Surround Structured Lighting for Full Object Scanning

Douglas Lanman, Daniel Crispell, and Gabriel Taubin
Department of Engineering, Brown University
{dlanman,daniel_crispell,taubin}@brown.edu

Abstract

This paper presents a new system for acquiring complete 3D surface models using a single structured light projector, a pair of planar mirrors, and one or more synchronized cameras. 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 be parallel, and 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. Using each real and virtual camera, we then recover a dense 3D point cloud spanning the entire object surface using traditional structured light algorithms. As we demonstrate, this configuration overcomes a major hurdle to achieving full 360 degree reconstructions using a single structured light sequence by eliminating the need for merging multiple scans or multiplexing several projectors.

1. Introduction

Although there exist a variety of structured light 3D scanning systems, previously none were able to capture a full 360 degree object scan using a single camera position. In this paper, we present such a system. 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. We discuss our implementation of the system and how we plan to leverage the design into a full 360 degree 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 walk through the implemented reconstruction algorithm, and present the resulting 3D models in Section 6. Finally, in Sections 7 and 8 we discuss our future plans for the system and conclude.

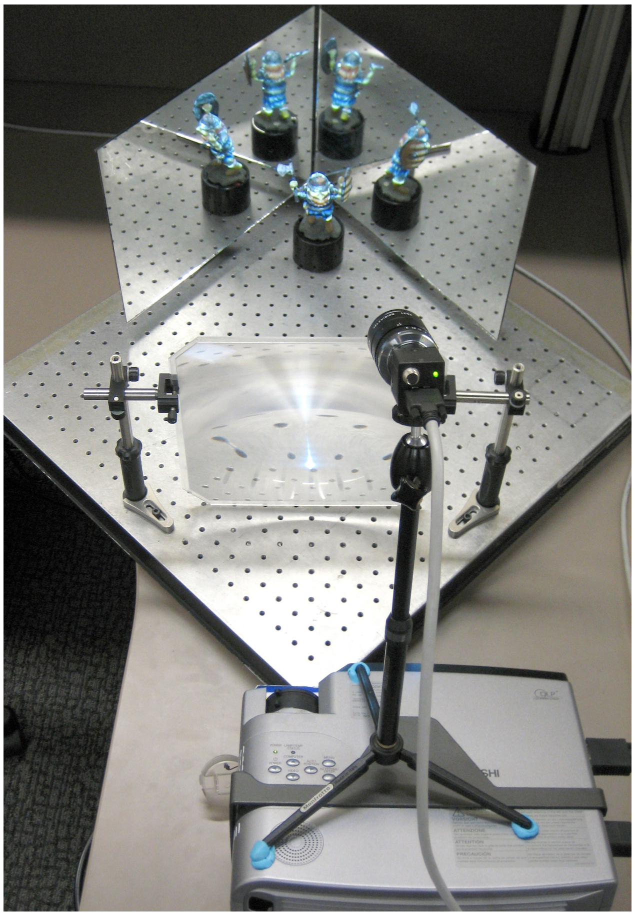


Figure 1. System architecture: DLP projector, Fresnel lens, planar mirrors, and a digital camera.

2. Related Work

This paper draws on three areas of active research within the computer vision and graphics communities: (1) structured lighting for 3D 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. [10]. As described in the recent survey article by Salvi et al. [13], coded structured light is a reliable, inexpensive method for for recovering 3D 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. [6], in which each pattern is composed of a sequence of black and white stripes oriented along the projector scanlines. By illuminating the object with a temporally-multiplexed sequence of increasingly-fine Gray code patterns, the corresponding projector scanline can be identified for each camera pixel. Afterwards, the appropriate ray-plane intersection can be used to recover a 3D point on the surface.

The idea of using planar mirrors to create virtual structured light projectors was first presented by Epstein et al. [2]. In their system, one or more planar mirrors are illuminated by a projector 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 [3, 4, 5, 9]. As discussed by Gluckman and Nayar [4], 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. [3] have explored precisely the 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. In their original work, the authors obtained a complete 3D surface model by estimating the visual hull [7] 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 scanlines and camera pixels.

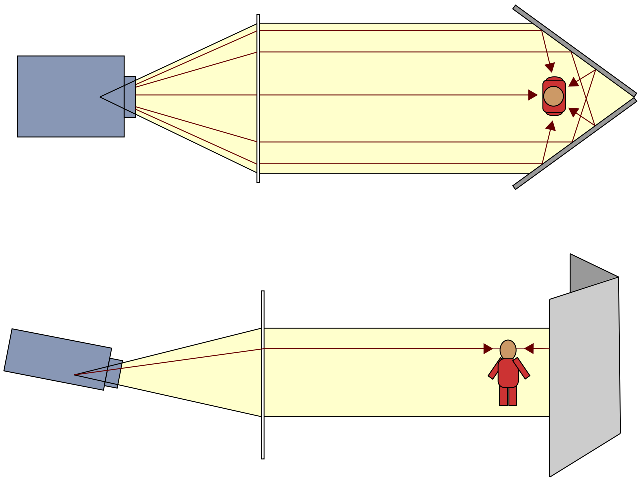


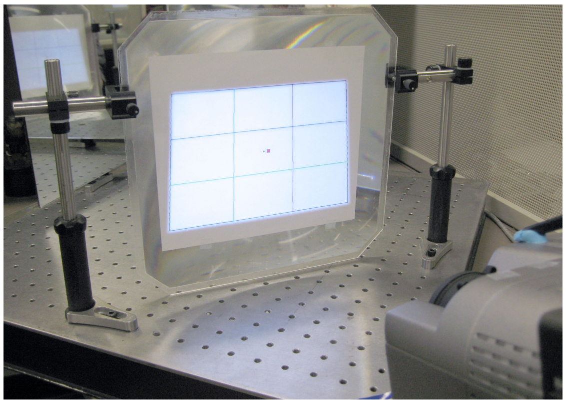
Figure 2. Diagram showing the orthographic projection system and multiple rays from a single projected scanline illuminating the object from different angles. The top and bottom diagrams show top and side views of the system, respectively.

Finally, we note that Nayar and Anand [10] 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.

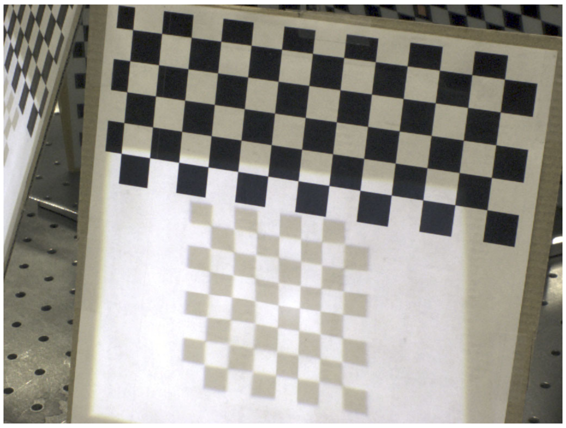
3. System Design and Construction

The key components of the proposed scanning system are an orthographic projector and two planar mirrors aligned such that their normal vectors are contained within the plane of light created by each projector scanline. 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 (Figure 2). A camera containing the object and both mirrors in its field of view can then obtain five views of the illuminated object: one direct view, two first reflections, and two second reflections [3], as can be seen in Figure 1. The full structured light pattern combined with the five views provides sufficient information for a 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 [10] 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 inch square Fresnel lens with 200 grooves per inch and a focal length of 24.0 inches, 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 focus 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 value 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 optical grade first-surface mirrors. The mirrors are positioned such that their surface normals are roughly 72 degrees apart, as per Forbes et al. [3] in order to provide the five equally spaced views, and such that their normals are perpendicular to the normals of the projected light planes. The mirrors are mounted on gimbals with fine tuning knobs in order to facilitate precise positioning.



(a) Aligning the projector with the Fresnel lens



(b) Vertically-spaced illumination plane calibration

Figure 3. Orthographic projector calibration.

4. Calibration

Because of the unique design of the scanning system, calibration of the multiple components is a non-trivial task. Our calibration procedure is divided into two stages: (1) configuration of the orthographic projector and (2) calibration of the camera/mirror system.

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 3D. To date, a wide variety of approaches for general projector calibration have been proposed, some of which have been specifically tailored to structured light configurations [8, 16]. For our system, we follow the well-established method of utilizing a calibrated camera to subsequently determine the intrinsic and extrinsic projector calibration [11]. We first estimate the intrinsic camera parameters, as well as a fourth-order lens distortion model, using the *Camera Calibration Toolbox for Matlab* [1]. Afterwards, 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 3D position of each projected checkerboard corner. Given the set of correspondences between projector pixels and 3D 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.

The second step involves the correct placement of the DLP projector with respect to the Fresnel lens as described in Section 3. 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 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. 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 scanlines of projected images (e.g., Gray codes) to planes in 3D (as defined in the camera coordinate system). If the projection system is close to orthographic, these planes should all be approximately parallel. To recover an estimate of the projected light planes, we use a calibration board that contains a printed checkerboard pattern on the upper half, and blank whitespace on the lower half (see Figure 3b) and project a known checkerboard pattern onto the blank half. 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 can determine a projected line in 3D for each scanline 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 scanline 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 degrees 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 the alignment procedure.

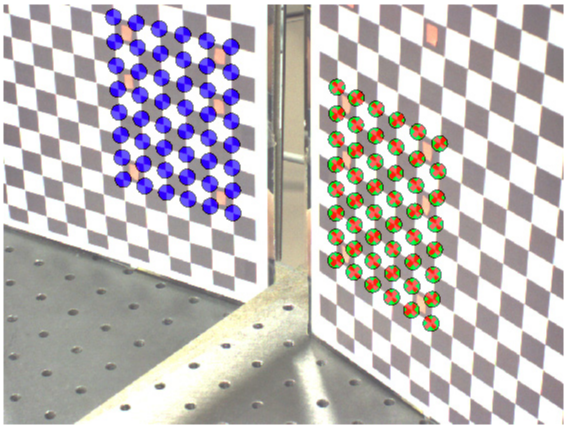


Figure 4. Planar mirror calibration. The real and reflected checkerboard corners are denoted by green and blue circles, respectively. The predicted real checkerboard corners, obtained by ray-tracing after bundle adjustment, are shown in red.

4.2. 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 methods for calibrating and exploiting multi-view systems containing planar mirrors [3, 4, 5, 9]. Briefly, these methods include: obtaining mirror calibration from object silhouettes extracted from the real and reflected images [3], exploiting the epipolar geometry of a real and reflected image pair [4], and using a bundle adjustment procedure to align actual and predicted images [9]. Because our system requires precise mirror alignment and calibration, we utilize an accurate bundle adjustment procedure similar to Lin et al. [9].

To obtain an initial estimate of the left (M1) and right (M2) mirror calibration, we begin by recording a pair of images of a planar checkerboard held against each mirror surface. Afterwards, we recover the initial rotation $\{\mathbf{R}_{M1}, \mathbf{R}_{M2}\}$ and translation $\{\mathbf{T}_{M1}, \mathbf{T}_{M2}\}$ using the set of known 3D checkerboard corners and corresponding image coordinates. Given these estimates, a point $\mathbf{x}_{C0}$ in the camera coordinate system will map to points $\{\mathbf{x}_{M1}, \mathbf{x}_{M2}\}$ in the mirror coordinate systems as follows.

$$\mathbf{x}_{C0} = \mathbf{R}_{M1}\mathbf{x}_{M1} + \mathbf{T}_{M1}$$
$$\mathbf{x}_{C0} = \mathbf{R}_{M2}\mathbf{x}_{M2} + \mathbf{T}_{M2}$$

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 Figure 4). 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 Figure 4). 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 Figure 4). 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 Figure 4) should be as close as possible to 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 (or similarly about the right mirror) of a point $\mathbf{x}_{C0}$ in the camera coordinate system is given by

$$\mathbf{x}'_{C0} = \mathbf{Q}_{M1}\mathbf{x}_{M1} + (\mathbf{I} - \mathbf{Q}_{M1})\mathbf{T}_{M1}, \tag{1}$$

where

$$\mathbf{Q}_{M1} = \mathbf{R}_{M1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \mathbf{R}^T_{M1}.$$

In addition, the reflections $\{\mathbf{v}_{M1}, \mathbf{v}_{M2}\}$ of an optical ray $\mathbf{v}_{C0}$, defined in the camera coordinate system, are given by the following expressions.

$$\mathbf{v}_{M1} = \mathbf{Q}_{M1}\mathbf{v}_{C0} \tag{2}$$
$$\mathbf{v}_{M2} = \mathbf{Q}_{M2}\mathbf{v}_{C0} \tag{3}$$

Using Equations 1-3 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 Figure 4, respectively).

5. Reconstruction Algorithm

Our reconstruction algorithm is similar to that used in conventional structured light scanners. As shown in Figure 6b, 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 [13]. By determining which patterns illuminated the object at each image pixel, we can uniquely identify the corresponding projector scanline. Typical results are shown in Figure 6c. 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 5 pixels was applied.)

After recovering the per-pixel projector scanline correspondences, we reconstruct a 3D 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 Figure 5, 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 Equation 1.

$$\begin{aligned} \mathbf{c}_1 &= (\mathbf{I} - \mathbf{Q}_{M1})\mathbf{T}_{M1} \\ \mathbf{c}_2 &= (\mathbf{I} - \mathbf{Q}_{M2})\mathbf{T}_{M2} \\ \mathbf{c}_{21} &= \mathbf{Q}_{M2}\mathbf{c}_1 + (\mathbf{I} - \mathbf{Q}_{M2})\mathbf{T}_{M2} \\ \mathbf{c}_{12} &= \mathbf{Q}_{M1}\mathbf{c}_2 + (\mathbf{I} - \mathbf{Q}_{M1})\mathbf{T}_{M1} \end{aligned}$$

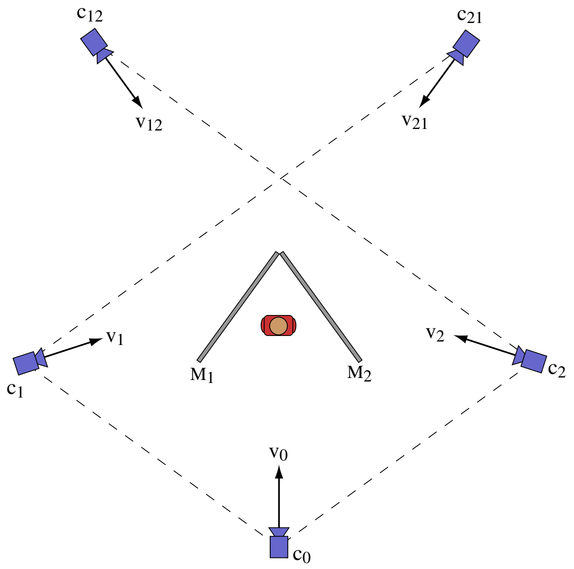


Figure 5. The position of the real (c_0) and virtual (c_1, c_2, c_{12}, c_{21}) cameras with respect to the planar mirrors. Dashed lines are drawn connecting the cameras with their reflected counterparts.

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 Equations 2 and 3.

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{Q}_{M1}\mathbf{v}_0 \\ \mathbf{v}_2 &= \mathbf{Q}_{M2}\mathbf{v}_0 \\ \mathbf{v}_{21} &= \mathbf{Q}_{M2}\mathbf{Q}_{M1}\mathbf{v}_0 \\ \mathbf{v}_{12} &= \mathbf{Q}_{M1}\mathbf{Q}_{M2}\mathbf{v}_0 \end{aligned}$$

To complete our reconstruction, we manually select five regions of interest within the projector row-estimate image (e.g., Figure 6c). 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 scanline plane in order to reconstruct a dense 3D 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 Figure 7, preliminary 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 (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.

7. 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 41×31 inch lens with a focal length of 39.4 inches (which would enable the creation of a scanning volume of roughly 20 inches cubed with our current projector).

Most structured light systems require the accumulation of several independent scans to cover the full object surface; as a result, existing systems are not well-suited for capturing moving objects when a complete scan is required. Our system has the potential to overcome this obstacle. Adding hardware synchronization would enable real-time capture at the maximum camera frame rate, which would be sufficient for reconstructing relatively slow-moving objects [15]. Currently there exist several systems that utilize synchronization to enable real-time capture, including those presented by Rusinkiewicz et al. [12], Zhang and Huang [15], as well as Zhang et al. [14]. We plan on experimenting with all of these methods in order to achieve rapid and reliable 360 degree real-time reconstructions with an improved system.

8. Conclusion

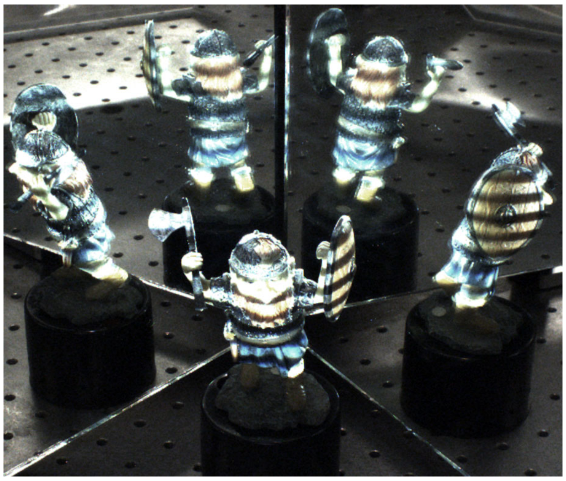
We have presented a novel scanning system which is capable of obtaining a complete 360 degree object reconstruction using a single sequence of structured light patterns. By eliminating the need for additional projectors or scanning passes, this system marks an important step towards obtaining complete real-time reconstructions. Encouraged by these preliminary results, we plan on refining the prototype system in the near term, with the ultimate goal of achieving real-time human motion capture in the round.

References

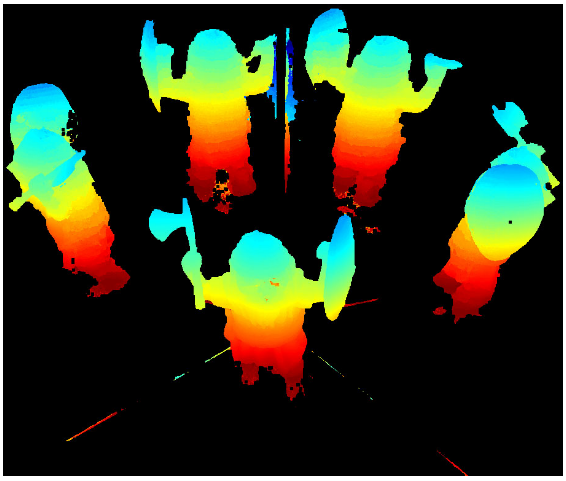
[1] J.-Y. Bouguet. Complete camera calibration toolbox for matlab. http://www.vision.caltech.edu/bouguetj/calib_doc.

[2] E. Epstein, M. Granger-Piché, and P. Poulin. Exploiting mirrors in interactive reconstruction with structured light. In *Vision, Modeling, and Visualization 2004*, pages 125–132, 2004.

(a) Ambient illumination used for texture mapping



(b) Projected Gray code pattern



(c) Projector rows recovered from Gray code sequence

Figure 6. Example of orthographic horizontal Gray code pattern and recovered projector rows.

Figure 7. Summary of preliminary reconstruction results. From left to right: the input images used for texture mapping and four views of the 3D point cloud recovered using the proposed method with a single camera.

[3] K. Forbes, F. Nicolls, G. de Jager, and A. Voigt. Shape-from-silhouette with two mirrors and an uncalibrated camera. In *ECCV 2006*, pages 165–178, 2006.

[4] J. Gluckman and S. Nayar. Planar Catadioptric Stereo: Geometry and Calibration. In *CVPR 1999*, June 1999.

[5] B. Hu, C. Brown, and R. Nelson. Multiple-view 3-d reconstruction using a mirror. Technical report, May 2005.

[6] S. Inokuchi, K. Sato, and F. Matsuda. Range imaging system for 3-d object recognition. In *Proceedings of the International Conference on Pattern Recognition*, 1984.

[7] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE TPAMI*, 16(2):150–162, 1994.

[8] R. Legarda-Sáenz, T. Bothe, and W. P. Jüptner. Accurate procedure for the calibration of a structured light system. *Optical Engineering*, 43(2):464–471, 2004.

[9] I.-C. Lin, J.-S. Yeh, and M. Ouhyoung. Extracting realistic 3d facial animation parameters from multiview video clips. *IEEE Computer Graphics and Applications*, 2002.

[10] S. K. Nayar and V. Anand. Projection Volumetric Display Using Passive Optical Scatterers. Technical report, 2006.

[11] R. Raskar and P. Beardsley. A self-correcting projector. *CVPR 2001*, 2:504–508, 2001.

[12] S. Rusinkiewicz, O. Hall-Holt, and M. Levoy. Real-time 3d model acquisition. In *SIGGRAPH 2002*, pages 438–446. ACM Press, 2002.

[13] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. In *Pattern Recognition*, volume 37, pages 827–849, April 2004.

[14] L. Zhang, B. Curless, and S. M. Seitz. Rapid shape acquisition using color structured light and multi-pass dynamic programming. *3DPVT 2002*, page 24, 2002.

[15] S. Zhang and P. S. Huang. High-resolution, real-time three-dimensional shape measurement. *Optical Engineering*, 45(12), 2006.

[16] S. Zhang and P. S. Huang. Novel method for structured light system calibration. *Optical Engineering*, 45(8), 2006.