

Shape-from-Silhouette with Two Mirrors and an Uncalibrated Camera

Keith Forbes¹, Fred Nicolls¹, Gerhard de Jager¹, and Anthon Voigt²

¹ Department of Electrical Engineering,
University of Cape Town, South Africa

{kforbes, nicolls, gdj}@dip.ee.uct.ac.za

² Automation and Informatics Research, De Beers Group Services,
Johannesburg, South Africa

Anthon.Voigt@debeersgroup.com

Abstract. Two planar mirrors are positioned to show five views of an object, and snapshots are captured from different viewpoints. We present closed form solutions for calculating the focal length, principal point, mirror and camera poses directly from the silhouette outlines of the object and its reflections. In the noisy case, these equations are used to form initial parameter estimates that are refined using iterative minimisation. The self-calibration allows the visual cones from each silhouette to be specified in a common reference frame so that the visual hull can be constructed. The proposed setup provides a simple method for creating 3D multimedia content that does not rely on specialised equipment. Experimental results demonstrate the reconstruction of a toy horse and a locust from real images. Synthetic images are used to quantify the sensitivity of the self-calibration to quantisation noise. In terms of the silhouette calibration ratio, degradation in silhouette quality has a greater effect on silhouette set consistency than computed calibration parameters.

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* [7] 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 Fig. 1). Two or more images of the object and its reflections are captured from different camera positions (without altering the internal parameters) to obtain a well-distributed set of silhouettes. 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. The user provides segmented images as input, and our software provides a visual hull model of the



Fig. 1. Two images of a double mirror setup positioned so that five views of the object can be seen. Note that the camera has moved between shots, but the mirrors and object have not moved.

object. Other methods [10, 8] typically require specialist equipment such as turntables, calibration objects, or multi-camera setups.

We provide closed form solutions for the focal length, principal point, 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, each five-view image is considered separately. Four epipoles are computed from the silhouette outlines. Each image constrains the principal point to lie on a line. The intersection of these lines yields the camera's principal point. The positions of the epipoles provide constraints that allow the focal length of the camera to be computed. The mirror normals are a function of the focal length, principal point, and positions of the epipoles. Once the mirror normals are known, the orientation associated with each silhouette view is computed with respect to the camera. Next, the positional component is computed using the epipolar tangency constraint.

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 single large silhouette set. We show how multiple five-view silhouette sets can be specified in a common reference frame using closed form solutions. This allows visual hulls to be computed from an arbitrary number of well-distributed views of an object.

A refined solution is obtained by treating the closed form solutions as initial estimates and then adjusting parameters to minimise the sum-of-square distances between epipolar tangencies and corresponding projected epipolar tangents. The various parameters are decoupled so that iterative refinement is applied at several steps using small numbers of parameters at each step, thus limiting the dimensionality of the search space.

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 focal length, principal point, mirror normals and positions from the silhouette outlines observed in images 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. Experimental results using real and synthetic data are presented in Section 7. Section 8 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 can be computed. Several approaches use the silhouettes themselves to estimate camera parameters.

Wong and Cipolla [13] 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.

Sinha et al. [12] make use of outer epipolar tangents to calibrate a network of cameras using silhouettes. Random sampling is used to identify consistent corresponding epipolar tangencies to use for computing initial parameter estimates.

Okatani and Deguchi [11] 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.

Many works describe the use of mirrors for generating multiple views of a scene. For example, Gluckman and Nayar [5] discuss the geometry and calibration of a two mirror system using point correspondences. Hu et al. [6] describe a setup similar to ours, however they use constraints imposed by both the silhouette outlines and point correspondences for calibration.

In earlier work [4], 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 dense search of parameter space to determine initial 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. [9] 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.

Fig. 2 shows an example of a camera observing a real object and its reflection in a mirror. The virtual camera is also shown. 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,

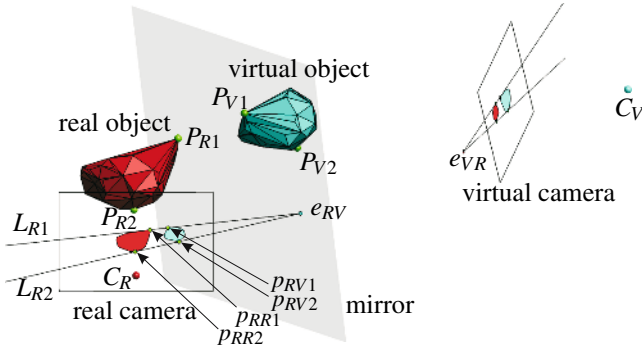


Fig. 2. A camera viewing an object and its reflection. The epipole e_{RV} corresponding to the virtual camera can be computed from the silhouette bitangent lines L_{R1} and L_{R2} .

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* [3]. 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. 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 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; it 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

Fig. 3 shows 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 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.

Our method requires six virtual cameras to be considered. The virtual cameras are reflections of the real camera C_R . The virtual cameras C_{V1} , C_{V2} , C_{V12} , and C_{V21} are

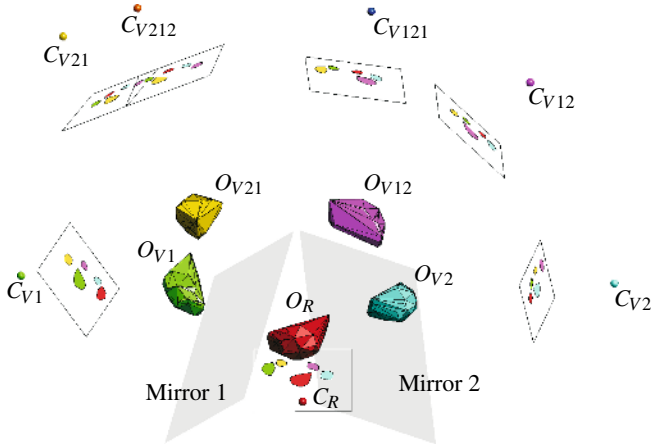


Fig. 3. Mirror setup showing one real and four virtual objects, and one real and six virtual cameras

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 can be used to compute the five-view visual hull of the object.

The virtual cameras C_{V121} (the reflection of C_{V12} in Mirror 1), and C_{V212} (the reflection of C_{V21} in Mirror 2) 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 and principal point of the camera and the poses of the virtual cameras relative to the pose of the real camera for the five camera views in an image. Next, a method for determining camera motion between snapshots is presented. This allows all silhouettes from all images to be specified in a common reference frame. Closed form solutions in which the required parameters are determined from the silhouette outlines alone are provided. Silhouette outlines are represented by polygons, and pixels are assumed to be square.

First, we show how lines that are tangent to pairs of silhouettes can be used to calculate the positions of four epipoles corresponding to four virtual cameras. The principal point is constrained by the epipoles to a line in each image; the intersection of the lines is the principal point. Next, we show how the focal length is a function of the relative positions of the four epipoles. Once the focal length is known, we show that mirror and camera orientation is easily determined from the positions of two epipoles. The positional component of the poses is computed using the epipolar tangency constraint.

Finally, we show how the camera poses between shots are constrained by the constant positions of the mirrors with respect to the object.

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

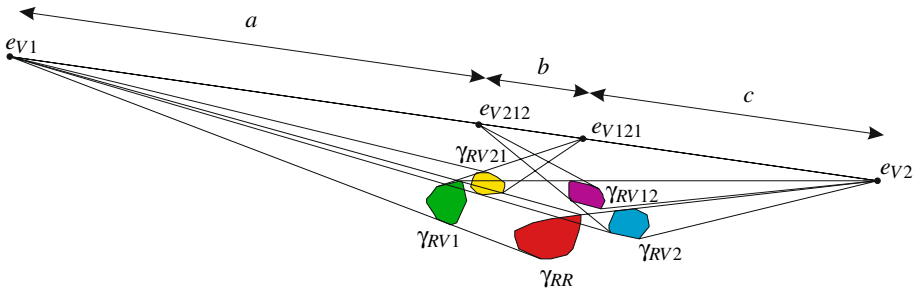


Fig. 4. Computing epipoles e_{V1} , e_{V2} , e_{V121} , and e_{V212} from the silhouette outlines in an image

Consider C_R observing O_{V1} . Object O_{V21} is related to O_{V1} through three reflections. Object O_{V1} must be reflected by Mirror 1 (to get O_R) and then Mirror 2 (to get O_{V2}) and then again by Mirror 1 to get O_{V21} . 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_{RV21} 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_{RV12} 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 and Principal Point from Epipoles

We now show how the focal length is computed from the positions of the four epipoles. This will be done by considering the positions of the camera centres in the plane Π_C .

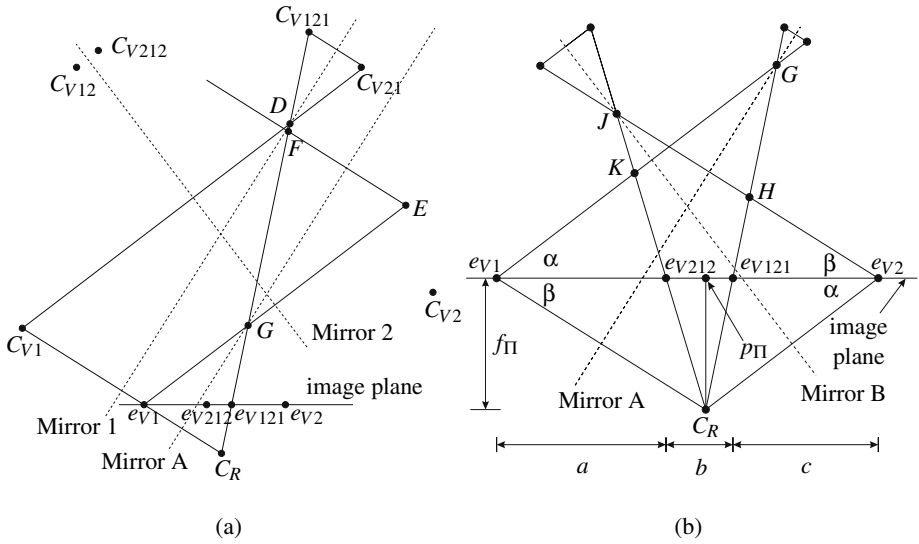


Fig. 5. Diagrams showing (a) the intersections of Mirror 1, Mirror A and Mirror 2 with Π_C along with the positions of the cameras and epipoles, all of which lie in Π_C , and (b) computing f_π and p_π from the four epipoles e_{V1} , e_{V2} , e_{V121} , and e_{V212}

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 Fig. 5a 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 because triangles $\triangle C_R C_{V1} D$ and $\triangle C_R e_{V1} G$ are similar.

Fig. 5b 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, 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 $d(e_{V1}, G) = d(C_R, e_{V2})(a + b)/c$.

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 $\triangle C_R e_{V1} G$:

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

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:

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

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

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

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

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

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

The line passing through p_Π and perpendicular to the line containing the epipoles passes through the principal point p_0 . The principal point can therefore be computed as the intersection of two such lines from two images of the scene. (If the principal point is assumed to lie at that the image centre, then a single snapshot could be used.)

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

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} :

$$\mathbf{m}_1 = - \left[\begin{array}{c} \mathbf{e}_{V1} - \mathbf{p}_0 \\ f \end{array} \right], \quad \mathbf{m}_2 = - \left[\begin{array}{c} \mathbf{e}_{V2} - \mathbf{p}_0 \\ f \end{array} \right]. \quad (5)$$

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:

$$R = \begin{pmatrix} -m_x^2 + m_y^2 + m_z^2 & -2m_x m_y & -2m_x m_z \\ -2m_x m_y & m_x^2 - m_y^2 + m_z^2 & -2m_y m_z \\ -2m_x m_z & -2m_y m_z & m_x^2 + m_y^2 - m_z^2 \end{pmatrix}. \quad (6)$$

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} , the epipolar tangency constraint is used: a tangent to a silhouette that passes through the epipole must be tangent to the corresponding point in its projection onto the image plane of the opposite view. The relationship between the views of 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:

$$M = \begin{pmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix}, \quad (7)$$

where the translational component \mathbf{t} is given by $\mathbf{t} = 2(m_x p_x + m_y p_y + m_z p_z)(m_x, m_y, m_z)^T$ and $(p_x, p_y, p_z)^T$ is a point on the mirror.

The matrix $M_1 M_2^{-1}$ 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

$$(\mathbf{p}_{V1V2} \times \mathbf{e}_{V1V2}) \cdot \mathbf{p}_{V1} = 0, \quad (8)$$

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 8 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} can be determined by exhaustive search, by finding the polygon vertex pair that fulfils the epipolar tangency constraint. Instead, we remove the need for an exhaustive search by using a parallel projection approximation to determine approximate correspondences. The tangencies are selected as the support points for outer tangent pairs that are parallel to the projected viewing direction. Unless the camera is very close to the object, this method selects either the same vertices, or vertices very close to the true tangencies under a perspective projection.

5.5 Combining Five-View Silhouette Sets

The calibration procedure described above allows five silhouette views from one image to be specified in a common reference frame. The pose and internal parameters of the four virtual cameras and one real camera are known. The silhouettes observed by these cameras are also known: the silhouettes observed by the virtual cameras are those observed by the real camera of the corresponding virtual object.

The next step is to specify the silhouette sets from two or more images in a common reference frame. This is easily achieved, since the mirror poses are known with respect to the real camera for each image. The five-view silhouette sets are aligned by aligning the mirrors across sets. There are two additional degrees of freedom that the mirrors do not constrain: a translation along the join of the mirrors, and an overall scale factor. These are approximated using the epipolar tangency constraint and a parallel projection model (as for computing w): each five-view silhouette set is scaled and translated along the mirror join so that outer epipolar tangents coincide with the projected tangents from silhouettes in the other silhouette set. Each silhouette pair between different sets gives an estimate of translation and scale. The average result over all pairings is used.

6 The Refined Self-calibration Procedure

The method described in Section 5 provides a means for computing all calibration parameters. However, better results are obtained if parameter estimates are refined at several steps. This is done by adjusting the parameters to minimise the sum-of-of square distances between epipolar tangencies and corresponding projected tangents using the Levenberg-Marquardt method. The geometry of the problem naturally allows for parameters to be decoupled from one another, allowing minimisation to be applied to small numbers of parameters at a time.

The first step of the procedure is to determine which silhouettes correspond to which camera views for each of the five silhouettes in the image. This is done by ordering the five silhouettes along the convex hull of the five silhouettes, and then considering the five possible arrangements. The four epipoles e_{V1} , e_{V2} , e_{V121} , and e_{V212} are computed for each of the five possible arrangements. The lowest sum-of-square distances between silhouette tangents passing through the epipoles and tangents on the corresponding silhouettes is used to select the correct arrangement.

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, so that the epipoles are constrained to be collinear. The sum-of-square distances from tangency points to the corresponding tangent lines generated by the opposite silhouette is minimised. The tangent lines pass through the appropriate epipole and touch the silhouette. To form a starting point for the minimisation, the tangent line intersections are computed, and the points closest to an orthogonal regression line through the intersection points are used.

Focal length and principal point values are then computed for each image, averaged, and adjusted to minimise reprojection error. The unknown positional component

is computed next for each image. Parameters are then adjusted by minimising reprojection error using all possible silhouette pairings between silhouettes within each set.

Finally, the five view sets are merged into a single large set as described in Section 5.5. A final minimisation adjusts all parameters simultaneously to minimise the sum-of-square distances across all silhouette pairings.

7 Experimental Results

Experiments were performed using real data to obtain qualitative results, and synthetic data to quantify the calibration performance degradation in the presence of noise.

7.1 Real Image Data

The proposed method was tested using a toy horse. Two 2592×1944 images captured from two viewpoints are shown in Fig. 1. The five silhouettes in each image were determined using an intensity threshold.

The resultant visual hull model is shown the third column of Fig. 6. The figure also shows visual hull models created using only the five silhouettes from each of the images. This demonstrates the improvement in the quality of the model obtained by merging the silhouette sets. Note that both five-view visual hulls have regions of extra volume that are not present in the ten-view visual hull.

The angle between the mirrors was computed to be 73.1 degrees. The focal length was computed to be 2754 pixels and the principal point located at (1306,981). This compares with values of 2875 and (1297,958) computed using a checkerboard calibration method (see www.vision.caltech.edu/bouguetj). Note, however, that a direct comparison of individual parameters does not necessarily provide a good indication of the quality of the calibration parameters. The calibration parameters should provide an accurate mapping from 2D image points to 3D rays *in the volume of interest*. The interplay between the different parameters can result in different parameter

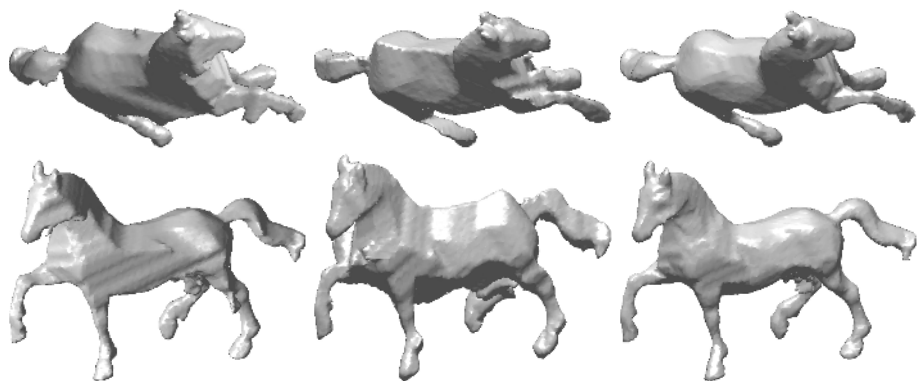


Fig. 6. Two views of the visual hull of the horse formed from the silhouettes in image 1 (*first column*), the silhouettes in image 2 (*second column*), and all ten silhouettes (*third column*)



Fig. 7. Two input images and resultant visual hull model of a toy locust

sets varying to some degree in magnitude, yet still providing a good mapping in the volume of interest. A difference in principal point location can largely be compensated for by a difference in translation parameters, for instance. A more meaningful measure of calibration parameter quality using the *silhouette calibration ratio* is described in Section 7.2.

Fig. 7 shows another example: a visual hull model of a toy locust.

7.2 Synthetic Image Data

Synthetic images were used to investigate the sensitivity of the method to noise. To ensure realistic parameter values were considered, the synthetic images were based on the real images of the toy horse. Exact polygonal projections of the ten-view polyhedral visual hull of the horse were generated using the output provided by the real images. This provides an exactly consistent set of silhouettes.

Quantisation noise was introduced by rendering the polygonal silhouettes, firstly at the original image resolution, and then at successively lower resolutions. Visual hulls

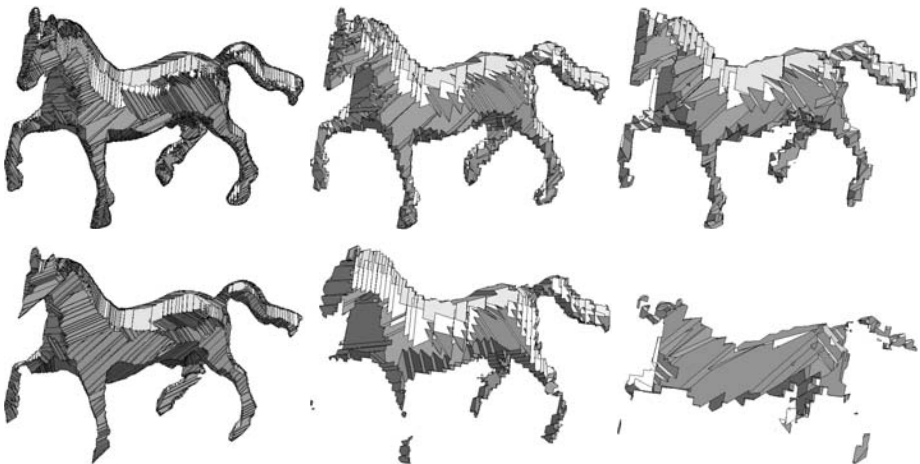


Fig. 8. Visual hull models created at 2, 8, and 15 times reduction of the original resolution (*left to right*), with iterative refinement (*top row*), and without iterative refinement (*bottom row*)

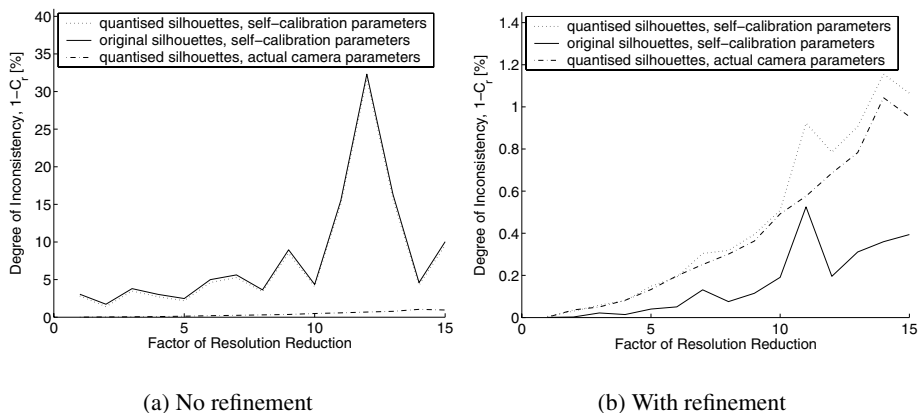


Fig. 9. Plots of image resolution versus silhouette inconsistency measured using the silhouette calibration ratio for self-calibration (a) without, and (b) with refinement

computed with and without iterative refinement are shown in Fig. 8 for three resolution levels. Note that without refinement, the computed calibration parameters cause the common volume of the visual cones to reduce substantially as the noise is increased.

Boyer [2] introduced the silhouette calibration ratio C_r as a measure of the combined quality of silhouettes and camera parameters. His reasoning is as follows. Ideally, some point on any viewing ray in a silhouette must intersect all $n - 1$ other visual cones of an n -view silhouette set. The ratio of the actual maximum number of intersections for points on the ray to $n - 1$ is a measure of consistency; C_r is the mean value for all rays from all silhouettes. We use $1 - C_r$ as a measure of inconsistency.

Fig. 9 shows plots of $1 - C_r$ versus the degree of resolution reduction. Results are also shown with the computed camera parameters and exact silhouettes, as well as quantised silhouettes and exact camera parameters. The plots show that without refinement, the poor accuracy of the camera parameters is a greater contributor to inconsistency than the quantisation of the silhouettes alone. However, for the refined camera parameters, the quantised silhouettes and exact camera parameters are more inconsistent than the exact silhouettes and the computed camera parameters, demonstrating the accuracy of the refined calibration method.

8 Summary

We have presented a method for creating 3D models from real world objects for 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 so that five views of the object can be seen, and capturing two or more images of the scene, we have shown how the internal parameters and poses associated with each silhouette can be computed from the silhouette outlines alone.

In the noisy case, closed form solutions are used for initial parameter estimates that are refined by Levenberg-Marquardt minimisation of sum-of-square reprojection error.

Experimental results demonstrating the quality of models created using real images have been presented. Synthetic images have been used to demonstrate the computed camera parameters have less of an effect on quality as measured by the silhouette calibration ratio than the noisy silhouettes from which they are computed.

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. Edmond Boyer. Camera calibration using silhouettes. Technical Report 5559, INRIA, 2005.
3. Roberto Cipolla and Peter Giblin. *Visual Motion of Curves and Surfaces*. Cambridge University Press, 2000.
4. K. Forbes, A. Voigt, and N. Bodika. Visual hulls from single uncalibrated snapshots using two planar mirrors. In *Proc. 15th South African Workshop on Pattern Recognition*, 2004.
5. J.M. Gluckman and S.K. Nayar. Planar catadioptric stereo: Geometry and calibration. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1999.
6. B. Hu, C.M. Brown, and R.C. Nelson. Multiple-view 3-D reconstruction using a mirror. Technical Report TR863, Computer Science Dept., U. Rochester, May 2005.
7. Aldo Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, 1994.
8. W. Matusik, C. Buehler, and L. McMillan. Polyhedral visual hulls for real-time rendering. In *Proceedings of Twelfth Eurographics Workshop on Rendering*, 2001.
9. 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.
10. A. Mülhayim, U. Yilmaz, and V. Atalay. Silhouette-based 3D model reconstruction from multiple images. *IEEE Transactions on Systems, Man and Cybernetics*, 33(4), 2003.
11. T. Okatani and K. Deguchi. Recovering camera motion from image sequence based on registration of silhouette cones. In *IAPR Workshop on Machine Vision Applications*, 2000.
12. S. Sinha, M. Pollefeys, and L. McMillan. Camera network calibration from dynamic silhouettes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2004.
13. 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.