Visual Hulls from Single Uncalibrated Snapshots Using Two Planar Mirrors

Keith Forbes[1] Anthon Voigt[2] Ndimi Bodika[2]

[1]Digital Image Processing Group
Department of Electrical Engineering
University of Cape Town
Private Bag, Rondebosch, 7701
kforbes@dip.ee.uct.ac.za

[2]Automation and Informatics Group
GTS Technology
De Beers Group Services
P. O. Box 82851, Southdale, 2135
{Anthon.Voigt, Ndimi.Bodika}@debeersgroup.com

Abstract

Two mirrors are used to create five views of an object: a view onto the object, two reflections and two reflections of reflections. The five views are captured in a single snapshot. Epipolar geometry of the object's five silhouettes is determined directly from the image without knowing the poses of the camera or the mirrors. The epipolar geometry provides constraints on the pose of each silhouette, allowing the pose of each silhouette to be computed in a common reference frame using only the silhouette outlines. Once the pose associated with each silhouette has been computed, a five-view visual hull of the object can be computed from the five silhouettes. By capturing several images of a rigid object in different poses, sets of five silhouettes can be combined into a single silhouette set in which the pose of each silhouette is known in a common reference frame. This allows visual hulls of an arbitrary number of views to be computed if more than one image is used. The method is applied to an ornamental cat, and experimental results are shown.

1 Introduction

The *visual hull* provides a relatively simple means for creating an approximate 3D model of an object: it is the largest object that is consistent with a set of silhouettes of the actual object whose poses (positions and orientations) are known in a common reference frame. Silhouettes of an object are typically captured by multiple calibrated cameras positioned at different viewpoints that are well-spaced about the viewing hemisphere. The pose associated with each camera (and its silhouette) is usually pre-computed using a calibration object [2].

If the poses of the silhouettes are not known, then it is not, in general, possible to compute the poses of the observed silhouettes from the silhouettes themselves [1]. However, if some information about the poses is known, then in some cases it is possible to compute the poses of the silhouettes using only the silhouettes as input. This is done by computing a set of poses that is *consistent* with the silhouettes: if one silhouette implies that a volume of space is empty, another silhouette in the set cannot indicate that some part of the volume contains the object.

The computer vision literature holds several examples of computing silhouette poses from silhouettes under certain constraints. For instance, Mendonça et al. [5] use a turntable to create silhouettes with circular motion. The poses associated with each silhouette are computed from the silhouettes using the constraint of circular motion. Okatani and Deguchi [6] use a camera with a gyro sensor so that the orientation component of each silhouette pose is known. The translational component is then computed from the silhouettes.

In this work, we use two mirrors to create a scene consisting of a real object and four virtual objects (see Figure 1). Two of the virtual objects are reflections of the real object, and the other two virtual objects are reflections of these reflections. Five views of the object are captured in a single snapshot of the scene. Segmenting the image yields five silhouette views of the object. Without knowing the poses of either the camera or the mirrors in advance, the poses associated with each of the five silhouettes are computed from the silhouettes. This allows us to to create a five-view visual hull model of the object from a single snapshot. If the object is rigid, then we can obtain several five-view silhouette sets of the object in different poses. This can be done by moving the object, the camera, or the mirrors. We then use our previously described method [3] to combine the silhouette sets into a single silhouette set in which each silhouette's pose is known in a common reference frame. This allows us to create a visual hull model from the single set that is a better approximation to the 3D shape of the object than visual hulls created from any of the original five-view sets.

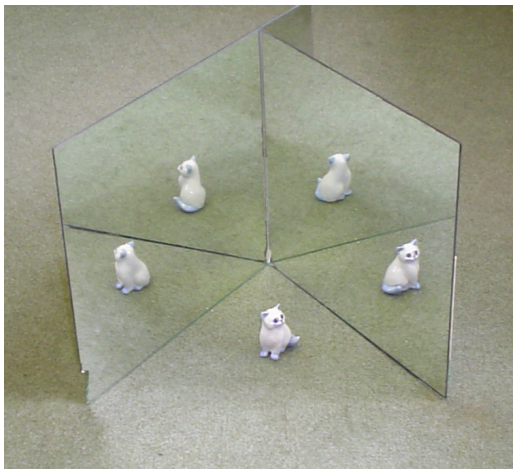


Figure 1: Two mirrors are used so that five views of the object can be seen.

Unlike other methods for computing visual hulls, our method does not require specialised equipment (no turntables, calibration objects, or synchronised cameras are required), and does not require special care to be taken in positioning any of the apparatus, since silhouette poses are computed from the image. We use two 300 mm × 300 mm × 3 mm off-the-shelf, bathroom-style mirrors and a four megapixel consumer-grade digital camera. (The increase in the number of pixels of digital cameras over the last few years has made the proposed method possible: the resolution of each of the five silhouettes within a single image is reasonably high.) The mirrors are positioned so that two of the sides are next to each other, and the angle between the mirrors is approximately 72° (as in Figure 1). (In principle, one could create more than five views by having a smaller angle between the mirrors, but in this work we restrict ourselves to the five-view case.) The object is placed in front of the mirrors so that five views of it can be seen. To aid segmentation, a background is chosen that contrasts with the colour of the object.

The computation of the silhouette poses is based on the observation that the epipolar geometry* relating two silhouette views of an object and its reflection in a single image can be computed without knowing the relative poses of the silhouettes, and without knowing point correspondences. We use an orthographic projection model for each of the silhouettes, since the variation in depth associated with each silhouette is small with respect to the distance to the camera. The pose parameters are divided into several components which are solved one by one: viewing direction, silhouette roll, and translation.

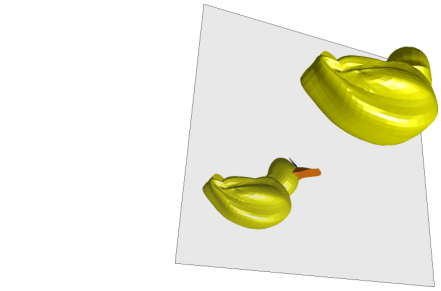
The remainder of this paper is organised as follows. Section 2 describes how the epipolar geometry relating silhouette views of an object and its reflection can be computed directly from the silhouette outlines. This observation is the basis for our method of determining silhouette poses for a double mirror setup. The geometry of the double mirror setup is described in Section 3, and terminology and view relationships used in the remainder of the paper are introduced. Section 4 shows how silhouettes can be scaled to create a good approximation to the silhouette image that would be observed by an orthographic camera. Section 5 explains how the poses of each of the five silhouettes visible in an image of the scene can be determined from the silhouette outlines. Experimental results are presented in Section 6, and the paper is summarised in Section 7.

2 Silhouette Epipolar Geometry with One Mirror

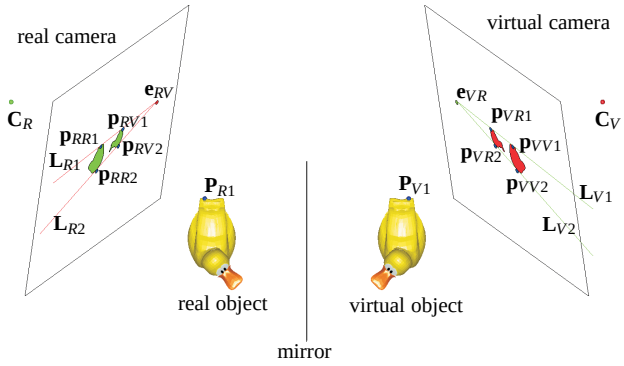
In this section, we show that the epipolar geometry relating the silhouette views of an object and its reflection can be determined by considering only the outline of the object and the outline of its reflection in a single image; neither the pose of the camera with respect to the mirror, nor point correspondences are required. This discussion uses perspective projections, however we will assume orthographic projections (a special case in which the cameras are at infinity) for the pose estimation described in Section 5.

Figure 2(a) shows a camera's view of a duck and its reflection.

*Definitions of terms relating to the geometry of multiple silhouette views are omitted in this paper because of space constraints. Please refer to our previous paper for these definitions [3].



(a)



(b)

Figure 2: Reflection of a duck in a mirror: (a) shows the image seen by the real camera, (b) shows the silhouette views seen by the real camera and by the virtual camera that is the reflection of the real camera.

Figure 2(b) shows the camera and the observed silhouette views of the real object and the virtual object (the reflection of the real object). The figure also shows a virtual camera (the reflection of the real camera) and the silhouettes that such a camera would observe. Note that the image observed by the virtual camera is simply a reflection of the image observed by the real camera. Although the virtual camera does not exist, we can determine the images that it would observe from the real camera's image: the virtual camera's view of the real object is simply the mirror image of the real camera's view of the virtual object. We can therefore consider the two available silhouettes (captured by the real camera) to be two views of the real object: one from the real camera's viewpoint, and one from the virtual camera's viewpoint.

Consider a plane π_1, perpendicular to the mirror, passing through the camera centre $\mathbf{C}_R$ and tangent to the top of the real object at $\mathbf{P}_{R1}$. Since π_1 is perpendicular to the mirror, it will also contain the reflections of $\mathbf{C}_R$ and $\mathbf{P}_{R1}$: $\mathbf{C}_V$ and $\mathbf{P}_{V1}$. Since the plane contains both camera centres, it will also contain both epipoles $\mathbf{e}_{RV}$ and $\mathbf{e}_{VR}$. The projections of $\mathbf{P}_{R1}$ and $\mathbf{P}_{V1}$ will lie on the silhouette outlines for both cameras, since π_1 is tangent to both objects. The real camera's projections of $\mathbf{P}_{R1}$ and $\mathbf{P}_{V1}$ are the points $\mathbf{p}_{RR1}$ and $\mathbf{p}_{RV1}$. Since $\mathbf{p}_{RR1}$, $\mathbf{p}_{RV1}$ and the epipole $\mathbf{e}_{RV}$ lie in both the image plane and π_1, they are collinear.

Now consider a second plane π_2 that is perpendicular to the

mirror, passes through the camera centre $\mathbf{C}_R$, but is tangent to the *bottom* of the real object at $\mathbf{P}_{R2}$. ($\mathbf{P}_{R2}$ is not shown in Figure 2(b) since it is on a hidden surface of the duck.) The plane π_2 can be used to show that $\mathbf{p}_{RR2}$, $\mathbf{p}_{RV2}$ and $\mathbf{e}_{RV}$ are collinear.

The points $\mathbf{p}_{RR1}$, $\mathbf{p}_{RV1}$ can be determined in an image by simply finding the line $\mathbf{L}_{R1}$ that is tangent to both silhouettes in the image. The epipole need not be known. The line $\mathbf{L}_{R2}$ is the tangent line on the other side of the silhouettes. The intersection of $\mathbf{L}_{R1}$ and $\mathbf{L}_{R2}$ is the epipole.

In the case of an orthographic projection, the epipoles lie at infinity. In this case the slope of the tangent lines is a projection of the camera's direction.

3 Double Mirror Scene Geometry

We use a scene with two mirrors for two reasons: (1) to obtain multiple silhouette views of an object, and (2) to obtain multiple silhouette tangent lines that provide sufficient information about the pose of the silhouettes that the poses of the silhouettes can be computed. Figure 3(a) shows the camera's view of a scene with two mirrors. The figure introduces some of the terminology that will be used later in this paper to determine the silhouette poses. Four virtual objects are also shown in the figure: V_1 is the reflection of R in Mirror 1; V_2 is the reflection of R in Mirror 2; V_{12} is the reflection of V_2 in Mirror 1; and V_{21} is the reflection of V_1 in Mirror 2. The normal vectors $\mathbf{m}_1$ and $\mathbf{m}_2$ for each of the two mirrors are shown.

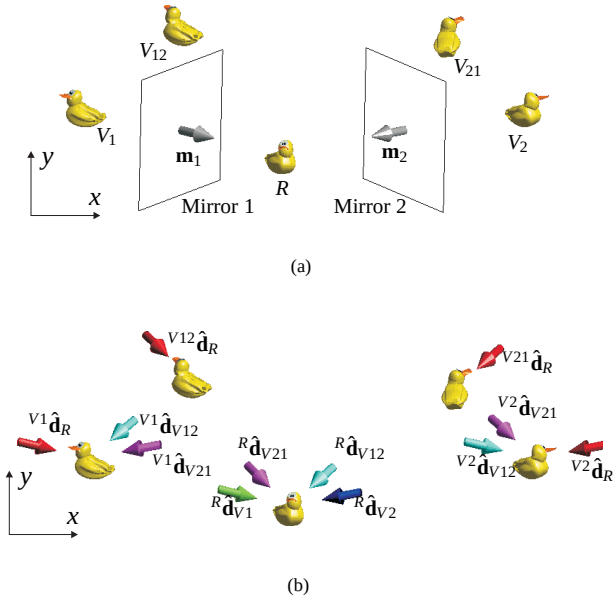


Figure 3: A double mirror scene in which five objects can be seen: (a) shows the positions of the two mirrors, and (b) shows various viewing direction vectors that are used in this work.

In Section 5, each of the five objects will be used to define a reference frame, and the mirror normals will be used to transform directions between reference frames. To determine the reflected direction $\mathbf{d}_r$ from a unit direction $\hat{\mathbf{d}}$ using unit mirror normal $\hat{\mathbf{m}}$ the following equation is used.

$$\mathbf{d}_r = \mathbf{d} - 2\hat{\mathbf{m}}(\hat{\mathbf{d}} \cdot \hat{\mathbf{m}}) \tag{1}$$

Figure 3(b) shows viewing directions in different reference frames. In each object's reference frame there are five views corresponding to the five available silhouettes (the silhouettes segmented from the image captured by the real camera). There are therefore 25 viewing directions that correspond to available silhouette views. The figure shows only the viewing directions that are used for computing the silhouette poses in Section 5. (The five directions of the form $[0,0,1]^T$ are omitted since they point directly into the page.) Note that the silhouettes observed in the reference frames of V_1 and V_2 are mirror images of the available silhouettes. Since we will be dealing with orthographic projections, the viewing direction $-\mathbf{d}$ will observe the non-mirrored silhouettes observed by viewing direction $\mathbf{d}$. We can therefore select viewing directions so that we only deal with non-mirrored silhouettes.

We use the superscript notation used by Forsyth and Ponce [4] to indicate the reference frame associated with a vector so that ${}^A d_B$ indicates the viewing direction of camera B onto object A (i.e. in the reference frame of A). The camera view that captures the observed silhouette of an object defines the reference frame so that ${}^A d_A = [0,0,1]^T$ for all A. The reference frame x- and y-axes are aligned with the image axes.

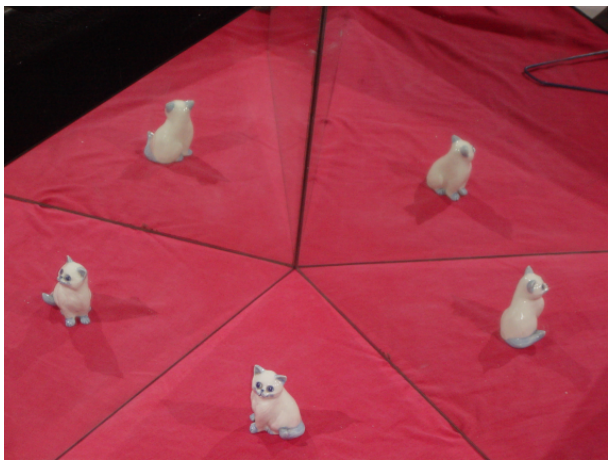
4 Approximating Orthographic Views

Figure 4(a) shows an example of an image of a scene. The segmented image consisting of five silhouette outlines is shown in Figure 4(b). Figure 4(c) shows an approximation to an orthographic projection of the scene that is derived from the perspective projection shown in Figure 4(b). Epipolar tangency lines are shown in Figures 4(b) and (c).

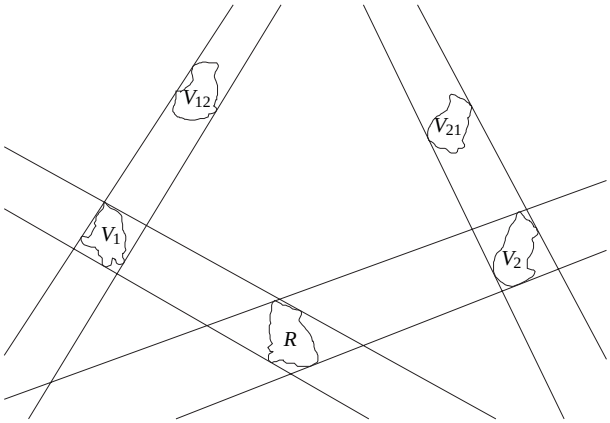
In order to compute the pose associated with each silhouette, we treat the silhouettes as if they were produced by an orthographic projection. This is done for the purpose of simplicity; the orthographic imaging model yields tractable equations without any significant affect on accuracy.

Consider moving the camera backwards from the scene and zooming so that the size of the silhouette of the real object R stays the same. As the camera is moved backwards and zoomed, the silhouettes of the virtual objects will become larger. The change in shape of the silhouettes is negligible, since the depth variation of the objects is small compared with the distances to the initial camera position. When the camera is moved back to infinity, we have orthographic imaging conditions. A very good approximation to the silhouettes that would be seen by such an orthographic camera is created by scaling the silhouettes of the virtual objects in the image captured by the real camera.

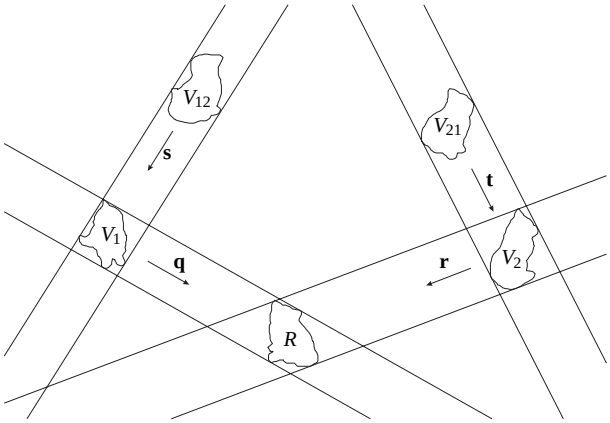
The results of scaling to create an orthographic projection are shown in Figure 4(c). The epipolar tangent lines are adjusted so that they are parallel and tangent to the silhouette. The direction of the new epipolar lines is that of the sum of two unit vectors parallel to each of the original epipolar lines. The silhouette of the reflection is then scaled so that it is tangent to both new epipolar



(a)



(b)



(c)

Figure 4: Images of a scene: (a) shows the raw image, (b) shows the segmented image with silhouette outlines and epipolar tangency lines, and (c) shows the derived orthographic image that would be seen by an orthographic camera.

lines. This process is applied to the four silhouettes of the virtual objects.

The 2D vectors **q**, **r**, **s**, and **t** indicated in Figure 4(c), are parallel to the tangent lines. These 2D vectors are determined from the tangent lines which are in turn determined directly from the silhouette outlines.

Note that **q** is a projection of ${}^R\mathbf{d}_{V1}$ onto the *xy*-plane. Once **q** is known, only the *z*-component of ${}^R\mathbf{d}_{V1}$ is unknown. Similarly, **r** is the projection of ${}^R\mathbf{d}_{V2}$; **s** is the projection of ${}^{V1}\mathbf{d}_{V12}$; and **t** is the projection of ${}^{V2}\mathbf{d}_{V21}$.

5 Computing Silhouette Poses

To determine the silhouette poses we exploit the constraints imposed by the directions of **q**, **r**, **s**, and **t** which are measured directly from the silhouette outlines. This allows us to compute the viewing directions for all five views of the real object. Once the viewing directions are known, the remaining component of orientation, the silhouette roll, can be computed. The translational component can then be determined for a given silhouette by using two other silhouettes to enforce the epipolar tangency constraint.

5.1 Viewing Direction

Here, we aim to compute the five viewing directions in the reference frame of *R*. To determine the five viewing directions we set up equations that specify the directions of **s** and **t** in terms of the direction vectors ${}^R\mathbf{d}_{V1}$ and ${}^R\mathbf{d}_{V2}$. Since only the *z*-components of ${}^R\mathbf{d}_{V1}$ and ${}^R\mathbf{d}_{V2}$ are unknown (the *x*- and *y*-components are determined from **q** and **r**), we will have two equations in two unknowns. This will allow us to solve for the *z*-components, and then to solve for the remaining direction vectors and mirror normals whose values are given in terms of the elements of ${}^R\mathbf{d}_{V1}$ and ${}^R\mathbf{d}_{V2}$.

The viewing direction ${}^R\mathbf{d}_{V1}$ consists of three components:

$$ {}^R\mathbf{d}_{V1} = \begin{pmatrix} {}^Rd_{xV1} \\ {}^Rd_{yV1} \\ {}^Rd_{zV1} \end{pmatrix}. \tag{2} $$

Note that ${}^Rd_{xV1}$ and ${}^Rd_{yV1}$ are selected so that $({}^Rd_{xV1}, {}^Rd_{yV1})^T = \mathbf{q}$. The magnitude of **q** is unimportant. Unit magnitude may be chosen, for instance.

The mirror normal $\mathbf{m}_1$ is given by the equation

$$ \mathbf{m}_1 = 1/2({}^R\hat{\mathbf{d}}_{V1} - {}^R\hat{\mathbf{d}}_R), \tag{3} $$

where ${}^R\mathbf{d}_R = [0,0,1]^T$, and the hat symbol indicates unit norm so that $\hat{\mathbf{v}} = \mathbf{v}/\|\mathbf{v}\|$. Similar equations are formulated for ${}^R\mathbf{d}_{V2}$ and $\mathbf{m}_2$.

In order to determine an expression for **s**, ${}^{V1}\hat{\mathbf{d}}_{V12}$ must be determined in terms of the elements of ${}^R\mathbf{d}_{V1}$ and ${}^R\mathbf{d}_{V2}$. This is done by starting with the vector ${}^{V12}\hat{\mathbf{d}}_{V12} = [0,0,1]^T$ and reflecting it first with Mirror 1, then with Mirror 2 and then again with Mirror 1, so that it is in the appropriate reference frame. This can be done, since the mirror normals are known in terms of ${}^R\mathbf{d}_{V1}$ and ${}^R\mathbf{d}_{V2}$. The viewing direction ${}^{V2}\hat{\mathbf{d}}_{V12}$ is the reflection of ${}^{V12}\hat{\mathbf{d}}_{V12}$ in Mirror 1:

$$ {}^{V2}\hat{\mathbf{d}}_{V12} = {}^{V12}\hat{\mathbf{d}}_{V12} - 2\hat{\mathbf{m}}_1({}^{V12}\hat{\mathbf{d}}_{V12} \cdot \hat{\mathbf{m}}_1). \tag{4} $$

The viewing direction ${}^{R}\hat{\mathbf{d}}_{V12}$ is the reflection of ${}^{V2}\hat{\mathbf{d}}_{V12}$ in Mirror 2:

$$ {}^{R}\hat{\mathbf{d}}_{V12} = {}^{V2}\hat{\mathbf{d}}_{V12} - 2\hat{\mathbf{m}}_2({}^{V2}\hat{\mathbf{d}}_{V12} \cdot \hat{\mathbf{m}}_2), \quad (5) $$

The viewing direction ${}^{V1}\hat{\mathbf{d}}_{V12}$ is the reflection of ${}^{R}\hat{\mathbf{d}}_{V12}$ in Mirror 1:

$$ {}^{V1}\hat{\mathbf{d}}_{V12} = {}^{R}\hat{\mathbf{d}}_{V12} - 2\hat{\mathbf{m}}_1({}^{R}\hat{\mathbf{d}}_{V12} \cdot \hat{\mathbf{m}}_1). \quad (6) $$

Similar equations are derived for ${}^{V1}\hat{\mathbf{d}}_{V21}$, ${}^{R}\hat{\mathbf{d}}_{V21}$, and ${}^{V2}\hat{\mathbf{d}}_{V21}$ to create an expression for $\mathbf{t}$.

Equation (6) can be used to express ${}^{V1}\hat{\mathbf{d}}_{V12}$ in terms of the components of ${}^{R}\mathbf{d}_{V1}$ and ${}^{R}\mathbf{d}_{V2}$. The unknowns are ${}^{R}d_{zV1}$ and ${}^{R}d_{zV2}$: ${}^{R}d_{V1x}$, ${}^{R}d_{V1y}$, ${}^{R}d_{V2x}$, and ${}^{R}d_{V2y}$ are computed from the epipolar tangency lines. Similarly, ${}^{V2}\mathbf{d}_{V21}$ can be formulated in terms of ${}^{R}\mathbf{d}_{V1}$ and ${}^{R}\mathbf{d}_{V2}$. Since the ratios ${}^{V1}d_{yV12}/{}^{V1}d_{xV12}$ and ${}^{V2}d_{yV21}/{}^{V2}d_{xV21}$ can be formulated in terms of the components of ${}^{R}\mathbf{d}_{V1}$ and ${}^{R}\mathbf{d}_{V2}$, and can also be measured from the silhouettes, two equations can be set up to solve for the two unknowns ${}^{R}d_{zV1}$ and ${}^{R}d_{zV2}$. The Matlab Symbolic Toolbox failed to find a solution to these two equations. However, solutions were found for ${}^{R}d_{zV1}$ in terms of ${}^{R}d_{zV2}$, and for ${}^{R}d_{zV2}$ in terms of ${}^{R}d_{zV1}$. (The equations are too large to reproduce here.) This allows us to set up an equation in terms of ${}^{R}d_{zV1}$. A value for ${}^{R}d_{zV1}$ can be determined by uniformly sampling direction vectors with appropriate x- and y-components so that the unknown angle varies between 0° and 360°. A few iterations of the bisection method can be applied in the vicinity of a root to obtain a result that is accurate to machine precision. When more than one root exists, we choose the solution that results in a consistent set of silhouettes.

5.2 Silhouette Roll

The roll is specified using the up vector, which is the direction of the silhouette's y-axis in the common reference frame. The up vector is $[0,1,0]^T$ in the view's reference frame. Once the mirror normals are known, the up vectors in the reference frame of R can easily be computed by appropriately reflecting the up vectors, starting in each view's reference frame.

5.3 Translational Component

Once the orientation of the silhouettes is known, the translational component of pose can be computed using the epipolar tangency constraint. Since the orientation is known, the epipolar directions are known. The translational component must be selected so that the silhouette is tangent to the projections of the epipolar tangency lines of the other silhouettes. The translational component is overconstrained, since each other silhouette provides two constraints (corresponding to the two outer epipolar tangencies). To estimate the translational component, we start by determining the translational component for the silhouette of V_1, using R as a reference frame. The mean of the translational components indicated by each of the two outer epipolar tangencies is used. The translational components for the remaining silhouettes are computed one by one by selecting the translational component so that the silhouette is tangent to projections of outer epipolar tangencies from silhouettes of R and V_1.

6 Experimental Results

Six images of an ornamental cat were used to test the proposed method. The cat was placed in six different poses.

6.1 Five-View Visual Hulls

Figure 5 shows the six five-view visual hulls computed from the six images. Silhouette poses were computed using the proposed method. The visual hulls give a reasonable gross approximation to the 3D shape of the cat.

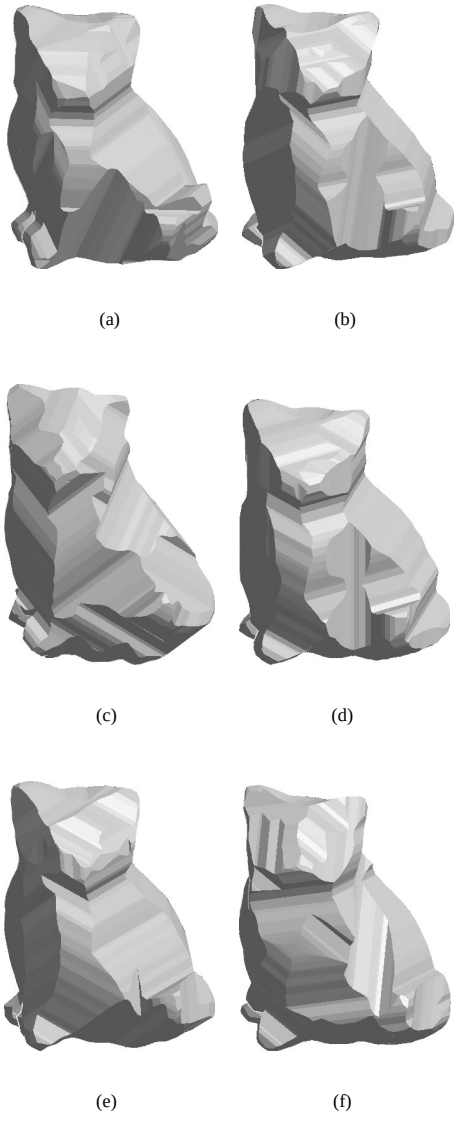


Figure 5: Six five-view visual hull models of an ornamental cat. The cat was placed in six different poses, but the visual hulls are shown aligned to aid comparison.

Figure 6 shows five silhouettes of the cat that were segmented from one of the images. Epipolar lines computed using the silhouette poses are shown in the figure. The epipolar lines are approxi-

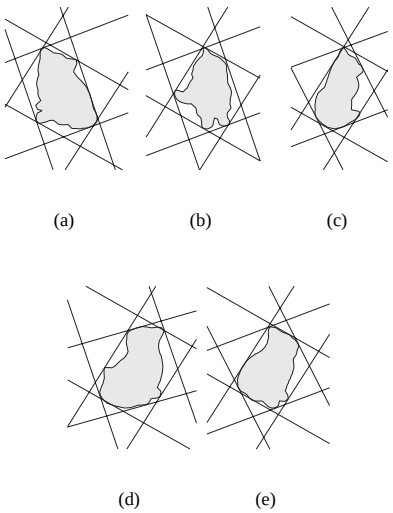


Figure 6: Five silhouettes showing epipolar tangency lines corresponding to the other four views in the set.

mately tangent to the silhouettes indicating that the silhouettes are consistent with the computed poses.

6.2 Visual Hulls from Merged Sets

The six silhouette sets were merged into a single silhouette set using our previously described method [3]. The silhouette sets were initially scaled so that the resultant visual hulls were of unit volume and the principal axes were origin-aligned. Silhouette inconsistency was then minimised by simultaneously adjusting the pose and scale parameters associated with each silhouette.

Three views of the 30-view visual hull with and without texture are shown in Figure 7. The texture was obtained by mapping regions of the input images to the faces of the polyhedral visual hull.

7 Summary

A novel method for forming visual hulls from single images has been presented. The method requires only easily obtainable apparatus: a digital camera and two planar mirrors. The mirrors are used to create multiple views of an object that are captured in a single image.

Neither the mirror poses, nor the camera pose are known in advance: silhouette poses are determined directly from the silhouette outlines in the image. No iterative multi-dimensional searches are required to determine the pose parameters. A simple, bounded, one-parameter search is required to solve one equation. Closed form solutions are given by the remaining equations.

Qualitative results have been presented showing visual hull models created using the proposed method. In addition, multiple silhouette sets obtained from several images of the object in different poses have been combined by determining the poses of all silhouettes in a common reference frame. Experimental results showing the visual hull model from such a merged silhouette set have been presented.

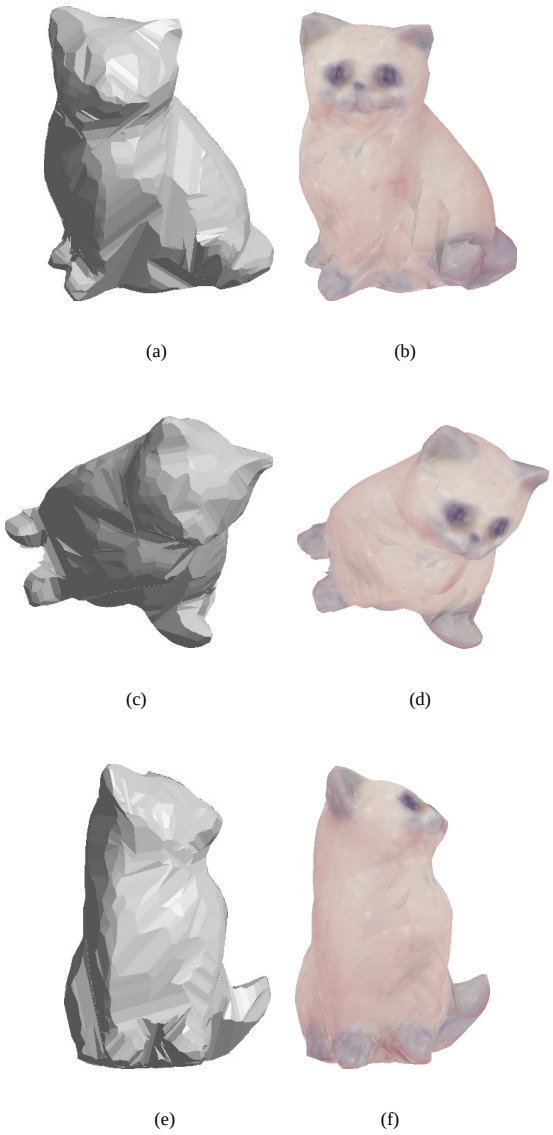


Figure 7: Three views of the 30-view visual hull: the first column shows the 3D shape and the second column shows the textured model.

References

[1] A. Bottino and A. Laurentini. Introducing a new problem: Shape-from-silhouette when the relative positions of the viewpoints is unknown. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11), November 2003.

[2] K. Forbes, A. Voigt, and N. Bodika. An inexpensive, automatic and accurate camera calibration method. In *Proceedings of the Thirteenth Annual South African Workshop on Pattern Recognition*, 2002.

[3] K. Forbes, A. Voigt, and N. Bodika. Using silhouette consistency constraints to build 3D models. In *Proceedings of the Fourteenth Annual South African Workshop on Pattern Recognition*, 2003. Available at `www.prasa.org`.

[4] David A. Forsyth and Jean Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2003.

[5] P. Mendonça, K.-Y. Wong, and R. Cipolla. Epipolar geometry from profiles under circular motion. *IEEE Trans. Pat. Anal. and Mach. Intel*, 23(6), 2001.

[6] T. Okatani and K. Deguchi. Recovering camera motion from image sequence based on registration of silhouette cones. In *Proceedings of the 6th IAPR Workshop on Machine Vision Applications (MVA2000)*, pages 451–454, 2000.