Creating Visual Hull Models Using Only Two Mirrors and an Uncalibrated Perspective Camera

Keith Forbes
Digital Image Processing Group
Department of Electrical Engineering
University of Cape Town
kforbes@dip.ee.uct.ac.za

Abstract

We present closed form solutions for the focal length and pose of a camera and two mirrors that are computed directly from the silhouette outlines that appear in a single image. In the noisy case, we show how these equations can be used to form an initial parameter estimate that can be refined with a nonlinear iterative minimisation. This allows five-view visual hulls to be constructed from a single image. We show how these five-view visual hulls can be used to form initial estimates of the similarity transforms that relate multiple five-view silhouette sets of a rigid object in different poses. A nonlinear iterative minimisation is then used to refine the solution so that all silhouettes are specified in a common reference frame, and visual hulls can be constructed from an arbitrary number of silhouettes. Experimental results demonstrating the reconstruction of a toy horse are presented.

1. Introduction

Shape-from-silhouette is a popular technique for creating 3D models of real world objects; silhouettes can often easily be extracted from images in a controlled environment. If camera pose and internal parameters are known, then the *visual hull* [4] can be computed by intersecting the visual cones corresponding to silhouettes captured from multiple viewpoints. The visual hull is often a good approximation to the 3D shape of the object and is useful for tasks such as 3D multimedia content creation.

We propose a simple setup for capturing images of an object from multiple well-distributed viewpoints. Two mirrors are used to create five views of an object: a view directly onto the object, two reflections, and two reflections of reflections (see Figure 1). Since the method requires only readily available equipment (two bathroom-style mirrors and a digital camera) it provides the non-specialist user with a simple low-cost means for creating 3D multimedia content from real objects. Other methods [8, 6] typically require specialist equipment such as turntables, calibration objects, or multiple camera systems.

We provide closed form solutions for the focal length and pose associated with each silhouette view. These values are computed directly from the silhouette outlines: no calibration markers or point correspondences are required. First, four epipoles are computed from the silhouette outlines. The positions of the epipoles provide constraints that allow the focal length of the camera to be computed. This can then be used to determine the direction of the mirror normals. Once the mirror normals are known, the orientation associated with each silhouette view is computed. Next, the positional component is computed using the epipolar tangency constraint. The solution can then be refined with an iterative nonlinear minimisation. Once the pose and focal length are computed, five-view visual hulls can be computed from the five silhouette views captured in a single image.

In some cases, five-view visual hulls provide a reasonable representation of the 3D shape of the object. However, the visual hull model can be improved by merging multiple five-view silhouette sets of the same rigid object into a

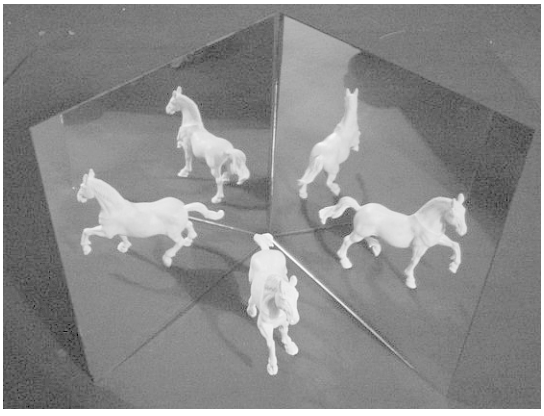

Figure 1: The double mirror setup used to capture five views of an object in a single image.

single large silhouette set. We show how five-view silhouette sets can be used to provide a good initial estimate of the similarity transform between different sets. The initial estimate is then refined using an iterative nonlinear minimisation. This allows visual hulls to be computed from an arbitrary number of views of an object.

The remainder of the paper is organised as follows. Section 2 provides a brief overview of related work. In Section 3 we demonstrate how a silhouette image of an object and its reflection can be used to compute the epipole from the silhouette outlines; this result will be used in computing the calibration parameters. Section 4 describes the geometry of the double mirror setup that is used to capture multiple views of an object. Section 5 presents closed form solutions for calculating camera focal length, mirror normals and position from the silhouette outlines observed in an image captured using our setup. In Section 6 we show how a nonlinear iterative minimisation can be used to refine the solution given by the closed form solutions in the presence of noise. Section 7 describes how multiple five-view silhouette sets of a rigid object in different poses can be specified in a common reference frame so that visual hull models can be constructed from an arbitrary number of views using the double mirror setup. Experimental results are presented in Section 8. Section 9 discusses the method and possible extensions, and Section 10 summarises the paper.

2. Related Work

The computer vision literature describes various approaches for capturing silhouettes of an object from multiple viewpoints so that the visual hull of the object can be computed. Several approaches have used the silhouettes themselves to estimate camera pose and internal parameters.

Wong and Cipolla [10] describe a system that is calibrated from silhouette views using the constraint of circular motion. Once an initial visual hull model is constructed from an approximately circular motion sequence, additional views from arbitrary viewpoints can be added to refine the model. The user must manually provide an approximate initial pose for each additional view which is then refined using an iterative optimisation. Their method of minimising the sum-of-square reprojection errors corresponding to all outer epipolar tangents is used in our work to provide a refined solution.

Okatani and Deguchi [9] use a camera with a gyro sensor so that the orientation component associated with each silhouette view is known. An iterative optimisation method is then used to estimate the positional component from the silhouettes by enforcing the epipolar tangency constraint.

Bottino and Laurentini [1] provide methods for determining viewpoints from silhouettes for the case of orthographic viewing directions parallel to the same plane. This type of situation applies to observing a vehicle on a planar surface, for instance.

In previous work [3], we describe a similar method to the one we describe in this paper. However, the previous work assumes an orthographic projection model and requires a one dimensional iterative search to determine initial parameter estimates. In this paper, we improve on the method by providing closed form solutions for the initial parameter estimates using a perspective camera model.

Moriya et al. [7] describe a related idea. Epipoles are computed from the silhouette outlines of three shadows of a solid cast onto a plane, and are shown to be collinear.

3. Epipoles from Bitangent Lines

This section deals with the case in which a camera views an object and its reflection. We show how the epipole corresponding to the virtual camera (the reflection of the real camera) can be computed directly from the silhouette outlines of the real object and the virtual object in the image captured by the real camera. This result will be used to calculate the positions of epipoles for the double mirror setup.

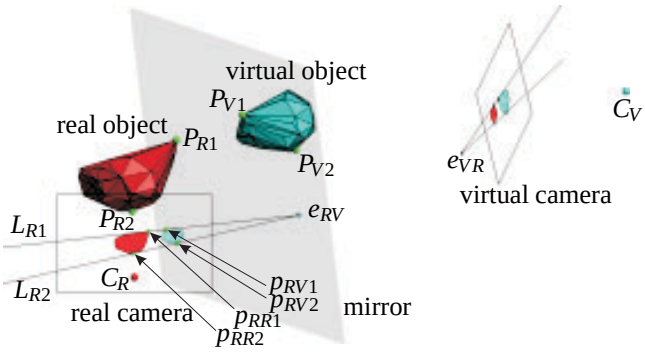

Figure 2: A camera viewing an object and its reflection. The epipole corresponding to the virtual camera can be computed from the silhouette bitangent lines L_{R1} and L_{R2}.

Figure 2 shows an example of a camera observing a real object and its reflection in a mirror. The virtual camera is also shown in the figure. Consider a plane Π_1 that passes through the camera centres C_R and C_V and touches the real object at the point P_{R1}. By symmetry, Π_1 will touch the virtual object at the point P_{V1} which is the reflection of P_{R1}. Since Π_1 is tangent to both objects and contains the camera centres C_R and C_V, P_{R1} and P_{V1} are *frontier points* [2]. They project onto the silhouette outlines on the real image at points p_{RR1} and p_{RV1}. The points p_{RR1}, p_{RV1} and the epipole e_{RV} (the projection of C_R into the real image) are therefore collinear, since they lie in both Π_1 and the real image plane. Observe that the bitangent line L_{R1} passing through these three points can be computed directly from

the silhouette outlines: it is simply the line that is tangent to both silhouettes. Another bitangent line L_{R2} passes through the epipole and touches the silhouettes on the opposite side to L_{R1}. These tangency points lie on a plane Π_2 that is tangent to the opposite side of the object and passes through both camera centres. Provided that the object does not intersect the line passing through both camera centres, there will always be two outer epipolar tangent lines L_{R1} and L_{R2} that touch the silhouettes on either side.

The position of the epipole e_{RV} can therefore be computed by determining L_{R1} and L_{R2} from the silhouette outlines. The epipole is located at the intersection of L_{R1} and L_{R2}. Note that the epipole is computed without requiring knowledge of the camera pose and without requiring any point correspondences.

We also note that by symmetry, the real camera's silhouette view of the virtual object is a mirror image of the virtual camera's silhouette view of the real object. The silhouette view observed by a reflection of a camera is therefore known if the camera's view of the reflection of the object is known.

4. Double Mirror Setup

Figure 3 shows an example of a double mirror setup that is used to capture five silhouette views of an object in a single image. The camera is centred at C_R and observes a real object O_R. The camera also captures the image of each of four virtual objects O_{V1}, O_{V2}, O_{V12}, and O_{V21}. Object O_{V1} is the reflection of O_R in Mirror 1; O_{V2} is the reflection of O_R in Mirror 2; O_{V12} is the reflection of O_{V1} in Mirror 2; and O_{V21} is the reflection of O_{V2} in Mirror 1.

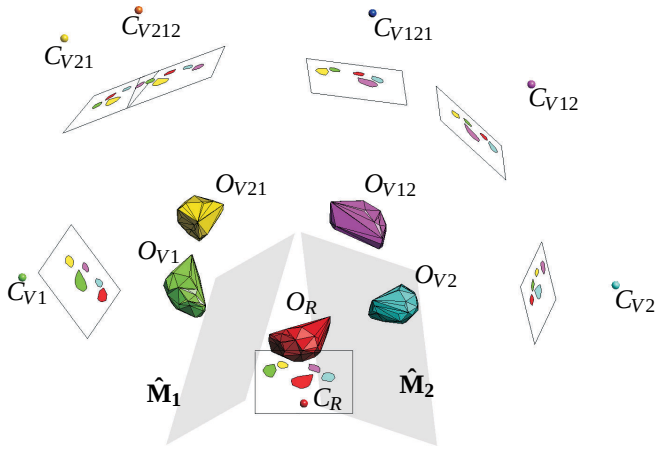

Figure 3: Double mirror setup showing one real object, one real camera, four virtual objects, and six virtual cameras.

Our method requires six virtual cameras to be considered. The virtual cameras are reflections of the real camera. The virtual cameras C_{V1}, C_{V2}, C_{V12}, and C_{V21} are required, as their silhouette views of the real object are the same as the silhouettes observed by the real camera (or reflections thereof). Since we have access to the silhouettes from the real camera, we can determine the silhouettes observed by the four virtual cameras. Each of the five cameras' silhouette views of the real object will then be used in computing the five-view visual hull of the object.

The virtual cameras C_{V121}, and C_{V212} (triple reflections of C_R) are to be considered too, since it turns out that their epipoles can be computed directly from the five silhouettes observed by the real camera. These epipoles, together with the epipoles from the virtual cameras C_{V1} and C_{V2} can then be used to calculate the focal length of the camera.

5. Analytical Solution

This section presents a method to calculate the focal length of the camera and the poses of the virtual cameras relative to the pose of the real camera. Closed form solutions in which the required parameters are determined from the silhouette outlines alone are provided. Silhouette outlines are represented by polygons. The camera model assumes square pixels and a principal point that is positioned at the image centre.

First, we show how lines that are tangent to pairs of silhouettes can be used to calculate the position of four epipoles corresponding to four virtual cameras. Next, we show how the focal length is a function of the relative positions of these four epipoles. Once the focal length is known, we show that mirror and camera orientation is easily determined from the positions of two epipoles. Finally, it is shown how the positional component of the poses can be computed using the epipolar tangency constraint.

5.1. Four Epipoles from Five Silhouettes

Here, we show how the epipoles are computed from pairs of silhouettes using the result explained in Section 3: the epipole corresponding to a camera's reflection can be computed from the camera's silhouette image of an object and its reflection by finding the intersection of the two outer bitangent lines. Figure 4 shows how the epipoles e_{V1}, e_{V2}, e_{V121}, and e_{V212} are computed from the outlines of the five silhouettes observed by the real camera. The distances a, b, and c between the epipoles will be used for computing the focal length. The outline γ_{RR} corresponds to the object O_R, and γ_{RV1} corresponds to O_{V1} which is the reflection of O_R in Mirror 1. The intersection of the pair of lines that are tangent to both γ_{RR} and γ_{RV1} is therefore the epipole e_{V1}, since C_{V1} is the reflection of C_R in Mirror 1. The pair of lines that are tangent to both γ_{RV2} and γ_{RV12} also meet at e_{V1}, since O_{V12} is the reflection of O_{V2} in Mirror 1. Similarly, the pairs of lines that are tangent to both γ_{RR} and γ_{RV2},

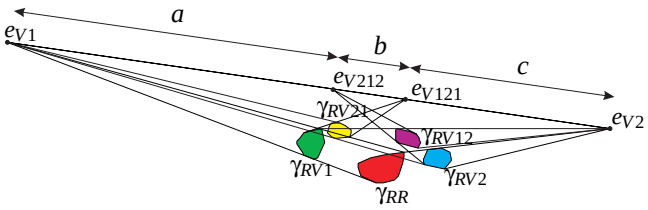

Figure 4: Computing epipoles e_{V1}, e_{V2}, e_{V121}, and e_{V212} from the five silhouette outlines in the image.

and to γ_{RV1} and γ_{RV21} meet at e_{V2}.

Consider C_R observing O_{V1}. Object O_{V12} is related to O_{V1} through three reflections. Object O_{V1} must be reflected by Mirror 1 and then Mirror 2 and then again by Mirror 1 to get O_{V12}. The effect of these three reflections can be considered to be a single reflection. Applying the triple reflection to C_R gives C_{V121}. The pair of lines that are tangent to both γ_{RV1} and γ_{RV12} therefore meet at e_{V121}. This is again because a camera (C_R) is observing silhouettes of an object (O_{V1}) and its reflection (O_{V12}), so the projection of the camera's reflection (C_{V121}) can be computed from the silhouette bitangent lines. Similarly, the pair of lines that are tangent to both γ_{RV2} and γ_{RV21} meet at e_{V212}.

Note that the epipoles e_{V1}, e_{V2}, e_{V121}, and e_{V212} are collinear, since they all lie in both the image plane of the real camera and in the plane Π_C in which all camera centres lie.

5.2. Focal Length from Four Epipoles

We now show how the focal length is computed from the positions of the four epipoles e_{V1}, e_{V2}, e_{V121}, and e_{V212}. This will be done by considering the positions of the camera centres in the plane Π_C.

First we introduce two new mirrors, Mirrors A and B, that do not correspond to physical mirrors in the scene. This approach makes the problem of calculating the focal length tractable. Mirror A has the same orientation as Mirror 1, but is positioned so that it passes midway between e_{V1} and C_R (see Figure 5 in which the positions of points in Π_C are shown). The point e_{V1} is therefore the reflection of C_R in Mirror A. Point E is the reflection of e_{V1} in Mirror 2, and F is the reflection of E in Mirror A. Note that F lies on the ray passing through e_{V121} and C_R. Also note that F will stay on this line if the position (but not the orientation) of Mirror 2 changes. This is so because triangles $\triangle C_R C_{V1} D$ and $\triangle C_R e_{V1} G$ are similar.

Figure 6 shows the positions of the epipoles and C_R in Π_C. The distances a, b, and c between the epipoles (as shown in the figure) are used to compute the distance f_Π between C_R and the image plane in the plane Π_C. The distance f_Π is then used to calculate the focal length. The figure also shows Mirror B which has the same orientation as Mirror 2,

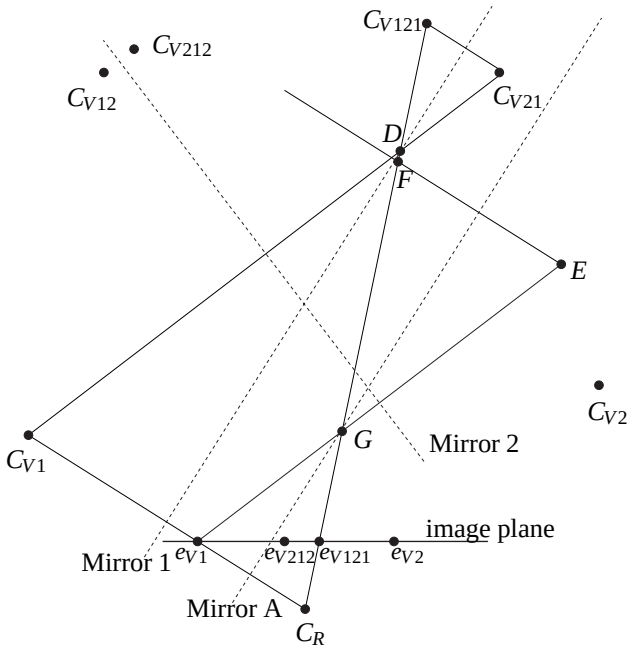

Figure 5: The intersections of Mirror 1, Mirror A and Mirror 2 with Π_C along with the positions of the cameras and epipoles, all of which lie in Π_C.

and is positioned midway between C_R and e_{V2}. The line joining e_{V2} to its reflection in Mirror B meets Mirror B at point J which projects onto e_{V212}.

The triangle $\triangle H e_{V2} C_R$ is similar to $\triangle C_R e_{V1} G$, the line segment from e_{V121} to e_{V2} is of length c, and the line segment from e_{V1} to e_{V121} is of length $a+b$. This indicates that the ratio of the sides of $\triangle H e_{V2} C_R$ to $\triangle C_R e_{V1} G$ is $c : (a+b)$. This means that $d(e_{V1}, G) = d(C_R, e_{v2})(a+b)/c$. (The notation $d(x,y)$ indicates the distance between x and y.)

Similarly, the triangle $\triangle K e_{V1} C_R$ is similar to $\triangle C_R e_{V2} J$, the line segment from e_{V1} to e_{V212} is of length a, and the line segment from e_{V212} to e_{V2} is of length $b+c$. This indicates that the ratio of the sides of $\triangle K e_{V1} C_R$ to $\triangle C_R e_{V2} J$ is $a : (b+c)$. Therefore $d(e_{V2}, J) = d(C_R, e_{V1})(b+c)/a$.

This allows us to write $d(C_R, e_{V1})$ in terms of $d(C_R, e_{V2})$, since $\triangle C_R e_{V2} J$ is similar to $\triangle C_R e_{V1} G$:

$$d(C_R, e_{V1}) = \frac{\sqrt{c\,(c+b)\,a\,(a+b)}}{c\,(c+b)} d(C_R, e_{V2}). \qquad (1)$$

We now know the sides of $\triangle C_R e_{V1} G$ up to a scale factor.

The angle $\angle C_R e_{V1} G = \alpha + \beta$ can be computed using the cosine rule:

$$\cos(\alpha+\beta) = 1/2\, \frac{\sqrt{c\,(c+b)\,a\,(a+b)}}{(c+b)\,(a+b)}. \qquad (2)$$

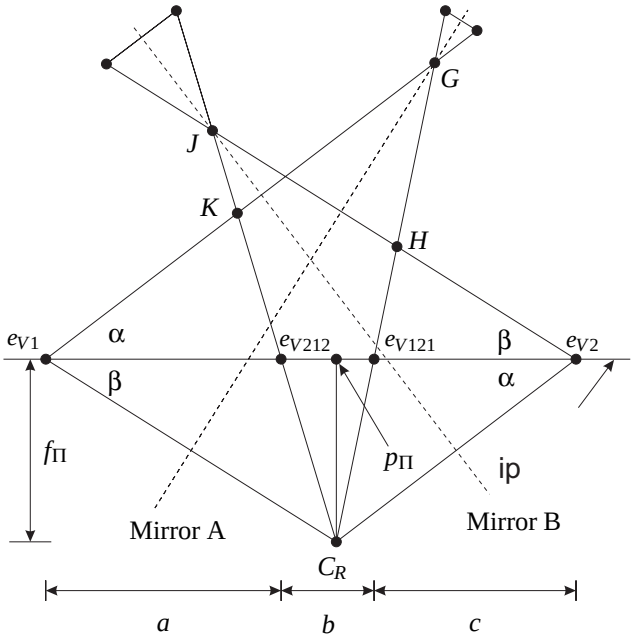

Figure 6: Computing the focal length from the four epipoles e_{V1}, e_{V2}, e_{V121}, and e_{V212}.

The cosine rule can be used to determine the sides of $\triangle e_{V1}C_R e_{V2}$. (The angle $\angle e_{V1}C_R e_{V2} = 180° - \alpha - \beta$.)

We can now (with the help of the Matlab Symbolic Toolbox for simplification) state f_Π in terms of a, b, and c:

$$f_\Pi = 1/2\,\frac{\sqrt{3ac+4ab+4cb+4b^2}\,(a+b+c)\,\sqrt{a}\sqrt{c}}{a^2+ab+c^2+cb+ac}. \tag{3}$$

The point closest to C_R on the line containing the epipoles, is

$$\mathbf{p}_\Pi = \mathbf{e}_{V1} + 1/2\,\frac{(2a+2b+c)a\,(a+b+c)}{a^2+ab+c^2+cb+ac}\,\frac{\mathbf{e}_{V2}-\mathbf{e}_{V1}}{||\mathbf{e}_{V2}-\mathbf{e}_{V1}||}. \tag{4}$$

The focal length (the distance from C_R to the image plane) can now be calculated from $\mathbf{p}_\Pi$, the principal point $\mathbf{p}_0$ and f_Π:

$$f = \sqrt{f_\Pi^2 - ||\mathbf{p}_0 - \mathbf{p}_\Pi||^2}. \tag{5}$$

5.3. View Orientations

Once the focal length of the camera has been calculated, the view orientation can be computed relatively easily. The mirror normal directions $\mathbf{m}_1$ and $\mathbf{m}_2$ are computed from the focal length, the principal point $\mathbf{p}_0$ and the epipoles $\mathbf{e}_{V1}$ and $\mathbf{e}_{V2}$:

$$\mathbf{m}_1 = -\begin{bmatrix} \mathbf{e}_{V1} - \mathbf{p}_0 \\ f \end{bmatrix} \tag{6}$$

$$\mathbf{m}_2 = -\begin{bmatrix} \mathbf{e}_{V2} - \mathbf{p}_0 \\ f \end{bmatrix}. \tag{7}$$

A 3×3 matrix R that represents a reflection by a mirror with unit normal $\hat{\mathbf{m}} = [m_x, m_y, m_z]^T$ is used to calculate view orientation:

$$R = \begin{pmatrix} -m_x^2+m_y^2+m_z^2 & -2m_xm_y & -2m_xm_z \\ -2m_xm_y & m_x^2-m_y^2+m_z^2 & -2m_ym_z \\ -2m_xm_z & -2m_ym_z & m_x^2+m_y^2-m_z^2 \end{pmatrix}. \tag{8}$$

5.4. View Positions

The point C_{V1} is constrained to lie on the line passing through e_{V1} and C_R. Similarly, the point C_{V2} is constrained to lie on the line passing through e_{V2} and C_R. Since absolute scale cannot be inferred from the image (if the scene were scaled, the image would not change), we fix C_{V1} at unit distance from C_R. The only positional unknown across the entire setup is now the position of C_{V2} on the line passing through e_{V2} and C_R.

To solve for w, the distance from C_R to C_{V2}, we use the epipolar tangency constraint: a tangent to a silhouette outline that passes through the epipole must be tangent to the corresponding point in its projection into the image plane of the opposite view. The relationship between the silhouette views of cameras C_{V1} and C_{V2} is used to enforce this constraint.

The poses of the cameras C_{V1} and C_{V2} are specified by 4×4 rigid transform matrices from the reference frame of the real camera:

$$M = \begin{pmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix}, \tag{9}$$

where the translational component $\mathbf{t}$ is given by

$$\mathbf{t} = 2(m_xp_x + m_yp_y + m_zp_z)\begin{pmatrix} m_x \\ m_y \\ m_z \end{pmatrix}, \tag{10}$$

and $(p_x, p_y, p_z)^T$ is a point on the mirror.

The matrix $M_1M_2^{-1}$ represents the rigid transform from the reference frame of C_{V2} to that of C_{V1}.

The point p_{V2} is one of two outer epipolar tangencies formed by lines passing through e_{V2V1} (the projection of C_{V1} onto the image plane of camera C_{V2}) and tangent to the silhouette observed by the camera C_{V2}.

The point p_{V1V2} is the projection of p_{V2} into camera C_{V1}. It must correspond to p_{V1}, one of two outer epipolar tangencies formed by lines passing through e_{V1V2} (the projection of C_{V2} onto the image plane of camera C_{V1}).

The epipolar tangency constraint is expressed as

$$(\mathbf{p}_{V1V2} \times \mathbf{e}_{V1V2}) \cdot \mathbf{p}_{V1} = 0, \tag{11}$$

where $\mathbf{p}_{V1V2}$, $\mathbf{e}_{V1V2}$, and $\mathbf{p}_{V1}$ are represented by homogeneous coordinates. In other words, the line passing through $\mathbf{p}_{V1V2}$ and $\mathbf{e}_{V1V2}$ must also pass through $\mathbf{p}_{V1}$.

Equation (11) can be specified in terms of p_{V1}, p_{V2}, the computed orientation and camera internal parameters, and w. The Matlab Symbolic Toolbox was used to determine a solution for w (the equation is too large to reproduce here). Unfortunately, we do not know the values of either p_{V1} or p_{V2}, since the epipoles from which they may be computed are functions of the unknown w.

The values of p_{V1} and p_{V2} are determined by exhaustive search. Each vertex on the polygon representing γ_{V1} is considered to be p_{V1} in turn. For each trial of p_{V1}, each vertex of the polygon representing γ_{V2} is considered to be p_{V2} in turn. The validity of each candidate solution is then tested by considering the epipolar tangency constraint across all silhouette pairings for the computed value of w. Since there are $n = 5$ silhouettes, there are $n(n-1) = 20$ frontier points, because each silhouette pairing generates two frontier points from the two pairs of outer tangencies. However, the epipolar tangencies used to compute the epipoles e_{V1}, e_{V2}, e_{V121}, and e_{V212} are invariant to the value of w. This leaves eight frontier points that are affected by the value of w. These eight frontier points correspond to the four pairs of frontier points generated by the pairs (γ_{V1}, γ_{V2}), (γ_R, γ_{V12}), (γ_R, γ_{V21}), and (γ_{V12}, γ_{V21}). The values of p_{V1} and p_{V2} that fulfil the epipolar tangency constraint imposed by the eight frontier points are selected. In the presence of noise, the values corresponding to the lowest sum-of-square distances between tangency points and corresponding projected tangent lines may be selected.

6. Refined Self-Calibration

The method described in Section 5 is used as a means for determining initial parameter estimates for pose and focal length that are refined using an iterative nonlinear optimisation. First, a method of estimating the positions of the four epipoles e_{V1}, e_{V2}, e_{V121}, and e_{V212} is presented. Next we show how the pose and focal length is optimised over all silhouette pairings.

6.1. Refined Estimate of Epipoles

In the presence of noise, the tangent line intersections used to calculate the four epipoles will, in general, produce epipoles that are not collinear. The epipoles e_{V1} and e_{V2} each lie at the intersection of four tangent lines. In the presence of noise, four tangent lines will not intersect at a point. For a refined estimate, the positions of the four epipoles are parameterised using only six degrees of freedom (d.o.f.), so that the epipoles are constrained to be collinear. The Levenberg-Marquardt method is used to minimise the sum-of-square distances from tangency points to the corresponding tangent lines generated by the opposite silhouette. The tangent lines pass through the appropriate epipole and touch the silhouette. To form a starting point (initial estimate) for the minimisation, the tangent line intersections are computed, and the points closest to an orthogonal regression line through the intersection points are used.

6.2. Refined Estimate of View Parameters

The Levenberg-Marquardt method is used to adjust the parameters describing the focal length, the two mirror normals and the distance w by minimising the sum-of-square distances between epipolar tangencies and corresponding projected epipolar tangents (reprojection errors) across all silhouette pairings. Initial parameter values are estimated from the positions of the four epipoles using the closed form solutions described in Section 5.

7. Combining Silhouette Sets

In this section, we present a method for specifying the relative poses of multiple five-view silhouette sets in a common reference frame. This allows an arbitrary number of silhouettes to be captured, so that visual hulls can be computed that are a more accurate representation of the 3D shape of the object than five-view visual hulls.

7.1. An Initial Estimate

Since the relative poses of the silhouettes within each set are known, we can use the corresponding visual hull to form an initial estimate of the similarity transform relating the silhouette sets. The volume and the principal axes of a polyhedral representation of the visual hull are computed [5]. These are used to form an initial estimate of the similarity transform relating the different silhouette sets of the same object to the reference frame of the first silhouette set. The four-way ambiguity of aligning the principal axes is resolved by selecting the case that results in the lowest sum-of-square reprojection error across all silhouettes. For compact objects, the five-view visual hull may provide a poor estimate of the principal axes of the object. In this case, random rotations can be evaluated until a sufficiently low sum-of-square reprojection error is achieved.

7.2. A Refined Estimate

The Levenberg-Marquardt method is used to minimise the sum-of-square reprojection errors across all silhouettes within all silhouette sets. For k sets there are $5k$ silhouettes and $5k(5k-1)$ frontier points. Two projections of each frontier point are used, so there are $10k(5k-1)$ reprojection errors. The similarity transforms relating silhouette sets have seven d.o.f., but we over-parameterise the rotation component by using a quaternion representation, so there

are $8(k-1)$ parameters describing the similarity transforms between sets. (Similarity transforms are specified with respect to the first set.) The parameters describing the views within each set are also included in the optimisation. There are six d.o.f. within each set: two mirror normals, focal length, and the distance w. We over-parameterise the mirror normals using three-element vectors giving a total of eight parameters for describing the views within each set. A total of $8(2k-1)$ parameters are therefore adjusted during the minimisation process.

8. Experimental Results

The proposed method was tested using a toy horse. Three images were captured with the horse in different poses (see Figure 7). The five silhouettes in each image were determined using an intensity threshold. Figure 8 shows the four epipoles computed from the five silhouettes in one of the images. Figure 9 shows the viewpoints and the visual cones corresponding to the five silhouettes from the same image.

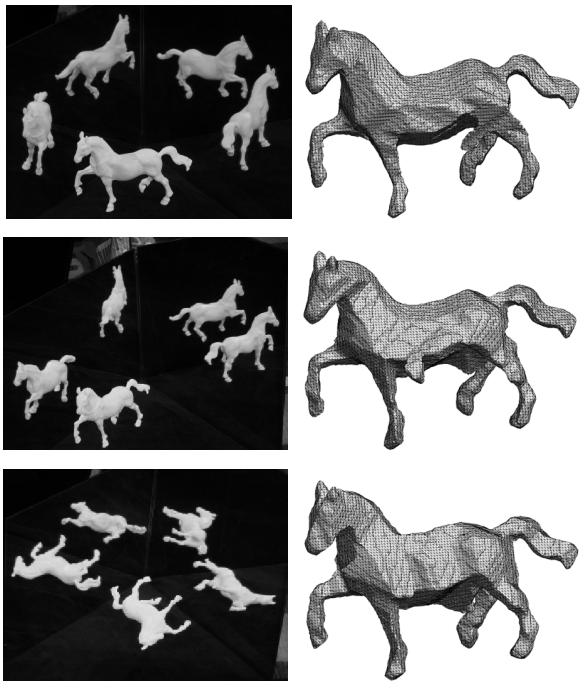

Figure 7: Input images and corresponding visual hulls.

The second column of Figure 7 shows the visual hulls computed from the five silhouette views in a single image. Each of the visual hulls has large regions of extra volume. The five-view visual hulls are therefore only useful as a coarse representation of the shape of the horse.

Better results are obtained when the fifteen-view visual hull is formed by specifying all silhouettes in a common reference frame as described in Section 7. Figure 10 shows the relative positions of the fifteen viewpoints and the corresponding visual cones. The fifteen-view visual hull is shown in Figure 11.

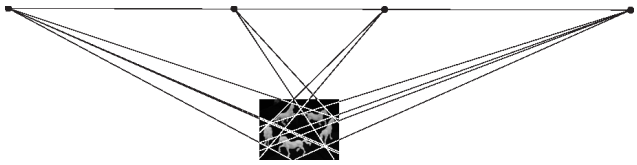

Figure 8: Four epipoles computed from bitangent lines.

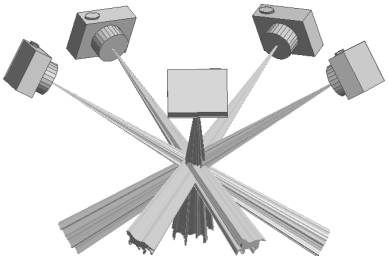

Figure 9: The relative position of the five silhouette views and visual cones from the first image.

9. Discussion

The proposed method of using two mirrors to capture multiple silhouette views of a rigid object provides a step towards moving 3D modelling of real objects out of the laboratory and into the hands of non-specialists. In this paper, we have restricted ourselves to five silhouette views, however, by reducing the angle between the mirrors, a larger number of silhouette views become available. As the pixel resolution of consumer-grade cameras continues to rapidly increase, the possibility of capturing entire turntable-like sequences in a single snapshot becomes increasingly attractive. The relatively small number of parameters that describe the view poses (four rotational d.o.f. and one positional d.o.f.) ensure that pose optimisation is computationally feasible. Moreover, this number of d.o.f. stays fixed as a larger number of views is obtained by decreasing the angle between the mirrors.

We also note that if an internally calibrated camera were used, the process of determining initial parameter estimates would be simplified and reflections of reflections would not be required. Mirror normals can be computed from the epipoles determined by bitangent line intersections using the silhouette image of the object and its reflection.

The usefulness of our method is not limited to 3D content creation: the coarse five-view visual hull created from

a single snapshot allows the 3D shape of a real object to be used to query a database of 3D models. The method can therefore be used as an aid for finding existing 3D models that are similar to a real object. For instance, using a five-view visual hull model of the toy horse as a query resulted in four horse models and one dinosaur model as the top five matches in the Princeton 3D Model Search Engine (see `http://shape.cs.princeton.edu`).

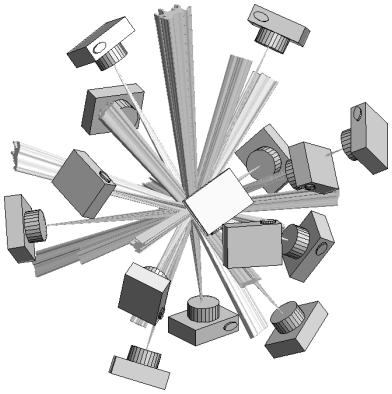

Figure 10: The relative position of the fifteen views in a common reference frame with the visual cones shown.

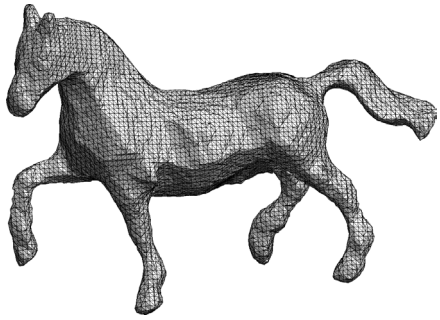

Figure 11: Visual hull formed by combining the three five-view silhouette sets into a common reference frame.

10. Summary

We have presented a method for creating 3D models from real world objects that is applicable to the non-specialist. The method requires only readily-available equipment: two off-the-shelf planar mirrors, and a digital camera. Once provided with the software, the non-specialist user will easily be able to create 3D multimedia content from real objects.

By positioning the mirrors and the object so that five views of the object can be seen, we have shown how the focal length and view poses associated with each silhouette can be computed from the silhouette outlines. Closed form solutions have been provided.

In the noisy case, these closed form solutions can be used as initial parameter estimates that can be further refined using an iterative nonlinear minimisation. This allows five-view visual hulls to be computed from silhouettes extracted from single snapshots without the need for calibration markers or point correspondences.

The five-view visual hulls of the same rigid object in different poses can be used to estimate the similarity transform relating the different silhouette sets. These estimates can then be refined using an iterative nonlinear minimisation so that all silhouettes are specified in a common reference frame. This allows us to construct visual hulls from an arbitrary number of views using the double mirror setup.

Experimental results demonstrating the use of the method to model a toy horse have been presented.

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] Roberto Cipolla and Peter Giblin. *Visual Motion of Curves and Surfaces*. Cambridge University Press, 2000.

[3] Keith Forbes, Anthon Voigt, and Ndimi Bodika. Visual hulls from single uncalibrated snapshots using two planar mirrors. In *Proceedings of the Fifeenth Annual South African Workshop on Pattern Recognition*, 2004.

[4] Aldo Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, 1994.

[5] S. Lien and J. Kajiya. A symbolic method for calculating the integral properties of arbitrary nonconvex polyhedra. *IEEE Computer Graphics and Applications*, 4(10):35–41, 1984.

[6] W. Matusik, C. Buehler, and L. McMillan. Polyhedral visual hulls for real-time rendering. In *Proceedings of Twelfth Eurographics Workshop on Rendering*, 2001.

[7] T. Moriya, F. Beniyama, K. Matsumoto, T. Minakawa, K. Utsugi, and H. Takeda. Properties of three shadows on a plane. In *Proceedings of the 12th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, 2004.

[8] A. Mülayim, U. Yilmaz, and V. Atalay. Silhouette-based 3D model reconstruction from multiple images. *IEEE Transactions on Systems, Man and Cybernetics*, 33(4), 2003.

[9] Takayuki Okatani and Koichiro 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.

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