# Coplanar Shadowgrams for Acquiring Visual Hulls of Intricate Objects

Shuntaro Yamazaki[*†] shuntaro@ni.aist.go.jp
Srinivasa Narasimhan[†] srinivas@cs.cmu.edu
Simon Baker[‡] sbaker@microsoft.com
Takeo Kanade[†*] tk@cs.cmu.edu

## Abstract

*Acquiring 3D models of intricate objects (like tree branches, bicycles and insects) is a hard problem due to severe self-occlusions, repeated thin structures and surface discontinuities. In theory, a shape-from-silhouettes (SFS) approach can overcome these difficulties and use many views to reconstruct visual hulls that are close to the actual shapes. In practice, however, SFS is highly sensitive to errors in silhouette contours and the calibration of the imaging system, and therefore not suitable for obtaining reliable shapes with a large number of views. We present a practical approach to SFS using a novel technique called coplanar shadowgram imaging, that allows us to use dozens to even hundreds of views for visual hull reconstruction. Here, a point light source is moved around an object and the shadows (silhouettes) cast onto a single background plane are observed. We characterize this imaging system in terms of image projection, reconstruction ambiguity, epipolar geometry, and shape and source recovery. The coplanarity of the shadowgrams yields novel geometric properties that are not possible in traditional multi-view camera-based imaging systems. These properties allow us to derive a robust and automatic algorithm to recover the visual hull of an object and the 3D positions of light source simultaneously, regardless of the complexity of the object. We demonstrate the acquisition of several intricate shapes with severe occlusions and thin structures, using 50 to 120 views.*

## 1. Introduction

Acquiring 3D shapes of objects that have numerous occlusions, discontinuities and repeated thin structures is challenging for vision algorithms. For instance, the wreath object shown in Figure 1(a) contains over 300 branch-lets each 1-3mm in diameter and 20-25mm in length. Covering the entire surface area of such objects requires a large number (dozens or even a hundred) of views. Thus, finding correspondences between views as parts of the object get occluded and "dis-occluded" becomes virtually impossible, often resulting in erroneous and incomplete 3D models.

If we only use the silhouettes of an object obtained from different views, it is possible to avoid the issues of correspondence and occlusion in the object, and reconstruct its *visual hull* [1]. The top row of Figure 2 illustrates the visual hulls

(a) An intricate object (b) Our reconstruction

**Figure 1:** Obtaining 3D models of intricate shapes such as in (a) is hard due to severe occlusions and correspondence ambiguities. (b) By moving a point source in front of the object, we capture a large number of shadows cast on a single fixed planar screen (122 views for this object). Applying our techniques to such *coplanar shadowgrams* results in accurate recovery of intricate shapes.

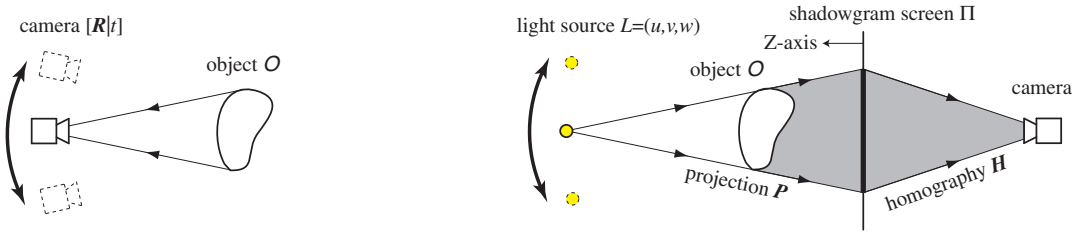
estimated using our technique from different numbers of silhouettes. While the visual hull computed using a few (5 or 10) silhouettes is too coarse, the reconstruction from a large number of views (50) is an excellent model of the original shape.

In practice, however, SFS algorithms are highly sensitive to errors in the geometric parameters of the imaging system (camera calibration) [15]. This sensitivity worsens as the number of views increases, resulting in poor quality models. The bottom row in Figure 2 shows the visual hulls of the wreath object obtained using a naïve SFS algorithm. This drawback must be addressed in order to acquire intricate shapes reliably.

In traditional SFS, a camera observes the object, and the silhouette is extracted from obtained images by matting [16]. Multiple viewpoints are captured by moving either the camera or the object (see Figure 3(a)). For each view, the relative pose between the object and the camera is described by six parameters (3D translation and 3D rotation). Savarese *et al*. [12] proposed a system that avoids silhouette matting. When an object is illuminated by a single point light source, the shadow cast onto a background plane (also known as a shadowgram [14]) is sharp and can be directly used as its silhouette. Silhouettes from multiple views are obtained by rotating the object. In terms of multi-view geometry, this is equivalent to traditional SFS, requiring six parameters per view.

In this paper, we present a novel approach to SFS called *coplanar shadowgram imaging*. We use a setup similar in spirit to that proposed by Savarese *et al*. [12] The key difference here is that the point source is moved, while the object, the

[*]National Institute of Advanced Industrial Science and Technology
[†]Carnegie Mellon University
[‡]Microsoft Research




(a) Traditional multi-view camera-based imaging

(b) Coplanar shadowgram imaging

**Figure 3:** (a) The object of interest is observed directly by a projective camera. The silhouette of the object is extracted from the captured image. Multiple views are obtained by moving the camera or the object. (b) A point source illuminates the object and its shadow cast on a planar rear-projection screen represents the silhouette of the object. Coplanar shadowgrams from multiple viewpoints are obtained by translating the light source. Note that the relative transformation between the object and the screen remains fixed across different views.

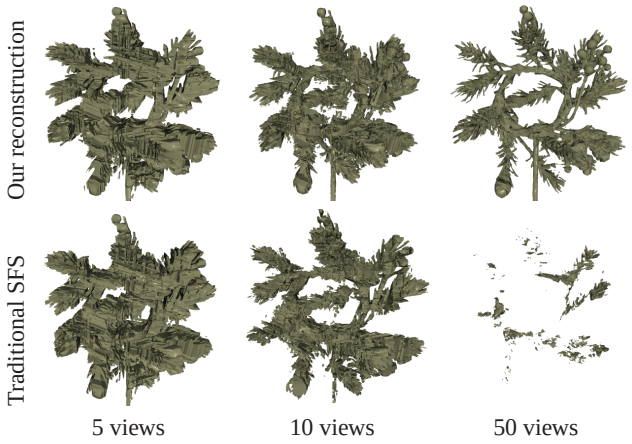



**Figure 2:** Sensitivity of SFS reconstruction. (Top) The visual hulls reconstructed using the light source positions estimated by our method. As the number of silhouettes increases, the visual hull gets closer to the actual shape. (Bottom) The reconstructions obtained from slightly erroneous source positions. As the number of views increases, the error worsens significantly.

camera and the background screen all remain stationary. The central focus of this work is the acquisition of visual hulls for intricate and opaque objects from a large number of coplanar shadowgrams. Our main contributions are described below.

**Multi-view Geometry of Coplanar Shadowgram Imaging:** Figure 3 shows the difference between the traditional camera-based and coplanar shadowgram imaging systems. Observe that the relative transformation between the object and screen remains fixed across different views. The image projection model is described by only three parameters per view (3D translation of the source) instead of six in the traditional system. Our geometry is similar in spirit to the parallax geometry [13, 3] where the homography between image planes is known to be an identity, which allows us to derive novel geometric properties that are not possible in the traditional multi-view camera-based imaging system. For instance, we show that epipolar geometry can be uniquely estimated from only the shadowgrams, without requiring any correspondences, and independent of the object's shape.

**Simple and Efficient Recovery of Source Positions:** When the shape of the object is unknown, the locations of all the point sources can be recovered from coplanar shadowgrams, only up to a four parameter linear transformation. In the technical report [18], we show how this transformation relates to the well-known *Generalized Perspective Bas-Relief* (GPBR) ambiguity [9] that is derived for a single viewpoint system. We break this ambiguity by simultaneously capturing the shadowgrams of two spheres.

**Robust Reconstruction of Visual Hull:** Even a small amount of blurring in the shadow contours may result in erroneous estimates of source positions that in turn can lead to erroneous visual hulls. We propose an optimization of the light source positions that can robustly reconstruct the visual hulls of intricate shapes. First, the large error in light source positions is corrected by enforcing the reconstructed epipolar geometry. We then minimize the mismatch between the acquired shadowgrams and those obtained by reprojecting the estimated visual hull. Undesirable local convergence in the non-linear optimization is alleviated using the convex polygons of the silhouette contours.

For the analogous camera-based imaging, a number of algorithms have been proposed to make SFS robust to errors in camera position and orientation. These techniques optimize camera parameters by exploiting either epipolar tangency [15, 2, 17] or silhouette consistency [19, 8], or assume orthographic projection [6]. However, they all require non-trivial parameter initializations and the knowledge of silhouette feature correspondences (known as frontier points [7]). This restricts the types of objects that one can reconstruct using these methods; silhouettes of simple objects such as spheres do not have enough features and intricate objects like branches have too many, making it hard to find correspondences automatically. As a result, previous approaches have succeeded in only acquiring the 3D shape of *reasonably complex* shapes like people and statues that can be modeled using a small number of views.

In contrast, our algorithm is effective for a large number of views (dozens to a hundred), does not require any feature cor-

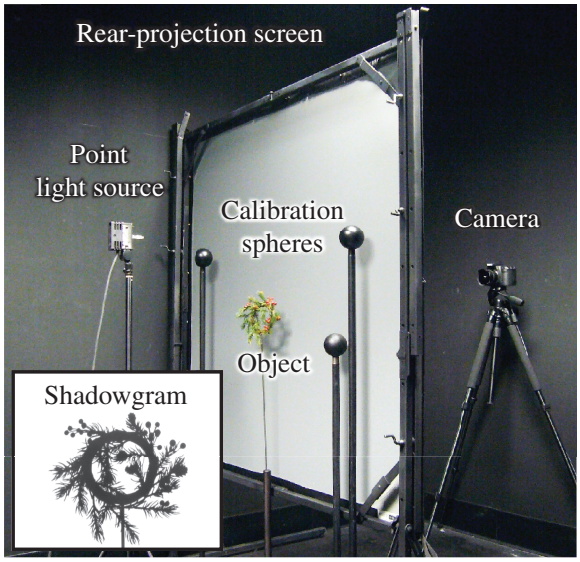



**Figure 4:** The setup used to capture coplanar shadowgrams includes a digital camera, a single point light source, and a rear-projection screen. The object is placed close to the screen to cover a large field of view. Two or more spheres are used to estimate the initial light source positions. (Inset) An example shadowgram obtained using the setup.

respondences and does not place any restriction on the shapes of the objects. The minimization of silhouette mismatch is also easier requiring optimization of source translation (3 DOF per view), instead of the harder (and sometimes ambiguous [7]) joint estimation of camera rotation and translation (6 DOF per view) in the traditional system. As a result, we achieve good quality reconstructions of real objects such as wreaths, wiry balls and palm trees, that show numerous occlusions, discontinuities and thin structures.

## 2. Coplanar Shadowgrams

We define shadowgrams as the shadows cast on a background plane by an object that occludes a point source. If the object is opaque, the shadowgram accurately represents the silhouette of the object. Henceforth, we shall use shadowgrams and silhouettes interchangeably. Coplanar shadowgram imaging is the process of acquiring several shadowgrams on *a single plane* by moving the light source. Our setup shown in Figure 4 includes a 6M pixel Canon EOS-20D digital camera, a 250 watt 4mm incandescent bulb, and a 4ft $\times$ 4ft translucent rear-projection screen.

Figure 3(b) illustrates the viewing and illumination geometry of coplanar shadowgram imaging. Without loss of generality, let the shadowgram plane $\Pi$ be located at $Z = 0$ and the $Z-$direction be perpendicular to $\Pi$. Suppose a point light source is at $L = (u, v, w)^T$, then the resulting shadowgram $S$ is obtained by applying a source dependent projective transformation $\boldsymbol{P}(L)$ to the object $O$ as:

**Table 1:** Comparison between the geometric parameters of silhouette projection. For $n$ views, the traditional multi-view system is described by $5 + 6n$ parameters. In comparison, the coplanar imaging system requires only $8 + 3n$ parameters.

| | View independent | View dependent |
|---|---|---|
| Projective cameras | 1 (focal length)<br>1 (aspect ratio)<br>1 (skew)<br>2 (image center) | 3 (rotation)<br>3 (translation) |
| Coplanar shadowgrams | 8 (homography $\boldsymbol{H}$) | 3 (translation $L$) |

$$S = \boldsymbol{P}(L)\,O \tag{1}$$

where, the projective transformation $\boldsymbol{P}(L)$ from 3D space to the 2D screen is (see the technical report [18] for the derivation):

$$\boldsymbol{P}(L) = \begin{pmatrix} -w & 0 & u & 0 \\ 0 & -w & v & 0 \\ 0 & 0 & 1 & -w \end{pmatrix}. \tag{2}$$

In Eq.(1), $S$ represents the set of 2D points (in homogeneous coordinates) within the shadowgram on the plane $\Pi$, and $O$ represents the 3D points on the object surface. The image $I$ captured by the camera is related to the shadowgram $S$ on the plane $\Pi$ by a 2D homography: $I = \boldsymbol{H}\,S$. This homography $\boldsymbol{H}$ is independent of the light source position and can be estimated separately using any computer vision algorithm (such as the four-point method [7]). In the following, we assume that the shadowgram $S$ has been estimated using $S = \boldsymbol{H}^{-1}\,I$.

Now let a set of shadowgrams $\{S_k\}$ be captured by moving the source to $n$ different locations $\{L_k\}$ $(k = 1, \cdots, n)$. Then, the visual hull $V$ of the object is obtained by the intersection:

$$V = \bigcap_k \boldsymbol{P}(L_k)^{-1}\,S_k. \tag{3}$$

Thus, given the 3D locations $L_k$ of the light sources, the visual hull of the object can be estimated using Eq.(2) and Eq.(3). Table 1 summarizes and contrasts the geometric parameters that appear in the traditional multi-view camera-based and coplanar shadowgram imaging systems.

## 3. Source Recovery using two Spheres

When the shape of the object is unknown, it is not possible to uniquely recover the 3D source positions using only the coplanar shadowgrams. In the technical report [18], we discuss the nature of this ambiguity and show that the visual hull and the source positions can be computed up to a 4 parameter linear transformation. This transformation is similar in spirit to the 4 parameter *Generalized Perspective Bas-Relief* (GPBR) transformation [9] with one difference: in the context of coplanar shadowgrams, the GPBR transformation is separately defined with respect to the local coordinate frame defined at each

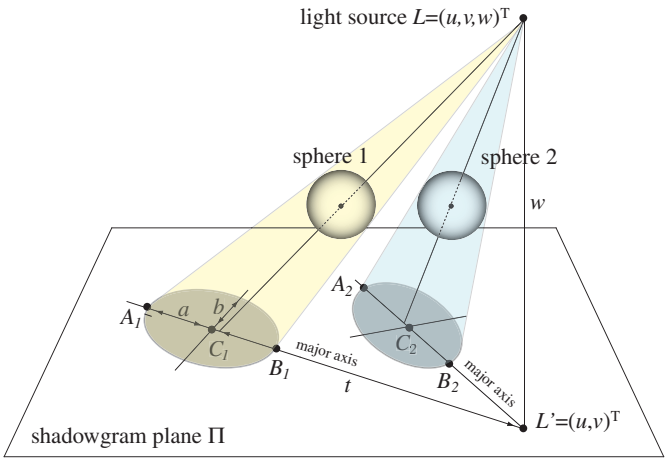



**Figure 5:** Source position $L = (u, v, w)^T$ is recovered using the elliptical shadowgrams of two spheres. The radii and positions of the spheres are unknown. The major axes of the ellipses intersect the screen at $L' = (u, v)^T$. The $w$ component is obtained using Eq.(4).

source location, whereas our ambiguity transformation is defined with respect to a global coordinate frame defined on the screen. We also derive a relationship between the two transformations.

We now present a simple calibration technique to break this ambiguity. The 3D location $L = (u, v, w)^T$ of a light source is directly estimated by capturing shadowgrams of two additional spheres that are placed adjacent to the object of interest. Figure 5 illustrates the coplanar elliptical shadowgrams cast by the two spheres. The ellipses are localized using a constrained least squares approach [4]. The intersection of the major axes $\overline{A_1B_1}$ and $\overline{A_2B_2}$ of the two ellipses yields two (out of three) coordinates $L' = (u, v)^T$ of the light source. The third coordinate $w$ is obtained as:

$$w = \sqrt{\frac{b^2t^2}{a^2 - b^2} - b^2} \tag{4}$$

where $a$ and $b$ are the semimajor and semiminor axes of one of the ellipses, and $t$ is the length between $L'$ and the center of the ellipse (see the technical report [18] for the derivation). Note that more than two spheres may be used for a robust estimate of the source position. The above method is completely automatic and does not require the knowledge of the radii of the spheres, the exact locations at which they are placed in the scene, or point correspondences.

**Sensitivity to silhouette blurring:** This technique of estimating the source position can be sensitive to errors in measured silhouettes especially when the cast shadow of the sphere is close to a right circle (i.e. $a^2 \approx b^2$) . Due to the finite size of the light bulb, the shadowgram formed may be blurred, making it hard to localize the boundary of the silhouette. The extent of blurring depends on the relative distances of the screen and source from the object. To show the sensitivity of the technique, we performed simulations with three spheres. We

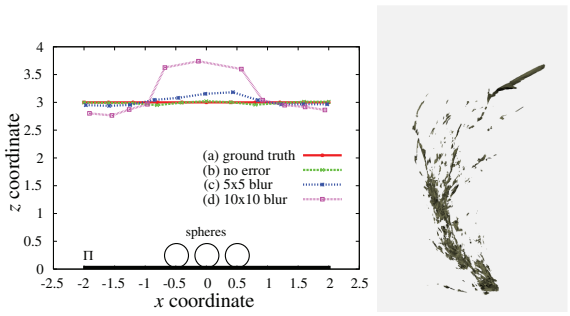



**Figure 6:** Source positions $(u, w)$ are estimated using three calibration spheres. The sizes and positions of the spheres and screen are shown in the plot. Each plot shows 11 source positions obtained from (a) ground truth, (b) accurate shadowgrams, and (c)-(d) shadowgrams blurred using $5 \times 5$ and $10 \times 10$ averaging filters. On the right is the visual hull of a branch reconstructed from 50 light sources. The poor result demonstrates the need for better algorithms for reconstructing intricate shapes.

blurred the simulated silhouettes (effective resolution 480×360 pixels) with $5 \times 5$ and $10 \times 10$ averaging kernels, and estimated the 3D coordinates of the light source using two elliptic shadows for which $a^2 - b^2$ are the largest. Figure 6 presents $u$ and $w$ components of the source positions reconstructed using three spheres. Observe that the estimation becomes poor when the source position is right above the spheres. In turn, the visual hull of a tree branch computed from the erroneous source positions is woefully inadequate. Thus, better algorithms for improving the accuracy of light source positions are crucial for obtaining 3D models of intricate shapes.

## 4. Epipolar Geometry

Analogous to the scenario of binocular stereo, we define the epipolar geometry between a pair of shadowgrams that are generated by placing the point source in two locations ($L_1$ and $L_2$ in Figure 7). Here, the locations of the point source are analogous to the centers-of-projection of the stereo cameras. The baseline connecting the two light sources $L_1$ and $L_2$ intersects the shadowgram plane Π at the epipole $E_{12}$. When the light sources are equidistant from the shadowgram plane Π, the epipole is at infinity. Based on these definitions, we make two key observations that do not hold for binocular stereo: since the shadowgrams are coplanar, (a) they share the *same epipole* and (b) the points on the two shadowgrams corresponding to the same scene point lie on the *same epipolar line*.

Let $L_i = (u_i, v_i, w_i)^T$ and $L_j = (u_j, v_j, w_j)^T$ be the 3D coordinates of the two light sources, and $E_{ij}$ be the homogeneous coordinate of the epipole on the plane Π, defined by $L_i$ and $L_j$. Then, the observations (a) and (b) are written as:

$$\boldsymbol{M}_{ij}\, E_{ij} = 0 \tag{5}$$

$$\boldsymbol{m}_i^T\, \boldsymbol{F}_{ij} \boldsymbol{m}_j = 0. \tag{6}$$

In Eq.(5), $\boldsymbol{M}_{ij}$ is a $2 \times 3$ matrix composed of two plane equa-

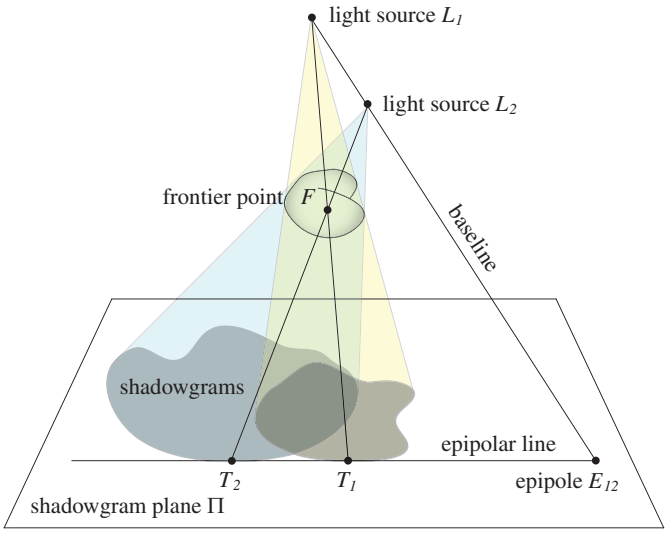



**Figure 7:** Epipolar geometry of two shadowgrams. The baseline connecting the two sources $L_1$ and $L_2$ intersects the shadowgram plane Π at an epipole $E_{12}$. Suppose an epipolar plane that is tangent to the surface of an object at a frontier point $F$, then the intersection of the epipolar plane and the shadowgram plane Π is an epipolar line. The epipolar line can be estimated as a line that is co-tangent to the shadowgrams at $T_1$ and $T_2$.

tions in the rows

$$\boldsymbol{M}_{ij} = \begin{pmatrix} -\Delta v & \Delta u & u_i v_j - u_j v_i \\ -\Delta u \Delta w & -\Delta v \Delta w & (u_i \Delta u + v_i \Delta v)\Delta w - w_i(\Delta u^2 + \Delta v^2) \end{pmatrix} \tag{7}$$

where, $\Delta u = u_j - u_i$, $\Delta v = v_j - v_i$, and $\Delta w = w_j - w_i$. In Eq.(6),

$$\boldsymbol{F}_{ij} = [E_{ij}]_\times \tag{8}$$

is the *fundamental matrix* that relates two corresponding points $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$ between shadowgrams. $[E_{ij}]_\times$ is the $3 \times 3$ skew symmetric matrix for which $[E_{ij}]_\times \boldsymbol{x} = E_{ij} \times \boldsymbol{x}$ for any 3D vector $\boldsymbol{x}$.

The camera geometry in coplanar shadowgram is similar in spirit to the parallax geometry [13, 3] where the image deformation is decomposed into a planar homography and a residual image parallax vector. In our system, however, the homography is exactly known to be an identity, which allows us to recover the epipolar geometry *only* from acquired images accurately regardless of the number of views or the complexity of the shadowgram contours.

### 4.1. Algorithm for estimating epipolar geometry

Suppose we have the plane in Figure 7 that includes the baseline and is tangent to the surface of an object at a *frontier point* $F$. The intersection of this plane and the shadowgram plane Π forms an epipolar line, which is also known as an *epipolar bitangent* [3], that can be estimated as one that is cotangent to the two shadowgrams (at $T_1$ and $T_2$ in Figure 7). Two such epipolar lines can then be intersected to localize the epipole.

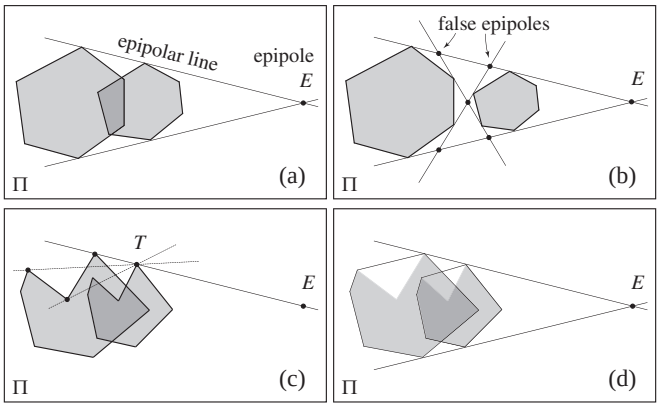



**Figure 8:** Localization of the epipole. (a),(b) If two shadowgrams are convex, a maximum of four co-tangent lines and six intersections are possible. Considering that the object and the light source are on the same side with respect to the screen, the epipole can be chosen uniquely out of the six intersections. (c),(d) If the shadowgrams are non-convex, the epipole is localized by applying the technique in (a) or (b) to the convex polygons of the original shadowgrams.

Figure 8(a) illustrates the simplest case of two convex shadowgrams partially overlapping each other. There are only two cotangent lines that touch the shadowgrams at the top and bottom region, resulting in a unique epipole $E$. When the convex shadowgrams do not overlap each other, four distinct cotangent lines are possible, generating six candidate epipoles, as shown by dots in Figure 8(b). We can detect actual epipolar lines by choosing the cotangent lines where the epipole does not appear between the two points of shadowgram tangency.

When shadowgrams are non-convex, the number of cotangent lines can be arbitrarily large depending on the complexity of the shadowgram contours. Figure 8(c) illustrates the multiple candidates of cotangent lines at the point of tangency $T$. In this case, we compute the convex polygon surrounding the silhouette contour as shown in Figure 8(d) and prove the following proposition (see the technical report [18] for the proof):
**Proposition** *The convex hulls of the silhouettes generated by an object are the silhouettes generated by the convex hull of the object.*

Using this proposition, the problem of estimating epipolar lines for concave silhouettes is reduced to the case of either (a) or (b). Thus, epipolar geometry can be reconstructed uniquely and automatically from only the shadowgrams. This capability of recovering epipolar geometry is independent of the shape of silhouette, and hence, the 3D shape of the object. Even when the object is a sphere, we can recover the epipolar geometry without any ambiguity. In traditional multi-view camera-based imaging, epipolar reconstruction requires at least seven pairs of correspondences [7]. Table 2 summarizes the differences between traditional imaging and coplanar shadowgrams in terms of recovering epipolar geometry.

For the special case where the baseline intersects a convex object, one convex silhouette lies completely within the other

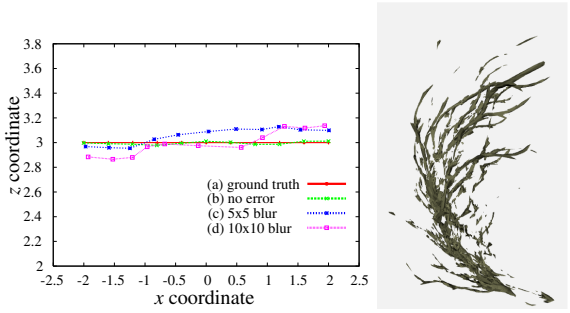



**Figure 9:** Initial light source positions in Figure 6 were improved by epipolar constraints in Eq.(9). On the right is the visual hull reconstructed from the improved source positions.

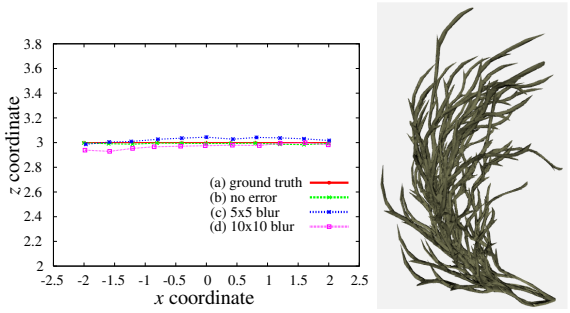



**Figure 10:** The light source positions reconstructed using epipolar constraint in Figure 9 were optimized by maximizing the shadowgram consistency in Eq.(14). On the right is the visual hull reconstructed from the optimized source positions.

and hence the epipole lies within the silhouettes. In this case, there are no frontier points formed (and hence no cotangent lines for convex silhouettes). We can avoid this case by placing the sources such that the baselines do not always intersect the object.

### 4.2. Improving accuracy of source locations

The error in the light source positions reconstructed using spheres can be arbitrarily large depending on the localization of the elliptical shadowgram for each sphere. This error can be reduced by relating different light source positions through the epipolar geometry. Let the set of epipoles $E_{ij}$ be estimated from all the source pairs $L_i$ and $L_j$. The locations of the sources are improved by minimizing the expression in Eq.(5) for each pair of light sources using least squares:

$$\{L_k^*\} = \underset{L_k}{\operatorname{argmin}} \sum_{i \neq j} \left\| \boldsymbol{M}_{ij}\, E_{ij} \right\|_2^2 \tag{9}$$

where $\|\cdot\|_2$ is the L2-norm of a vector. The source positions reconstructed from the shadowgrams of spheres are used as initial estimates. We evaluate this approach using the simulated silhouettes described in Figure 6. Figure 9 shows considerable improvement in accuracy obtained by enforcing the epipolar constraint in Eq.(5). Compared to the result in Figure 6, collinearity in the positions of light sources is better recovered in this example.

## 5. Using Shadowgram Consistency

While the epipolar geometry improves the estimation of the light source positions, the accuracy of estimate can still be insufficient for the reconstruction of intricate shapes (Figure 9). In this section, we present an optimization algorithm that improves the accuracy of all the source positions even more significantly.

### 5.1. Optimizing light source positions

Let $V$ be the visual-hull obtained from the set of captured shadowgrams $\{S_k\}$ and the estimated projection matrices $\{\boldsymbol{P}(L_k)\}$. When $V$ is re-projected back onto the shadowgram plane, we obtain the silhouettes $S_k^V$:

$$S_k^V = \boldsymbol{P}(L_k)\, V\,. \tag{10}$$

Due to the nature of the intersection operator, the re-projected silhouettes $S_k^V$ always satisfy:

$$\forall k : S_k^V \subseteq S_k\,. \tag{11}$$

Only when the source positions are perfect, will the reprojected silhouettes match the acquired silhouettes. Thus, we can define a measure of silhouette mismatch by the sum of squared difference:

$$E^2_{reprojection} = \sum_k \sum_{\boldsymbol{x}} \left| S_k^V(\boldsymbol{x}) - S_k(\boldsymbol{x}) \right|^2 \tag{12}$$

where $\boldsymbol{x}$ is a pixel coordinate in silhouette image. We minimize the above mismatch by optimizing for the locations of the light sources. Unfortunately, optimizing Eq.(12) solely is known to be inherently ambiguous owing to 4 DOF transformation mentioned in Section 3. To alleviate this issue, we simultaneously minimize the discrepancy between the optimized light source positions $L_k$ and the initial source positions $L_k^*$ estimated from the spheres (Section 3) and epipolar geometry (Section 4):

$$E^2_{initial} = \sum_k \left\| L_k - L_k^* \right\|_2^2\,. \tag{13}$$

**Table 2:** Differences between traditional multi-view camera-based imaging and coplanar shadowgrams in epipolar reconstruction. The traditional multi-view images require at least 7 point correspondences between the silhouette contours. Coplanar shadowgrams allow unique epipolar reconstruction irrespective of the shape of the 3D object. We assume that the baselines do not always intersect the object.

| Silhouette complexity | Convex | Non-convex | | |
|---|---|---|---|---|
| #correspondences | 2 | $< 7$ | $\geq 7$ | $\gg 7$ |
| Traditional multi-camera | impossible | impossible | not always | hard |
| Coplanar shadowgrams | possible | possible | possible | possible |

possible — The epipolar geometry can be reconstructed uniquely.
not always — Possible if seven correspondences are found.
hard — Hard to find the correct correspondences in practice.
impossible — Impossible because of the insufficient constraints.

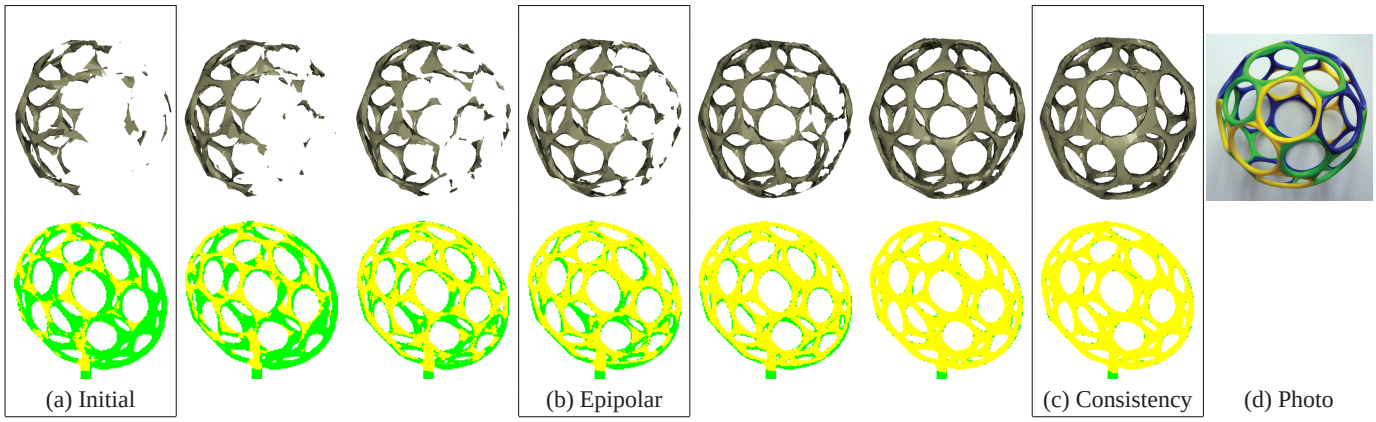



**Figure 11:** Reconstructed shape of a thin wire-frame object is improved with each iteration from left to right. (Top) Reconstructed visuals hull at the end of each iteration. (Bottom) The reprojection of the reconstructed visual hulls onto one of captured silhouette images. The reprojection and silhouettes are consistent at yellow pixels, and inconsistent at green. The boxed figures show the reconstruction from the light source positions (a) estimated from spheres, (b) improved by epipolar geometry, and (c) optimized by maximizing shadowgram consistency.

The final objective function is obtained by a linear combination of the two errors:

$$E_{total} = E^2_{reprojection} + \alpha E^2_{initial} \tag{14}$$

where, $\alpha$ is a user-defined weight. While the idea of minimizing silhouette discrepancy is well known in the traditional multi-view camera-based SFS [15, 19, 17, 8], the key advantage over prior work is the reduced number of parameters our algorithm needs to optimize (three per view for the light source position, instead of six per view for rotation and translation of the camera). In turn, this allows us to apply our technique to a much larger number of views than possible before.

### 5.2. Implementation

We use the signed Euclidean distances as the scalar-valued functions $S^V_k(\boldsymbol{x})$ and $S_k(\boldsymbol{x})$ in Eq.(12). The intersection of silhouettes is computed for each 3D ray defined by a pixel in $S_k$, and then projected back to the silhouette to obtain $S^V_k$. This is a simplified version of image-based visual hull [10] and has been used in silhouette registration methods [8]. Eq.(14) is minimized using Powell's gradient-free technique [11].

Due to the intricate shapes of the silhouettes, the error function in Eq.(14) can be complex and may have numerous local minima. We alleviate this issue using the convex polygons of the silhouette contours described in Section 4. Given the proposition shown in Section 4.1, we minimize Eq.(14) using the convex silhouettes with $\{L^*_k\}$ as initial parameters. The resulting light source positions are in turn used as starting values to minimize Eq.(14) with the original silhouettes. Using convex silhouettes, in practice, also speeds up convergence.

We evaluate this approach using the simulated silhouettes described in Figure 6 and 9. Compare the results in Figure 6 (using spheres to estimate source positions) and Figure 9 (enforcing epipolar constraints) with those in Figure 10. The final reconstruction of the tree branch is visually accurate highlighting the performance for our technique.

## 6. Results

In this section, we demonstrate the accuracy of our techniques using real objects. All results of 3D shape reconstructions shown in this paper are generated by exact polyhedral visual hull method proposed by Franco and Boyer [5]. We have quantitatively evaluated our algorithms using objects with known ground truth structure in the technical report [18].

The shadowgrams of real objects were acquired using our experimental setup shown in Figure 4. We selected objects that have intricate structures — the wreath object with numerous thin needles (Figure 1), the thin wiry polyhedral object (Figure 11), and two palm trees with flat leaves and severe occlusions (Figure 12). For each object, we captured a large number of silhouettes (122, 45 and 56 respectively) by moving the light source to different locations.

Figure 11 illustrates the convergence properties of our optimization algorithm. Figure 11(a) shows the visual hull of the wiry polyhedral object obtained using the initial source positions estimated from the calibration spheres. The reprojection of the visual hull shows poor and incomplete reconstruction. By optimizing the light source positions, the quality of the visual hull is noticeably improved in only a few iterations. Figure 12(d) also shows noticeable improvement in reconstructed shape of the palm trees from the acquired silhouettes.

The convergence of the reconstruction algorithm is quantitatively evaluated in Figure 13. The error in light source positions estimated by the algorithm proposed in Section 5 is shown in the left plot. The vertical axis shows L2 distance between the ground truth and the current estimate of light source positions. After convergence, the errors in the light source positions are less than 1% of the sizes of the objects. The silhouette mismatch defined in Eq.(12) is plotted on the right. On average, the silhouettes cover on the order of $10^5$ pixels. The error in the reprojection of the reconstructed visual hulls is less than 1% of the silhouette pixels for the real objects.

(a) Photo (b) Acquired shadowgrams (9 out of 56) (c) Initial reconstruction (d) final reconstruction

**Figure 12:** A palm-tree object model with a large number of thin structures. The real photograph of the object is shown in (a). (b) Fifty six shadowgrams with average resolution $520 \times 425$ pixels are acquired by the setup shown in Figure 4. (c) Initial and (d) final reconstruction of the object by coplanar shadowgrams.

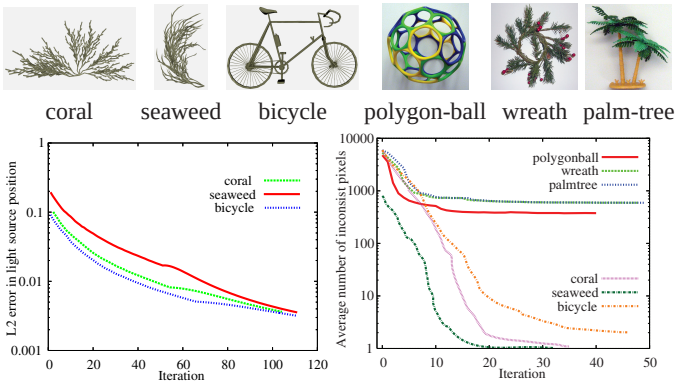



**Figure 13:** Convergence of error: (Left) Error in light source positions computed using ground truth for simulation models. (Right) Error in shadowgram consistency. Both plots are in logarithmic scale.

## 7. Discussion of Limitations

A single planar screen cannot be used to capture the complete $360° \times 360°$ view of the object. For instance, it is not possible to capture the silhouette observed in the direction parallel to a shadowgram plane. This limitation can be overcome by augmenting our system with more than one shadowgram screen (or move one screen to different locations). Another drawback of SFS techniques is the inability to model concavities on the object's surface. Combining our approach with other techniques, such as photometric stereo or multi-view stereo can overcome this limitation, allowing us to obtain appearance together with a smoother shape of the object. Finally, using multiple light sources of different spectra to speed up acquisition, and the analysis of defocus blur due to a light source of finite area are our directions of future work.

## Acknowledgements

This research was supported by ONR award N00014-05-1-0188 and ONR DURIP award N00014-06-1-0762. Narasimhan is also supported by NSF CAREER award IIS-0643628. Kanade is supported by NSF grant EEEC-540865.

## References

[1] B. G. Baumgart. *Geometric modeling for computer vision*. PhD thesis, Stanford University, 1974.

[2] R. Cipollaand, K. Åström, and P. Giblin. Motion from the frontier of curved surfaces. In *Proc. International Conference on Computer Vision '95*, pages 269–275, 1995.

[3] G. Cross, A. W. Fitzgibbon, and A. Zisserman. Parallax geometry of smooth surfaces in multiple views. In *Proc. International Conference on Computer Vision '99*, pages 323–329, 1999.

[4] A. Fitzgibbon, M. Pilu, and R. Fisher. Direct least squares fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):476–480, 1999.

[5] J.-S. Franco and E. Boyer. Exact polyhedral visual hulls. In *Proc. the 15th British Machine Vision Conference*, pages 329–338, 2003.

[6] Y. Furukawa, A. Sethi, J. Ponce, and D. Kriegman. Robust structure and motion from outlines of smooth curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):302–315, 2006.

[7] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2004.

[8] C. Hernández, F. Schmitt, and R. Cipolla. Silhouette coherence for camera calibration under circular motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(2):343–349, 2007.

[9] D. J. Kriegman and P. N. Belhumeur. What shadows reveal about object structure. *Journal of the Optical Society of America*, 18(8):1804–1813, 2001.

[10] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan. Image-based visual hulls. In *Proc. SIGGRAPH 2000*, pages 369–374, 2000.

[11] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.

[12] S. Savarese, M. Andreetto, H. Rushmeier, F. Bernardini, and P. Perona. 3d reconstruction by shadow carving: Theory and practical evaluation. *International Journal of Computer Vision*, 71(3):305–336, 2005.

[13] H. S. Sawhney. Simplifying motion and structure analysis using planar parallax and image warping. In *Proc. International Conference of Pattern Recognition*, pages 403–408, 1994.

[14] G. S. Settles. *Schlieren & Shadowgraph Techniques*. Springer-Verlag, 2001.

[15] S. N. Sinha, M. Pollefeys, and L. McMillan. Camera network calibration from dynamic silhouettes. In *Proc. Computer Vision and Pattern Recognition 2004*, volume 1, pages 195–202, 2004.

[16] A. R. Smith and J. F. Blinn. Blue screen matting. In *Proc. SIGGRAPH '96*, pages 259–268, 1996.

[17] K.-Y. K. Wong and R. Cipolla. Reconstruction of sculpture from its profiles with unknown camera positions. *IEEE Transactions on Image Processing*, 13(3):381–389, 2004.

[18] S. Yamazaki, S. Narasimhan, S. Baker, and T. Kanade. On using coplanar shadowgrams for visual hull reconstruction. Technical Report CMU-RI-TR-07-29, Carnegie Mellon University, Aug. 2007.

[19] A. J. Yezzi and S. Soatto. Stereoscopic segmentation. *International Jornal of Computer Vision*, 1(53):31–43, 2003.