

# Reconstruction of Sculpture From Its Profiles With Unknown Camera Positions

Kwan-Yee Kenneth Wong and Roberto Cipolla, *Member, IEEE*

**Abstract**—Profiles of a sculpture provide rich information about its geometry, and can be used for shape recovery under known camera motion. By exploiting correspondences induced by epipolar tangents on the profiles, a successful solution to motion estimation from profiles has been developed in the special case of circular motion. The main drawbacks of using circular motion alone, namely the difficulty in adding new views and part of the object always being invisible, can be overcome by incorporating arbitrary general views of the object and registering its new profiles with the set of profiles resulted from the circular motion. In this paper, we describe a complete and practical system for producing a three-dimensional (3-D) model from uncalibrated images of an arbitrary object using its profiles alone. Experimental results on various objects are presented, demonstrating the quality of the reconstructions using the estimated motion.

**Index Terms**—Shape from profiles, structure and motion, 3-D model acquisition.

## I. INTRODUCTION

**P**ROFILES (also known as outlines, apparent contours, and 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 [1]–[5] for model reconstruction and motion estimation, however, rely on point or line correspondences, and 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.

A traditional approach to the problem of motion estimation from profiles is to search for *epipolar tangencies* [6], [7]. However, such a search involves a nonlinear optimization with non-trivial initialization, and requires the presence of seven or more epipolar tangencies. In [8], we exploited a compact parameterization of the fundamental matrix specific to circular motion and developed a practical solution in the special case of complete motion with dense image sequence, which requires only two epipolar tangencies. In [9], we extended the algorithm presented in [8] to cope with incomplete circular motion and more widely space images. However, there are two main drawbacks of using circular motion:

- 1) new views cannot be added easily at a later time;
- 2) part of the structure will always remain invisible under circular motion.

In this paper, we present a complete and practical system for generating high quality three-dimensional (3-D) models using two-dimensional (2-D) profiles from uncalibrated views taken under both circular and arbitrary general motion. The system first employs our previous technique [9] for estimating the circular motion of an object from its profiles alone, and builds an initial 3-D model of the object by an octree carving technique [10] using the profiles and the estimated camera positions. The system then allows the 3-D model thus obtained to be refined incrementally by adding new arbitrary general views of the object and estimating the corresponding camera positions. This is achieved by registering the profile in the new view with the set of profiles generated by the now estimated circular motion [11]. The incorporation of arbitrary general views reveals information which is concealed under circular motion, and overcomes the drawbacks of using circular motion alone. Besides, we introduce the use of the two *outer epipolar tangents*, which simplifies the correspondence problem in matching epipolar tangent points, making the algorithms more robust. The system described here needs no corner detection nor matching, and is practical in virtually all situations. It is particular suitable for reconstructing sculptures, which have sparse surface texture.

This paper is organized as follows. Section II reviews existing techniques in the literature for motion estimation from profiles, and discusses their shortcomings. Section III studies the epipolar constraint between profiles from distinct view-points and introduces the use of the outer epipolar tangents that simplifies the correspondence problem. Section IV addresses the problem of estimating the motion of a rotating object and presents a useful parameterizations of the fundamental matrix specific to circular motion. The general motion case is then tackled in Section V. The reconstruction of 3-D models using the profiles and the estimated motion is described in Section VI. Implementation details are described in Section VII, followed by experimental results in Section VIII. Finally, conclusions are given in Section IX.

## II. PREVIOUS WORKS

The study of motion estimation from profiles was pioneered by Rieger [12], who showed that camera motion can be recovered from three *fixed points* of a deforming profile under orthographic projection. He also set forth the idea that under perspective projection, the epipole is constrained to the line spanned by the tangent vector to the profile at the fixed point (i.e., epipolar tangency constraint). In [6], it was noted that the intersection of

Manuscript received February 26, 2003; revised September 5, 2003. This work was supported in part by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China, under Project HKU 7155/03E. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Alessandro Piva.

K.-Y. K. Wong is with the Department of Computer Science and Information Systems, The University of Hong Kong, Hong Kong, China (e-mail: kykwong@csis.hku.hk).

R. Cipolla is with the Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, U.K. (e-mail: cipolla@eng.cam.ac.uk).

Digital Object Identifier 10.1109/TIP.2003.821113

two contour generators from two distinct viewpoints generates a point that is visible in both images as a fixed point. This point was identified as a *frontier point* in [13], where Giblin *et al.* developed an algorithm for motion estimation from the profiles of a rotating surface under orthographic projection.

The methods mentioned so far deal with the motion recovery problem under orthographic projection, which is a rather restrictive situation. The use of frontier points and epipolar tangents for motion recovery under perspective projection was introduced in [7], [14]. These techniques require the presence of at least seven pairs of corresponding epipolar tangents in the image pair, which are localized by iterative methods. By using an affine approximation, a similar technique that only requires four pairs of corresponding epipolar tangents was developed in [15]. In [16], a noniterative method was presented in the case of linear camera motion, where common tangents are used to determine both the images of the frontier points and the epipoles. By combining the ideas in [14] and [16], Cross *et al.* implemented a parallax-based technique [17] in which images are registered using a reference plane to “undo” the effect of rotation. Related works also include [18] and [19].

This paper tackles the problem of structure and motion from profiles observed under perspective projection using a single camera. The approach here is to first constrain the motion to be circular. This allows a trivial initialization of the parameters which all bear physical meaning (e.g., image of the rotation axis, the horizon and the rotation angles). When there are three or more images in the circular motion sequence, a solution is possible by using only the two outer epipolar tangents to the profiles. The drawbacks of using circular motion alone are then overcome by incorporating new views from arbitrary general motion. The initialization of the general motion can be done relatively easily by using the partial model built from the estimated circular motion. By registering the profile in the new view with the set of profiles resulted from the circular motion, the camera motion can again be estimated using only the two outer epipolar tangents.

### III. EPIPOLAR CONSTRAINT BETWEEN PROFILES

Profiles are projections of contour generators [20], [21] which divide the visible part from the occluded part of a surface [22], and depend on both surface geometry and viewpoint (i.e., camera position). In general, two contour generators on a surface, associated with two distinct viewpoints, will be two different space curves. Hence the corresponding profiles on the images do not readily provide point correspondences which are needed in estimating the camera motion. The only exception is the *frontier point* [7], [13], [22] which is the intersection of two 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 [6], [7], [12]. Epipolar tangencies thus provide point correspondences on profiles, and can be exploited for motion estimation.

Theoretically, if seven or more pairs of corresponding epipolar tangent points are available, the epipolar geometry

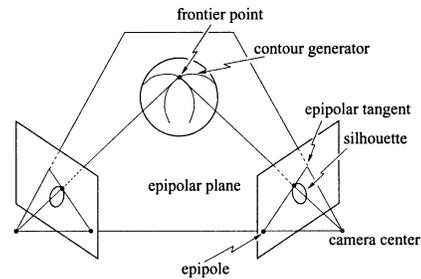


Fig. 1. Frontier point is the intersection of two contour generators and lies on an epipolar plane which is tangent to the surface. It follows that a frontier point will project to a point on the profile which is also on an epipolar tangent.

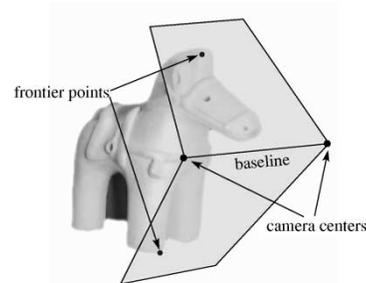


Fig. 2. Outer epipolar tangents correspond to the two epipolar tangent planes which touch the object, and are always available in any pair of views except when the baseline passes through the object.

between the two views can be estimated, and the camera intrinsic parameters can then be used to recover the relative motion [23], [24]. However, when the epipolar geometry is not known, the localization of the epipolar tangents involves a nonlinear optimization. Examples of this iterative approach can be found in [7], [14]. The need for a good but nontrivial initialization, the unrealistic demand for a large number of epipolar tangent points, and the presence of local minima all make this approach impractical.

In the next two sections, two motion estimation algorithms which only require the two *outer epipolar tangents* are presented. The outer epipolar tangents correspond to the two epipolar tangent planes which touch the object (see Fig. 2). Except when the baseline passes through the object, the two outer epipolar tangents are always available in any pair of views and are guaranteed to be in correspondence. The use of the outer epipolar tangents avoids false matches due to self-occlusions and greatly simplifies the matching problem. This is illustrated in Fig. 3 which shows two profiles from two distinct viewpoints. The profile in the left image has eleven epipolar tangents, whereas the profile in the right image has only six epipolar tangents. A careful examination will show that not all six epipolar tangents in the right image have a correspondence in the left image. There are actually only four pairs of corresponding epipolar tangents, which are the two outer epipolar tangents and another two tangents at the front and back left legs of the horse sculpture. By considering only the outer epipolar tangents, possible false matches are eliminated and the problem is reduced to matching only two epipolar tangents against another two, leaving only two possible cases.

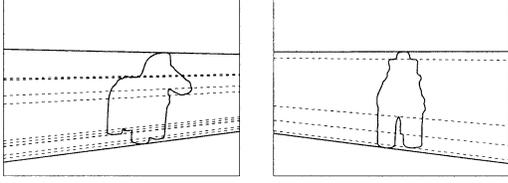


Fig. 3. Two discrete views showing seventeen epipolar tangents in total, of which only four pairs are in correspondence. The use of the two outer epipolar tangents (in solid lines), which are guaranteed to be in correspondence, avoids false matches due to self-occlusions, and greatly simplifies the matching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

#### IV. CIRCULAR MOTION

##### A. Fixed Image Features Under Circular Motion

Consider a pinhole camera rotating about a fixed axis. Let  $\Pi_s$  be the plane that contains the rotation axis and the camera center, and  $\Pi_h$  be the plane that contains the trajectory of the camera center. Note that the projections of  $\Pi_s$  and  $\Pi_h$  on the image plane are the image of the rotation axis  $\mathbf{l}_s$  and the horizon  $\mathbf{l}_h$ , respectively. Let  $\mathbf{v}_x$  be the vanishing point corresponding to the normal direction  $\mathbf{N}_x$  of  $\Pi_s$ . Note that  $\mathbf{N}_x$  is parallel to  $\Pi_h$ , and hence it follows that  $\mathbf{v}_x$  must lie on  $\mathbf{l}_h$  (i.e.,  $\mathbf{l}_h^T \mathbf{v}_x = 0$ ). It has been shown in [25] that there exists a pole-polar relationship, with respect to the absolute conic, between the vanishing point  $\mathbf{v}_x$  and vanishing line  $\mathbf{l}_s$  of  $\Pi_s$ , given by

$$\boldsymbol{\omega} \mathbf{v}_x = \mathbf{l}_s \quad (1)$$

where  $\boldsymbol{\omega} = \mathbf{K}^{-T} \mathbf{K}^{-1}$  is the *image of the absolute conic* and  $\mathbf{K}$  is the  $3 \times 3$  camera calibration matrix.

If the intrinsic parameters of the camera are kept fixed, due to symmetry in the configuration,  $\mathbf{l}_s$ ,  $\mathbf{l}_h$  and  $\mathbf{v}_x$  will be fixed throughout the image sequence (see Fig. 4). The fundamental matrix associated with any pair of views in the circular motion sequence can be parameterized explicitly in terms of these fixed features [4], [26], and a simple derivation of this parameterization is given in the next section.

##### B. Parameterization of the Fundamental Matrix

Consider two pinhole cameras  $\hat{\mathbf{P}}_1$  and  $\hat{\mathbf{P}}_2$ , given by

$$\hat{\mathbf{P}}_1 = [\mathbf{l}_3 \ \mathbf{t}], \quad \text{and} \quad (2)$$

$$\hat{\mathbf{P}}_2 = [\mathbf{R}_y(\theta) \ \mathbf{t}] \quad (3)$$

where  $\mathbf{t} = [0 \ 0 \ 1]^T$  and  $\mathbf{R}_y(\theta)$  is a rotation by an angle  $\theta \neq 0$  about the  $y$ -axis, given by

$$\mathbf{R}_y(\theta) = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix}. \quad (4)$$

Under this camera configuration, the image of the rotation axis, the horizon and the special vanishing point are given by

$$\hat{\mathbf{l}}_s = [1 \ 0 \ 0]^T \quad (5)$$

$$\hat{\mathbf{l}}_h = [0 \ 1 \ 0]^T \quad (6)$$

$$\hat{\mathbf{v}}_x = [1 \ 0 \ 0]^T \quad (7)$$

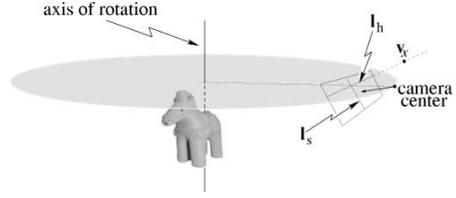


Fig. 4. Under circular motion the image of the rotation axis  $\mathbf{l}_s$ , the horizon  $\mathbf{l}_h$  and a special vanishing point  $\mathbf{v}_x$  are fixed throughout the sequence.

respectively. The fundamental matrix associated with  $\hat{\mathbf{P}}_1$  and  $\hat{\mathbf{P}}_2$  is given by

$$\begin{aligned} \hat{\mathbf{F}} &= \begin{bmatrix} 0 & \tan \frac{\theta}{2} & 0 \\ \tan \frac{\theta}{2} & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}_{\times} + \tan \frac{\theta}{2} \left( \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} [0 \ 1 \ 0] + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} [1 \ 0 \ 0] \right). \quad (8) \end{aligned}$$

By substituting (5), (6) and (7) into (8), the fundamental matrix can be rewritten in terms of the fixed image features under circular motion, and is given by [8]

$$\hat{\mathbf{F}} = [\hat{\mathbf{v}}_x]_{\times} + \tan \frac{\theta}{2} (\hat{\mathbf{l}}_s \hat{\mathbf{l}}_h^T + \hat{\mathbf{l}}_h \hat{\mathbf{l}}_s^T). \quad (9)$$

Consider now a pair of camera  $\mathbf{P}_1$  and  $\mathbf{P}_2$  obtained by introducing the intrinsic parameters represented by the camera calibration matrix  $\mathbf{K}$  to  $\hat{\mathbf{P}}_1$  and  $\hat{\mathbf{P}}_2$ , respectively, and by applying the rotation  $\mathbf{R}$  to  $\hat{\mathbf{P}}_1$  and  $\hat{\mathbf{P}}_2$  about their optical centers, respectively. Hence  $\mathbf{P}_1 = \mathbf{H} \hat{\mathbf{P}}_1$  and  $\mathbf{P}_2 = \mathbf{H} \hat{\mathbf{P}}_2$ , where  $\mathbf{H} = \mathbf{K} \mathbf{R}$ . The fundamental matrix associated with  $\mathbf{P}_1$  and  $\mathbf{P}_2$  is then given by [8]

$$\begin{aligned} \mathbf{F} &= \mathbf{H}^{-T} \hat{\mathbf{F}} \mathbf{H}^{-1} \\ &= \frac{1}{\det(\mathbf{H})} [\mathbf{v}_x]_{\times} + \tan \frac{\theta}{2} (\mathbf{l}_s \mathbf{l}_h^T + \mathbf{l}_h \mathbf{l}_s^T) \\ &= \frac{1}{\det(\mathbf{K})} [\mathbf{v}_x]_{\times} + \tan \frac{\theta}{2} (\mathbf{l}_s \mathbf{l}_h^T + \mathbf{l}_h \mathbf{l}_s^T) \quad (10) \end{aligned}$$

where  $\mathbf{l}_s = \mathbf{H}^{-T} \hat{\mathbf{l}}_s$ ,  $\mathbf{l}_h = \mathbf{H}^{-T} \hat{\mathbf{l}}_h$  and  $\mathbf{v}_x = \mathbf{H} \hat{\mathbf{v}}_x$ . Note that  $\mathbf{l}_s$ ,  $\mathbf{l}_h$ , and  $\mathbf{v}_x$  are the image of the rotation axis, the horizon and the special vanishing point, respectively, under this new camera configuration.

Equation (10) gives a simple parameterization of the fundamental matrix relating any pair of views in the circular motion sequence. This parameterization allows a trivial initialization of the parameters which all bear physical meaning, and greatly reduces the dimension of the search space for the optimization problem in motion estimation. When the intrinsic parameters of the camera are fixed and known, two parameters are enough to fix  $\mathbf{l}_s$  and  $\mathbf{v}_x$ . Since  $\mathbf{v}_x$  must also lie on  $\mathbf{l}_h$ , only one further parameter is needed to fix  $\mathbf{l}_h$ . As a result, a sequence of  $N$  images taken under circular motion can be described by  $N + 2$  motion parameters (two for  $\mathbf{l}_s$  and  $\mathbf{v}_x$ , one for  $\mathbf{l}_h$  and the  $N - 1$  rotation angles). By exploiting the two outer epipolar tangents, the  $N$  images will provide  $2N$  (or two when  $N = 2$ ) independent constraints on these  $N + 2$  parameters, and a solution will be possible when  $N$  is greater than or equal to three.

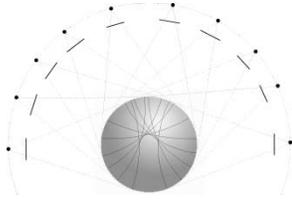


Fig. 5. The circular motion will generate a web of contour generators around the object, which can be used for registering any new arbitrary general view.

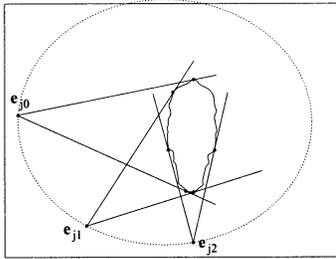


Fig. 6. Three views from circular motion provide six outer epipolar tangents to the profile in the new general view for estimating its pose.

## V. GENERAL MOTION

### A. Web of Contour Generators

The circular motion will generate a *web of contour generators* around the object (see Fig. 5), 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 six motion parameters of the new view (three for rotation and three for translation) can be fixed if there are six or more frontier points on the associated contour generator. This corresponds to having a minimum of three views under circular motion, each providing two outer epipolar tangents to the profile in the new general view (see Fig. 6). The motion parameters of the arbitrary general view can then be estimated by minimizing the reprojection errors of the two outer epipolar tangents resulting from each view in the estimated circular motion sequence.

### B. Initialization Problem

The difficulty of nontrivial initialization, which exists in every algorithm for general motion estimation from profiles, is overcome by exploiting the 3-D model built from the circular motion. After the estimation of the circular motion, a volumetric model of the object can be constructed by an octree carving technique [10] using the resulting camera positions and the profiles, and a triangulated mesh can then be extracted from the octree using the marching cubes algorithm [27] (see Section VI). The vertices in the mesh of the model are projected onto the new view whose pose is to be estimated. A very good initialization can be obtained by manually rotating and translating the camera (i.e., changing the six extrinsic parameters of the camera) until the projection of the initial 3-D model roughly matches the profile in the new view (see Fig. 7).

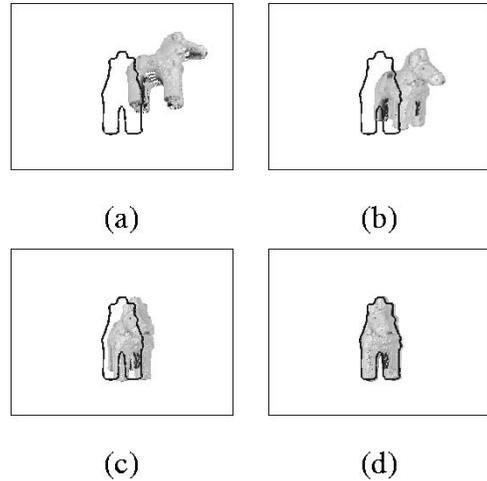


Fig. 7. The arbitrary general motion can be initialized by manually rotating and translating the camera until the projection of the initial 3-D model, built from the estimated circular motion, roughly matches the profile in the new view.

## VI. 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 its profiles. For continuous camera motion and simple smooth surfaces, a surface representation can be obtained from the profiles using differential techniques. Alternatively, for discrete motion and objects with more complex geometry, a volume intersection technique can be employed to produce a volumetric representation.

The surface reconstruction of smooth objects from profiles was pioneered by Giblin and Weiss [28]. Under the assumption of orthographic projection, they demonstrated that a surface can be reconstructed from the envelope of all its tangent planes computed directly from the family of profiles of the surface under planar viewer motion. Cipolla and Blake [21] extended the studies of Giblin and Weiss to curvilinear viewer motion under perspective projection, and developed the *osculating circle method* by introducing the *epipolar parameterization*. Vaillant and Faugeras [29] developed a similar technique in which the surface is parameterized by the *radial curves* instead of the *epipolar curves*. Based on the osculating circle method, Szeliski and Weiss [30] used a linear smoother to compute epipolar curves on the whole surface together with an estimate of uncertainty, and reported improvements in the reconstruction. In [31], Boyer and Berger derived a depth formulation from a local approximation of the surface up to order two for discrete motion.

The volume intersection technique for constructing volumetric descriptions of objects from multiple views was first proposed by Martin and Aggarwal [32], who introduced the *volume segment* representation. In [33], Chien and Aggarwal presented an algorithm for generating an octree of an object from three orthogonal views under orthographic projection. Their work was further developed by Ahuja and Veenstra [34], who extended the algorithm to handle images from any subset of thirteen standard viewing directions. In [35], Hong and Shneier introduced a technique for generating an octree from multiple arbitrary views under perspective projection.

Their approach first constructs an octree for each image by projecting the octree cubes onto the image and intersecting their projections with the profiles, and the final octree of the object is given by the intersection of the octrees obtained from all images. In [36], Potmesil described a similar approach in which the images are represented by *quadrees* to facilitate the intersection of the projections of the cubes with the profiles. Other similar approaches also include [37] and [38], where the octree for each image is constructed by intersecting, in 3-D space, the octree cubes with the polyhedral cone formed from the back-projection of the profile. In [10], Szeliski introduced an efficient algorithm which constructs an octree in a hierarchical coarse-to-fine fashion. His approach is similar to that of [36], except that only a single octree is constructed using all the images simultaneously.

In this paper, the volume intersection approach is chosen due to its ability to describe objects with more complex topologies (e.g., object with holes). Based on [10], an algorithm for generating an octree using profiles from multiple views is developed. The main difference between the work presented in [10] and the technique developed here is that instead of using a *background subtraction* technique as described in [10], the object/background binary images are computed directly from the B-spline snakes which are used to extract and represent the profiles during motion estimation. The sub-pixel accuracy of the B-spline snakes allows a binary image to have a resolution higher than the original image, and this may help to improve the cube classification when the object is relatively small in the image. For the sake of display, a triangulated mesh is extracted from the octree, and the colors of the vertices in mesh are estimated from the original images.

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 [39], [40] could be used instead.

## VII. IMPLEMENTATIONS

In this paper, the cubic B-spline snake [21] is chosen for the automatic extraction of profiles from the image sequence. Cubic B-spline snakes provide a compact representation of profiles of various complexity, and can achieve sub-pixel localization accuracy. Its parameterization also facilitates the localization of epipolar tangents.

The motion estimation proceeds as an optimization which minimizes the reprojection errors of epipolar tangents. Given a pair of views  $i$  and  $j$ , the associated fundamental matrix  $\mathbf{F}_{ij}$  is formed and the epipoles  $\mathbf{e}_{ij}$  and  $\mathbf{e}_{ji}$  are obtained from the right and left nullspaces of  $\mathbf{F}_{ij}$ , respectively. The outer epipolar tangent points  $\mathbf{u}_{ij0}$ ,  $\mathbf{u}_{ij1}$  and  $\mathbf{u}_{ji0}$ ,  $\mathbf{u}_{ji1}$  are located in view  $i$  and view  $j$  (see Fig. 8). The reprojection errors are then given by the geometric distances between the epipolar tangent points and their epipolar lines [41]

$$d_{ijk} = \frac{\mathbf{u}_{jik}^T \mathbf{F}_{ij} \mathbf{u}_{ijk}}{\sqrt{(\mathbf{F}_{ij}^T \mathbf{u}_{jik})_1^2 + (\mathbf{F}_{ij}^T \mathbf{u}_{jik})_2^2}} \quad (11)$$

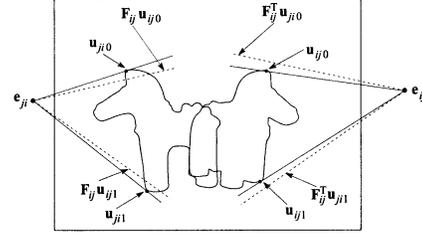


Fig. 8. The motion parameters can be estimated by minimizing the reprojection errors of epipolar tangents, which are given by the geometric distances between the epipolar tangent points and their epipolar lines.

$$d_{jik} = \frac{\mathbf{u}_{jik}^T \mathbf{F}_{ij} \mathbf{u}_{ijk}}{\sqrt{(\mathbf{F}_{ij}^T \mathbf{u}_{ijk})_1^2 + (\mathbf{F}_{ij}^T \mathbf{u}_{ijk})_2^2}} \quad (12)$$

where  $(\mathbf{F}_{ij}^T \mathbf{u}_{jik})_1$  and  $(\mathbf{F}_{ij}^T \mathbf{u}_{jik})_2$  indicate the 1st and 2nd coefficients of  $(\mathbf{F}_{ij}^T \mathbf{u}_{jik})$ , respectively. Similarly,  $(\mathbf{F}_{ij} \mathbf{u}_{ijk})_1$  and  $(\mathbf{F}_{ij} \mathbf{u}_{ijk})_2$  indicate the 1st and 2nd coefficients of  $(\mathbf{F}_{ij} \mathbf{u}_{ijk})$ , respectively.

### A. Circular Motion Estimation

For a sequence of  $N$  images taken under circular motion, the  $N - 1$  rotation angles are arbitrarily initialized. Usually by just inspecting the image sequence, a very good initialization for the image of the rotation axis  $\mathbf{l}_s$  can be obtained manually. The horizon  $\mathbf{l}_h$  is initialized manually by having a rough idea of the camera setup. Nonetheless, experimental results show that even with a poor initialization of  $\mathbf{l}_s$  and  $\mathbf{l}_h$ , the algorithm always converges to the same solution. As a result,  $\mathbf{l}_s$  and  $\mathbf{l}_h$  can be conveniently initialized as the vertical and horizontal lines through the image center, respectively.

During each iteration of the optimization, a fundamental matrix  $\mathbf{F}_{ij}$  between views  $i$  and  $j$  is computed from the current estimate of the motion parameters using (10) and the reprojection errors  $d_{ij1}(\mathbf{m})$ ,  $d_{ij2}(\mathbf{m})$ ,  $d_{ji1}(\mathbf{m})$  and  $d_{ji2}(\mathbf{m})$  are determined. The cost function for the circular motion is then given by

$$\text{Cost}_{\text{cm}}(\mathbf{m}) = \sqrt{\frac{\sum_{i=1}^N \sum_{j=i+1}^{\min(i+2, N)} \sum_{k=1}^2 (d_{ijk}(\mathbf{m})^2 + d_{jik}(\mathbf{m})^2)}{4(2N - 3)}} \quad (13)$$

where  $\mathbf{m}$  consists of the  $N + 2$  motion parameters. Note that the cost to be minimized here is the rms reprojection error of the epipolar tangents. The cost is minimized using the *conjugate gradient method* [42], with the gradient vector computed by finite differences. Typically, the cost is less than 0.2 (pixels) at the end of the optimization.

### B. Registration of the General Motion

The six motion parameters for the optimization of the general motion are initialized by observing the projection of the 3-D model built from the circular motion, as described in Section V. This is achieved by using a user-friendly interface in which the rotation and translation of the camera are controlled by the mouse movement. Usually, an initialization obtained by this method gives a very small rms reprojection error of just a

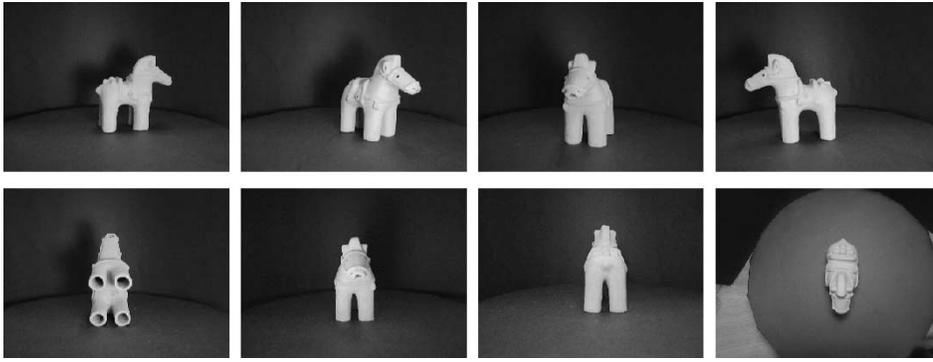


Fig. 9. Top: four images from an uncalibrated sequence of a Hanuwa under circular motion. Bottom: four additional images of the Hanuwa under general motion, featuring the bottom, front, rear and top views.

few pixels, and is good enough to avoid local minima and allow convergence to the global minimum in a few iterations.

During each iteration of the optimization, the projection matrix  $\mathbf{P}_i$  of the arbitrary general view  $i$  is formed using the current estimate of the motion parameters. For each view  $j$ , with projection matrix  $\mathbf{P}_j$ , in the estimated circular motion sequence, a fundamental matrix  $\mathbf{F}_{ij}$  is computed from  $\mathbf{P}_i$  and  $\mathbf{P}_j$  and the reprojection errors  $d_{ij1}(\mathbf{m}')$ ,  $d_{ij2}(\mathbf{m}')$ ,  $d_{ji1}(\mathbf{m}')$  and  $d_{ji2}(\mathbf{m}')$  are determined. The cost function of general motion for view  $i$  is then given by

$$\text{Cost}_{\text{gm}}(\mathbf{m}') = \sqrt{\frac{\sum_{j=1}^N f_{ij} \sum_{k=1}^2 (d_{ijk}(\mathbf{m}')^2 + d_{jik}(\mathbf{m}')^2)}{4 \sum_{j=1}^N f_{ij}}} \quad (14)$$

where  $N$  is the number of views in the estimated circular motion sequence, and  $\mathbf{m}'$  consists of the six motion parameters for the arbitrary general view  $i$  whose pose is to be estimated. The coefficient  $f_{ij}$  is determined by the availability of the two outer epipolar tangents between views  $i$  and  $j$  (see Section III). It is 0 if the baseline between views  $i$  and  $j$  passes through the object, otherwise it is 1. Similar to the circular motion case, the cost here is the rms reprojection error of the epipolar tangents, and is minimized using the conjugate gradient method with the gradient vector computed by finite differences.

### C. Octree Carving

Given the profiles and the estimated camera positions, a volumetric model can be obtained by an octree carving algorithm [10]. Initially, the octree is initialized as a single cube in space which is large enough to enclose the model to be reconstructed. This root cube is projected onto the images and classified as either (a) completely outside one or more profiles, (b) completely inside all the profiles, or (c) ambiguous (i.e., partially outside some profiles). If the cube is classified as type (c), it is subdivided into 8 sub-cubes (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. For the sake of rendering, surface triangles are then extracted from type (c) cubes using

a marching cubes algorithm [27], [43]. The octree carving technique is summarized in algorithm 1.

```

Algorithm 1 Octree Carving From Profiles
initialize a cube that enclose the whole
model;
while max level not reached do
  for each cube in the current level do
    project the cube onto each image;
    classify the cube as either:
      (a) completely outside one 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

```

## VIII. EXPERIMENTS AND RESULTS

The first experimental sequence consisted of fifteen images of a Hanuwa (large hollow baked clay sculpture placed on ancient Japanese burial mounds), of which the first eleven images were taken under unknown circular motion of the Hanuwa, and the last four were taken under unknown general motion (see Fig. 9). The circular motion was first estimated using the circular motion algorithm, and the 3-D model built from the estimated circular motion alone is shown in Fig. 10. Note that the gaps between the legs were not carved away since they never appeared as part of the profiles, and textures were missing in areas (top and bottom) which were invisible under circular motion. The last four views were then registered using the general motion algorithm, and Fig. 11 shows the refined model after incorporating the four arbitrary general views. The model was now fully covered with



Fig. 10. 3-D model of the Hanuwa built from the estimated circular motion alone. The gaps between the legs were not carved away since they never appeared as part of the profiles, and textures were missing in areas (top and bottom) which were invisible under circular motion.



Fig. 11. Refined model of the Hanuwa after incorporating the four 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.

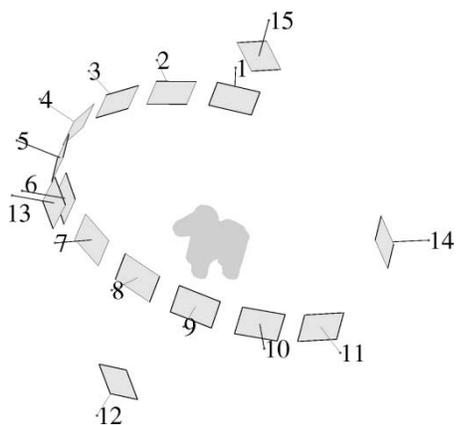


Fig. 12. Camera poses estimated from the Hanuwa sequence.

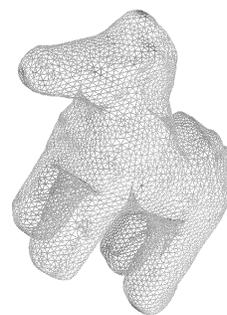


Fig. 13. Triangulated mesh of the Hanuwa model. This model was composed of 12 028 triangles.

textures and showed great improvements in shape, especially in the front, back and top views. The estimated camera configuration used in the reconstruction is shown in Fig. 12, and Fig. 13 shows the triangulated mesh of the Hanuwa model.

The second experimental sequence consisted of fourteen images of an outdoor sculpture acquired by a hand-held camera

under approximate circular motion (see Fig. 14). The circular motion algorithm was first applied to obtain a rough estimate of the camera configuration [44]. Each camera pose in the resulting configuration was then refined in turn by applying the general motion algorithm against the other camera poses, and this process was repeated until no further improvement could be made. Note that after the refinement, the camera was no longer constrained to a circular motion. The final camera configuration estimated from the outdoor sequence is shown

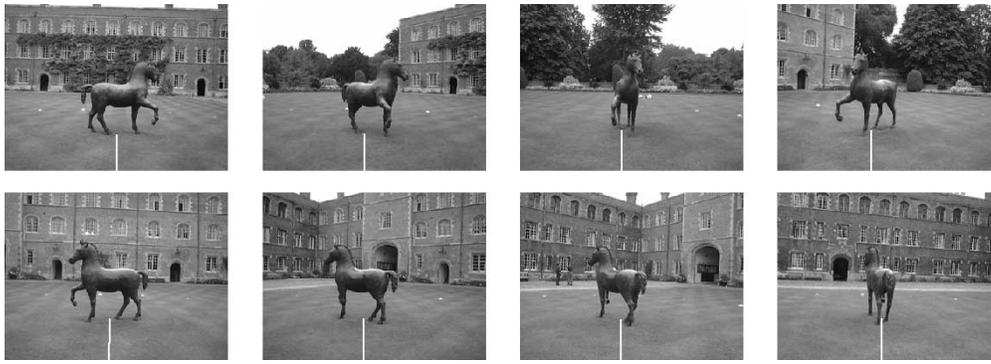


Fig. 14. Eight images of an outdoor sculpture acquired by a hand-held camera under approximate circular motion.

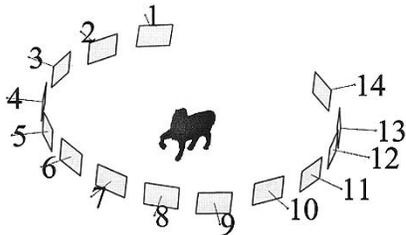


Fig. 15. Camera poses estimated from the outdoor sculpture sequence.

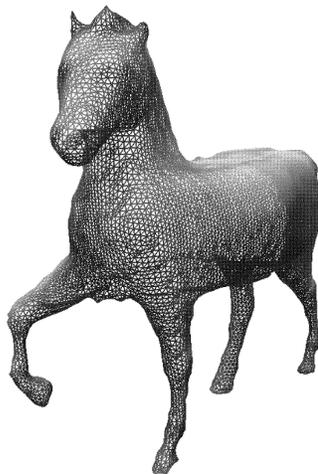


Fig. 16. Triangulated mesh of the reconstructed outdoor sculpture. This model was composed of 29 672 triangles.

in Fig. 15, and Fig. 16 shows the triangulated mesh of the reconstructed model.

## IX. CONCLUSIONS

In this paper, we have presented a complete and practical system for generating high quality 3-D models from 2-D profiles. The input to the system is a sparse, uncalibrated image sequence of an object under both circular and general motion. The incorporation of arbitrary general views reveals information which is concealed under circular motion, and greatly improves both the shape and textures of the 3-D models. It also allows incremental refinement of the 3-D models by adding new views at any time, without the need of setting up the exact, identical scene carefully. The registration of general motion using circular motion (or alternatively three or

more known views) avoids the problem of local minima which exists in every general motion estimation using profiles. The introduction of the use of outer epipolar tangents, which are always available and are guaranteed to be in correspondence, avoids false matches due to self-occlusions and greatly simplifies the corresponding problem. Since only profiles have been used in both the motion estimation and model reconstruction, no corner detection nor matching is necessary. This means that the system is capable of reconstructing any kind of objects, including most sculptures which are composed of *smooth* and *textureless* surfaces. Experiments on various objects have produced convincing 3-D models, demonstrating the practicality of the system.

## REFERENCES

- [1] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *Int. J. Comput. Vis.*, vol. 9, no. 2, pp. 137–154, Nov. 1992.
- [2] P. A. Beardsley, A. Zisserman, and D. W. Murray, "Sequential updating of projective and affine structure from motion," *Int. J. Comput. Vis.*, vol. 23, no. 3, pp. 235–259, June 1997.
- [3] R. Koch, M. Pollefeys, and L. van Gool, "Multi viewpoint stereo from uncalibrated video sequences," in *Proc. 5th Eur. Conf. on Computer Vision*, vol. 1406, H. Burkhardt and B. Neumann, Eds., Freiburg, Germany, June 1998, pp. 55–71.
- [4] A. W. Fitzgibbon, G. Cross, and A. Zisserman, "Automatic 3D model construction for turn-table sequences," in *3D Structure From Multiple Images of Large-Scale Environments, European Workshop SMILE'98*, vol. 1506, R. Koch and L. Van Gool, Eds., Freiburg, Germany, June 1998, pp. 155–170.
- [5] O. D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*. Cambridge, MA: MIT Press, 1993.
- [6] J. Porrill and S. B. Pollard, "Curve matching and stereo calibration," *Image Vis. Comput.*, vol. 9, no. 1, pp. 45–50, Feb. 1991.
- [7] R. Cipolla, K. E. Åström, and P. J. Giblin, "Motion from the frontier of curved surfaces," in *Proc. 5th Int. Conf. on Computer Vision*, Cambridge, MA, June 1995, pp. 269–275.
- [8] P. R. S. Mendonça, K.-Y. K. Wong, and R. Cipolla, "Epipolar geometry from profiles under circular motion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 604–616, June 2001.
- [9] K.-Y. K. Wong, P. R. S. Mendonça, and R. Cipolla, "Structure and motion estimation from apparent contours under circular motion," *Image Vis. Comput.*, vol. 20, no. 5–6, pp. 441–448, Apr. 2002.
- [10] R. Szeliski, "Rapid octree construction from image sequences," *Comput. Vis., Image Process.*, vol. 58, no. 1, pp. 23–32, July 1993.
- [11] K.-Y. K. Wong and R. Cipolla, "Structure and motion from silhouettes," in *Proc. 8th Int. Conf. Computer Vision*, vol. II, Vancouver, BC, Canada, July 2001, pp. 217–222.
- [12] J. H. Rieger, "Three dimensional motion from fixed points of a deforming profile curve," *Opt. Lett.*, vol. 11, no. 3, pp. 123–125, Mar. 1986.
- [13] P. J. Giblin, F. E. Pollock, and J. E. Rycroft, "Recovery of an unknown axis of rotation from the profiles of a rotating surface," *J. Opt. Soc. Amer. A*, vol. 11, no. 7, pp. 1976–1984, July 1994.

- [14] K. Åström, R. Cipolla, and P. Giblin, "Generalized epipolar constraints," *Int. J. Comput. Vis.*, vol. 33, no. 1, pp. 51–72, Sept. 1999.
- [15] P. R. S. Mendonça and R. Cipolla, "Estimation of epipolar geometry from apparent contours: affine and circular motion cases," in *Proc. Conf. Computer Vision and Pattern Recognition*, vol. I, Fort Collins, CO, 1999, pp. 9–14.
- [16] J. Sato and R. Cipolla, "Affine reconstruction of curved surfaces from uncalibrated views of apparent contours," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 11, pp. 1188–1198, November 1999.
- [17] G. Cross, A. W. Fitzgibbon, and A. Zisserman, "Parallax geometry of smooth surfaces in multiple views," in *Proc. 7th Int. Conf. Computer Vision*, Corfu, Greece, Sept. 1999, pp. 323–329.
- [18] T. Joshi, N. Ahuja, and J. Ponce, "Structure and motion estimation from dynamic silhouettes under perspective projection," *Int. J. Comput. Vis.*, vol. 31, no. 1, pp. 31–50, Feb. 1999.
- [19] A. A. Grattarola, "Volumetric reconstruction from object silhouettes: A regularization procedure," *Signal Process.*, vol. 27, no. 1, pp. 27–35, 1992.
- [20] D. Marr, "Analysis of occluding contour," in *Proc. R. Soc. Lond. B*, vol. 197, 1977, pp. 441–475.
- [21] R. Cipolla and A. Blake, "Surface shape from the deformation of apparent contours," *Int. J. Comput. Vis.*, vol. 9, no. 2, pp. 83–112, Nov. 1992.
- [22] R. Cipolla and P. J. Giblin, *Visual Motion of Curves and Surfaces*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
- [23] H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, vol. 293, pp. 133–135, September 1981.
- [24] O. D. Faugeras, "What can be seen in three dimensions with an uncalibrated stereo rig," in *Proc. 2nd Eur. Conf. Computer Vision*, vol. 588, G. Sandini, Ed., Santa Margherita Ligure, Italy, May 1992, pp. 563–578.
- [25] K.-Y. K. Wong, P. R. S. Mendonça, and R. Cipolla, "Camera calibration from surfaces of revolution," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 2, pp. 147–161, Feb. 2003.
- [26] T. Vieville and D. Lingrand, "Using singular displacements for uncalibrated monocular visual systems," in *Proc. 4th Eur. Conf. Computer Vision*, vol. 1065, B. Buxton and R. Cipolla, Eds., Cambridge, U.K., April 1996, pp. 207–216.
- [27] W. E. Lorensen and H. E. Cline, "Marching cubes: a high resolution 3D surface construction algorithm," *ACM Comput. Graph.*, vol. 21, no. 4, pp. 163–169, July 1987.
- [28] P. J. Giblin and R. S. Weiss, "Reconstructions of surfaces from profiles," in *Proc. 1st Int. Conf. Computer Vision*, London, U.K., June 1987, pp. 136–144.
- [29] R. Vaillant and O. D. Faugeras, "Using extremal boundaries for 3D object modeling," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 157–173, Feb. 1992.
- [30] R. Szeliski and R. Weiss, "Robust shape recovery from occluding contours using a linear smoother," *Int. J. Comput. Vis.*, vol. 28, no. 1, pp. 27–44, June 1998.
- [31] E. Boyer and M. O. Berger, "3D surface reconstruction using occluding contours," *Int. J. Comput. Vis.*, vol. 22, no. 3, pp. 219–233, Mar. 1997.
- [32] W. N. Martin and J. K. Aggarwal, "Volumetric descriptions of objects from multiple views," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 5, no. 2, pp. 150–158, March 1983.
- [33] C. H. Chien and J. K. Aggarwal, "Volume/surface octrees for the representation of three-dimensional objects," *Comput. Vis., Graph., Image Process.*, vol. 36, no. 1, pp. 100–113, Oct. 1986.
- [34] N. Ahuja and J. Veenstra, "Generating octrees from object silhouettes in orthographic views," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 2, pp. 137–149, Feb. 1989.
- [35] T. H. Hong and M. O. Shneier, "Describing a robot's workspace using a sequence of views from a moving camera," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-7, pp. 721–726, Nov. 1985.
- [36] M. Potmesil, "Generating octree models of 3D objects from their silhouettes in a sequence of images," *Comput. Vis., Graph., Image Process.*, vol. 40, no. 1, pp. 1–29, Oct. 1987.
- [37] H. Noborio, S. Fukuda, and S. Arimoto, "Construction of the octree approximating three-dimensional objects by using multiple views," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, no. 6, pp. 769–782, Nov. 1988.
- [38] S. K. Srivastava and N. Ahuja, "Octree generation from object silhouettes in perspective views," *Comput. Vis., Graph., Image Process.*, vol. 49, no. 1, pp. 68–84, Jan. 1990.
- [39] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," *Int. J. Comput. Vis.*, vol. 38, no. 3, pp. 197–216, July 2000.
- [40] A. Broadhurst, T. W. Drummond, and R. Cipolla, "A probabilistic framework for space carving," in *Proc. 8th Int. Conf. Computer Vision*, vol. I, Vancouver, BC, Canada, July 2001, pp. 388–393.
- [41] Q. T. Luong and O. D. Faugeras, "The fundamental matrix: theory, algorithm, and stability analysis," *Int. J. Comput. Vis.*, vol. 17, no. 1, pp. 43–75, Jan. 1996.
- [42] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, Jan. 1993.
- [43] C. Montani, R. Scateni, and R. Scopigno, "A modified look-up table for implicit disambiguation of marching cubes," *Vis. Comput.*, vol. 10, no. 6, pp. 353–355, 1994.
- [44] K.-Y. K. Wong and R. Cipolla, "Reconstruction of outdoor sculptures from silhouettes under approximate circular motion of an uncalibrated hand-held camera," in *Proc. IAPR Workshop on Machine Vision Applications*, Nara, Japan, Dec. 2002, pp. 459–462.



**Kwan-Yee Kenneth Wong** received the B.Eng. degree in computer engineering, with first class honors, from the Chinese University of Hong Kong in 1998. He received the M.Phil. and Ph.D. degrees, both in computer vision, from the University of Cambridge, Cambridge, U.K., in 2000 and 2001, respectively.

Since 2001, he has been with the Department of Computer Science and Information Systems at the University of Hong Kong, where he is now an Assistant Professor. His research interests are in computer vision and image processing, including camera calibration, motion tracking, model reconstruction and representation, and motion estimation from image sequences.



**Roberto Cipolla** received the B.A. degree in engineering from the University of Cambridge, Cambridge, U.K., in 1984, and the M.S.E.E. degree in electrical engineering from the University of Pennsylvania, Philadelphia, in 1985. From 1985 to 1988, he studied at Osaka University of Foreign Studies (Japanese language) and the Electro-technical Laboratory, Tsukuba (Visiting Scientist), and received the M.Eng. degree in robotics from the University of Electro-Communications, Tokyo, Japan, in 1988. In 1991, he received the D.Phil. degree in computer vision from the University of Oxford, Oxford, U.K.

From 1991 to 1992, he was a Toshiba Fellow and Engineer at the Toshiba Corporation Research and Development Center, Kawasaki, Japan. He joined the Department of Engineering, University of Cambridge, in 1992 as a Lecturer and a Fellow of Jesus College. He became a Reader in 1997 and a Professor of information engineering in 2000. His research interests are in computer vision and robotics.