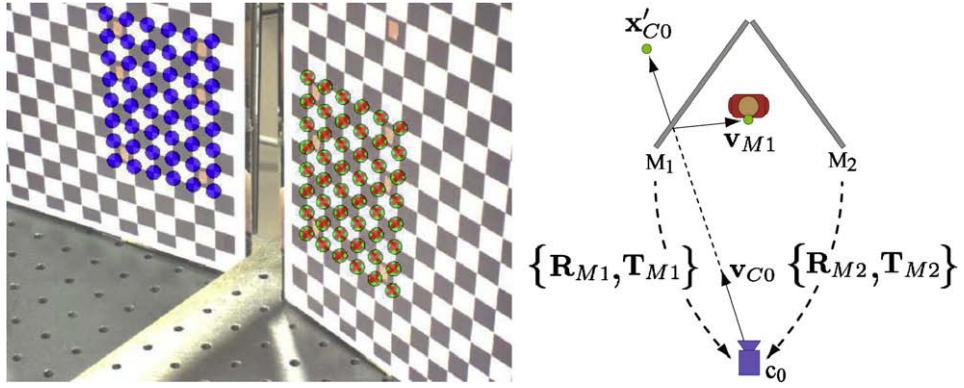


Fig. 11. Planar mirror calibration. (Left) The real and reflected checkerboard corners are denoted by green and blue circles, respectively. (Right) The translation and rotation for the left and right mirrors, given by $\{\mathbf{T}_{M1}, \mathbf{R}_{M1}\}$ and $\{\mathbf{T}_{M2}, \mathbf{R}_{M2}\}$, are optimized by bundle adjustment. At each iteration, the last mirror positions are used to transform the optical rays for each reflected checkerboard corner, given by \mathbf{v}_{C0} , into the corresponding reflected rays (e.g., \mathbf{v}_{M1}). Afterwards, the reflected rays are intersected with the checkerboard plane to produce an estimate of the 3-D position of each real checkerboard corner (denoted in red on the left). (For interpretation of the references to colours in this figure legend, the reader is referred to the web version of this paper.)

though the center of projection and a given camera pixel, are provided by the following expressions.

$$\mathbf{v}_{M1} = \mathbf{Q}_{M1} \mathbf{v}_{C0} \quad (5)$$

$$\mathbf{v}_{M2} = \mathbf{Q}_{M2} \mathbf{v}_{C0}. \quad (6)$$

Using Eqs. 3, 5, 6 we determine the reflection of each ray defined by the real camera center and the reflected checkerboard corners. Using the Levenberg–Marquardt algorithm, we simultaneously optimize the mirror parameters to minimize the sum of squared errors between the measured and predicted checkerboard corners (i.e., the green and red markers in Fig. 11, respectively). Experimentally, we find that the optimized solutions of the mirror parameters are robust to small perturbations in their initial estimates provided by the printed checkerboard pattern.

5. Reconstruction algorithm

Our reconstruction algorithm is similar to that used in conventional structured light scanners. As shown in Fig. 5, we begin by displaying 10 (horizontal) Gray code patterns and record their

appearance using a single camera. We then project an additional 10 patterns composed of the inverses of the regular Gray codes in order to improve the decoding accuracy [7]. By determining which patterns illuminated the object at each camera pixel, we can uniquely identify the corresponding projector row. Typical results are shown in Fig. 5. Note that additional filtering can be applied to the recovered row-estimate image to reduce noise and eliminate outliers. (In this example a morphological erosion by five pixels was applied.)

After recovering the per-pixel projector row correspondences, we reconstruct a 3-D point for each camera pixel as the intersection of the corresponding real (or virtual) camera ray with the appropriate calibrated light plane. As shown in Fig. 12, the virtual camera centers $\{\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_{21}, \mathbf{c}_{12}\}$ can be defined with respect to the real camera center $\mathbf{c}_0 = (0, 0, 0)^T$ using Eq. (3).

$$\mathbf{c}_1 = (\mathbf{I} - \mathbf{Q}_{M1}) \mathbf{T}_{M1} \quad (7)$$

$$\mathbf{c}_2 = (\mathbf{I} - \mathbf{Q}_{M2}) \mathbf{T}_{M2} \quad (8)$$

$$\mathbf{c}_{21} = \mathbf{Q}_{M2} \mathbf{c}_1 + (\mathbf{I} - \mathbf{Q}_{M2}) \mathbf{T}_{M2} \quad (9)$$

$$\mathbf{c}_{12} = \mathbf{Q}_{M1} \mathbf{c}_2 + (\mathbf{I} - \mathbf{Q}_{M1}) \mathbf{T}_{M1}. \quad (10)$$

Similarly, the virtual camera rays $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_{21}, \mathbf{v}_{12}\}$ can be defined in terms of \mathbf{v}_0 using Eqs. (5) and (6).

$$\mathbf{v}_1 = \mathbf{Q}_{M1} \mathbf{v}_0 \quad (11)$$

$$\mathbf{v}_2 = \mathbf{Q}_{M2} \mathbf{v}_0 \quad (12)$$

$$\mathbf{v}_{21} = \mathbf{Q}_{M2} \mathbf{Q}_{M1} \mathbf{v}_0 \quad (13)$$

$$\mathbf{v}_{12} = \mathbf{Q}_{M1} \mathbf{Q}_{M2} \mathbf{v}_0. \quad (14)$$

To complete our reconstruction, we manually select five regions of interest within the projector row-estimate image. For each region we apply the previous expressions to construct the optical rays corresponding to the appropriate real or virtual camera center. We then intersect the rays with their associated projector plane (corresponding to an individual orthographic projector row) in order to reconstruct a dense 3-D point cloud.

6. Experimental results

The proposed system was used to scan a variety of small objects with varying material and topological properties. As shown in Fig. 13, the current results are encouraging. For each example, we find that the five reconstructions originating from the one real and four virtual cameras are in close alignment—validating our proposed calibration procedure. These results also clearly verify the basic system concept, since nearly the entire object surface

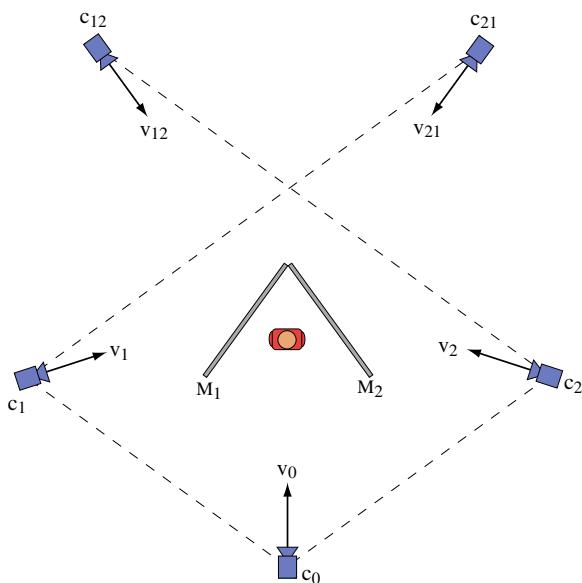


Fig. 12. The position of the real (c_0) and virtual (c_1, c_2, c_{12}, c_{21}) cameras with respect to the planar mirrors. Dashed lines are drawn connecting the cameras with their reflected counterparts.



Fig. 13. Summary of reconstruction results. From left to right: the input images used for texture mapping and four views of the 3-D point cloud recovered using the proposed method with a single camera and orthographic projector.

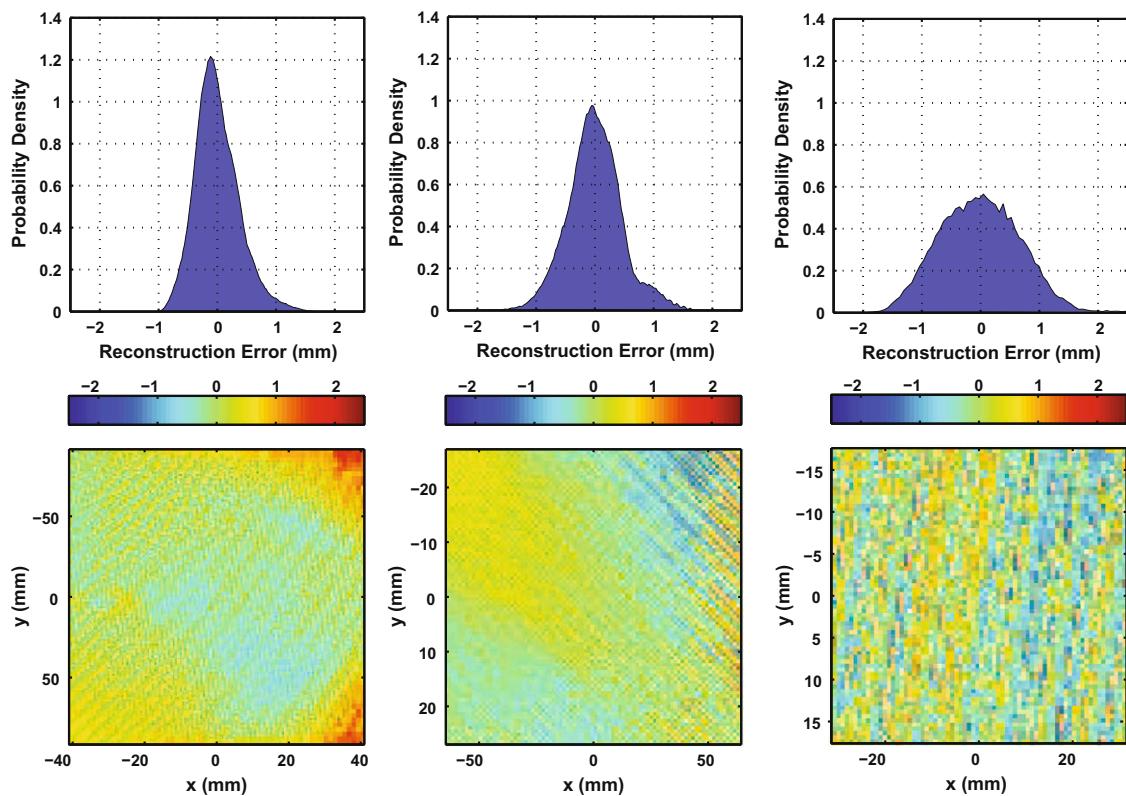


Fig. 14. Analysis of reconstruction errors. (Top row, from left to right) Estimated probability densities for the out-of-plane reconstruction errors, with: direct imaging, imaging with one mirror reflection, and imaging with two mirror reflections. (Bottom row, from left to right) Plot of the out-of-plane reconstruction errors (in millimeters) as a function of position within the calibration plane, for: direct imaging, imaging with one mirror reflection, and imaging with two mirror reflections.

(excluding the bottom) has been reconstructed from a single vantage point. In order to eliminate extreme outliers, the reconstructed point cloud was clipped by a coarse bounding volume. We believe that the remaining artifacts can be eliminated by improving both the alignment and quality of the optical components. The inclusion of additional post-processing should significantly reduce existing outliers and improve the color-blending between multiple views. For example, we note that the current results are shown as point-based renderings with no additional post-processing applied after reconstruction—the inclusion of which should significantly reduce existing outliers and improve the color-blending. Aside from outliers, the most significant artifacts we observe are: (1) visible distortion near the bottom and top of each object and (2) decreased reconstruction accuracy using the reflected images. We expect that both of these effects can be minimized by utilizing higher-quality optical components and better mechanical stabilization.

In order to quantify the reconstruction accuracy, we scanned a diffuse calibration plane at various orientations within the reconstruction volume. An optically-flat surface was obtained by affixing paper vellum to a section of the same type of front-surface mirror used in our prototype apparatus [15]. A plane was fit using least-squares to the recovered point cloud and the distance from each reconstructed point to the best-fit plane was used to quantify the reconstruction error. As shown in Fig. 14, our prototype achieves a reconstruction error of less than 1 mm in practical situations. The reconstruction errors tend to increase as a function of the number of mirror reflections that occur along a ray connecting a point on the surface to a given camera pixel. Specifically, we report standard deviations for the reconstruction errors of 0.3743 mm, 0.4646 mm, and 0.6981 mm for direct, one-bounce, and two-bounce imaging conditions, respectively. As described in Section 3, the orthographic projector has a finite depth of field. Since the rays which undergo multiple reflections have a longer optical path, we expect that their defocus would be greater than direct rays. In practice, we focus the projector such that the center of the depth of field is located at the average optical path length. As a result, we primarily attribute the increased reconstruction error for multiple reflections to mechanical alignment errors.

By examining the reconstruction error as a function of position within the calibration plane (as shown in the bottom row of Fig. 14), we find that the errors are not uniformly distributed. We observe a general trend in which the errors are greater near the image periphery. This observation can be attributed to the inability of the Fresnel lens to create an ideal orthographic projector. This claim is further confirmed by Fig. 9, where the estimated plane coefficients exhibit significant variation near the image periphery. In addition, we note that the reconstruction errors show a periodic spatial variation, manifesting as a high-frequency pattern in the error images in Fig. 14. We attribute this observation to the quantization noise due to assigning a given camera pixel to a discrete projector row. In the future, we believe these errors could be reduced by using a structured light sequence which achieves sub-pixel accurate assignment of projector rows to camera pixels [7].

Finally, we observe that the preceding analysis has only quantified the reconstruction error as a function of the number of reflections occurring within the imaging system. In practice, we found it challenging to isolate and identify how many reflections occurred before a given illuminating ray reached the object surface. Practically, however, the prior analysis spans the worst-case and best-case scenarios. Specifically, the illumination for the direct imaging test was also direct (i.e., no mirror reflections occurred between the Fresnel lens and the object surface). Similarly, the two-bounce results predominantly contain illumination rays which underwent multiple-bounces. As a result, we can confidently report that the

reconstruction errors remain within about 1 mm for our current prototype—regardless of the number of reflections occurring along either the illuminating or imaging paths.

7. Limitations and future work

There are several design issues we plan on addressing in the near term. Foremost, we plan on increasing the size of the scanning volume and incorporating higher-quality optical components. Currently, the reconstruction volume is limited by the dimensions and focal length of the Fresnel lens. Large, inexpensive Fresnel lenses with longer focal lengths are commercially available. As a result, we plan on incorporating a 44.5 × 31 in. lens (Model NT46-659 from Edmund Optics, Inc. [24]) with a focal length of 39.4 in. (which would enable the creation of a scanning volume of roughly 20 in. cubed with our current projector). In addition, this lens has 226 grooves per inch—a critical factor which should further reduce the artifacts observed with our current implementation by producing a closer-approximation of orthographic illumination.

The proposed system uses only a single camera to observe five views of the object. This scheme divides the camera pixels among multiple image regions. Compared to a temporally-multiplexed design, such as using multiple cameras and projectors or a turntable, our system requires a higher resolution camera. In contrast, the benefit of our design over these alternatives include automatic frame synchronization, color calibration, and identical intrinsic calibration parameters [11]. In practice, a designer would have to weigh these benefits and limitations when considering the proposed design versus other alternatives.

While our goal was to acquire a complete 360° reconstruction, practical issues limit the reconstruction volume beyond that simply defined by the common field of view of the orthographic projector and the real and virtual cameras. In particular, the object can block certain direct and single-reflection rays that would otherwise illuminate the object surface on a subsequent first or second reflection, respectively. Experimentally, these occlusion events did not appear to significantly impact the final reconstruction. This situation only impacts points that are solely illuminated by reflected rays. If we consider a point (imaged by at least one real or virtual camera) that cannot be directly illuminated, then there exist four other optical rays, corresponding to each virtual orthographic projector, that could illuminate the point. In this case, reconstruction would fail only if the object occluded all four rays. In the future we plan on characterizing the optimum placement of an object in order to minimize the impact of such object-dependent occlusions.

In general, the primary benefit of the proposed system is that it allows rapid acquisition of complete 3-D surfaces without the need for turntables or multiple camera/projector pairs. In contrast to existing structured light systems, our design is well-suited for capturing moving objects when a complete scan is required. Adding hardware synchronization between the camera and projector would enable real-time capture at the maximum camera frame rate, which would be sufficient for reconstructing relatively slow-moving objects [25]. Currently there exist several structured light systems that utilize synchronization to enable real-time capture, including those presented by Rusinkiewicz et al. [26], Zhang and Huang [25], as well as Zhang et al. [27]. We plan on experimenting with these methods in order to achieve rapid and reliable 360° real-time reconstructions with an improved system. In fact, one promising extension would be to use the “single-shot” pattern (i.e., single projected image) proposed by Zhang et al. [27]. Although the use of such a pattern would further reduce the effective resolution of the final surface

model, we note that our system could then be used to produce a “single-shot” full 3-D model. As a result, we could replace our digital projector with a simple slide projector—significantly reducing the system cost.

Aside from real-time reconstruction, we note that our proposed architecture is well-suiting for rapid prototyping systems. In effect, our current prototype is the 3-D equivalent of a flatbed scanner. By coupling our design with a rapid prototyping machine, we could create a simple 3-D fax machine [28]. Such devices could prove beneficial for telecollaboration, where preliminary designs could be rapidly sculpted, scanned, and reconstructed remotely. In the future we plan on enhancing the post-processing pipeline of our prototype to produce the watertight models that would be necessary for such rapid prototyping applications.

8. Conclusion

In this work, we have presented a novel extension to traditional structured light systems that is capable of obtaining a complete 360° object reconstruction using a single sequence of structured light patterns. By eliminating the need for additional projectors or scanning passes, this system marks an important step towards obtaining complete real-time reconstructions. While the current prototype can only scan relatively small volumes, we believe this system has already demonstrated practical benefits for telecollaboration applications. In addition, we have presented a complete set of calibration and alignment procedures for creating orthographic projectors using digital projectors and low-cost Fresnel lenses. We hope that our methods will inspire similar work on applying orthographic illumination to outstanding problems in active imaging and 3-D reconstruction.

References

- [1] A. Laurentini, The Visual Hull Concept for Silhouette-based Image Understanding, *IEEE TPAMI* 16 (2) (1994) 150–162.
- [2] K.N. Kutulakos, S.M. Seitz, A theory of shape by space carving, *Int. J. Comput. Vis.* 38 (3) (2000) 199–218.
- [3] S. Seitz, B. Curless, J. Diebel, D. Scharstein, R. Szeliski, A comparison and evaluation of multi-view stereo reconstruction algorithms, in: *CVPR 2006*, 2006.
- [4] F. Blais, Review of 20 years of range sensor development, *J. Electron. Imaging* 13 (1) (2004) 231–240.
- [5] Z. Zhang, Iterative point matching for registration of free-form curves and surfaces, *Int. J. Comput. Vis.* 13 (2) (1994) 119–152.
- [6] S.K. Nayar, V. Anand, Projection Volumetric Display Using Passive Optical Scatterers, Tech. rep., 2006.
- [7] J. Salvi, J. Pages, J. Batlle, Pattern codification strategies in structured light systems, in: *Pattern Recognition*, vol. 37, 2004, pp. 827–849.
- [8] S. Inokuchi, K. Sato, F. Matsuda, Range imaging system for 3-D object recognition, in: *Proceedings of the International Conference on Pattern Recognition*, 1984.
- [9] E. Epstein, M. Granger-Piché, P. Poulin, Exploiting mirrors in interactive reconstruction with structured light, in: *Vision, Modeling, and Visualization 2004*, 2004, pp. 125–132.
- [10] K. Forbes, F. Nicolls, G. de Jager, A. Voigt, Shape-from-silhouette with two mirrors and an uncalibrated camera, in: *ECCV 2006*, 2006, pp. 165–178.
- [11] J. Gluckman, S. Nayar, Planar catadioptric stereo: geometry and calibration, in: *CVPR 1999*, 1999.
- [12] B. Hu, C. Brown, R. Nelson, Multiple-view 3-D reconstruction using a mirror, Tech. rep., May 2005.
- [13] I.-C. Lin, J.-S. Yeh, M. Ouhyoung, Extracting realistic 3D facial animation parameters from multiview video clips, *IEEE Comput. Graph. Appl.* 22 (6) (2002) 72–80.
- [14] Fresnel Technologies, Inc., Fresnel lenses. Available from: <<http://www.fresneltech.com/>>.
- [15] HighReflectiveMirrors.com, HR94SC digital 3 mm front surface mirror. Available from: <<http://www.highreflectivemirrors.com/>>.
- [16] Point Grey Research, Inc., Grasshopper IEEE-1394b digital camera line. Available from: <<http://www.ptgrey.com/>>.
- [17] J.C. Lee, P.H. Dietz, D. Maynes-Aminzade, R. Raskar, S.E. Hudson, Automatic projector calibration with embedded light sensors, in: *UIST’04: Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, ACM, New York, NY, USA, 2004, pp. 123–126.
- [18] R. Raskar, P. Beardsley, A self-correcting projector, *CVPR 2001* 2 (2001) 504–508.
- [19] R. Sukthankar, R.G. Stockton, M.D. Mullin, N.W. Labs, Smarter presentations: exploiting homography in camera-projector systems, in: *CVPR 2001*, 2001, pp. 247–253.
- [20] R.J. Surati, Scalable self-calibrating display technology for seamless large-scale displays, Ph.D. thesis, supervisor-Thomas F. Knight, Jr., 1999.
- [21] S. Zhang, P.S. Huang, Novel method for structured light system calibration, *Opt. Eng.* 45(8).
- [22] R. Legarda-Sáenz, T. Bothe, W.P. Jüptner, Accurate procedure for the calibration of a structured light system, *Opt. Eng.* 43 (2) (2004) 464–471.
- [23] J.-Y. Bouguet, Complete camera calibration toolbox for matlab. Available from: <http://www.vision.caltech.edu/bouguetj/calib_doc>.
- [24] Edmund Optics, Inc., Fresnel lenses. Available from: <<http://www.edmundoptics.com/>>.
- [25] S. Zhang, P.S. Huang, High-resolution, real-time three-dimensional shape measurement, *Opt. Eng.* 45(12).
- [26] S. Rusinkiewicz, O. Hall-Holt, M. Levoy, Real-time 3D model acquisition, in: *SIGGRAPH 2002*, ACM Press, 2002, pp. 438–446.
- [27] L. Zhang, B. Curless, S.M. Seitz, Rapid shape acquisition using color structured light and multi-pass dynamic programming, *3DPVT 2002* (2002) 24.
- [28] B. Curless, New methods for surface reconstruction from range images, Tech. rep., 1997.