Shape from Profiles

By Roberto Cipolla and Kwan-Yee K. Wong

Department of Engineering, University of Cambridge,
Trumpington Street, Cambridge CB2 1PZ, UK.
http://svr-www.eng.cam.ac.uk/∼cipolla

Profiles of a sculpture provide rich information about its geometry, and can be used for model reconstruction under known camera motion. By exploiting correspondences induced by epipolar tangents on the profiles, a successful solution to motion estimation has been developed for the case of circular motion. Arbitrary general views can then be incorporated to refine the model built from circular motion.

Keywords: Structure and motion, shape from silhouette, model acquisitio

1. Introduction

Profiles (also known as outlines, apparent contours or silhouettes) are often a dominant feature in images. They can be extracted relatively easily and reliably from the images, and provide rich information about both the shape and motion of an object. Classical techniques Faugeras (1993) for model reconstruction and motion estimation depend on point and/or line correspondences, and hence cannot be applied directly to profiles, which are viewpoint dependent. This calls for the development of a completely different set of algorithms specific to profiles. This paper will give a brief review of some of the state-of-art algorithms for model building and motion estimation from profiles.

2. Profiles of Surfaces

Profiles are projections of contour generators which divide the visible from the occluded part of the surface Cipolla and Giblin (1999) and hence depend on both the surface geometry and camera position. In general, 2 contour generators on a surface, associated with 2 different camera positions, will be 2 distinct space curves, and thus the corresponding profiles on the images do not readily provide point correspondences. A *frontier point* Cipolla and Giblin (1999) is the intersection of 2 contour generators and lies on an epipolar plane tangent to the surface (see fig. 1). It follows that a frontier point will project to a point on the profile which is also on an epipolar tangent Porrill and Pollard (1991). Epipolar tangencies thus provide point correspondences on profiles, and can be exploited for motion estimation.

3. Model Reconstruction

The image profiles of an object provide rich information about its shape. Under known camera motion, it is possible to reconstruct a model of the object from

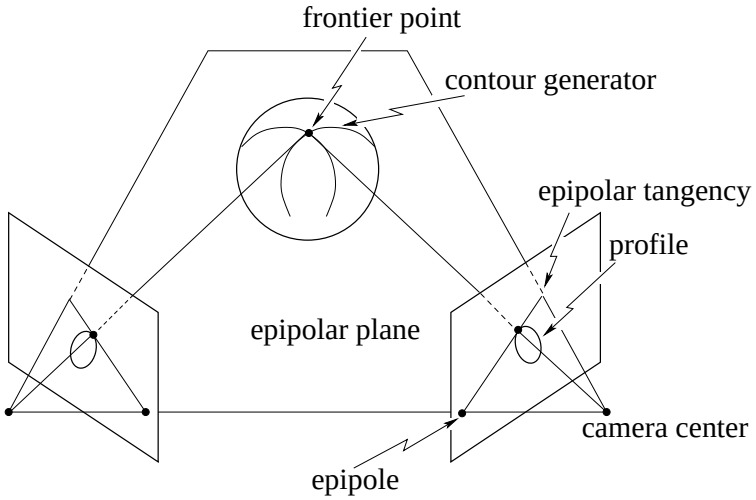

Figure 1. A frontier point is the intersection of two contour generators and lies on an epipolar plane which is tangent to the surface. It follows that a frontier point will project to a point on the profile which is also on an epipolar tangent.

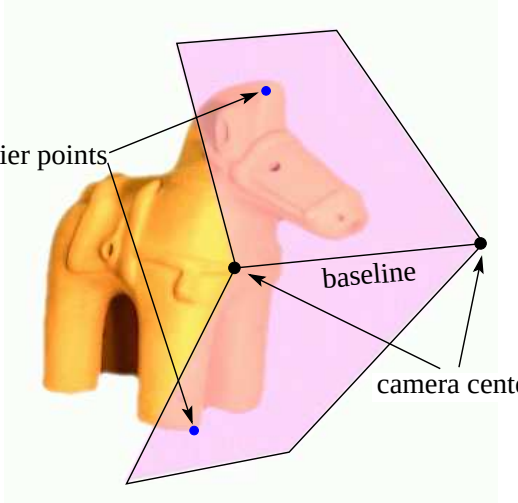

Figure 2. The outer epipolar tangents correspond to the 2 epipolar tangent planes which touch the object, and are always available in any pair of views except when the baseline passes through the object.

its profiles. For continuous camera motion and simple smooth surfaces, a surface representation can be obtained from the profiles using the epipolar parameterization Cipolla and Blake (1992). Cipolla and Blake Cipolla and Blake (1992) developed a simple numerical method for estimating depth from a minimum of 3 discrete views by determining the osculating circle on each epipolar plane. Vaillant and Faugeras Vaillant and Faugeras (1992) developed a similar algorithm which

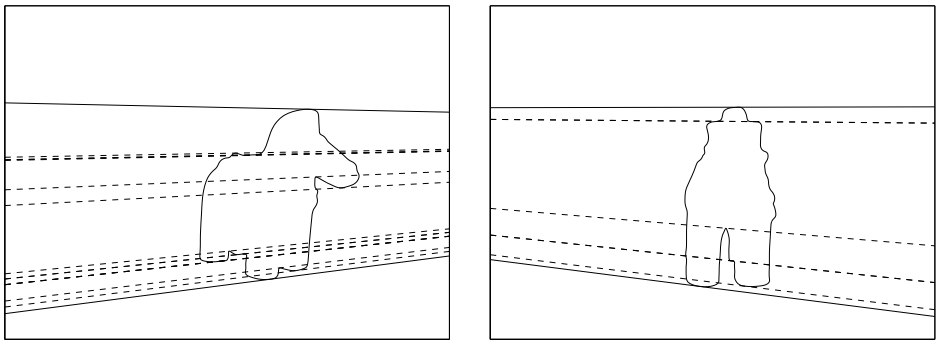

Figure 3. Two discrete views showing 17 epipolar tangents in total, of which only 4 pairs are in correspondence. The use of the 2 outer epipolar tangents (in solid lines), which are guaranteed to be in correspondence, avoids false matches due to self-occlusions, and greatly simplifies the matching problem.

uses the radial plane instead of the epipolar plane. Boyer and Berger Boyer and Berger (1997) derived a depth formulation from a local approximation of the surface up to order 2, which allows the local shape to be estimated from 3 consecutive views by solving a pair of simultaneous equations.

Alternatively, for discrete motion and objects with more complex geometry, a volumetric model can be obtained by an octree carving algorithm Szeliski (1993). This technique is chosen in Section 4 for illustrating the results of reconstruction using the motion estimated from profiles, and thus is described in more details here. Initially, the octree consists of a single cube in space which encloses the model to be reconstructed. The cube is projected onto each images and classified as either (a) completely outside 1 or more profiles, (b) completely inside all the profiles, or (c) ambiguous. If the cube is classified as type (c), it is subdivided into 8 subcubes (hence the name octree) each of which is again projected onto the images and classified. This process is repeated until a preset maximum level (resolution) is reached. Cubes classified as type (a) are thrown away, leaving type (b) and (c) cubes which constitute the volumetric model of the object. Surface triangles, if needed, can be extracted from type (c) cubes using a marching cubes algorithm Lorensen and Cline (1987). The octree carving technique is summarized in algorithm 1, and the implemented algorithm for octree carving can be downloaded from:
http://svr-www.eng.cam.ac.uk/∼cipolla.

It is worth noting that both the surface and volumetric models, estimated only from the profiles of the object, correspond to the visual hull of the object with respect to the set of camera positions. Concavities cannot be recovered as they never appear as part of the profiles. In order to "carve" away the concavities, methods like space carving Kutulakos and Seitz (2000); Broadhurst et al. (2001) could be used instead.

4. Motion Estimation

A practical algorithm for motion estimation from profiles, in the case of complete circular motion, was introduced in Mendonça et al. Mendonça (Wong and Cipolla).

Algorithm 1 Octree Carving from Profiles

```
initialize a cube that enclose the model;
while max level not rearched do
  for each cube in the current level do
    project the cube onto each image;
    classify the cube as either:
      (a) completely outside 1 or more profiles,
      (b) completely inside all the profiles, or
      (c) ambiguous;
    if the cube is classified as type (c) then
      subdivide the cube into 8 sub-cubes;
      add the sub-cubes to the next level;
    end if
  end for
  increase the level count;
end while
```

In Wong and Cipolla (2001), the profiles from (incomplete) circular motion was exploited for the registration of any arbitrary general view, and a complete system for model acquisition from uncalibrated profiles under both circular and general motion was developed. A summary of the techniques first reported in Mendonça et al. Mendonça (Wong and Cipolla); Wong and Cipolla (2001) is given below.

The 3 main image features in circular motion, namely the image of the rotation axis $\mathbf{l}_\mathrm{s}$, the horizon $\mathbf{l}_\mathrm{h}$ and a special vanishing point $\mathbf{v}_x$ (see Mendonça et al. Mendonça (Wong and Cipolla) for details), are fixed throughout the sequence and satisfy

$$\mathbf{v}_x \cdot \mathbf{l}_\mathrm{h} = 0, and \tag{4.1}$$

$$\mathbf{v}_x = \mathbf{K}\mathbf{K}^\mathrm{T}\mathbf{l}_\mathrm{s}, \tag{4.2}$$

where $\mathbf{K}$ is the 3×3 camera calibration matrix. The fundamental matrix can be parameterized explicitly in terms of these features Vieville and Lingrand (1996); Mendonça et al. Mendonça (Wong and Cipolla), and is given by

$$\mathbf{F} = [\mathbf{v}_x]_\times + \kappa \tan\frac{\theta}{2}(\mathbf{l}_\mathrm{s}\mathbf{l}_\mathrm{h}^\mathrm{T} + \mathbf{l}_\mathrm{h}\mathbf{l}_\mathrm{s}^\mathrm{T}), \tag{4.3}$$

where θ is the angle of rotation, and κ is a constant which can be determined from the camera intrinsic parameters. A sequence of N images taken under circular motion, with known camera intrinsic parameters, can hence be described by $N + 2$ motion parameters. By using the 2 outer epipolar tangents Wong and Cipolla (2001), the N images will provide $2N$ (or 2 when $N = 2$) independent constraints on these parameters, and a solution will be possible when $N \geq 3$.

The circular motion will generate a *web of contour generators* around the object, which can be exploited for registering any new arbitrary general view. Given an arbitrary general view, the associated contour generator will intersect with this web and form frontier points. If the camera intrinsic parameters are known, the 6 motion parameters of the new view can be fixed if there are 6 or more frontier points on the associated contour generator. This corresponds to having a minimum

of 3 views under circular motion, each providing 2 outer epipolar tangents to the profile in the new general view (see fig. 5).

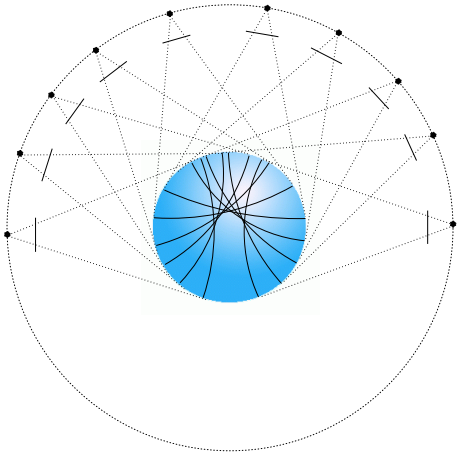

Figure 4. The circular motion will generate a web of contour generators around the object, which can be used for registering any new arbitrary general view.

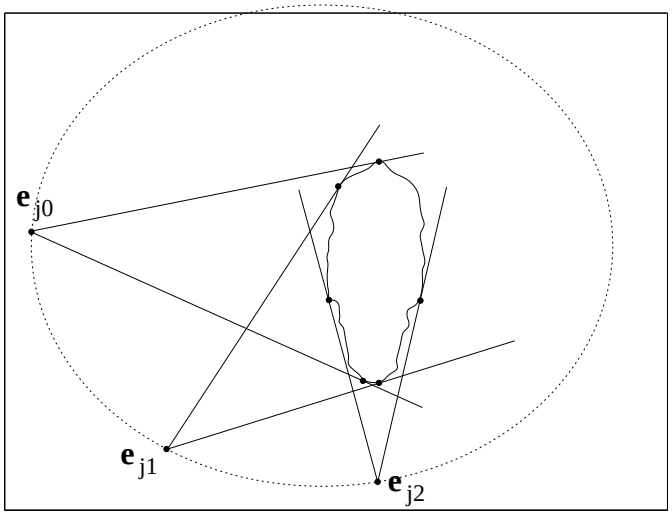

Figure 5. Three views from circular motion provide 6 outer epipolar tangents to the profile in the new general view for estimating its pose.

The motion estimation proceeds as an optimization which minimizes the reprojection errors of epipolar tangents. For view i and view j, a fundamental matrix $\mathbf{F}_{ij}$ is formed from the current estimate of the motion parameters, and the epipoles $\mathbf{e}_{ij}$ and $\mathbf{e}_{ji}$ are obtained from the right and left nullspaces of $\mathbf{F}_{ij}$. The outer epipolar tangent points $\mathbf{t}_{ij0}$, $\mathbf{t}_{ij1}$ and $\mathbf{t}_{ji0}$, $\mathbf{t}_{ji1}$ are located in view i and j respectively (see

fig. 6). The reprojection errors are then given by the geometric distances between the epipolar tangent points and their epipolar lines,

$$d_{ijk} = \frac{\mathbf{t}_{jik}^{\mathrm{T}} \mathbf{F}_{ij} \mathbf{t}_{ijk}}{\sqrt{(\mathbf{F}_{ij}\mathbf{t}_{ijk})_1^2 + (\mathbf{F}_{ij}\mathbf{t}_{ijk})_2^2}}, \tag{4.4}$$

$$d_{jik} = \frac{\mathbf{t}_{jik}^{\mathrm{T}} \mathbf{F}_{ij} \mathbf{t}_{ijk}}{\sqrt{(\mathbf{F}_{ij}^{\mathrm{T}}\mathbf{t}_{jik})_1^2 + (\mathbf{F}_{ij}^{\mathrm{T}}\mathbf{t}_{jik})_2^2}}. \tag{4.5}$$

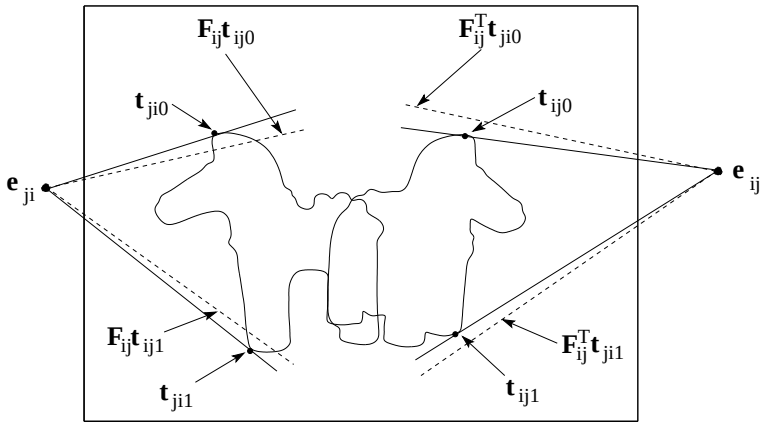

Figure 6. The motion parameters can be estimated by minimizing the reprojection errors of epipolar tangents, which are given by the geometric distances between the epipolar tangent points and their epipolar lines.

For a sequence of N images taken under circular motion, the image of the rotation axis and the horizon are initialized approximately, and the angles are arbitrarily initialized. The cost function is given by

$$C_{cm}(\mathbf{x}) = \frac{1}{4} \sum_{i=1}^{N} \sum_{j=i+1}^{min(j+3,N)} \sum_{k=1}^{2} d_{ijk}(\mathbf{x})^2 + d_{jik}(\mathbf{x})^2, \tag{4.6}$$

where $\mathbf{x}$ consists of the $N+2$ motion parameters.

For arbitrary general motion, the 6 motion parameters can be initialized by roughly aligning the projection of the 3D model built from the estimated circular motion with the image. The cost function of general motion for view j is given by

$$C_j(\mathbf{x}') = \sqrt{\frac{\sum_{i=1}^{N} f_{ij} \sum_{k=1}^{2} d_{ijk}(\mathbf{x}')^2 + d_{jik}(\mathbf{x}')^2}{4 \sum_{i=1}^{N} f_{ij}}}, \tag{4.7}$$

where $\mathbf{x}'$ consists of the 6 motion parameters. f_{ij} is 0 if the baseline formed with view i passes through the object, otherwise it is 1.

Algorithm 2 Motion Estimation from profiles

extract the profiles using cubic B-spline snakes Cipolla and Blake (1992) ;
initialize $\mathbf{l}_s$, $\mathbf{l}_h$ and the $N-1$ angles for the circular motion;
while not converged **do**
  compute the cost for circular motion using (4.6);
  update the $N+2$ motion parameters to minimize the cost;
end while
Form the essential matrices from the fundamental matrices using the calibration matrix;
Decompose the essential matrices to obtain the projection matrices;
Build a partial model using an octree carving algorithm;
for each arbitrary general view **do**
  initialize the 6 motion parameters using the partial model from circular motion;
  while not converged **do**
    compute the cost for the general motion using (4.7);
    update the 6 motion parameters to minimize the cost;
  end while
end for
Refine the model using the now calibrated images from general motion.

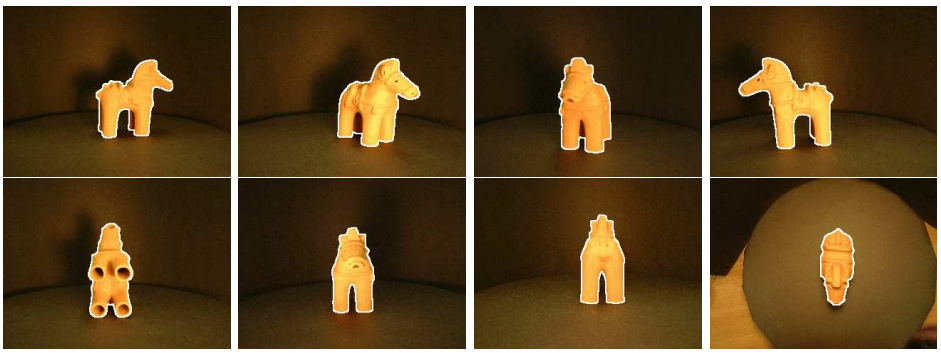

Figure 7. Top: 4 images from an uncalibrated sequence of a Haniwa taken under circular motion (views 1,4, 7 and 10 respectively). Bottom: 4 images from arbitrary camera positions (views 12,13,14 and 15 respectively).

The motion estimation procedure is summarized in algorithm 2. Fig. **??** and fig. **??** show some examples of reconstruction from the motion estimated using profiles (see fig. 7 for the image sequences used).

In the next experimental sequence 14 images of an outdoor sculpture were acquired by a hand-held camera (see figure 11). An approximate circular motion of the camera was achieved by using a string which was fixed to the ground by a peg at one end. A circular path on the ground was then obtained by rotating the free end of the string about its fixed end. Each image in the sequence was acquired by positioning the camera roughly at a fixed height above the free end of the (rotating) string, and pointing it towards the sculpture. Note that since the camera center, the string and the rotation axis were roughly coplanar, the image of the string in each

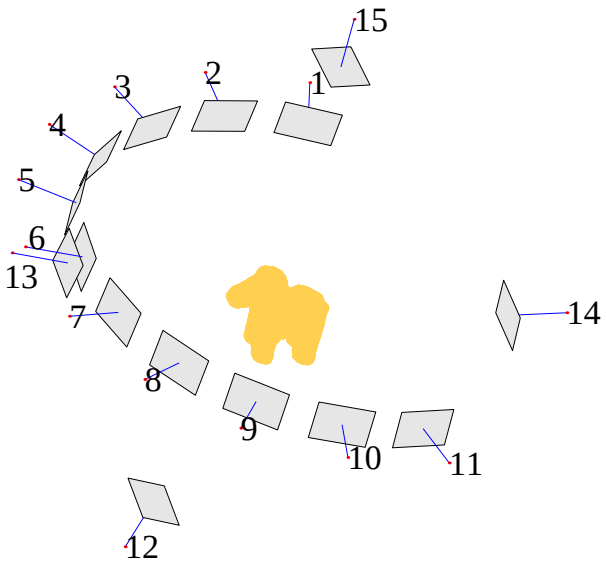

Figure 8. Camera poses estimated from the sequence of the haniwa.

image provided a very good estimate for the image of the rotation axis $\mathbf{l}_s$. Although the camera center roughly followed a circular path, the orientation of the camera was unconstrained and hence the image of the rotation axis $\mathbf{l}_s$ and the horizon $\mathbf{l}_h$ were not fixed throughout the image sequence. In order to allow the camera motion to be estimated using the circular motion algorithm described in Section **??**, the images were first rectified using the technique described in Section **??** so that the image of the string (i.e. the image of the rotation axis) became a fixed vertical line passing through the principal point throughout the sequence. A transformation induced by a rotation about the x-axis of the camera was then applied to each image so that the image of the fixed end of the string became a fixed point on $\mathbf{l}_s$ throughout the rectified sequence (see figure **??**). The resulting image sequence resembled a circular motion sequence, in which the horizon $\mathbf{l}_h$, the image of the rotation axis $\mathbf{l}_s$, and the special vanishing point $\mathbf{v}_x$ were fixed (see figure**??**). The algorithm for circular motion estimation was then applied to this rectified sequence, and the resulting camera poses were then iteratively refined by applying the general motion algorithm. The final camera poses estimated from the rectified sequence are shown in figure **??**, and the 3D model built from the estimated motion is shown in figure **??**.

5. Conclusions

The incorporation of arbitrary general views reveals information which is concealed under circular motion, and greatly improves both the shape and textures of the 3D models. Since only profiles have been used in both the motion estimation and octree carving, no corner detection nor matching is necessary. This means that the algorithm is capable of reconstructing any kind of objects, including *smooth* and

Figure 9. Triangulated mesh of the Haniwa model. This model was composed of 12,028 triangles.

textureless surfaces. Experiments on various objects have produced convincing 3D models, demonstrating the practicality of the algorithm.

References

Boyer, E. and Berger, M. O. (1997) 3d surface reconstruction using occluding contours. *Int. Journal of Computer Vision*, **22**, 219–233.

Broadhurst, A., Drummond, T. and Cipolla, R. (2001) A probabilistic framework for space carving. In *Proc. 8th Int. Conf. on Computer Vision*. Vancouver, Canada.

Figure 10. Refined model of the Haniwa after incorporating the 4 arbitrary general views. The model was now fully covered with textures and showed great improvements in shape, especially in the front, back and top views.

Cipolla, R. and Blake, A. (1992) Surface shape from the deformation of apparent contours. *Int. Journal of Computer Vision*, **9**, 83–112.

Cipolla, R. and Giblin, P. J. (1999) *Visual Motion of Curves and Surfaces.* Cambridge, UK: Cambridge University Press.

Faugeras, O. D. (1993) *Three-Dimensional Computer Vision: a Geometric Viewpoint.* MIT Press.

Kutulakos, K. N. and Seitz, S. M. (2000) A theory of shape by space carving. *Int. Journal of Computer Vision*, **38**, 197–216.

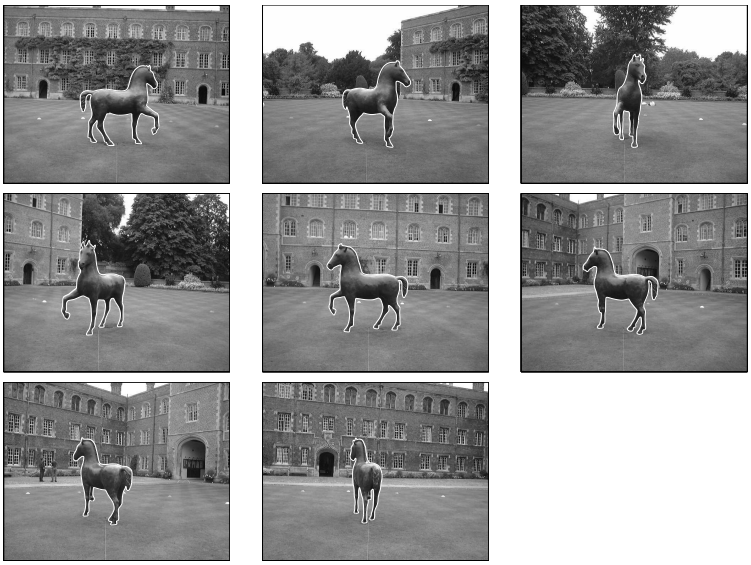

Figure 11. Eight images of an outdoor sculpture acquired by a hand-held camera. Although the camera center roughly followed a circular path, the orientation of the camera was unconstrained and hence the image of the rotation axis and the horizon were not fixed throughout the image sequence.

Lorensen, W. E. and Cline, H. E. (1987) Marching cubes: a high resolution 3D surface construction algorithm. In *Proc. SIGGRAPH 87, Computer Graphics Proceedings, Annual Conference Series* (ed. M. C. Stone), 163–170. Anaheim, CA.

Mendonça, P. R. S., Wong, K.-Y. K. and Cipolla, R. (to appear in 2001) Epipolar geometry from profiles under circular motion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*.

Porrill, J. and Pollard, S. B. (1991) Curve matching and stereo calibration. *Image and Vision Computing*, **9**, 45–50.

Szeliski, R. (1993) Rapid octree construction from image sequences. *CVGIP: Image Understanding*, **58**, 23–32.

Vaillant, R. and Faugeras, O. D. (1992) Using extremal boundaries for 3D object modelling. *IEEE Trans. on Pattern Analysis and Machine Intell.*, **14**, 157–173.

Vieville, T. and Lingrand, D. (1996) Using singular displacements for uncalibrated monocular visual systems. In *Proc. 4th European Conf. on Computer Vision* (eds. B. Buxton and R. Cipolla), vol. II, 207–216. Cambridge, UK: Springer–Verlag.

Wong, K.-Y. K. and Cipolla, R. (2001) Structure and motion from silhouettes. In *Proc. 8th Int. Conf. on Computer Vision*. Vancouver, Canada.

Figure 12. Triangulated mesh of the outdoor sculpture of a horse. This model was composed of 29,672 triangles.