Shape from Depth Discontinuities under Orthographic Projection

Douglas Lanman, Daniel Cabrini Hauagge, and Gabriel Taubin
Division of Engineering, Brown University
Providence, RI 02912 USA

Abstract

We present a new method for reconstructing the 3-D surface of an opaque object from the motion of its depth discontinuities, when viewed under orthographic projection as the object undergoes rigid rotation on a turntable. A novel shape completion scheme is introduced to fill in gaps in the recovered surface, which would otherwise be impossible to reconstruct from the motion of the depth discontinuities alone. To verify the basic theory, we construct a large-format orthographic multi-flash camera capable of recovering the depth discontinuities using controlled illumination. We analyze the limitations of multi-flash depth edge detection using orthographic imaging with both point sources and directional illumination. We conclude by considering future applications for the shape completion scheme and the specialized hardware introduced in this work.

1. Introduction

Piecewise smooth surfaces can be used to describe the exterior boundary of solid objects, and have found widespread use in industrial design and manufacturing. These descriptions are composed of smooth surface patches meeting along piecewise smooth boundary curves called *feature lines*, across which surface normals can be discontinuous. A wide variety of scanning methods, generally referred to as *shape-from-X*, have been proposed to recover 3-D shapes. Many active methods (including structured lighting [20], laser striping [4], and photometric stereo [28]) and passive methods (such as multi-view stereo [22]) recover points located on smooth surface patches, yet are unable to directly sample feature lines. Instead, post-processing must be used to detect feature lines from sampled point clouds [17]. Unfortunately, reconstructing feature lines in this manner is technically impossible, since the corresponding space curves are not band-limited signals.

Relatively few shape capture methods are tailored for direct detection of feature lines. One notable exception is the shape-from-silhouette algorithm introduced by Cipolla and Giblin [6]. While the visual hull [15] can be estimated from sparse viewpoints, Cipolla and Giblin describe a differential formulation, where an object is slowly rotated on a turntable and a dense set of frames is recorded. The manner in which silhouettes deform between images is used to reconstruct surface points that are tangent to the camera viewing direction and located along silhouette boundaries. More recently, Crispell *et al.* [7] used this algorithm to recover additional points located on general depth discontinuities, not just silhouettes, using the multi-flash imaging method proposed by Raskar *et al.* [19]. While locally convex points inside concavities can be estimated from these additional samples, locally concave points at the bottom of concavities cannot be recovered, leaving holes in the recovered surface.

In this paper we present a shape capture method, building on prior systems, that is able to directly detect feature lines. We use an *orthographic* multi-flash camera to measure the motion of depth discontinuities as an object rotates on a turntable. We describe a new method to detect and fill local concavities, exploiting imaging symmetries unique to orthographic projection. We note that our method is well-suited for industrial inspection and reverse engineering of manufactured parts containing numerous feature lines.

1.1. Contributions

The primary contributions of this paper are as follows.

- We present a new shape completion algorithm that can be used to detect and fill local concavities in the surface recovered from the visual motion of depth discontinuities viewed under orthographic projection.
- We analyze the properties of orthographic multi-flash cameras for depth edge detection, using either near-field point sources or directional illumination.
- We describe a calibration method for orthographic cameras using at least four images of a planar pattern augmented with a single point above its surface.
- We present and analyze the performance of an experimental prototype, which is the first to exploit the unique properties of orthographic multi-flash imaging to reconstruct the 3-D shape of solid surfaces.

2. Related Work

2.1. Epipolar-Plane Image Analysis

The study of image slices over time and the structures observed in such imagery is considered in the prior work on *epipolar-plane images* (EPI). One of the earliest studies was published by Bolles [5], in which he considers the case of linear motion for parallel cameras. In this case, a single scene point maps to a line in the EPI, with a slope corresponding to the distance of the point from the camera. Lines corresponding to points closer to the camera overlap those for points further away, allowing reconstruction without explicit feature matching [22]. This model is extended by Baker and Bolles [2] to deal with non-parallel cameras. Feldmann *et al.* [10] describe the properties of EPI curves for a circular motion path due to camera rotation. Their parameterized curves cannot be applied to our system, as they model texture features rather than depth discontinuities. Apostoloff and Fitzgibbon [1] detect T-junctions for general camera motion. In a closely related work, Berent and Dragotti [3] achieve scene segmentation by extracting level sets from a collection of epipolar-plane images.

2.2. Optical Shape Capture

In this work we propose a shape capture method inspired by the work of Crispell *et al.* [7]. In contrast to photometric stereo [28], in which lights are placed far from a camera, Raskar *et al.* [19] propose placing light sources *close* to the center of projection to estimate the set of visible depth discontinuities. Crispell *et al.* [7] show that such multi-flash cameras can be used to measure the visual motion of depth discontinuities as an object undergoes rigid rotation, allowing surface reconstruction using the differential method of Cipolla and Giblin [6]. Unlike these systems, in which a single perspective camera is used, we study the visual motion of depth discontinuities under orthographic projection.

2.3. Orthographic Imaging and Illumination

Orthographic projection can be achieved using telecentric lenses and has found widespread use in the machine vision community due to its lack of perspective distortion when inspecting machined parts. As demonstrated by Watanabe and Nayar [27], a telecentric lens can be fashioned from a conventional lens by placing an aperture at a specific location (*e.g.*, at a focal point for a thin lens). Orthographic imaging conditions are also typical in remote surveillance, when the camera is sufficiently far from the scene. Similarly, the properties of orthographic illumination (*e.g.*, directional light sources) have been applied for scene understanding [16]. In this paper, we propose a method for calibrating telecentric lenses, and describe the extension of multi-flash imaging to the orthographic case.

2.4. Curve and Surface Completion

The proposed surface reconstruction method cannot recover locally concave points in deep concavities. As a result, a shape completion method is required to fill remaining gaps. Shape completion has been extensively studied in 2-D and 3-D. Numerous 2-D curve completion schemes have been proposed [25, 14, 13, 12]; generally, two position and tangent constraints are specified. As an infinite number of curves satisfy such boundary conditions, additional constraints have been proposed to obtain a unique solution. Ullman [25] proposes a curve of minimum total curvature formed by two circular arcs, tangent at both ends, meeting in the center. Horn [12] further analyzes the curvature energy when the number of circular arcs increases, proving that the internal energy is smaller than that of an Euler spiral or a simple circle. Knuth [14] proposes a set of properties for visually appealing and scalable curves, arguing that cubic splines possess the desired properties. Kimia *et al.* [13] propose minimizing variation of curvature, yielding completions based on the Euler spiral. In 3-D, diffusion-based mesh inpainting [9, 26] has been proposed, where the gaps are filled inwards from the border while preserving some measure of smoothness. Pauly *et al.* [18] complete meshes by detecting and replicating repeated patterns. In a closely-related work, Crispell *et al.* [7] fill gaps using the implicit surface defined by an oriented point cloud. Finally, Curless and Levoy [8] propose a volumetric method for fitting a signed distance function to a set of range images.

3. Properties of Depth Discontinuities under Orthographic Projection

The properties of apparent contours under perspective projection have been used to recover 3-D shapes from turntable sequences using silhouettes [6] and depth discontinuities [7]. In this paper we extend these results to the case of orthographic imaging. Specifically, we consider the scenario in which an object undergoes rigid rotation about a fixed axis, such that the rotation axis lies in a plane parallel to image (see Figure 5). In this situation, a point $\boldsymbol{p}$ located on a depth discontinuity can be represented as $\boldsymbol{p} = \boldsymbol{q} + \lambda \boldsymbol{v}$, where $\boldsymbol{q}$ is the orthographic projection of the point onto the image plane along the tangent ray $\boldsymbol{v}$, and λ is the scale factor corresponding to scene depth. Cipolla and Giblin [6] show that the 3-D surface can be recovered such that

$$\lambda = \frac{-\boldsymbol{n}^{\mathsf{T}} \dot{\boldsymbol{q}}}{\boldsymbol{n}^{\mathsf{T}} \dot{\boldsymbol{v}}}, \text{ for } \boldsymbol{n}^{\mathsf{T}} \dot{\boldsymbol{v}} \neq 0, \tag{1}$$

where $\boldsymbol{n}$ is the surface normal at $\boldsymbol{p}$, and $\dot{\boldsymbol{q}}$ and $\dot{\boldsymbol{v}}$ are the derivatives with respect to the turntable angle θ. In this section, we demonstrate that (1) can be used to reconstruct both visible and hidden surfaces, using unique properties of depth discontinuities under orthographic projection.

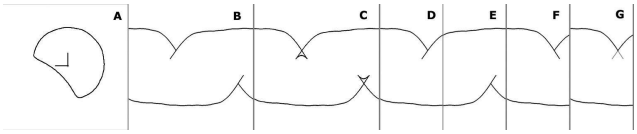


Figure 1. Depth discontinuities under orthographic projection. (a) A simple 2-D shape with one concavity is rotated 360 degrees about a fixed axis. (b) The corresponding epipolar-plane image with the set of visible depth edges shown in black. (c) Hidden depth discontinuities are included in the EPI. (d) The first 180 degrees of B. (e) The second 180 degrees of B, (f) E flipped vertically. (f) "Fishtail" ends are matched by overlapping D and F. Note that the original T-junction in B denotes a local concavity.

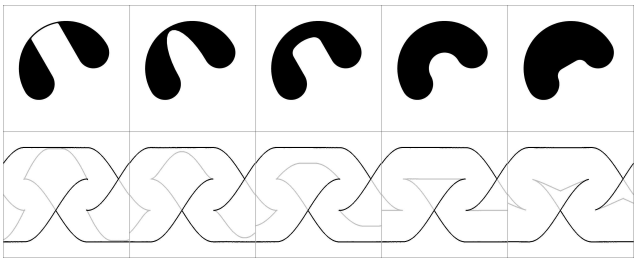


Figure 2. Different shapes that produce the same depth edges in the epipolar-plane image. Black and gray lines denote the visible and hidden depth discontinuities, respectively. Note the variability in the shape and location of the curve joining the "fishtail" ends.

3.1. Orthographic Projection Model

The orthographic projection $\boldsymbol{p} = [u, v]^{\mathsf{T}}$ of a point $\boldsymbol{P} = [X, Y, Z]^{\mathsf{T}}$, defined in world coordinates, can be modeled as

$$\widetilde{\boldsymbol{p}} = \boldsymbol{K}\boldsymbol{E}\widetilde{\boldsymbol{P}}, \tag{2}$$

where $\widetilde{\boldsymbol{x}}$ denotes the homogeneous vector obtained by appending 1 to the last element, and where the camera intrinsic matrix $\boldsymbol{K}$ and extrinsic parameters $\boldsymbol{E}$ are given by

$$\boldsymbol{K} = \begin{bmatrix} \alpha & \gamma & 0 & 0 \\ 0 & \beta & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{E} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{T} \\ \boldsymbol{0} & 1 \end{bmatrix}.$$

The elements of $\boldsymbol{K}$ include the image skew factor $\gamma \in \mathbb{R}$ and the scale factors $\{\alpha, \beta\} \in \mathbb{R}^+$. The matrix $\boldsymbol{E}$ represents the rigid transformation that brings the world coordinate system to that of the camera, composed of a rotation $\boldsymbol{R} \in \mathbb{R}^{3\times3}$ and translation $\boldsymbol{T} \in \mathbb{R}^3$. The details of estimating the parameters of this model, for a given orthographic imaging system, are described in Section 5.

For this model, consider the case of a cylinder of radius r separated from the rotation axis by a distance R (with its symmetry axis parallel to the rotation axis). The EPI for image row v will consist of two depth edge contours $u_+(\theta)$ and $u_-(\theta)$, given by

$$u_\pm(\theta) = \alpha R \cos(\theta) \pm \alpha r, \tag{3}$$

assuming $\gamma = 0$. Locally, the surface boundary can be approximated by an osculating circle; as a result, we expect the EPI image to consist of intersecting sinusoidal contours (as evidenced by the EPI image shown in Figure 3).

3.2. Basic Concepts

Epipolar-plane images produced using a setup composed of an orthographic camera and a turntable, as described above, show interesting symmetries which are easy to describe. There is a one-to-one correspondence between tangent rays (*i.e.*, depth edges) sampled in images taken from opposite points of view with respect to the rotation axis. Two corresponding tangent rays are supported by the same straight line, and have opposite directions. Thus, the depth discontinuities detected in opposite images are spaced 180 degrees apart in epipolar-plane images; since they are viewed from opposite directions, their projection in the EPI is related by an additional mirror symmetry. Furthermore, points on the visual hull are visible from both directions. As a result, they produce two visible depth discontinuities and two corresponding points in the EPI. Other depth discontinuities are generated by tangent rays visible from only one direction, yielding only one point in the EPI. These effects are illustrated in Figure 1.

3.3. Detecting and Filling Local Concavities

As shown in Figure 1, a set of cusps will be present in the EPI containing both visible and hidden depth discontinuities. These cusps correspond to positions where locally convex visible points on the surface transition into locally concave hidden points. Every concavity in the primal curve maps to a "fishtail" structure in the EPI. In other words, for a simple concavity, a T-junction will be present in the EPI, corresponding to a point of bitangency between the viewing ray and the surface [1]. By first shifting each EPI by 180 degrees and reflecting it about the rotation axis, we can recover the two sides of the concavity (up to the points of inflection on either side). As a result, T-junctions can be used to isolate corresponding points on either side of a concavity. We note that higher-order junctions, while improbable, correspond to points of multiple-tangency and can be processed in a similar manner.

We propose reconstructing the shape of the hidden surface within the concavity by first matching the "fishtail" ends, and then connecting them with a curved segment tangent to the two ends. As described in Section 2.4, such "hallucinated" curves have multiple solutions (see Figure 2). In our implementation we chose to use a piecewise cubic Hermite interpolating polynomial [11], however this choice is subject to the application and underlying properties of the object being scanned.

4. Orthographic Multi-Flash Photography

To verify the basic theory, we constructed a large-format orthographic multi-flash camera capable of recovering depth discontinuities using controlled illumination. In this section we describe the properties of multi-flash depth edge detection under orthographic projection. We begin by analyzing the illumination and imaging conditions required to detect depth discontinuities. Consider the case when point lights are used with an orthographic camera, with a point source located at $\boldsymbol{L} = [X_L, Y_L, Z_L]^\mathsf{T}$, a depth discontinuity at $\boldsymbol{D} = [X_D, Y_D, Z_D]^\mathsf{T}$, and a backdrop in the plane $Z = Z_B$. The separation Δ_X of the outside edge of the cast shadow from the depth discontinuity, along the X-axis, is

$$\Delta_X = \frac{-(X_L - X_D)(Z_B - Z_D)}{Z_D - Z_L}.$$

Unlike the case of point light sources under perspective projection [19], several limitations are imposed by orthographic imaging. First, the light source must be located to the right of the depth edge, in order to cast a shadow to the left (*i.e.*, $X_L > X_D$). As a result, the point light sources must be located outside the region of interest; otherwise, shadows will not be visible. Second, the width of the cast shadow increases with the separation of the depth edge from the light source. This leads to a third limitation; disconnected shadows will be more likely further from the light source, with the following constraint on the minimum width Δ_D of a fronto-parallel plane to the camera.

$$\Delta_D > \frac{(X_L - X_D)(Z_B - Z_D)}{Z_B}$$

A simple solution to overcome these limitations is to use *directional* lights rather than point sources. In this case, the width of the shadow is invariant to the position of the depth discontinuity. As for perspective projection [19], disconnected shadows can be eliminated by using a small angular variation between directional lights. We observe that directional lights achieve the original design goals of multi-flash imaging; specifically, the effective center of projection is now located at $Z = -\infty$. Since multi-flash edge detection requires point sources close to the center of projection, one naturally concludes that directional lights are the appropriate sources for the case of orthographic imaging.

4.1. Data Capture and Depth Edge Estimation

Our data capture process follows Crispell *et al.* [7]. An opaque object is placed on a turntable and a sequence of n_θ orthographic images, each with a resolution of $n_v \times n_u$, is recorded using point light sources, located outside the entrance aperture of the telecentric lens (see Section 6.1). The axis of rotation is manually-adjusted to be located in a plane parallel to the image plane. Each image is rectified so the projection of the rotation axis is aligned with the central image column. As shown in Figure 3, each multi-flash sequence is decoded to obtain a *depth edge confidence image* [19], estimating the likelihood of the projection of a depth discontinuity being located in any given pixel $\boldsymbol{p} = [u, v]^\mathsf{T}$. A set of n_v epipolar-plane images is extracted by concatenating the confidence values for each image row v over the turntable sequence. Each EPI consists of an $n_\theta \times n_u$ slice of the confidence volume.

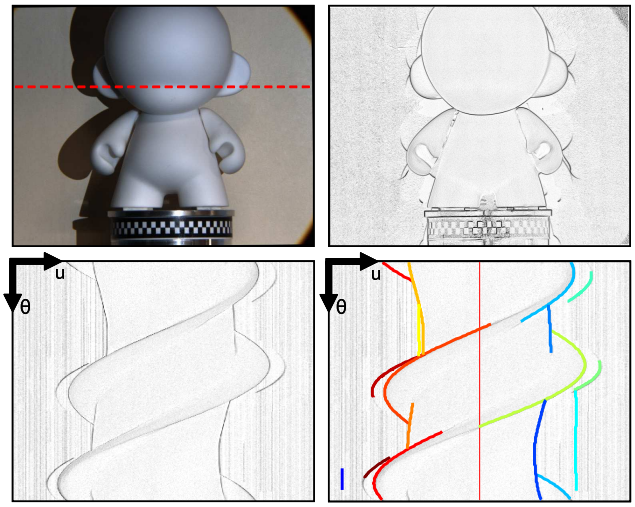


Figure 3. Orthographic multi-flash depth edge detection. (Top left) Image acquired with a flash located to the right of the scene. Note the cylindrical calibration pattern affixed to the turntable surface. (Top right) Depth edge confidence image estimated using the method of Raskar *et al.* [19]. (Bottom left) An epipolar-plane image extracted along the dashed red line shown above. (Bottom right) Detected ridges modeled as trigonometric polynomials. All images are rectified so the rotation axis is the center column. Additional detection examples are provided in the supplement.

4.2. Epipolar-Plane Image Analysis

As described in Section 3, each EPI can be processed in parallel to estimate the depth of each discontinuity, to identify local concavities, and to find smooth curves that fill remaining gaps. To accomplish these tasks, we require a parametric model for EPI ridges (see Figure 3). Cipolla and Giblin [6] use B-spline snakes to track silhouette boundaries over turntable sequences. We note, however, that the ridges observed in depth edge confidence images contain high-frequency features and numerous junctions. In contrast, the ridges in the EPI are generally low in frequency and contain a small number of isolated junctions. As a result, we propose tracking ridges in the EPI domain directly. Specifically, we use a modified tracking procedure previously proposed for this problem [7]. We refer the reader to that paper for the specific details of their tracking algorithm.

Rather than modeling EPI ridges with quadratic polynomials [7], we use real trigonometric polynomials of order N ($N = 3$ in our implementation), such that

$$u_i(\theta) = a_{i0} + \sum_{n=1}^{N} a_{in}\cos(n\theta) + \sum_{n=1}^{N} b_{in}\sin(n\theta), \quad (4)$$

where $u_i(\theta)$ models ridge i. Unlike polynomials, trigonometric polynomials capture 2π-periodicity in the orthographic turntable sequences. Furthermore, a 1^{st}-order trigonometric polynomial exactly models the EPI curve generated by a cylinder under constant rotation given by (3). For a piecewise-smooth surface, we expect both the curvature and distance of the external boundary from the rotation axis will be piecewise-smooth functions. Thus, the surface can be represented as a sequence of osculating circles, each generating a 1^{st}-order trigonometric polynomial for small rotations. In this manner, we expect a low-order trigonometric polynomial to closely-approximate EPI ridges.

Typical edge detection and linking results are shown in Figure 3. Note that the tracking algorithm fails to link contours across some junctions and only recovers the visible set of depth discontinuities. To compensate, the additional set $\{\bar{u}_i(\theta)\}$ of EPI curves is obtained by shifting each tracked contour by 180 degrees and reflecting about the rotation axis (as described in Section 3.2), such that

$$\bar{u}_i(\theta) = (-a_{i0} + 2u_0) + \sum_{n=1}^{N}(-1)^{n+1}a_{in}\cos(n\theta) + \sum_{n=1}^{N}(-1)^{n+1}b_{in}\sin(n\theta), \quad (5)$$

where u_0 is the column containing the projected axis of rotation. The superset $\{u'_i(\theta)\} = \{u_i(\theta)\} \cup \{\bar{u}_i(\theta)\}$ is processed to obtain an estimate of the apparent contours (*i.e.*, both visible and hidden depth discontinuities). Specifically, any pair of curves $u'_i(\theta)$ and $u'_j(\theta)$, for $i \neq j$, are joined if a significant region of overlap is detected. Typical refined apparent contour estimates are shown in Figure 4.

Following Section 3.3, local concavities correspond to T-junctions in the EPI. We first estimate a candidate set of local concavities using the pairwise-intersection of contours in $\{u'_i(\theta)\}$. Each contour pair produces a trigonometric polynomial whose coefficients are given by the difference of the two coefficients for each curve, the roots of which correspond to the points of intersection [21]. Only those points of intersection that are near T-junctions are retained. In our implementation we use steerable filters [23, 1] to estimate the T-junction likelihood at each point in the EPI.

4.3. Surface Reconstruction

The set of tracked EPI contours and calibration are used to recover the scene depth using (1). The analytic derivatives $\{\frac{\partial u'_i}{\partial \theta}(u'_i(\theta), \theta)\}$, required for reconstruction, are computed using $\{u'_i(\theta)\}$. The remaining task is to "hallucinate" curves connecting depth discontinuities on either side of each T-junction. As described, numerous completions exist; for instance, requiring continuity of position and tangents between matched EPI contours imposes four constraints, resulting in a one-parameter family of 2^{nd}-order trigonometric polynomials. In our implementation we use Matlab to fit a cubic Hermite interpolating polynomial to either side of each local concavity. Results are shown in Figure 4.

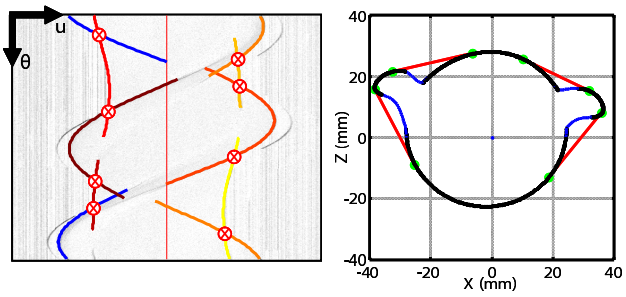


Figure 4. Epipolar-plane image analysis and curve completion. (Left) The tracked contours after merging detected segments (*e.g.*, those shown in Figure 3) with the segments given by (5). Segments due to disconnected shadows are rejected by eliminating any curves projecting outside the visual hull. T-junctions corresponding to local concavities are indicated with red crosses. (Right) Recovered external boundary for corresponding image row (shown in black). Points on either side of each concavity, corresponding to EPI T-junctions, are linked with red lines. The piecewise cubic Hermite polynomial completions are shown in blue. Additional EPI reconstruction results are included in the supplement.

5. Orthographic Camera Calibration

Surface reconstruction using (1) requires accurate camera calibration. In this section we describe a general method for estimating the parameters of the orthographic imaging model in Section 3.1. While Cipolla and Giblin [6] propose a suitable method, their approach involves simultaneous solution of the intrinsic and extrinsic parameters. Generalized methods have also been developed [24]. However, we propose a factorized method inspired by Zhang [29], in which intrinsic and extrinsic parameters are separately estimated from multiple images of a planar checkerboard pattern.

By convention the checkerboard lies in the plane $Z = 0$, eliminating the third row/column from $\boldsymbol{E}$ so (2) becomes

$$\tilde{\boldsymbol{p}} = \begin{bmatrix} \alpha & \gamma & 0 \\ 0 & \beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{12} & t_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \boldsymbol{K}_s \boldsymbol{E}_s \tilde{\boldsymbol{P}}_s,$$

where $\boldsymbol{R}_s \in \mathbb{R}^{2\times 2}$ and $\boldsymbol{T}_s \in \mathbb{R}^2$ represent the corresponding truncated rotation and translation matrices, respectively.

5.1. Estimation of Intrinsic Parameters

Assuming that the homography $\boldsymbol{H} \in \mathbb{R}^{3\times3}$ mapping the points on the checkerboard to those in the image [29] has already been computed, we have $\widetilde{\boldsymbol{p}} = \boldsymbol{H}\widetilde{\boldsymbol{P}}$ such that

$$\boldsymbol{H} = \boldsymbol{K}_s\boldsymbol{E}_s. \tag{6}$$

Multiplying both sides of this expression by $\boldsymbol{K}_s^{-1}$ and then multiplying both sides by their respective transposes yields

$$\boldsymbol{H}^{\mathsf{T}}(\boldsymbol{K}_s^{-1})^{\mathsf{T}}\boldsymbol{K}_s^{-1}\boldsymbol{H} = \begin{bmatrix} \boldsymbol{R}_s^{\mathsf{T}}\boldsymbol{R}_s & \boldsymbol{R}_s^{\mathsf{T}}\boldsymbol{T}_s \\ \boldsymbol{T}_s^{\mathsf{T}}\boldsymbol{R}_s & \boldsymbol{T}_s^{\mathsf{T}}\boldsymbol{T}_s \end{bmatrix}. \tag{7}$$

We note that the upper left submatrix $\boldsymbol{R}_s^{\mathsf{T}}\boldsymbol{R}_s$ can be used to solve for the intrinsic calibration parameters. First, the rotation matrix $\boldsymbol{R}$ is expressed in terms of $\boldsymbol{R}_s$ as

$$\begin{bmatrix} \boldsymbol{R}_s & \boldsymbol{A} \\ \boldsymbol{B}^{\mathsf{T}} & c \end{bmatrix} = \boldsymbol{R},$$

where $\{\boldsymbol{A}, \boldsymbol{B}\} \in \mathbb{R}^2$ and $c \in \mathbb{R}$. Next, both sides of this expression are multiplied, on the left, by their transposes and the upper left 2×2 submatrix is extracted to obtain

$$\boldsymbol{R}_s^{\mathsf{T}}\boldsymbol{R}_s + \boldsymbol{B}\boldsymbol{B}^{\mathsf{T}} = \boldsymbol{I}_2,$$

where $\boldsymbol{I}_2$ denotes the 2×2 identity matrix. Since $\boldsymbol{B}^{\mathsf{T}}\boldsymbol{B}$ is rank deficient, we find that

$$\det(\boldsymbol{R}_s^{\mathsf{T}}\boldsymbol{R}_s - \boldsymbol{I}_2) = 0. \tag{8}$$

Equations (8) and (7) can be combined to obtain

$$x_1 - x_2\left(h_{21}^2 + h_{22}^2\right) - x_3\left(h_{11}^2 + h_{12}^2\right) + 2x_4\left(h_{11}h_{21} + h_{12}h_{22}\right) = -(h_{12}h_{21} - h_{11}h_{22})^2,$$

where $x_1 = \alpha^2\beta^2$, $x_2 = \alpha^2 + \gamma^2$, $x_3 = \beta^2$, and $x_4 = \beta\gamma$. Since there is one homography for each image of the checkerboard, a minimum of four images are required to recover the unknowns $\{x_i\}$. The intrinsic parameters α, β, and γ can then be recovered from $\{x_i\}$.

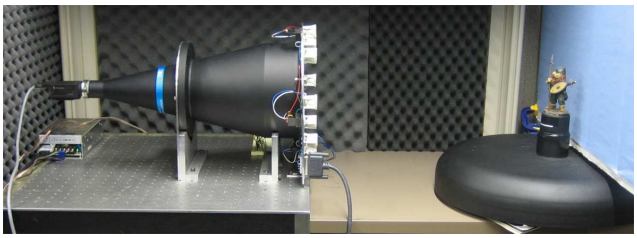
5.2. Estimation of Extrinsic Parameters

Following intrinsic calibration, the extrinsic parameters $\boldsymbol{R}_s$ and $\boldsymbol{T}_s$ can be recovered using (6). The full rotation matrix $\boldsymbol{R}$ can be recovered from $\boldsymbol{R}_s$ by requiring the columns of $\boldsymbol{R}$ to be orthogonal and unitary. The following set of constraints are given by ensuring orthonormality.

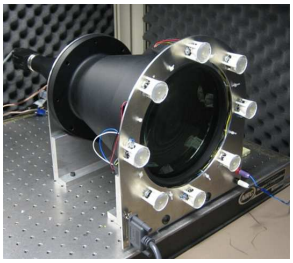
$$\boldsymbol{r}_3 = \boldsymbol{r}_1 \times \boldsymbol{r}_2 \tag{9}$$

$$r_{3i} = \pm\sqrt{1 - r_{1i}^2 - r_{2i}^2}, \text{ for } i = \{1, 2\} \tag{10}$$

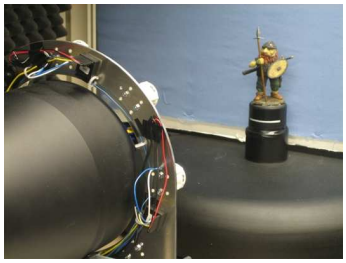
$$r_{31}r_{32} = -r_{11}r_{12} - r_{21}r_{22} \tag{11}$$



(a) system architecture for orthographic multi-flash photography



(b) multi-flash camera



(c) turntable, backdrop, and object

Figure 5. System architecture for multi-flash detection of depth discontinuities under orthographic projection. (a) The system contains an orthographic camera, a multi-flash LED array, and a computer-controlled turntable. (b) A digital camera is mounted to a large-format telecentric lens, with an LED array surrounding the entrance aperture. (c) The object is placed on a turntable, with a backdrop included to facilitate silhouette detection.

Equation (10) defines r_{31} and r_{32} up to an ambiguity in sign, with two degrees of freedom remaining in the solution. Equation (11) restricts the sign so that only one degree of freedom is left. The remaining ambiguity cannot be recovered and corresponds to the fact that the orthographic projection of a planar object is invariant when the object is reflected across the XY-plane. To recover the correct sign, extra information is required; in our implementation we modify the checkerboard plane with a pole of known height, perpendicular to the surface (see the supplement for more details). This pole is parallel to the third column of the rotation matrix and, by dividing the coordinates (in world units) of its projection by the physical length of the pole, we can recover the first two components of the third column of the rotation matrix. Finally, (9) is used to obtain the last elements of the first two columns of the rotation matrix.

$$r_{3i} = \frac{r_{1i}r_{13} + r_{2i}r_{23}}{\det(\boldsymbol{R}_s)}, \text{ for } i = \{1, 2\}.$$

5.3. Optimization of Model Parameters

We further refine the initial parameter estimates by minimizing the distance between the measured and predicted image coordinates of the checkerboard corners, given by

$$\sum_{j=1}^{n_{\mathcal{C}}}\sum_{i=1}^{n_{\mathcal{P}}}\left\|\widetilde{\boldsymbol{p}}_{ij} - \boldsymbol{K}_s\boldsymbol{E}'_{sj}\widetilde{\boldsymbol{P}}_{si}\right\|^2, \tag{12}$$

where $n_{\mathcal{C}}$ is the number of cameras, $n_{\mathcal{P}}$ is the number of checkerboard corners, and $\{\boldsymbol{E}'_{s1}, \boldsymbol{E}'_{s2}, \ldots, \boldsymbol{E}'_{sn_{\mathcal{C}}}\}$ are the extrinsic calibration parameters for each checkerboard pose. This error function is minimized using the Levenberg-Marquardt algorithm. Even if we do not use the modified checkerboard pattern, Equation (12) can still be used to refine our estimate of the intrinsic parameters; in this case, it suffices to arbitrarily choose the remaining sign for $\boldsymbol{R}$ (resulting in the same projected image).

6. Implementation and Results

6.1. Implementation

Following the design in Section 4, we constructed the prototype shown in Figure 5. The orthographic imaging system consists of a 1600×1200 24-bit color video camera from Point Grey (model GRAS-20S4M/C), mounted with a large-format bi-telecentric lens from Opto-Engineering (model TC 12 144) with a field of view of 81.0mm×60.7mm. The multi-flash illumination system is composed of an array of eight Philips White Luxeon III LEDs (model LXHL-LW3C), individually controlled via the parallel port using BuckPuck 1000mA DC LED drivers. The object is rotated using a Kaidan Magellan Desktop Turntable (model MDT-19). A backdrop is placed behind the object to aide in silhouette extraction.

A typical capture sequence consists of 670 viewpoints, separated by a rotation of approximated 0.527 degrees. For each viewpoint four images are recorded in which the scene is sequentially illuminated by the top, bottom, left, and right flashes. The intrinsic parameters of the camera are measured once, using the procedure described in Section 5. The extrinsic calibration is estimated by tracking the corners of the cylindrical checkerboard pattern shown in Figure 3. The various post-processing steps, including image rectification, depth edge estimation, epipolar-slice analysis, and reconstruction were implemented in Matlab and were evaluated on a PC cluster.The data capture process can be completed in under two hours, and the reconstruction pipeline requires several hours using the current implementation.

6.2. Results

To evaluate the overall system performance, the test object in Figure 3 was reconstructed using the procedure from Section 4. Each EPI, similar to Figure 4, was evaluated independently. Gaps were filled using the procedure from Section 4.3. The reconstruction results, with and without curve completion, are compared in Figure 6. Additional results are included in the supplementary material. Note that independent surface completions exhibit strong coherent across neighboring layers. While post-processing of the point cloud was not applied in this example, such processing would further improve inter-layer coherence.

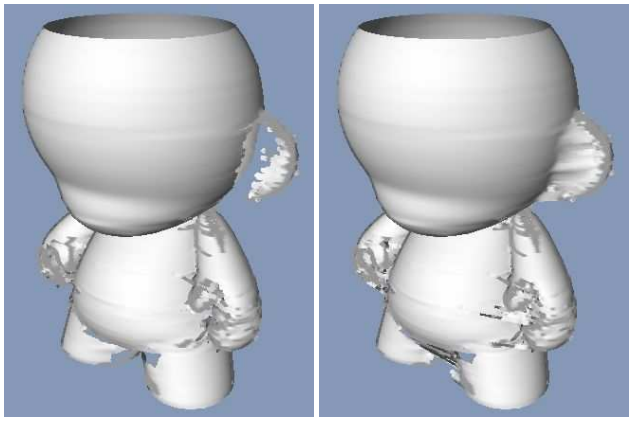


Figure 6. 3-D surface reconstructions without (left) and with (right) curve completion. Each EPI, similar to Figure 4, was reconstructed independently using the method in Section 4. All renderings were produced using Pointshop3D [30], with surface normals obtained using the Diffuser plug-in. Shading variation is due to surface normal estimation errors using the plug-in. Additional views of this model are included in the supplement.

7. Discussion and Future Work

7.1. Benefits and Limitations

Benefits: Recovered points are obtained along surface patches oriented orthogonal to the viewing direction, allowing a subset of feature lines to be directly sampled. The reconstruction and completion schemes are highly parallel. The unique properties of orthographic imaging allow a subset of apparent contours to be estimated, including both visible and hidden depth discontinuities. Because the sampling rate of reconstructed points is proportional to surface curvature, we recover points that are sparsely-sampled by many optical shape capture methods. Similarly, the proposed shape completion scheme can be evaluated independently for each EPI. While this can lead to artifacts, it also facilitates parallel processing.

Limitations: Recovering 3-D shapes using the proposed procedure has several significant drawbacks. Because an orthographic camera is required, only a small volume can be scanned, since large-format telecentric lenses are typically cumbersome and expensive. The properties described in Section 3 only hold under ideal orthographic projection. In practice it is challenging to ensure proper mechanical alignment of the rotation axis to achieve this condition. Because each EPI image is processed independently, the "hallucinated" completions may be inconsistent between layers. As described in Section 4, using point light sources with orthographic imaging is sensitive to disconnected shadows. The model-based detection and linking of EPI ridges is susceptible to errors and, as a result, certain T-junctions may not be detected, leaving gaps in the recovered surface.

7.2. Future Work

In the near term, several limitations can be addressed by a revised prototype. Detached shadows can be eliminated using directional illumination. We propose a simple solution, in which a beam-splitter is positioned at a 45 degree angle in front of the telecentric lens. A Fresnel lens can be placed on the new optical path, with a small-baseline LED array located at its focus, achieving directional lighting. In the long term, we are interested in exploring the problem of shape from depth discontinuities. We are encouraged that our approach acquires samples traditionally lost by many optical scanning methods. Future work will examine the traditional case of perspective imaging with algorithms that are less susceptible to EPI segmentation and tracking errors.

8. Conclusion

While shape-from-X is a crowded field, shape from depth discontinuities remains a promising direction of future research. Scanning the feature lines typical of machined parts remains a challenging problem. In this paper we have proposed several practical solutions, including characterizing the performance of multi-flash photography under orthographic projection and robust calibration of telecentric lenses. We have also made several theoretical contributions to the understanding of the visual motion of depth discontinuities, including procedures for modeling, linking, and completing gaps using properties unique to orthographic projection. We hope our methods and prototype inspire researchers to continue applying non-traditional cameras to the established field of shape-from-X.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. CCF-0729126.

References

[1] N. Apostoloff and A. Fitzgibbon. Learning spatiotemporal T-junctions for occlusion detection. In *CVPR*, 2005.

[2] H. Baker and R. Bolles. Generalizing epipolar-plane image analysis on the spatiotemporal surface. In *IJCV*, 1989.

[3] J. Berent and P. Dragotti. Segmentation of epipolar-plane image volumes with occlusion and disocclusion competition. In *IEEE Multimedia Signal Processing*, 2006.

[4] F. Blais. Review of 20 years of range sensor development. *Journal of Electronic Imaging*, 13(1), 2004.

[5] R. C. Bolles, H. H. Baker, and D. H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. In *IJCV*, 1987.

[6] R. Cipolla and P. Giblin. *Visual Motion of Curves and Surfaces*. Cambridge University Press, 2000.

[7] D. Crispell, D. Lanman, P. G. Sibley, Y. Zhao, and G. Taubin. Beyond silhouettes: Surface reconstruction using multi-flash photography. In *3DPVT*, 2006.

[8] B. Curless and M. Levoy. A volumetric method for building complex models from range images. In *SIGGRAPH*, 1996.

[9] J. Davis, S. Marschner, M. Garr, and M. Levoy. Filling holes in complex surfaces using volumetric diffusion. In *3DPVT*, 2002.

[10] I. Feldmann, P. Eisert, and P. Kauff. Extension of epipolar image analysis to circular camera movements. In *ICIP*, 2003.

[11] F. N. Fritsch and R. E. Carlson. Monotone piecewise cubic interpolation. *SIAM J. Num. Analysis*, 17:238–246, 1980.

[12] B. Horn. The curve of least energy. *ACM Trans. Math. Soft.*, 9(4), 1983.

[13] B. Kimia, I. Frankel, and A. Popescu. Euler spiral for shape completion. In *IJCV*, 2003.

[14] D. Knuth. Mathematical typography. *American Mathematical Society*, 1(2), 1979.

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

[16] Y. Li, S. Lin, H. Lu, and H.-Y. Shum. Multiple-cue illumination estimation in textured scenes. In *ICCV*, 2003.

[17] S. Mada, M. Smith, L. Smith, and P. Midha. Overview of passive and active vision techniques for hand-held 3D data acquistion. In *SPIE*, 2003.

[18] M. Pauly, N. J. Mitra, J. Wallner, H. Pottmann, and L. Guibas. Discovering structural regularity in 3D geometry. *ACM Trans. Graph.*, 27(3), 2008.

[19] R. Raskar, K.-H. Tan, R. Feris, J. Yu, and M. Turk. Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. *ACM Trans. Graph.*, 23(3), 2004.

[20] J. Salvi, J. Pagès, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37:827–849, 2004.

[21] A. Schweikard. Real zero isolation for trigonometric polynomials. *ACM Trans. Math. Soft.*, 18(3), 1992.

[22] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR*, 2006.

[23] E. P. Simoncelli and H. Farid. Steerable wedge filters for local orientation analysis. *IEEE TIP*, 5, 1996.

[24] P. Sturm, S. Ramalingam, and S. Lodha. On calibration, structure from motion and multi-view geometry for generic camera models. Computational Imaging and Vision. Springer, 2006.

[25] S. Ullman. Filling-in the gaps: The shape of subjective contours and a model for their generation. *Biological Cybernetics*, 25(1), 1976.

[26] J. Verdera, V. Caselles, M. Bertalmio, and G. Sapiro. Inpainting surface holes. In *ICIP*, 2003.

[27] M. Watanabe and S. K. Nayar. Telecentric optics for computational vision. In *ECCV*, 1996.

[28] R. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1), 1980.

[29] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *ICCV*, 1999.

[30] M. Zwicker, M. Pauly, O. Knoll, and M. Gross. Pointshop 3D: an interactive system for point-based surface editing. In *SIGGRAPH*, 2002.