The Theory and Practice of Coplanar Shadowgram Imaging for Acquiring Visual Hulls of Intricate Objects

Shuntaro Yamazaki

National Institute of Advanced Industrial Science and Technology

shun-yamazaki@aist.go.jp

Srinivasa Narasimhan

Carnegie Mellon University

srinivas@cs.cmu.edu

Simon Baker

Microsoft Research

sbaker@microsoft.com

Takeo Kanade

Carnegie Mellon University

tk@cs.cmu.edu

June 18, 2008

Revision 2

Abstract

Acquiring 3D models of intricate objects (like tree branches, bicycles and insects) is a challenging task due to severe self-occlusions, repeated thin structures, and surface discontinuities. In theory, a shape-from-silhouettes (SFS) approach can overcome these difficulties and reconstruct visual hulls that are close to the actual shapes, regardless of the complexity of the object. In practice, however, SFS is highly sensitive to errors in silhouette contours and the calibration of the imaging system, and has therefore not been used for obtaining accurate shapes with a large number of views. In this work, 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. A point light source is moved around an object and the shadows (silhouettes) cast onto a single background plane are imaged. 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 unique 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 the 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.

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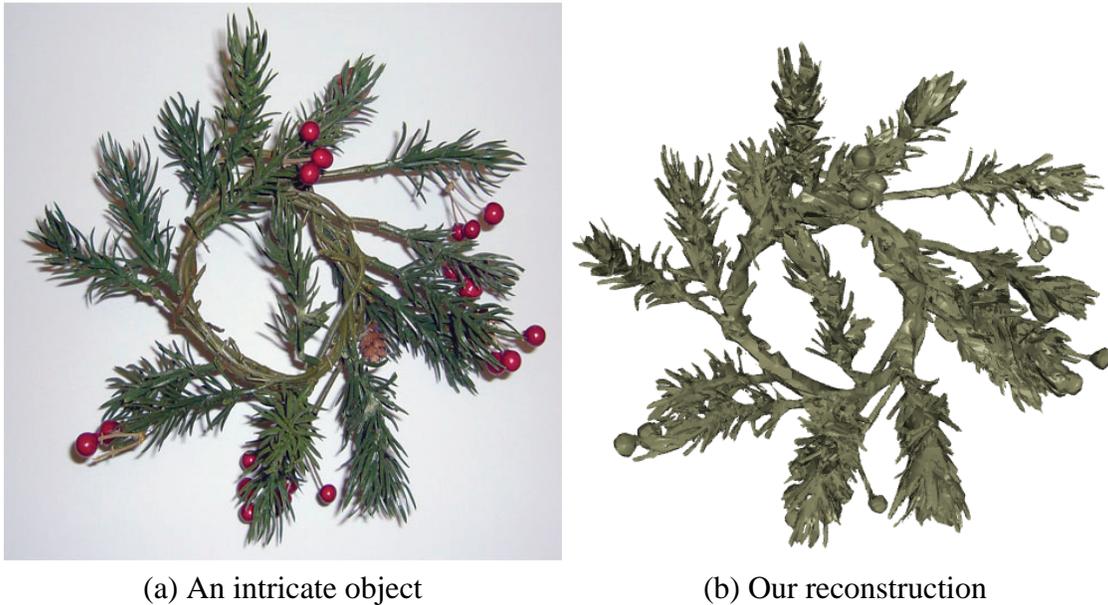
Figure 1: Obtaining 3D models of intricate shapes such as in (a) is hard due to severe occlusions and correspondence ambiguities. (b) By moving a point source in front of the object, we capture a large number of shadows cast on a single fixed planar screen (122 views for this object). Applying our techniques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

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.

The issues of correspondence and occlusion in the object can be avoided if we only use the silhouettes of an object obtained from different views and reconstruct its *visual hull* [2]. The top row of Figure 2 illustrates the visual hulls 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

5 views   10 views   50 views
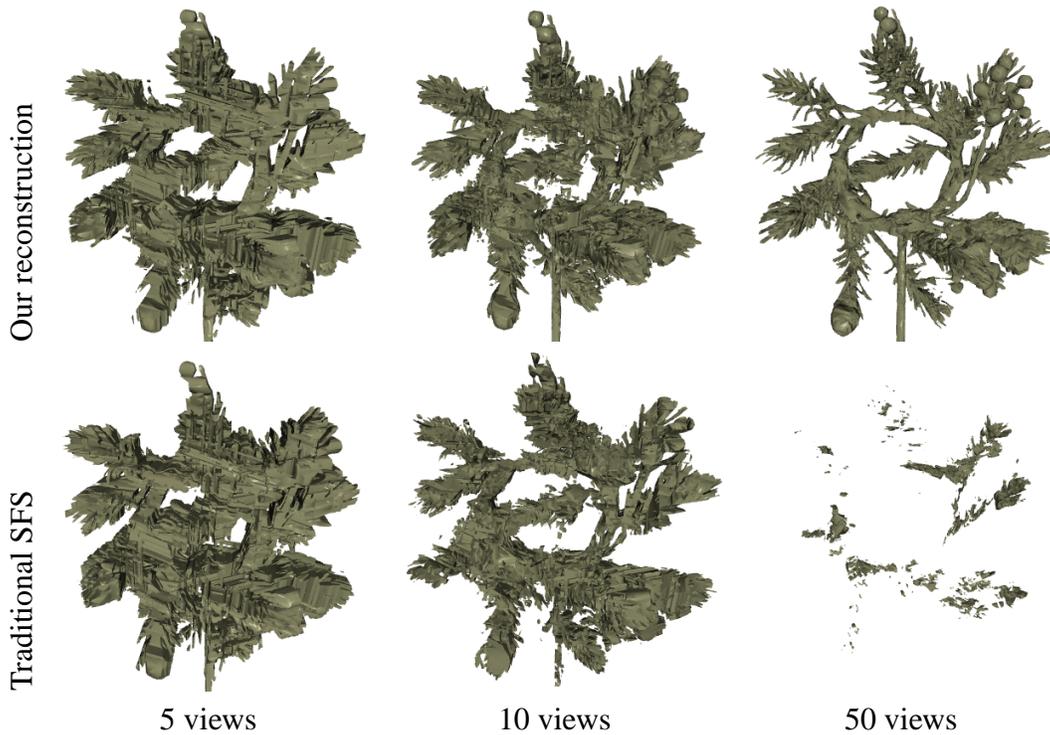
Figure 2: Sensitivity of SFS reconstruction: (Top) The visual hulls reconstructed using the light source positions estimated by our method. As the number of silhouettes increases, the visual hull gets closer to the actual shape. (Bottom) The reconstructions obtained from slightly erroneous source positions. As the number of views increases, the error worsens significantly.

excellent model of the original shape. Thus, a Shape-From-Silhouettes (SFS) approach may be used to recover accurate 3D models of intricate objects.

In practice, however, SFS algorithms are highly sensitive to errors in silhouette contours and the geometric parameters of the imaging system (camera calibration) [3]. 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 from slightly erroneous source positions. This drawback must be addressed in order to acquire intricate shapes reliably.

The traditional SFS system assumes that a camera observes the object, and the silhouette is extracted from the obtained image by image processing, such as, background subtraction [4, 5], image segmentation [6, 7] or manual segmentation [8]. Multiple view-

points 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.* [9] proposed a system that avoids the difficulty in silhouette extraction using cast shadows. When an object is illuminated by a single point light source, the shadow cast onto a background plane is sharp and can be directly used as its silhouette. Silhouettes from multiple views are obtained by rotating the object and capturing the shadow images (also known as *shadowgrams*[1]). Balan *et al.* [12] proposed a hybrid method that reconstructs the visual hull of an object using the shadows cast on the floor as well as the silhouettes captured by multiple cameras. In terms of multi-view geometry, these methods are equivalent to traditional SFS, requiring six parameters per view.

This paper proposes a novel approach to SFS called *coplanar shadowgram imaging*. We use a setup similar in spirit to that proposed by Savarese *et al.* [9] The key difference here is that the point source is moved, while the object, the camera and the background screen all remain stationary. The principal focus of this work is on acquiring visual hulls of intricate and opaque objects from a large number of coplanar shadowgrams. Our main contributions are described below.

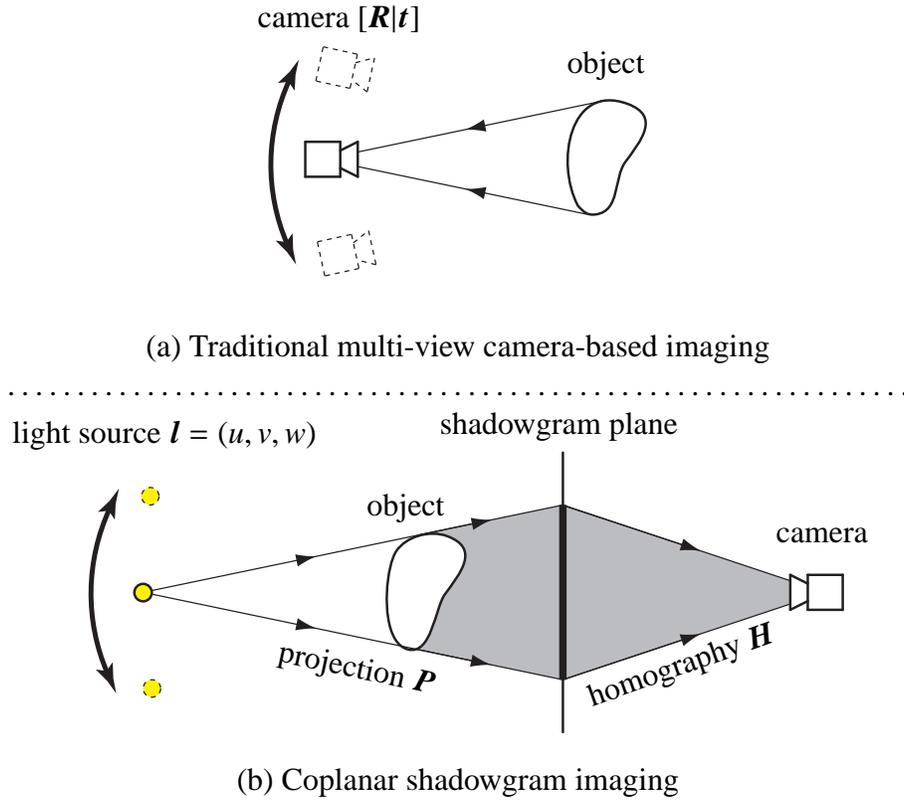**Multi-view geometry of coplanar shadowgram imaging:**

We propose a coplanar shadowgram imaging system, and clarify how it is different from traditional imaging system where the camera directly observes the silhouette of an object. Figure 3 compares the geometry of the two systems. The key observation is that the relative transformation between the object and screen remains fixed across different views in our coplanar shadowgram imaging system. The image projection model is hence described by only three parameters per view (3D translation of the source) instead of six in traditional systems. The geometry is similar in spirit to parallax geometry [13, 14] where the homography between image planes is known to be an identity, which allows the derivation of unique 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.

Recovery of shape and source positions:

When the shape of the object is unknown, the locations of all the point sources (and therefore the object's shape) can be recovered from coplanar shadowgrams, only

---

[1]Shadowgrams have also been widely used for visualizing the 3D structure of transparent objects such as glasses or fluids [10, 11].

camera $[\boldsymbol{R}|\boldsymbol{t}]$

object

(a) Traditional multi-view camera-based imaging

light source $\boldsymbol{l} = (u, v, w)$ shadowgram plane

object

camera

projection $\boldsymbol{P}$

homography $\boldsymbol{H}$

(b) Coplanar shadowgram imaging

Figure 3: (a) The object of interest is observed directly by a 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 moving the light source. Note that the relative transformation between the object and the screen remains fixed across different views.

up to a four parameter perspective transformation. We show how this transformation relates to the Generalized Perspective Bas-Relief (GPBR) ambiguity [15] that is derived for a single viewpoint system. We break this ambiguity by simultaneously capturing the shadowgrams of two or more spheres and using them as soft constraints.

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 due to the non-linear nature of the reconstruction algorithm. 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. In practice, the optimization on the convex silhouettes also leads to faster convergence.

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 or silhouette consistency.

When we exploit the epipolar tangency and reconstruct the projective geometry of imaging system, the localization of epipole and the correspondence between tangent lines have to be solved simultaneously. To solve this combinatorial optimization, existing methods assume either good initializations [16, 3] or an approximated camera model such as linear projection [8]. These methods also assume that the shape of the object is *reasonably complex*; 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. Wong and Cipolla propose a technique that can recover the epipolar geometry regardless of the complexity of the silhouette, using the outer epipolar tangency which is similar in spirit to our approach [17]. The camera motion, however, is limited to circular motion since they use the multi-view camera-based imaging system.

Another approach to the camera reconstruction from silhouettes is estimating the camera poses that can reconstruct the visual hull consistent to with the acquired silhouettes. Yezzi and Soatto [18] solved this highly non-linear optimization problem assuming a good initialization and simple silhouettes. Hernández *et al.*proposed an efficient optimization by maximizing the consistency only along the contour lines of the silhouettes. However, the methods based on the silhouette consistency also have difficulty in the initialization and the limitation on the complexity of the shape. As a result, these techniques have succeeded in only acquiring the 3D shapes of relatively simple objects like people and statues using a small number of views and/or assuming limited camera motion.

In contrast, our algorithm is effective for a large number of views (dozens to a hundred) from a wide range of view-points, does not require any feature correspondences 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 [19]) 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 a wreath, a wiry ball, a palm tree, and an octopus that show numerous occlusions, discontinuities and thin structures. In addition, we have also evaluated our techniques quantitatively using simulations with objects such as a coral, branches, a bicycle and an insect whose ground truth shapes are known beforehand.

Despite significant progress in optical scanning hardware [20, 21] and multi-view geometry [19, 22], reconstruction of intricate shapes remains an open problem. We believe this work is an initial step to solving this problem. Possible extensions of our work include multiple screens covering $360° \times 360°$ views of the objects and/or multiple light sources for dynamic objects, and combine our techniques with stereo and photometric stereo, to obtain reconstructions that are smoother than visual hulls, including concavities.
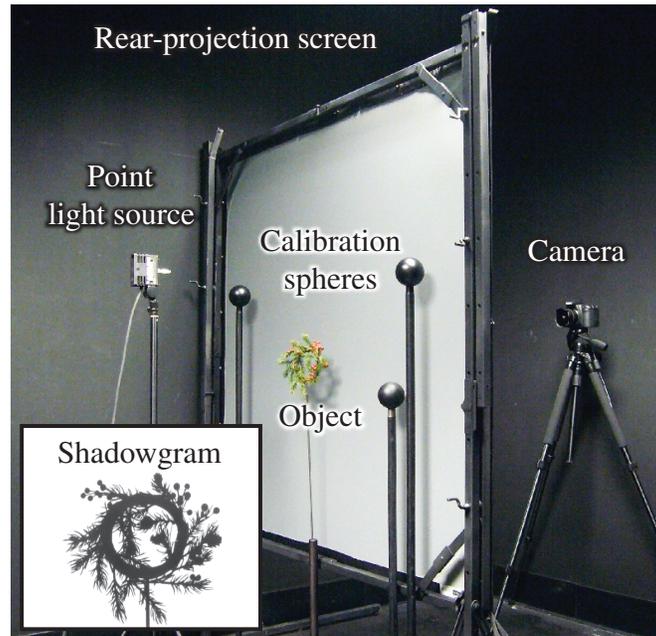
2 Coplanar Shadowgrams

We define *shadowgrams* as the shadow pattern cast on a background plane by an object that occludes a point source. If the object is opaque, the sharp shadowgram accurately represents the silhouette of the object. Henceforth, we shall use shadowgrams and silhouettes interchangeably. We also refer to the images that captures shadowgrams as shadowgram images. Coplanar shadowgram imaging is the process of acquiring several shadowgrams on *a single plane* by moving the light source while keeping the shadow plane fixed.

Our acquisition setup shown in Figure 4 includes a 6M-pixel Canon EOS-20D digital camera, a 250 watt 4mm incandescent bulb, and a 4ft × 4ft translucent rear-projection screen. This setup is similar to the one in [9] where the object is rotated and the light source is fixed. We prove in Section 4 that the fixed relative position between the image plane and the object is crucial to deriving a strong constraint on the shadowgrams. Another advantage of moving only the light source is that we can easily acquire the shadowgrams of non-rigid objects. In this section, we describe how the visual hull of the object is obtained using our setup.

2.1 Shadowgram Projection

Throughout this paper, we represent the locations of light sources, the surface of 3D objects, and the shadowgrams on a background plane in either a two or three dimensional real projective space ($\mathbb{RP}^2$ or $\mathbb{RP}^3$) unless defined explicitly. Sets are written in caligraph face as $\mathcal{S}$, operators are in sans-serif font as P, vectors and matrices are in bold face as $\boldsymbol{v}$ and $\boldsymbol{M}$, and scalars are in normal fonts as x. All the vectors without the transpose are column vectors.

Figure 4: The setup used to capture coplanar shadowgrams includes a single point light source, a rear-projection screen, and a digital camera. 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.

Figure 3(b) illustrates the viewing and illumination geometry of coplanar shadowgram imaging. Without loss of generality, we assume the shadowgram plane Π is located at $z = 0$. When an 3D object $\mathcal{O} \subset \mathbb{RP}^3$ is illuminated by a point light source at $\boldsymbol{l} \in \mathbb{RP}^3$, the shadowgram $\mathcal{S}^\mathcal{O} \subset \mathbb{RP}^2$ of the object on the shadowgram plane is obtained by a perspective projection:

$$\mathcal{S}^\mathcal{O} = \mathsf{P}(\boldsymbol{l}) \cdot \mathcal{O} . \tag{1}$$

The projection $\mathsf{P}(\boldsymbol{l}) : \mathbb{RP}^3 \rightarrow \mathbb{RP}^2$ can be represented as a 3×4 matrix $\boldsymbol{P}(\boldsymbol{l})$ (See Appendix A for the derivation):

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

The image $\mathcal{I}$ of the shadowgram acquired by a camera is related to the shadowgram

$\mathcal{S}^O$ on the plane Π by a 2D homography:

$$\mathcal{I} = \mathsf{H} \cdot \mathcal{S}^O \; . \tag{3}$$

This homography $\mathsf{H} : \mathbb{RP}^2 \to \mathbb{RP}^2$ is independent of the light source position and can be estimated separately using a standard computer vision algorithm. In the following, we assume that the shadowgram $\mathcal{S}^O$ has been estimated using

$$\mathcal{S}^O = \mathsf{H}^{-1} \cdot \mathcal{I} \; . \tag{4}$$

2.2 Visual Hull Reconstruction

Now let a set of shadowgrams $\mathcal{S}_i^O$ of an object O ($i = 1, \cdots, N$) be acquired by moving the source to different locations l_i. The visual cone $C_i \subset \mathbb{RP}^3$ associated with the shadowgram $\mathcal{S}_i^O$ is defined as

$$C_i \stackrel{\text{def}}{=} \left\{ \, p \in \mathbb{RP}^3 \mid \mathsf{P}(l_i) \cdot p \in \mathcal{S}_i^O \, \right\} \; . \tag{5}$$

Then, the visual hull $\mathcal{V} \subset \mathbb{RP}^3$ of the object is obtained as

$$\mathcal{V} \stackrel{\text{def}}{=} \bigcap_{i=1}^{N} C_i \; . \tag{6}$$

Given the 3D locations l_i of the light sources, the visual hull of the object can be estimated using Equation (2) and Equation (6). Due to the nature of the reconstruction algorithm, the reprojection $\mathcal{S}^{\mathcal{V}}$ of the visual hull $\mathcal{V}$ to the shadowgram plane reproduces the shadowgram identical to the input.

$$\mathcal{S}_i^{\mathcal{V}} \stackrel{\text{def}}{=} \mathsf{P}(l_i) \cdot \mathcal{V} \tag{7}$$
$$= \mathcal{S}_i^O \tag{8}$$

Table 1 summarizes and contrasts the geometric parameters that appear in the traditional multi-view camera-based and coplanar shadowgram imaging systems. In multi-view camera-based imaging system (Figure 3(a)), we must specify both the intrinsic camera parameters that define the relationship between pixels and ray directions, and the extrinsic camera parameters (rotation and translation) that define the ray directions in a world coordinate frame. As the camera moves, the intrinsic parameters remain the same, while the six extrinsic parameters vary. On the other hand, in coplanar shadowgram imaging (Figure 3(b)), the point source is defined only by its location, requiring three parameters per view as compared to the six in the traditional system. As we shall show, this difference is crucial to the success of the technique presented in this paper.

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)	3 (rotation)
	1 (aspect ratio)	3 (translation)
	1 (skew)	
	2 (image center)	
Coplanar shadowgrams	8 (homography H)	3 (translation *l*)

3   Source Recovery

When the shape of the object is unknown, it is not possible to uniquely recover the 3D source positions only using the coplanar shadowgrams. In this section, we discuss the nature of the ambiguity in the 3D reconstruction from our coplanar shadowgrams and propose a simple method to break it.

3.1   Ambiguity in 3D Reconstruction from Coplanar Shadowgrams

Suppose two sets of light source locations, $l_i = (u_i, v_i, w_i, 1)^T$ and $l'_i = (u'_i, v'_i, w'_i, 1)^T$ for $i = 1, \cdots, N$, are estimated from a given set of coplanar shadowgrams. We can reconstruct respective visual hulls $\mathcal{V}$ and $\mathcal{V}'$ using Equation (6). The reprojection of the visual hulls defined in Equation (7) satisfy:

$$\mathsf{P}(l_i) \cdot \mathcal{V} = \mathsf{P}(l'_i) \cdot \mathcal{V}' \ . \tag{9}$$

Let $\mathsf{A} : \mathbb{RP}^3 \to \mathbb{RP}^3$ be a transformation that deforms $\mathcal{V}$ into $\mathcal{V}'$:

$$\mathcal{V}' = \mathsf{A} \cdot \mathcal{V} \ . \tag{10}$$

Equation (9) and Equation (10) lead to the identical equation:

$$\mathsf{P}(l_i) = \mathsf{P}(l'_i) \cdot \mathsf{A} \ . \tag{11}$$

Here, A is a linear transformation because any 3D point along a ray from a point light source at l_i is transformed to the point on a ray from the source at l'_i. Comparing all

elements in both sides of Equation (11), we obtain the matrix A of the transformation (See Appendix B for the derivation):

$$A = \begin{pmatrix} 1 & 0 & a_1 & 0 \\ 0 & 1 & a_2 & 0 \\ 0 & 0 & a_3 & 0 \\ 0 & 0 & a_4 & 1 \end{pmatrix} . \tag{12}$$

The four parameters a_1, a_2, a_3, and a_4 take arbitrary numbers under the condition that

$$a_3 \neq 0 \qquad \text{and} \qquad a_3 a_4 \geq 0 . \tag{13}$$

Using Equation (11), we can prove that the transformation between the source locations is also described by the same matrix:

$$l'_i = A l_i . \tag{14}$$

Equation (10) and (12) show that we can recover the shape of an object only from coplanar shadowgrams, up to a four-parameter family of perspective transformations A.

This result is consistent with the theory on the ambiguity in the geometric reconstruction from shadow information. Kriegman and Belhumeur [15] prove that, when the shape of an object and the locations of light sources are all unknown, we can at best reconstruct the 3D structure from the shadow information, up to a four-parameter family of perspective transformation which they call the Generalized Perspective Bas-Relief (GPBR) transformation. In fact, our ambiguity transformation A can be viewed as the GPBR transformation if the origin of the world coordinate system is translated to the center of projection of a calibrated camera, and the shadowgram plane is regarded as a part of the object. In our imaging system, however, we use the coordinate system where the origin is on the shadowgram plane for the following reasons. Firstly, the estimation of the homography between the camera image plane and the shadowgram plane is much easier than the calibration of camera projection. We can reconstruct the shape of the object without knowing any camera parameter. Secondly, the ambiguity in the global scale of the reconstruction can be resolved using the physical dimension of the shadowgram screen.

3.2   Determining Source Position Using Spheres

Significant effort has been devoted to understanding when and how the ambiguity in the shape reconstruction using shadows can be resolved. If the surface reflection on the object of interest is observed by a camera, the ambiguity is resolved by exploiting relative albedo values [23], interreflections [24], specularities [25, 26, 27], or Helmholtz reciprocity [28].
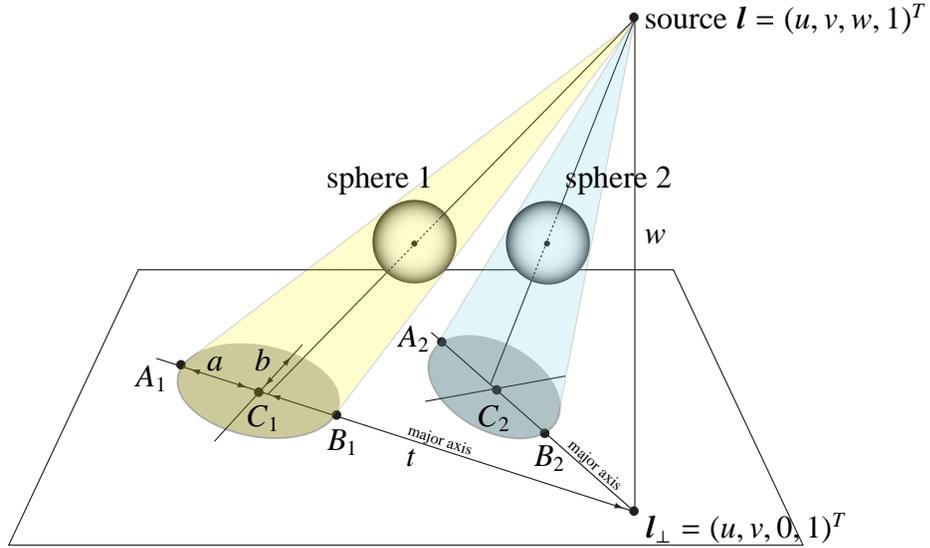
Figure 5: Source position $l = (u, v, w, 1)^T$ is recovered using the elliptical shadowgrams of two spheres. The radii and positions of the spheres are unknown. The major axes of the ellipses intersect the screen at $l_\perp = (u, v, 0, 1)^T$. The w component is obtained using Equation (18).

However, we cannot observe the surface reflection in coplanar shadowgram imaging system. The robustness of the analyses of surface reflection is also open to question when the object has an intricate shape.

We now propose a technique of breaking the ambiguity using calibration objects. The location $l = (u, v, w, 1)^T$ of a point light source is directly estimated by capturing shadowgrams of two additional spheres that are placed adjacent to the object of interest.

The first and second coordinates u and v can be estimated by analyzing the shadowgrams of the spheres. Figure 5 illustrates the coplanar elliptical shadowgrams cast by the two spheres [2]. The ellipses are localized using a constrained least squares approach [30]. The intersection of the major axes $\overline{A_1 B_1}$ and $\overline{A_2 B_2}$ of the two ellipses is the foot of the perpendicular line from l to Π, denoted by $l_\perp = (u, v, 0, 1)^T$ in Figure 5.

The third coordinate w is obtained as the intersection of hyperbolae in 3D space as shown below. Without loss of generality, consider the 3D coordinate system whose origin is at the center of the ellipse, and X and Y axes are respectively the major and minor axes

---

[2]Each sphere is placed so that the minimum distance between a light source and the rear-projection screen is larger than the distance between the center of the sphere and the screen. Under this configuration, the cast shadow of the sphere is always an ellipse [29].
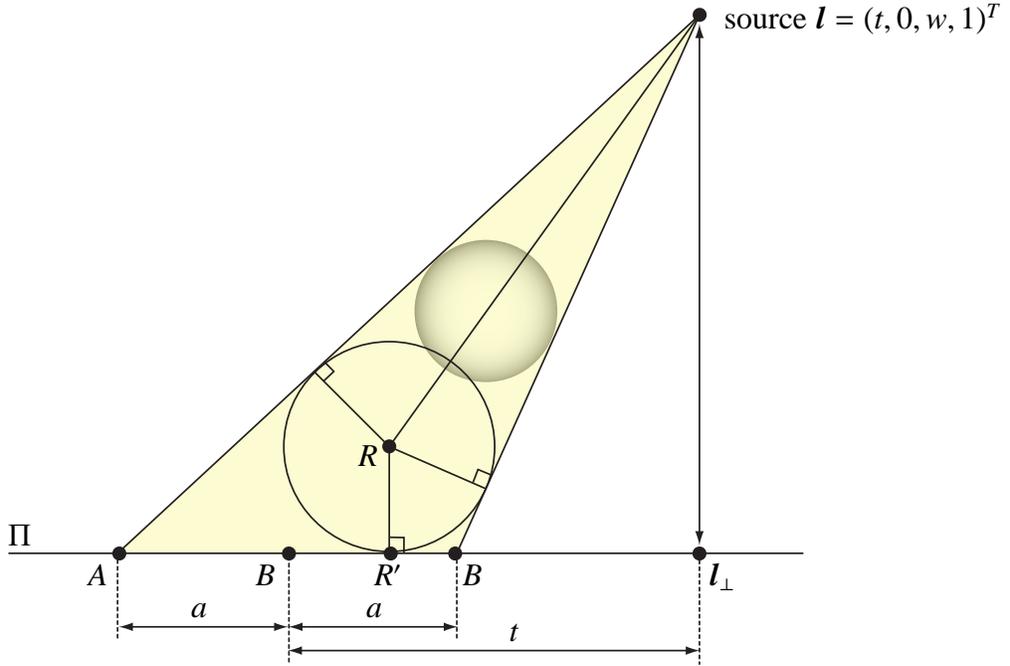
Figure 6: The cross section of a right circular conical surface formed by the light rays emanating from a point light source at $\boldsymbol{l}$ and tangent to a calibration sphere.

of the ellipse. Then, the ellipse is represented in the following form.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \quad (a > b) \tag{15}$$

In 3D space, there exists an inscribed sphere tangent to the conical surface and the plane, regardless of the position or the radius of the sphere. The cross section of the inscribed sphere by the plane that includes the apex of the cone and the major axis of the ellipse is shown in Figure 6. The center of the inscribed sphere is shown by R. The other symbols are corresponding to those in Figure 5. The center of the ellipse C is the origin of the coordinate system.

The inscribed sphere is tangent to XY-plane at a focus of the ellipse R', hence

$$\overline{CR'} = \sqrt{a^2 - b^2}. \tag{16}$$

Using the symmetry of triangles,

$$\overline{LA} - \overline{AR'} = \overline{LB} - \overline{BR'}. \tag{17}$$
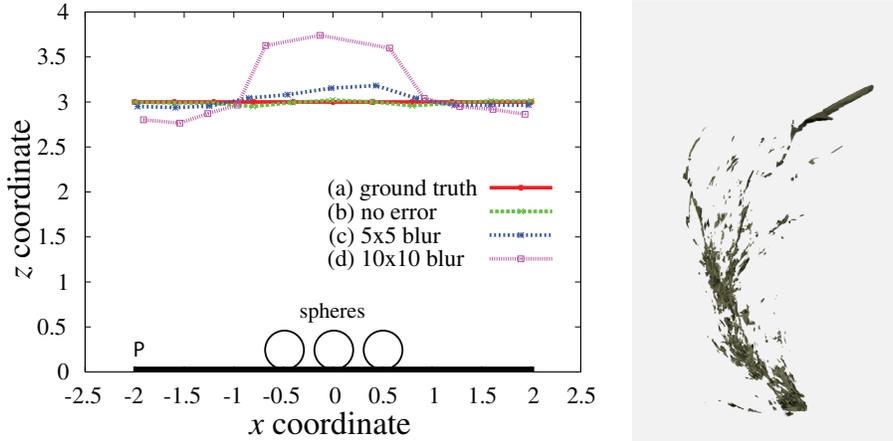
Figure 7: Source positions (u, w) are estimated using three calibration spheres. The sizes and positions of the spheres and screen are shown in the plot. Each plot shows 11 source positions obtained from (a) ground truth, (b) accurate shadowgrams, and (c)-(d) shadowgrams blurred using 5×5 and 10×10 averaging 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.

Let the position of the apex be $l = (t, 0, w, 1)^T$ in this coordinate system, then we can solve w with respect to t as:

$$w = \sqrt{\frac{b^2 t^2}{a^2 - b^2} - b^2},$$ (18)

where a and b are the semimajor and semiminor axes of one of the ellipses, and t is the length between $l_\perp$ and the center of the ellipse.

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.

This technique for estimating the source position can be sensitive to errors in measured silhouettes. 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 spheres. We blurred the simulated silhouettes (effective resolution 480×360 pixels) with 5×5 and 10×10 averaging

kernels, and estimated the 3D coordinates of the light source. Figure 7 presents u and w components of the source positions reconstructed using three spheres. Observe that the estimation becomes poor when the shadowgram is close to a right circle. In turn, the visual hull of a tree branch computed from the erroneous source positions is too erroneous even to perceive the 3D structure.

Due to the nature of visual hull reconstruction, a large error in one light source can blast the perfect reconstruction obtained by the other sources. Thus, better algorithms for averaging out the errors in individual light sources are crucial for obtaining accurate 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 ($l_i = (u_i, v_i, w_i, 1)^T$ and $l_j = (u_j, v_j, w_j, 1)^T$ in Figure 8). 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_i and l_j intersects the shadowgram plane Π at the epipole e_{ij}. 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*. These observations are respectively written as

$$M_{ij}e_{ij} = 0 \; , \tag{19}$$

$$m_i^T F_{ij} m_j = 0 \; . \tag{20}$$

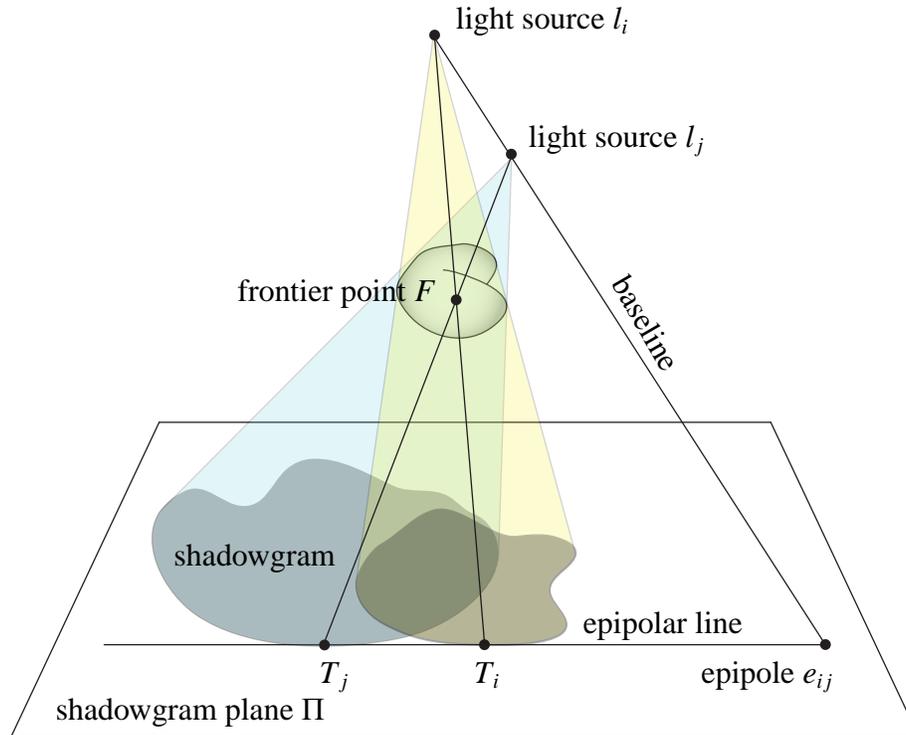In Equation (19), M_{ij} is a 2×3 matrix composed of two plane equations in the rows

$$M_{ij} = \begin{pmatrix} -\Delta v & \Delta u & u_i v_j - u_j v_i \\ \Delta u \Delta w & -\Delta - v\Delta w & (u_i \Delta u + v_i \Delta v)\Delta w - w_i(\Delta u^2 + \Delta v^2) \end{pmatrix} \tag{21}$$

where $\Delta u = u_j - u_i$, $\Delta v = v_j - v_i$, and $\Delta w = w_j - w_i$. In Equation (20),

$$F_{ij} = [e_{ij}]_\times \tag{22}$$

is the *fundamental matrix* that relates two corresponding points m_i and m_j between shadowgrams. $[e_{ij}]_\times$ is the 3×3 skew symmetric matrix:

$$[e_{ij}]_\times x = e_{ij} \times x \tag{23}$$

Figure 8: Epipolar geometry of two shadowgrams. The baseline connecting the two sources l_i and l_j intersects the shadowgram plane Π at an epipole e_{ij}. 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_i and T_j.

for any 3D vector x.

The camera geometry in coplanar shadowgram is similar in spirit to the parallax geometry [13, 14] 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

We now show that the above observations enable us to uniquely estimate the epipolar geometry from only the shadowgram images. Suppose we have the plane in Figure 8 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* [14], that can be estimated as one that is cotangent to the two shadowgrams (at T_i and T_j in Figure 8). Two such epipolar lines can then be intersected to localize the epipole. But the reverse need not hold; every cotangent line need not be an epipolar line. So, how do we localize the epipole and estimate the epipolar lines without knowing any frontier points?

Figure 9(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 9(b). Only two of these four cotangent lines pass through the actual epipole, hence, the other two are false detections. Indeed, the false detections correspond to infeasible cases where the light source is located between the object and the screen, or behind the screen. 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 9(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 9(d) and prove the following proposition (See Appendix C for the proof):

Proposition 1 *The silhouettes of the convex hull of an object are the convex hulls of the silhouettes.*

Using Proposition 1, 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. This geometric property is also observed in the parallel projection model where cameras are internally calibrated and moved without rotation [31]. In the coplanar shadowgram imaging, however, the same geometric structure is accomplished without camera calibration or any apparatus to control the cameras. Table 2 summarizes the dif-
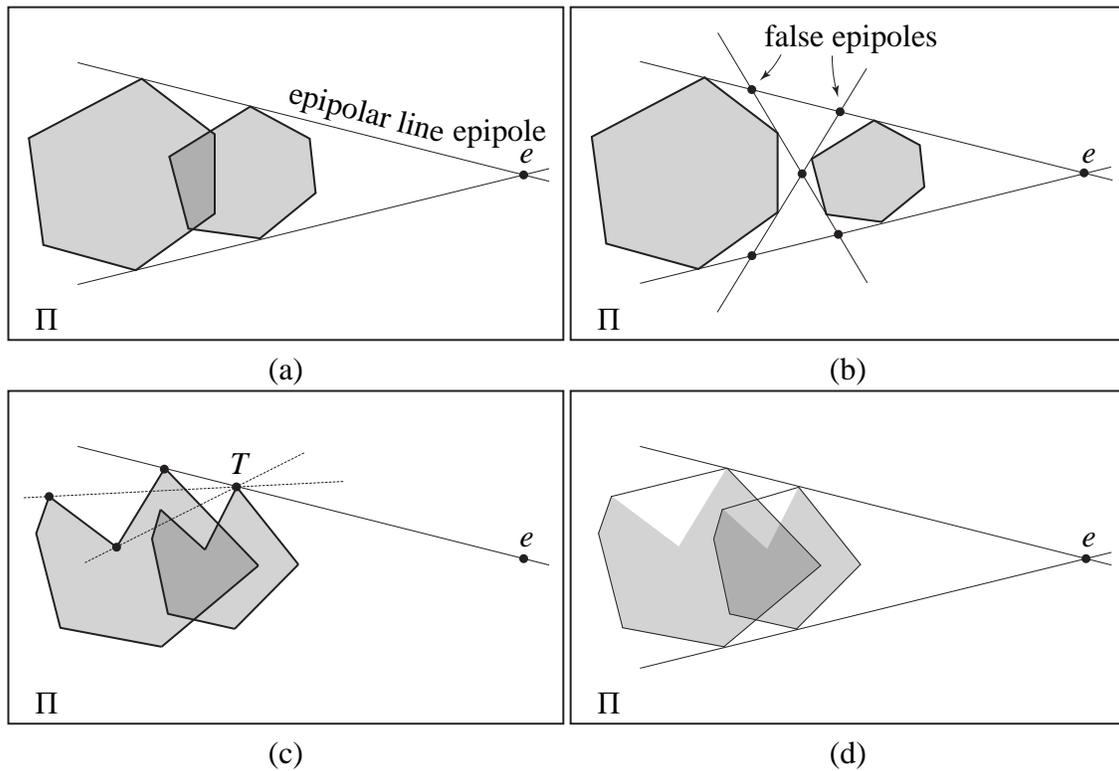
Figure 9: Localization of the epipole. (a),(b) If two shadowgrams are convex, a maximum of four co-tangent lines and six intersections are possible. Considering that the object and the light source are on the same side with respect to the screen, the epipole can be chosen uniquely out of the six intersections. (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.

ferences between traditional multi-view imaging by uncalibrated cameras 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 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.

Table 2: Comparison between traditional multi-view camera-based imaging and coplanar shadowgrams in epipolar reconstruction. The traditional multi-view images acquired by uncalibrated cameras require at least 7 point correspondences of the silhouette contours. Coplanar shadowgrams allow unique epipolar reconstruction irrespective of the shape of the 3D object.

Silhouette complexity #correspondences	Convex 2	Non-convex		
		< 7	≥ 7	≫ 7
Uncalibrated 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.

## 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 Equation (19) for each pair of light sources using least squares:

$$\{l_1, \cdots, l_N\} = \underset{\{l_1, \cdots, l_N\}}{\operatorname{argmin}} \sum_{i \neq j} \left\| M_{ij} e_{ij} \right\|_2^2 \tag{24}$$

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 7. Figure 10 shows considerable improvement in accuracy obtained by enforcing the epipolar constraint in Equation (19). Compared to the result in Figure 7, 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 (Fig-
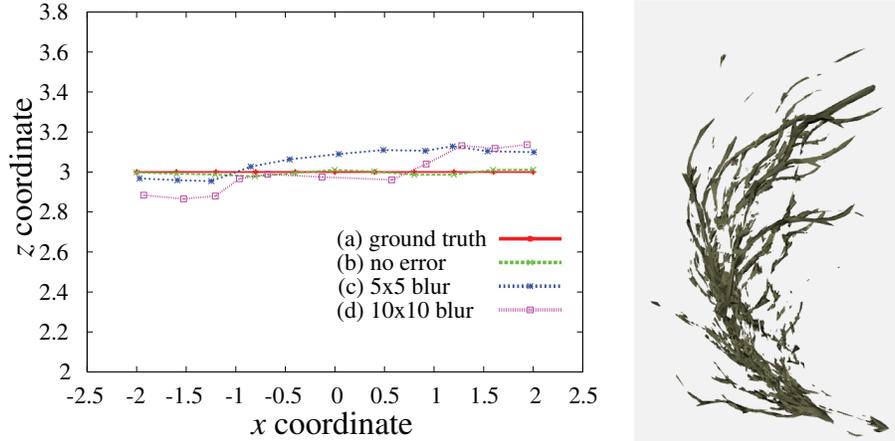
Figure 10: Initial light source positions in Figure 7 were improved by epipolar constraints in Equation (24). On the right is the visual hull reconstructed from the improved source positions.

ure 10). In this section, we present an optimization algorithm that improves the accuracy of all the source positions even more significantly. As we will show, combining the epipolar constraints and the optimization algorithm results in high quality models of intricate shapes.
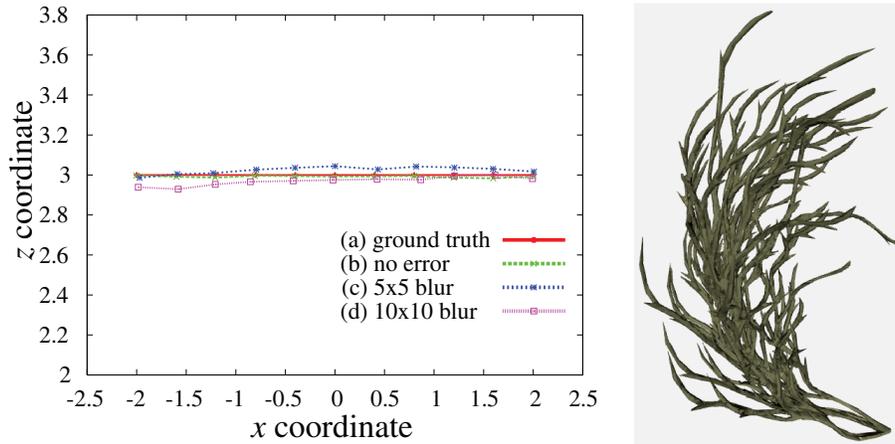
5.1 Optimizing light source positions

Let $\mathcal{V}$ be the visual-hull obtained from the set of captured shadowgrams $\mathcal{S}_i$ and the estimated projection $\mathsf{P}(\boldsymbol{l}_i)$ for $i = 1, \cdots, N$. Due to the nature of the intersection operator, the reprojections of the visual hull to the shadowgram plane satisfy:

$$\mathcal{S}_i^{\mathcal{V}} \subseteq \mathcal{S}_i^{O} . \tag{25}$$

The reprojections match the acquired silhouettes when the source positions are perfect. In other words, if they match, we cannot make any more improvement to the shape only by the silhouettes. We can define a measure of silhouette mismatch by the sum of squared difference:

$$\epsilon_{reprojection}^2 = \sum_{i=1}^{N} \mathsf{D}\left(\mathcal{S}_i^{\mathcal{V}}(\boldsymbol{x}), \mathcal{S}_i^{O}(\boldsymbol{x})\right) \tag{26}$$

where $\mathsf{D} : \mathbb{RP}^2 \times \mathbb{RP}^2 \to \mathbb{R}$ evaluates the difference between two sets and will be discussed in Section 5.2. We minimize the above mismatch by optimizing for the locations of the

Figure 11: The light source positions reconstructed using epipolar constraint in Figure 10 were optimized by maximizing the shadowgram consistency in Equation (28). On the right is the visual hull reconstructed from the optimized source positions.

light sources. Unfortunately, optimizing solely Equation (26) is known to be inherently ambiguous owing to the four-parameter transformation mentioned in Section 3. To alleviate this issue, we simultaneously minimize the discrepancy between the optimized light source positions l_i and the initial source positions l_i^0 estimated from the spheres (Section 3) and epipolar geometry (Section 4):

$$\epsilon_{initial}^2 = \sum_{i=1}^{N} \left\| l_i - l_i^0 \right\|_2^2 \tag{27}$$

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

$$\epsilon_{total} = \epsilon_{reprojection}^2 + \alpha \epsilon_{initial}^2. \tag{28}$$

where α is a user-defined weight. While the idea of minimizing silhouette discrepancy is well known in the traditional multi-view camera-based SFS [3, 18, 17, 32], 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.
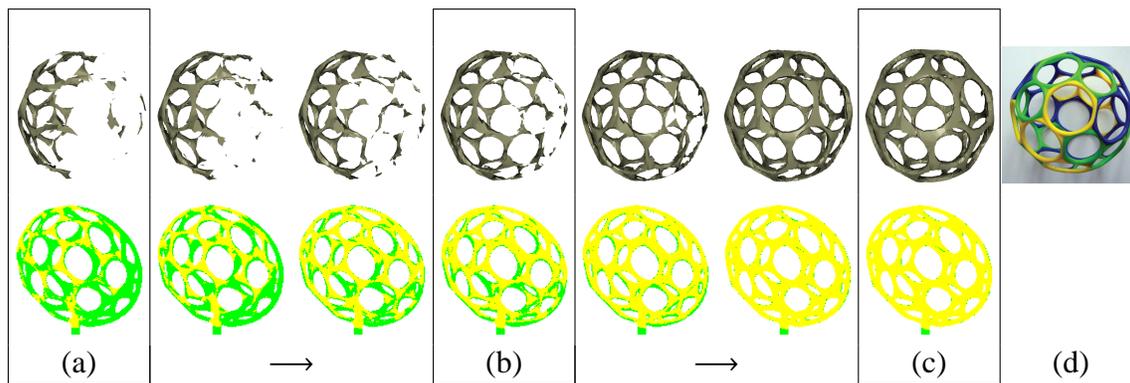
Figure 12: Reconstructed shape of a polyhedron object is improved with each iteration from left to right. (Top) Reconstructed visuals hull at the end of each iteration. (Bottom) The reprojection of the reconstructed visual hulls onto one of captured silhouette images. The reprojection and silhouettes 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. (d) Photograph of the object.

5.2 Implementation

How do we represent the scalar-valued function D that describes the *difference* between two silhouettes in Equation (26)? Suppose the silhouettes $\mathcal{S}_i^{\mathcal{V}}$ and $\mathcal{S}_i^{\mathcal{O}}$ are given as images in the same dimension, then it is natural to define D as the sum of square distances between the pixels values of the silhouettes. Then, the next question is how to define the pixels values of the silhouette images.

The simplest one is a binary function that returns 0 and 1 when a pixel is located respectively outside and inside the silhouette. This binary function, however, is not suitable for iterative minimization of Equation (26) since $\mathsf{D}\left(\mathcal{S}_i^{\mathcal{V}}, \mathcal{S}_i^{\mathcal{O}}\right)$ is non-zero at only a small number of pixels around mis-matched contour lines. Due to errors in the measured silhouette, the binary function causes undesirable bias.

We use the signed Euclidean distances to the contour of $\mathcal{S}_i^{\mathcal{V}}$ and $\mathcal{S}_i^{\mathcal{O}}$ as the pixels values in the silhouette images. The intersection of silhouettes is computed for each 3D ray defined by a pixel in $\mathcal{S}_i^{\mathcal{O}}$, and then projected back to the silhouette to obtain $\mathcal{S}_i^{\mathcal{V}}$. This is a simplified version of image-based visual hull [33] and has been used in silhouette registration methods [32]. Equation (28) is minimized using Powell's gradient-free technique [34].
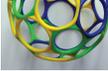
Due to the intricate shapes of the silhouettes, the error function in Equation (28) can be complex and may have numerous local minima. We alleviate this issue using the convex polygons of the silhouette contours described in Section 4. From Proposition 1, a following corollary is derived regarding the consistency between the visual hull and the silhouettes of an object (See Appendix D for the proof).

Proposition 2 *If silhouette contours are consistent in that they can be generated from a physical 3D object, then the convex polygons obtained from the silhouette contours are also consistent.*

Using Proposition 2, we minimize Equation (28) using the convex silhouettes with l_i^0 as initial parameters. The resulting light source positions are in turn used as starting values to minimize Equation (28) with the original silhouettes. In practice, using convex silhouettes also speeds up convergence.

We evaluate this approach using the simulated silhouettes described in Figure 7 and 10. Compare the results in Figure 7 (using spheres to estimate source positions) and Figure 10 (enforcing epipolar constraints) with those in Figure 11. The final reconstruction of the tree branch is visually accurate highlighting the performance for our technique.

Table 3: The models used in our experiments: We reconstructed visual hulls from coplanar shadowgrams using four simulation and four real data sets. The detail of each experiment is shown in the corresponding figure. The average resolution of shadow region in the shadowgrams is shown in the row of shadowgram size. The computation time in optimizing epipolar geometry and silhouette consistency is shown in minutes (See the main text for our computational environment). Reprojection error indicates the mismatch between the input shadowgrams and those generated by reprojecting the estimated visual hull. For simulation data, the volume ratio between the ground truth and the reconstructed visual hulls is presented.

model	simulation data				real data			
	coral	seaweed	bicycle	spider	polyhedron	wreath	palm-tree	octopus
Figure	13	14	15	16	17	18	19	20
# of views	84	49	61	76	45	122	56	53
shadowgram size	530×270	334×417	635×425	356×354	126×116	674×490	520×425	451×389
reconstruction time	532min	384min	429min	498min	235min	9312min	429min	402min
reprojection error	2.2%	3.2%	2.3%	1.3%	3.2%	5.2%	4.8%	4.6%
volume error	0.15%	0.21%	0.12%	0.08%	—	—	—	—

6   Results

In this section, we demonstrate the accuracy of our techniques using both simulated and real experimental data. Table 3 summarizes the data set used in the experiment and the performance of our reconstruction algorithms.

We first generated or captured a sufficiently large number of shadowgrams, and started our reconstruction algorithm using a small number of randomly-chosen images. The number of images used is increased until the reconstructed visual gets sufficiently close to the actual shape. We used a workstation equipped with four dual-core AMD Opteron 8218 2.6GHz and 16GB main memory for optimization of epipolar geometry and silhouette consistency explained in Section 4 and Section 5.

All results of 3D shape reconstructions shown in this paper are generated by the exact polyhedral visual hull method proposed by Franco and Boyer [35]. The acquired 3D shape is then rendered using Autodesk Maya rendering package.

6.1 Reconstruction of Visual Hulls and 3D Source Positions

6.1.1 Simulation data

We have chosen four objects with complex structure in our simulations — a coral, a seaweed (also used in Figure 7, 10, and 11 in the main paper), a bicycle, and a spider. The seaweed and coral objects have many thin sub-branches with numerous occlusions. The bicycle object is composed of very thin structures such as spokes, chains, and gears. The spider object is composed of both thick and thin structure. The simulation experiments with known ground truth shape and source positions are shown respectively in Figure 13, 14, 15, and 16.
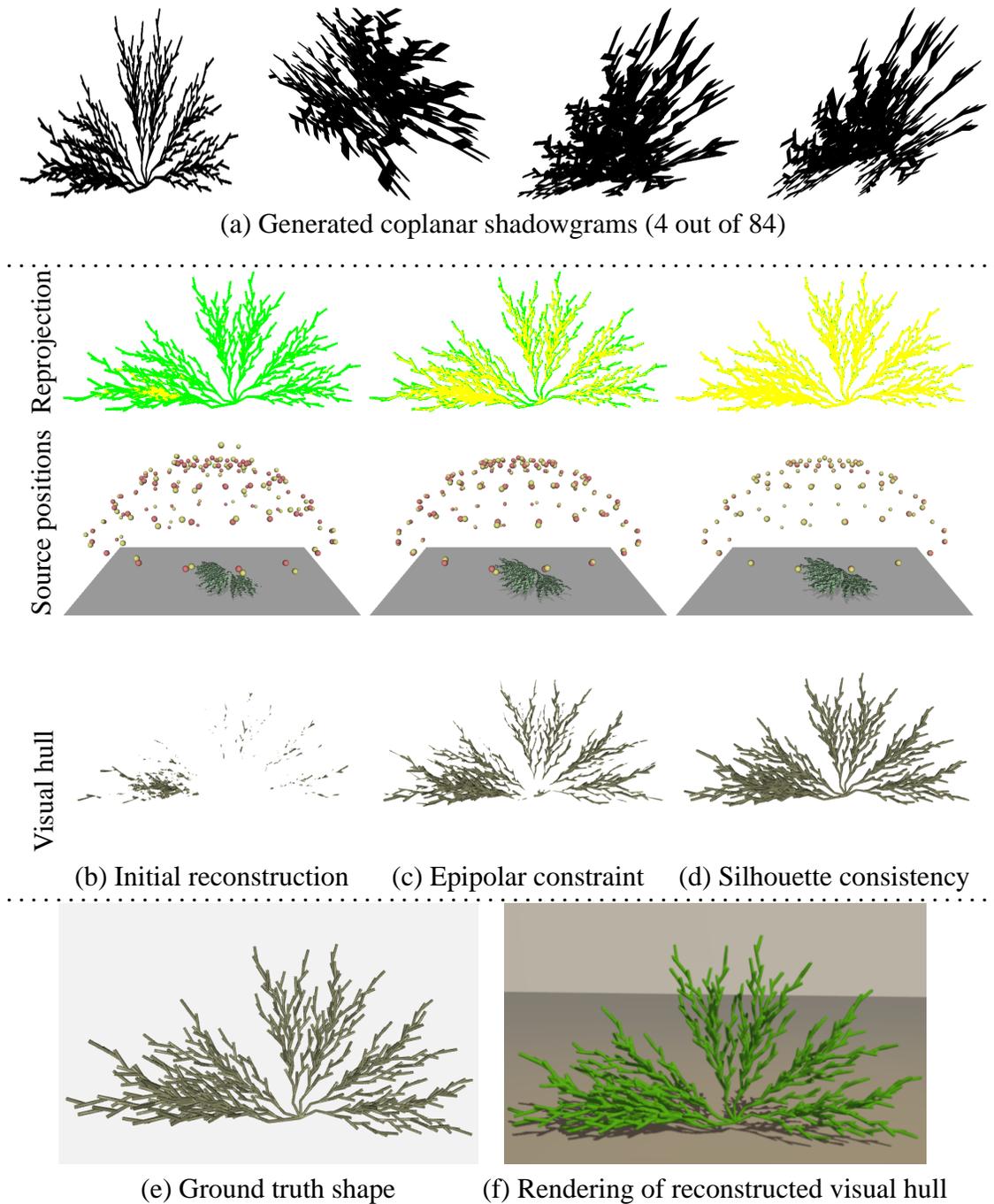
Each of the figures is organized as follows: (a) A set of coplanar shadowgrams of the object is generated by a shadow simulator implemented by Direct 3D graphics library. (b) The positions of light source are perturbed with random noise with $\sigma = 5\%$ of the object size, and the silhouettes are blurred by 3×3 averaging filters. (c) The positions of the light sources are recovered using epipolar geometry followed by the maximization of silhouette consistency in (d). For each of (b), (c), and (d), the top row shows one of captured silhouette images (in green), overlaid with the reprojection of the reconstructed visual hulls onto the silhouette (in yellow). The middle row shows the ground truth positions of light sources (in red) and the estimated positions (in yellow). The reconstructed 3D shape is shown at the bottom. Finally, (e) the ground truth 3D shape and (f) the reconstructed visual hull rendered by Maya is shown.
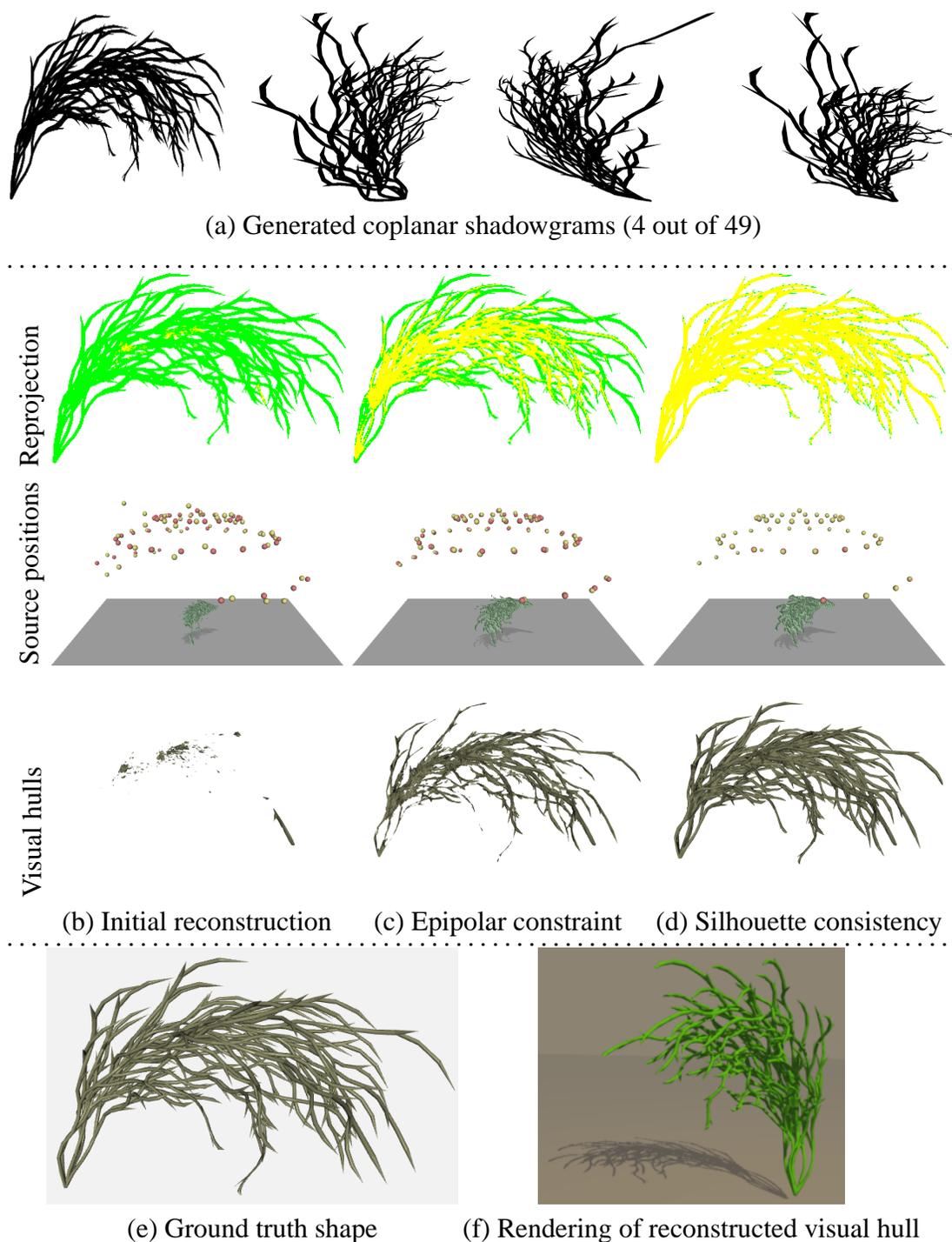
6.1.2 Real data

We show the 3D shape reconstruction of four different objects — a polyhedron (also used in the Figure 12 in the main paper), a wreath (Figure 1 and 2), a palm-tree object, and an octopus object. The wreath object has numerous thin needles which cause severe occlusions. The polyhedron is a thin wiry polyhedral object. The palm-tree object is a plastic object composed of two palm trees with flat leaves. The octopus is a plastic object that has complex surface reflection and large concavities. The results of reconstructing 3D shape are shown in Figure 17, 18, 19, and 20. Each figure is organized in the same way as those of simulation data, except that: The final reconstruction of source positions are presented in red in the middle row of (b), (c), and (d). The photograph of the object is shown in (e).

6.2 Convergence

Figure 12 illustrates the convergence properties of our optimization algorithm. Figure 12(a) shows the visual hull of the wiry polyhedral object obtained using the initial positions of

(a) Generated coplanar shadowgrams (4 out of 84)

(b) Initial reconstruction   (c) Epipolar constraint   (d) Silhouette consistency

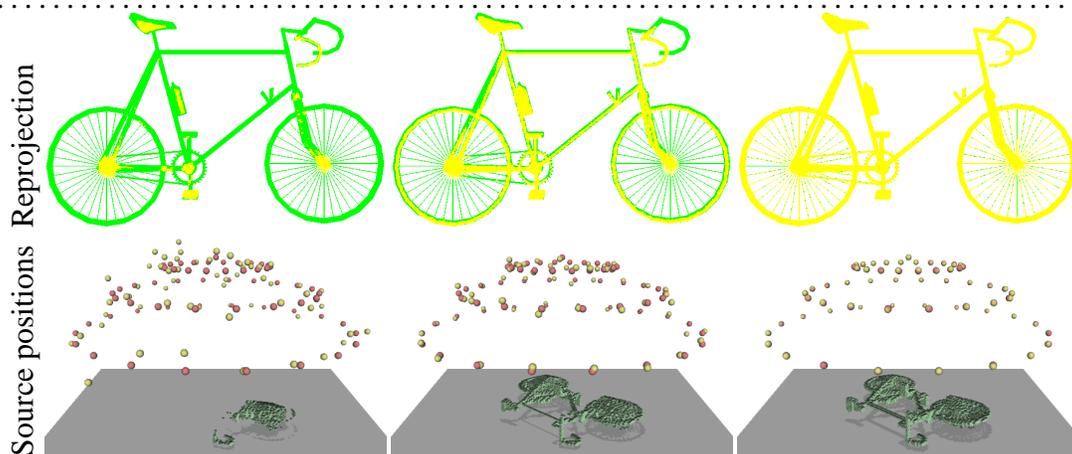(e) Ground truth shape   (f) Rendering of reconstructed visual hull

Figure 13: Simulation with a **coral** object: (a) Eighty four coplanar shadowgrams of the object are generated with average resolution 530×270 pixels. (b) Initial reconstruction. (c) The reconstruction using epipolar geometry. (d) The reconstruction using silhouette consistency. (e) The ground truth 3D shape. The volume difference is 0.15% of the volume of the ground truth 3D shape. (f) Rendering of the reconstructed shape. (Refer to main text for the detail of each figure.)

(a) Generated coplanar shadowgrams (4 out of 49)

Reprojection

Source positions

Visual hulls

(b) Initial reconstruction   (c) Epipolar constraint   (d) Silhouette consistency

(e) Ground truth shape   (f) Rendering of reconstructed visual hull

Figure 14: Simulation with a **seaweed** object: (a) Forty nine coplanar shadowgrams of the object are generated with average resolution 334×417 pixels. (b) Initial reconstruction. (c) The reconstruction using epipolar geometry. (d) The reconstruction using silhouette consistency. (e) The ground truth 3D shape. The volume difference is 0.21% of the volume of the ground truth 3D shape. (f) Rendering of the reconstructed shape. (Refer to main text for the detail of each figure.)

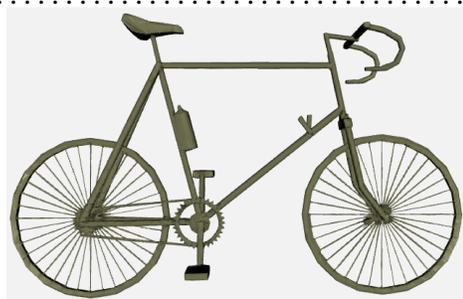(a) Generated coplanar shadowgrams (4 out of 61)

Reprojection

Source positions

Visual hulls

(b) Initial reconstruction   (c) Epipolar constraint   (d) Silhouette consistency
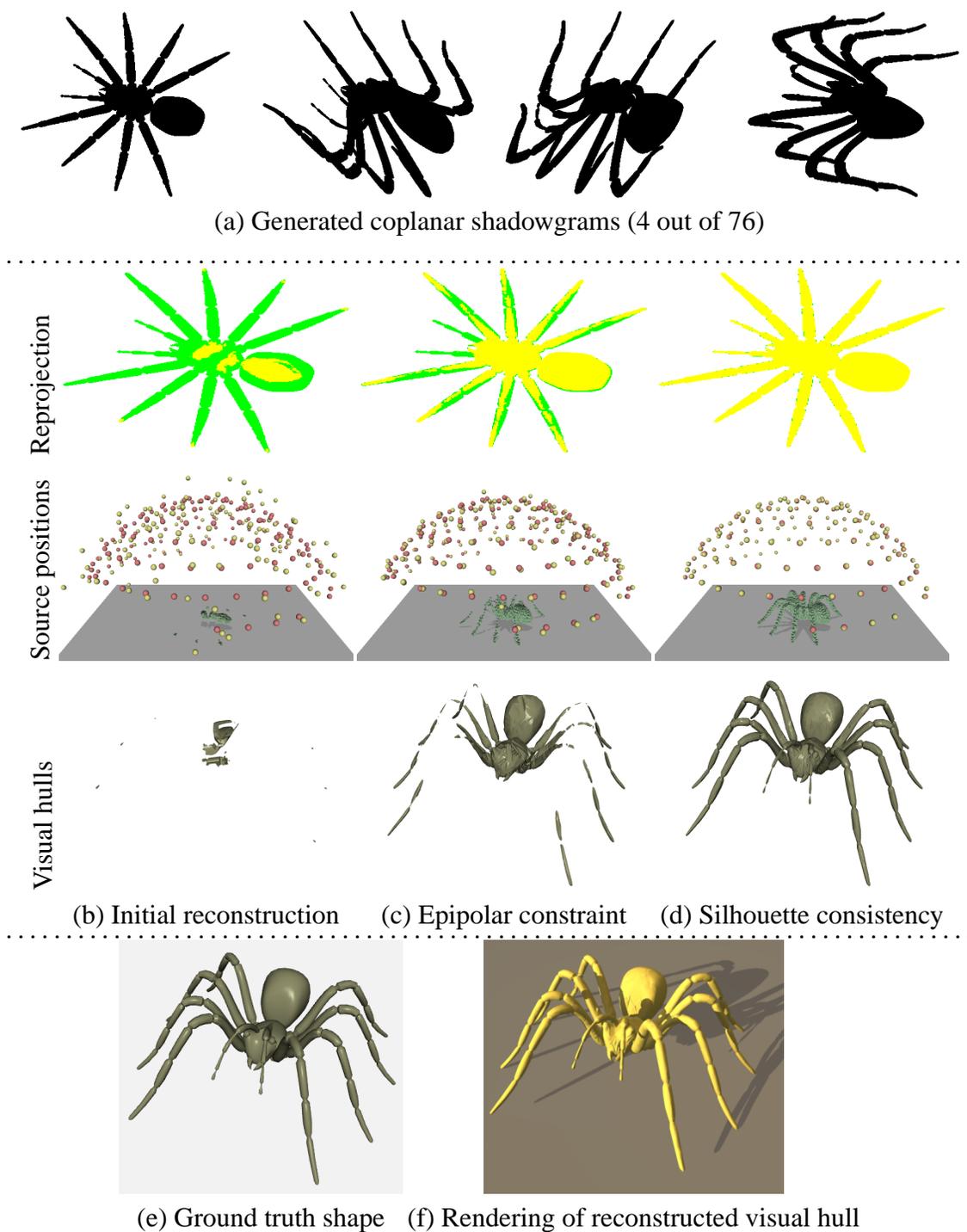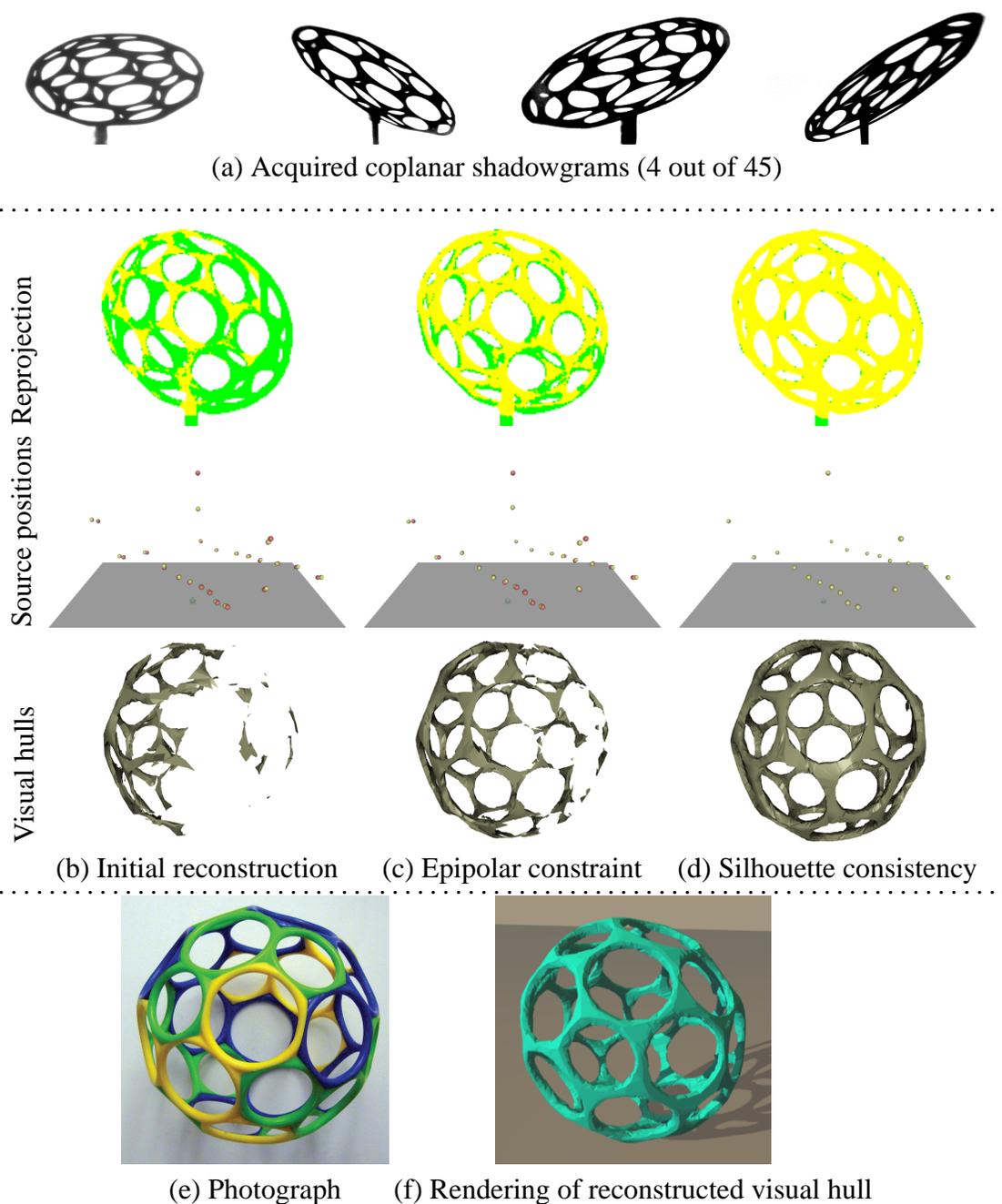
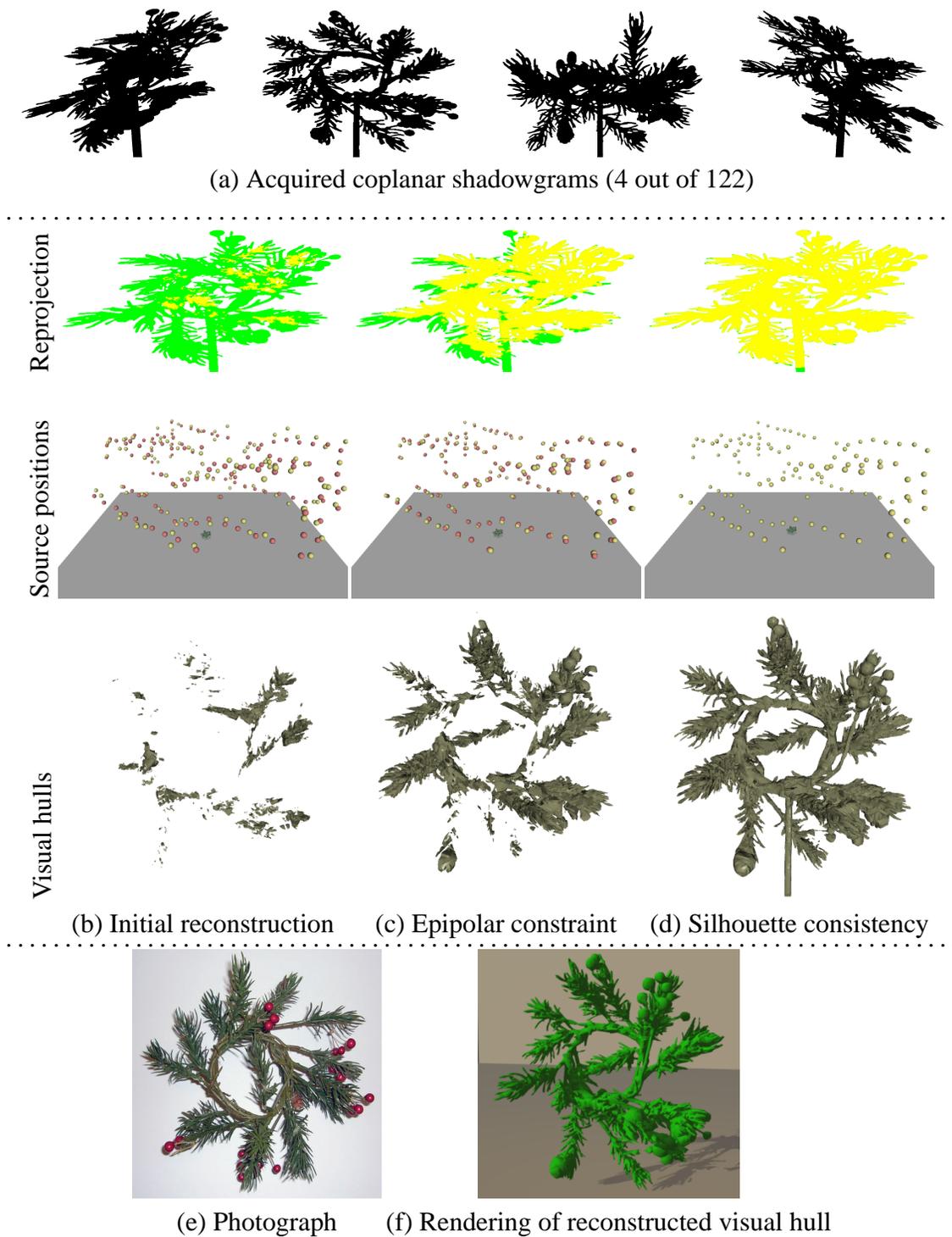(e) Ground truth shape   (f) Rendering of reconstructed visual hull

Figure 15: Simulation with a **bicycle** object: (a) Sixty one coplanar shadowgrams of the object are generated with average resolution 635×425 pixels. (b) Initial reconstruction. (c) The reconstruction using epipolar geometry. (d) The reconstruction using silhouette consistency. (e) The ground truth 3D shape. The volume difference is 0.12% of the volume of the ground truth 3D shape. (f) Rendering of the reconstructed shape. (Refer to main text for the detail of each figure.)

(a) Generated coplanar shadowgrams (4 out of 76)

Reprojection

Source positions

Visual hulls

(b) Initial reconstruction   (c) Epipolar constraint   (d) Silhouette consistency

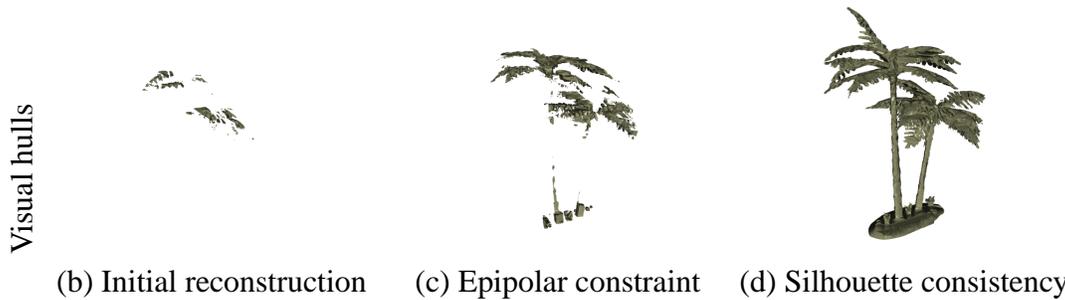(e) Ground truth shape   (f) Rendering of reconstructed visual hull

Figure 16: Simulation with a **spider** object: (a) Seventy six coplanar shadowgrams of the object are generated with average resolution 356×354 pixels. (b) Initial reconstruction. (c) The reconstruction using epipolar geometry. (d) The reconstruction using silhouette consistency. (e) The ground truth 3D shape. The volume difference is 0.08% of the volume of the ground truth 3D shape. (f) Rendering of the reconstructed shape. (Refer to main text for the detail of each figure.)
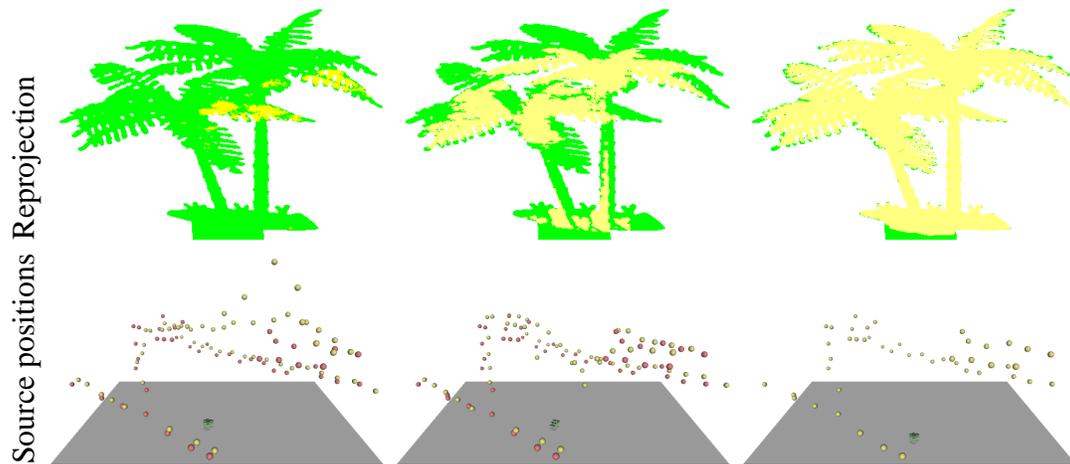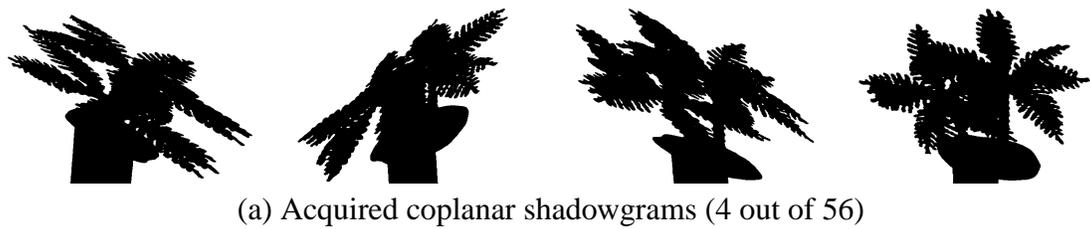
(a) Acquired coplanar shadowgrams (4 out of 45)

(b) Initial reconstruction (c) Epipolar constraint (d) Silhouette consistency

(e) Photograph (f) Rendering of reconstructed visual hull

Figure 17: Real experiment with a **polyhedron** object: (a) Forty five coplanar shadow-grams of the object are generated with average resolution 126×116 pixels. (b) Initial reconstruction. (c) The reconstruction using epipolar geometry. (d) The reconstruction using silhouette consistency. (e) Photograph of the object. (f) Rendering of the reconstructed shape. (Refer to main text for the detail of each figure.)

(a) Acquired coplanar shadowgrams (4 out of 122)

(b) Initial reconstruction   (c) Epipolar constraint   (d) Silhouette consistency

(e) Photograph   (f) Rendering of reconstructed visual hull

Figure 18: Real experiment with a **wreath** object: (a) 122 coplanar shadowgrams of the object are generated with average resolution 674×490 pixels. (b) Initial reconstruction. (c) The reconstruction using epipolar geometry. (d) The reconstruction using silhouette consistency. (e) Photograph of the object. (f) Rendering of the reconstructed shape. (Refer to main text for the detail of each figure.)

(a) Acquired coplanar shadowgrams (4 out of 56)

Reprojection

Source positions

Visual hulls

(b) Initial reconstruction   (c) Epipolar constraint   (d) Silhouette consistency

(e) Photograph   (f) Rendering of reconstructed visual hull

Figure 19: Real experiment with a **palm-tree** object: (a) Fifty six coplanar shadowgrams of the object are generated with average resolution 520×425 pixels. (b) Initial reconstruction. (c) The reconstruction using epipolar geometry. (d) The reconstruction using silhouette consistency. (e) Photograph of the object. (f) Rendering of the reconstructed shape. (Refer to main text for the detail of each figure.)
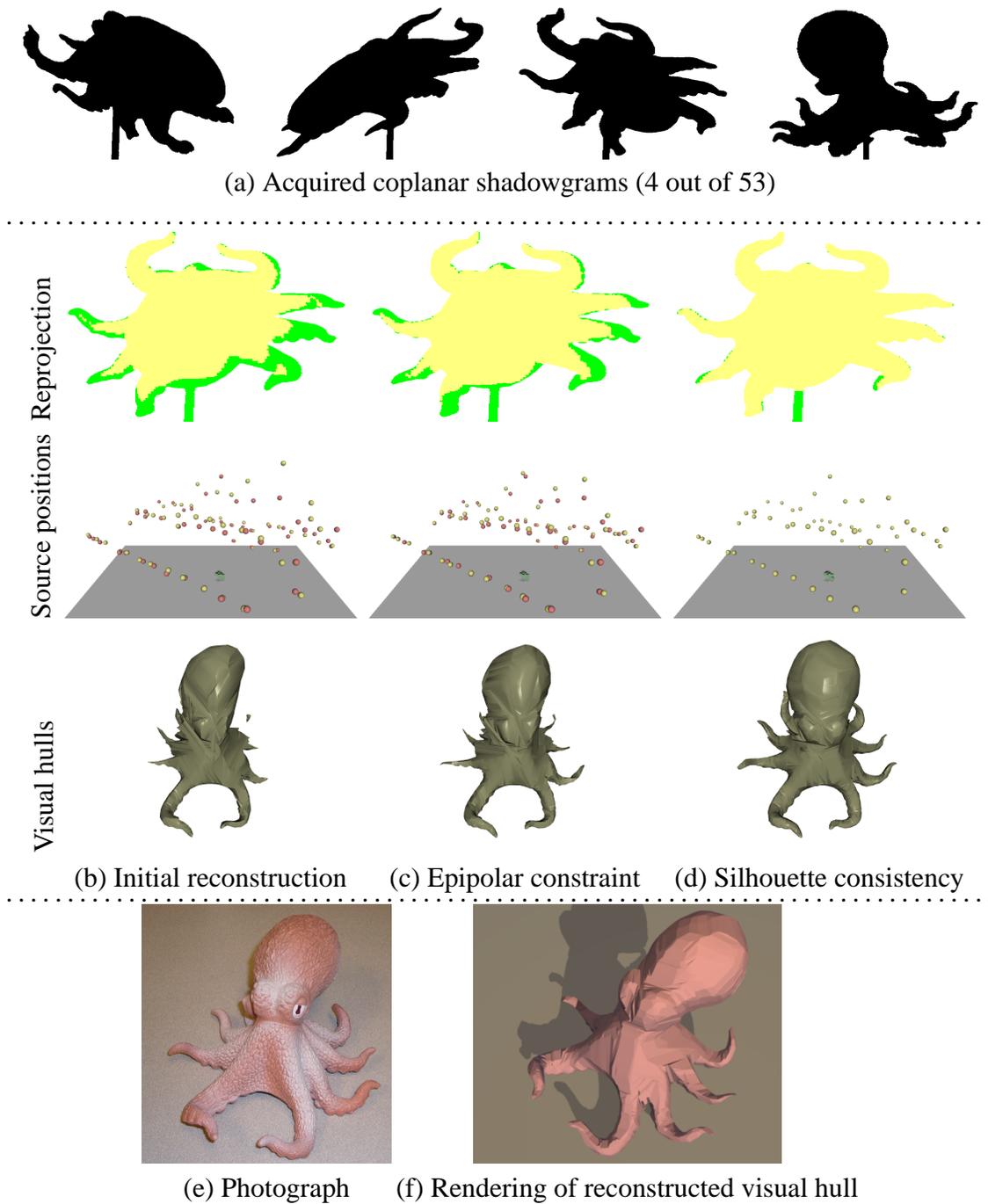
(a) Acquired coplanar shadowgrams (4 out of 53)

Reprojection

Source positions

Visual hulls

(b) Initial reconstruction (c) Epipolar constraint (d) Silhouette consistency

(e) Photograph (f) Rendering of reconstructed visual hull

Figure 20: Real experiment with an **octopus** object: (a) Fifty three coplanar shadowgrams of the object are generated with average resolution 451×389 pixels. (b) Initial reconstruction. (c) The reconstruction using epipolar geometry. (d) The reconstruction using silhouette consistency. (e) Photograph of the object. (f) Rendering of the reconstructed shape. (Refer to main text for the detail of each figure.)
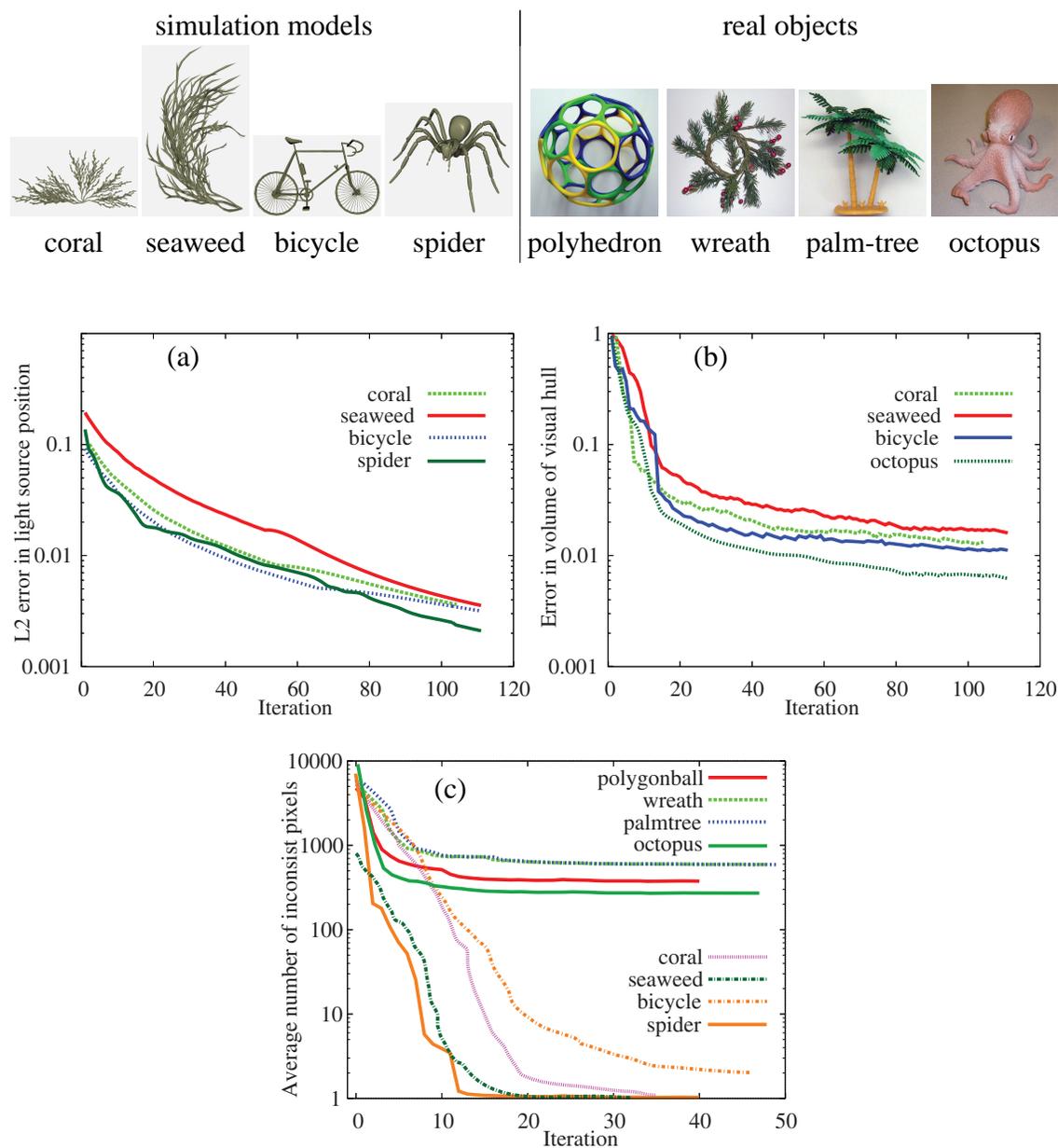
Figure 21: Convergence of error: (a,b) Errors in light source positions and visual hull are computed using ground truth for simulation models. (a) The error in the source position computed by L2 distance in meter. (b) The volume ratio between actual and erroneously-reconstructed visual hulls. (c) Error in shadowgram consistency computed in pixels for both simulation and real data. All plots are in logarithmic scale.

light sources 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.

The convergence of the reconstruction algorithm is quantitatively evaluated in Figure 21. The error in light source positions estimated by the algorithm proposed in Section 5 is shown in the top-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 volume ratio between actual and erroneously-reconstructed visual hulls is presented in the top-right. The silhouette mismatch defined in Equation (26) is plotted on the bottom. 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.

7   Conclusion

7.1   Summary

Coplanar shadowgram imaging is an technique to easily capture a large number of silhouettes of intricate objects from a wide range of viewpoints, with high accuracy. The setup shown in Figure 4 is inexpensive and requires only off-the-shelf components like a camera, a rear-projection screen and a small light source. However, traditional shape-from-silhouettes (SFS) approaches are highly sensitive to (even tiny) errors in acquired silhouettes or light source positions. The epipolar geometry of the shadowgrams and the silhouette consistency based optimization described in this work are crucial to obtaining extremely accurate reconstructions of intricate shapes. Most approaches for shape-from-silhouettes have acquired shapes of non-intricate objects (simple models of people, statues and figurines) that can be modeled with a small number of views. We believe this is the first time such accurate shapes have been estimated automatically using a large number of views (50 to 120) of objects with severe occlusions, discontinuities and thin structures.

7.2   Discussion

A single 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 the system with more than one shadowgram screen (or move one screen to different locations). The algorithm of the multi-screen coplanar shadowgram imaging can be divided into offline and online steps:

side-direction   frontal-direction
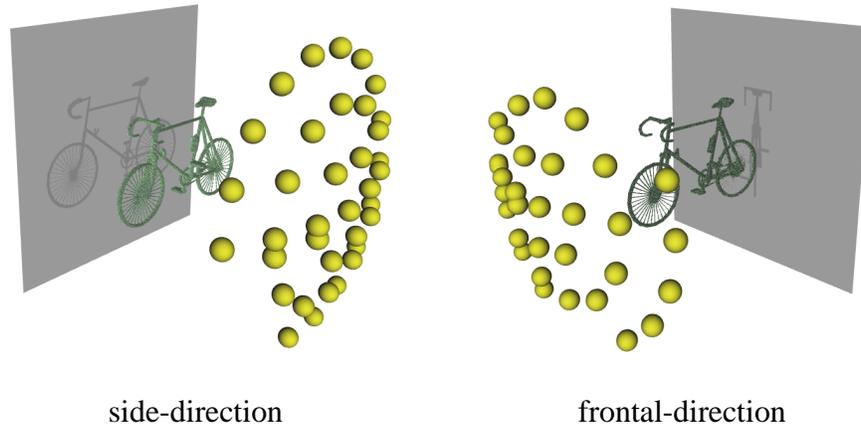
Figure 22: Two different configurations of coplanar shadowgrams of a bicycle object: Gray rectangle and yellow spheres indicate respectively a shadow screen and light source position. 36 light sources are used in both configurations. The screen is rotated by 90 degrees, while the object remains fixed. For the demonstration of the two-screen algorithm, a small number of light sources are used.

Off-line Calibration (one-time): This calibration can be done in several ways and we mention a simple one here. In the case of two-screen setup which is observed by a single camera, we only need to estimate the homography between each screen and image plane. The extra work required over the one-screen case is an additional homography estimation. The homographies in turn can be used to recover the relative transformation between the screens.

Online Calibration: In the two-screen setup, we can estimate the light source positions for each set of shadowgrams on one screen separately using the technique demonstrated in the paper. Finally, we merge the two sets of results using the relative orientation between the screens resulting from the off-line calibration.

In principle, it is possible to also optimize (minimize) the errors due to off-line calibration. However, the off-line intrinsic calibration of a camera and the screen-to-image homography can be done carefully. More importantly, it is independent of the complexity of the object and the number of source positions.

We have performed simulations with a bicycle object with two screen positions as shown in Figure 22. The bicycle was chosen since frontal and side views are both nec-

essary to carve the visual hull satisfactorily. Combining the two sets of shadowgrams enlarges the coverage of source positions, which successfully reduces the stretching artifact of the reconstructed shape in Figure 23.

7.3 Future Work

One 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 penumbra due to the finite size of a light source are our directions of future work.

Acknowledgements

References

[1] Shuntaro Yamazaki, Srinivasa Narasimhan, Simon Baker, and Takeo Kanade. Coplanar shadowgrams for acquiring visual hulls of intricate objects. In *Proc. International Conference on Computer Vision '07*, October 2007.

[2] Bruce Guenther Baumgart. *Geometric modeling for computer vision*. PhD thesis, Stanford University, 1974.

[3] Sudipta N. Sinha, Marc Pollefeys, and Leonard McMillan. Camera network calibration from dynamic silhouettes. In *Proc. Computer Vision and Pattern Recognition '04*, volume 1, pages 195–202, 2004.

[4] Kong Man Cheung, Simon Baker, and Takeo Kanade. Shape-from-silhouette across time, Part I: Theory and algorithms. *International Journal of Computer Vision*, 62(3):221–247, 2005.

[5] Alvy Ray Smith and James F. Blinn. Blue screen matting. In *Proc. SIGGRAPH '96*, pages 259–268, 1996.

(a) original shape

(b) side shadowgrams

(c) frontal shadowgrams

(d) side + frontal

reconstructed shape   close-up view of pedals

Figure 23: Comparison of shape reconstruction: We synthesized 36 coplanar shadow-grams of a 3D shape shown in (a). The visual hull of the object is reconstructed from: (b) the shadowgrams taken from side-direction (Figure 22 left) and (c) the shadowgrams taken from frontal-direction (Figure 22 right). The reconstructed shape is stretched into the direction perpendicular to a shadow screen due to the lack of views parallel to the screen. (d) Combining shadowgrams (b) and (c) enlarges the coverage of light source positions, which successfully reduces the stretching artifact in the reconstructed shape.

[6] Yuri Boykov and Gareth Funka-Lea. Graph cuts and efficient n-d image segmentation. *International Journal of Computer Vision*, 70(2):109–131, 2006.

[7] Neill Campbell, George Vogiatzis, Carlos Hernández, and Roberto Cipolla. Automatic 3d object segmentation in multiple views using volumetric graph-cuts. In *Proc. The British Machine Vision Conference '07*, pages 530–539, 2007.

[8] Yasutaka Furukawa, Amit Sethi, Jean Ponce, and David Kriegman. Robust structure and motion from outlines of smooth curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):302–315, 2006.

[9] Silvio Savarese, Marco Andreetto, Holly Rushmeier, Fausto Bernardini, and Pietro Perona. 3D reconstruction by shadow carving: Theory and practical evaluation. *International Journal of Computer Vision*, 71(3):305–336, 2005.

[10] Gary S. Settles. *Schlieren & Shadowgraph Techniques*. Springer-Verlag, 2001.

[11] Robert Hooke. *Micrographia*, chapter Observation LVIII. London, 1667.

[12] Alexandru Balan, Leonid Sigal, Michael Black, and Horst Haussecker. Shining a light on human pose: On shadows, shading and the estimation of pose and shape. In *Proc. International Conference on Computer Vision '07*, 2007.

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

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

[16] Roberto Cipolla, Kalle Åström, and Peter Giblin. Motion from the frontier of curved surfaces. In *Proc. International Conference on Computer Vision '95*, pages 269–275, 1995.

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

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

[19] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2004.

[20] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *Proc. SIGGRAPH '96*, pages 303–312, 1996.

[21] Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, Jonathan Shade, and Duane Fulk. The digital michelangelo project: 3D scanning of large statues. In *Proc. SIGGRAPH'00*, pages 131–144, 2000.

[22] Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. Computer Vision and Pattern Recognition '06*, volume 1, pages 519–526, 2006.

[23] Hideki Hayakawa. Photometric stereo under a light-source with arbitrary motion. *JOSA*, 11(11):3079–3089, 1994.

[24] Manmohan K. Chandraker, Fredrik Kahl, and David J. Kriegman. Reflections on the generalized bas-relief ambiguity. In *Proc. Computer Vision and Pattern Recognition '05*, volume 1, pages 788 – 795, 2005.

[25] Athinodoros S. Georghiades. Incorporating the torrance and sparrow model of reflectance in uncalibrated photometric stereo. In *Proc. International Conference on Computer Vision '03*, volume 2, pages 816–823, 2003.

[26] Ondrej Drbohlav and Radim Sara. Specularities reduce ambiguity of uncalibrated photometric stereo. In *Proc. the 7th European Conference on Computer Vision*, pages 46–62, 2002.

[27] Ondrej Drbohlav and Mike Chantler. Can two specular pixels calibrate photometric stereo? In *Proc. International Conference on Computer Vision '05*, pages 1850–1857, 2005.

[28] Ping Tan, Satya Mallick, David Kriegman, Long Quan, and Todd Zickler. Isotropy, reciprocity and the gbr ambiguity. In *Proc. Computer Vision and Pattern Recognition '07*, pages 1–8, 2007.

[29] William Henry Besant. *Conic sections, Treated geometrically*. Cambridge, 1890.

[30] Andrew Fitzgibbon, Maurizio Pilu, and Robert Fisher. Direct least squares fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(5):476–480, 1999.

[31] Kalle Åström, Roberto Cipolla, and Peter Giblin. Generalised epipolar constraints. *International Journal of Computer Vision*, 33(1):51–72, September 1999.

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

[33] Wojciech Matusik, Chris Buehler, Ramesh Raskar, Steven J. Gortler, and Leonard McMillan. Image-based visual hulls. In *Proc. SIGGRAPH '00*, pages 369–374, 2000.

[34] William Press, Brian Flannery, Saul Teukolsky, and William Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.

[35] Jean-Sébastien Franco and Edmond Boyer. Exact polyhedral visual hulls. In *Proc. the 15th British Machine Vision Conference*, pages 329–338, 2003.

Appendix

A   Derivation of Equation (2)

The projective transformation in coplanar shadowgram imaging is viewed as a perspective transformation in the translated coordinate system whose origin is at the location of a light source. Thus, the projection by the source located at $l = (u, v, w, 1)^T$ is written in matrix form as

$$P(l) = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{drop } z \,(\equiv 0)} \cdot \begin{pmatrix} I_3 & l \\ \mathbf{0}_3^T & 1 \end{pmatrix} \cdot \underbrace{\begin{pmatrix} -wI_3 & \mathbf{0}_3 \\ (0,0,1) & 0 \end{pmatrix}}_{\text{project to } \Pi} \cdot \begin{pmatrix} I_3 & -l \\ \mathbf{0}_3^T & 1 \end{pmatrix}$$

$$= \begin{pmatrix} -w & 0 & u & 0 \\ 0 & -w & v & 0 \\ 0 & 0 & 1 & -w \end{pmatrix}$$

where I_3 is a 3×3 identity matrix and $\mathbf{0}_3 = (0, 0, 0)^T$.

B   Derivation of Equation (12)

The linear transformation A is generally written in a 4×4 matrix as

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}. \tag{29}$$

Then, the elements in Equation (11) can be expanded into a 3×4 matrix using Equation (2) and Equation (29), which yields twelve equations. Using the six equations that are identical to 0 in the left side, we obtain

$$a_{21} = a_{31} = a_{41} = a_{12} = a_{32} = a_{42} = a_{14} = a_{24} = a_{34} = 0. \tag{30}$$

Similarly,

$$a_{11} = a_{22} = a_{44} \neq 0 \tag{31}$$

because $w \neq 0$ and $w' \neq 0$. Dividing A by $a_{11}(= a_{22} = a_{44} \neq 0)$, we can reparameterize A to Equation (12). By solving Equation (11) with respect to l', the transformation from l to l' can also be obtained as Equation (14). To guarantee that l and l' are in the same side with respect to the shadowgram plane, the additional constraints on the elements in A is derived as Equation (13).

C   Proof of Proposition 1

Proof

   Suppose the silhouette $\mathcal{S}_i^O$ of an object O is generated by perspective projection P_i $(i = 1, \cdots, N)$. The convex hull of the object and the silhouettes are respectively defined as

$$\hat{O} \overset{\text{def}}{=} \bigcup_{x_j, x_k \in O} \overline{x_j x_k} \qquad \text{and} \tag{32}$$

$$\hat{\mathcal{S}}_i^O \overset{\text{def}}{=} \bigcup_{m_j, m_k \in \mathcal{S}_i^O} \overline{m_j m_k} \tag{33}$$

where the over-line $\overline{p_j p_k}$ represents the line segment spanned by two points p_j and p_k.

Then, the shadowgram $\mathcal{S}_i^{\hat{O}}$ of the convex object $\hat{O}$ generated by the projection P_i is

$$\mathcal{S}_i^{\hat{O}} \stackrel{\text{def}}{=} \mathsf{P}_i \cdot \hat{O} \tag{34}$$

$$= \bigcup_{x_j, x_k \in O} \mathsf{P}_i \cdot \overline{x_j x_k} \qquad (\because \text{Equation (32)}) \tag{35}$$

$$= \bigcup_{m_j, m_k \in \mathcal{S}_i^O} \overline{m_j m_k} \qquad (\because \mathsf{P}_i \text{ is linear.}) \tag{36}$$

$$= \hat{\mathcal{S}}_i^O \; . \qquad (\because \text{Equation (33)}) \tag{37}$$

∎

D   Proof of Proposition 2

Proof   Suppose the visual hull $\mathcal{V}$ of an object O is reconstructed from the shadowgrams $\mathcal{S}_i^O$ and the projections P_i ($i = 1, \cdots, N$). Proposition 1 implies that the convex hull $\hat{O}$ of the object generates the the convex hulls $\hat{\mathcal{S}}_i^O$ of the shadowgrams:

$$\hat{\mathcal{S}}_i^O = \mathsf{P}_i \cdot \hat{O} \tag{38}$$

Hence, the convex shadowgrams $\hat{\mathcal{S}}_i^O$ are consistent by definition.   ∎