Title	Head model acquisition from silhouettes
Author(s)	Wong, KKY; Mendonca, PRS; Cipolla, R
Citation	The 4th International Workshop on Visual Form (IWVF-4), Capri, Italy, 28-30 May 2001. In Proceedings of the 4th International Workshop on Visual Form, 2001, p. 787-796
Issued Date	2001
URL	http://hdl.handle.net/10722/137127
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Head Model Acquisition from Silhouettes

Kwan-Yee K. Wong, Paulo R. S. Mendonça, and Roberto Cipolla

Department of Engineering
University of Cambridge
Cambridge, CB2 1PZ, UK
`[kykw2|prdsm2|cipolla]@eng.cam.ac.uk`
`http://svr-www.eng.cam.ac.uk/research/vision`

Abstract This paper describes a practical system developed for generating 3D models of human heads from silhouettes alone. The input to the system is an image sequence acquired from circular motion. Both the camera motion and the 3D structure of the head are estimated using silhouettes which are tracked throughout the sequence. Special properties of the camera motion and their relationships with the intrinsic parameters of the camera are exploited to provide a simple parameterization of the fundamental matrix relating any pair of views in the sequence. Such a parameterization greatly reduces the dimension of the search space for the optimization problem. In contrast to previous methods, this work can cope with incomplete circular motion and more widely spaced images. Experiments on real image sequences are carried out, showing accurate recovery of 3D shapes.

1 Introduction

The reconstruction of 3D head models has many important applications, such as video conferencing, model-based tracking, entertainment and face modeling [13]. Existing commercial methods for acquiring such models, such as laser scans, are expensive, time-consuming and cannot cope with low-reflectance surfaces. Image based systems can easily overcome these difficulties by tracking point features along video sequences [5]. However, this can be remarkably difficult for human faces, where there are not many reliable landmarks with long life span along the sequence.

In this paper we present a practical system for generating 3D head models from silhouettes alone. Silhouettes are comparatively easy to track and provide useful information for estimating the camera motion [1,10] and reconstruction [12,2,15]. Since they tend to concentrate around regions of high curvature, they provide a compact way of parameterizing the reconstructed surface. In our system, images are acquired by moving the camera along a circular path around the head. This imposes constraints on the fundamental matrix relating each pair of images, simplifying the motion estimation. The system does not require the motion to be a full rotation and the images can be acquired at more widely spaced positions around the subject, an advantage over the technique introduced in [10].

Section 2 presents the theoretical background of motion estimation from silhouettes. The algorithms for model building are described in Section 3, and Section 4 shows the experimental results. Conclusions are given in Section 5.

2 Theoretical Background

The fundamental difficulty in solving the problem of structure and motion from silhouettes is that, unlike point or line features, the silhouettes do not readily provide image correspondences that allow for the computation of the epipolar geometry, summarized by the fundamental matrix. The usual solution to this problem is the use of *epipolar tangencies* [11,3], as shown in Fig. 1. An epipolar tangent point is the projection of a *frontier point* [3], which is the intersection of two contour generators. If 7 or more
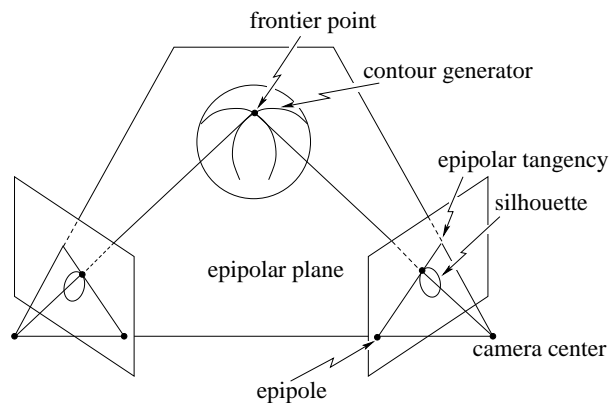

Figure 1. A frontier point is the intersection of two contour generators and is visible in both views. The frontier point projects onto a point on the silhouette which is also on an epipolar tangent

epipolar tangent points are available, the epipolar geometry can be estimated. The intrinsic parameters of the cameras can then be used to recover the motion [6,4]. However the unrealistic demand for a large number of epipolar tangent points makes this approach impractical. By constraining the motion to be circular, a parameterization of the fundamental matrix with only 6 degrees of freedom (dof) is possible [14,5,9]. This parameterization explicitly takes into account the main image features of circular motion, namely the image of the rotation axis, the horizon and a special vanishing point, which are fixed throughout the sequence. This makes it possible to estimate the epipolar geometry by using only 2 epipolar tangencies [9].

In [10], a practical algorithm has been introduced for the estimation of motion and structure from silhouettes of a rotating object. The image of the rotation axis and the vanishing point are first determined by estimating the harmonic homology associated with the image of surface of revolution spanned by the object. In order to obtain such an image, a dense image sequence from a complete circular motion is required. In this paper, the parameters of the harmonic homology and other motion parameters are estimated simultaneously by minimizing the reprojection errors of epipolar tangents. This algorithm does not require the image of such a surface of revolution and thus can cope

with incomplete circular motion and more widely spaced images, an advantage over the algorithm presented in [10].

2.1 Symmetry and Epipolar Geometry in Circular Motion

Consider a pinhole camera undergoing circular motion. If the camera intrinsic parameters are kept constant, the projection of the rotation axis will be a line l_s which is pointwise fixed on each image. This means that, for any point $\mathbf{x}$ on l_s, the equation $\mathbf{x}^T\mathbf{F}\mathbf{x} = 0$ is satisfied, where $\mathbf{F}$ is the fundamental matrix related to any image pair in the sequence. For circular motion, all the camera centers lie on a common plane. The image of this plane is a special line l_h, the *horizon*. Since the epipoles are the images of the camera centers, they must lie on l_h. In general, l_s and l_h are not orthogonal. Another feature of interest is the vanishing point $\mathbf{v}_x$ which corresponds to the normal direction of the plane defined by the camera center and the axis of rotation. The vanishing point and the horizon satisfy $\mathbf{v}_x^T l_h = 0$. A detailed discussion of the above can be found in [14,5,9].

Consider now a pair of cameras, denoted as $\mathbf{P}_1$ and $\mathbf{P}_2$, related by a rotation about an axis not passing through their centers, and let $\mathbf{F}$ be the fundamental matrix associated with this pair. It has been shown that corresponding epipolar lines associated with $\mathbf{F}$ are related to each other by a harmonic homology $\mathbf{W}$ [9], given by

$$\mathbf{W} = \mathbb{I} - 2\frac{\mathbf{v}_x l_s}{\mathbf{v}_x^T l_s}. \tag{1}$$

Note that $\mathbf{W}$ has 4 dof: 2 corresponding to the axis and 2 corresponding to the vanishing point. If $\mathbf{P}_1$ and $\mathbf{P}_2$ point towards the axis of rotation, $\mathbf{v}_x$ will be at infinity and $\mathbf{W}$ will be reduced to a skew symmetry $\mathbf{S}$ with only 3 dof. Besides, if the cameras also have zero skew and aspect ratio 1, the transformation will be further specialized to a bilateral symmetry $\mathbf{B}$ with only 2 dof. A pictorial description of these transformations can be seen in Fig. 2
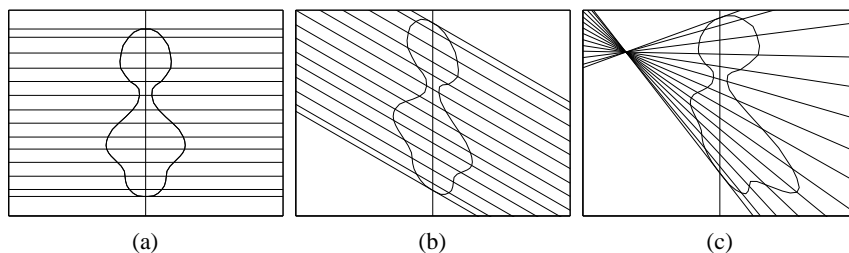

(a)	(b)	(c)

Figure 2. (a) A curve displaying bilateral symmetry. The horizon is orthogonal to the axis. (b) Same curve, distorted by an affine transformation. The horizon is no longer orthogonal to the axis, and each side of the curve is mapped to the other by a skew symmetry transformation. (c) The curve is now distorted by a special projective transformation (harmonic homology), and the lines of symmetry intersect at a point corresponding to the vanishing point

In [16], an algorithm has been presented for estimating the camera intrinsic parameters from 2 or more silhouettes of surfaces of revolution. For each silhouette, the associated harmonic homology $\mathbf{W}$ is estimated and this provides 2 constraints on the camera intrinsic parameters:

$$\mathbf{v}_x = \mathbf{K}\mathbf{K}^{\mathrm{T}}\mathbf{l}_{\mathrm{s}}, \tag{2}$$

where $\mathbf{K}$ is the 3×3 camera calibration matrix. Conversely, if the camera intrinsic parameters are known, (2) provides 2 constraints on $\mathbf{W}$ and as a result $\mathbf{W}$ has only 2 dof.

2.2 Parameterization of the Fundamental Matrix

In [8,17], it has been shown that any fundamental matrix $\mathbf{F}$ can be parameterized as $\mathbf{F} = [\mathbf{e}_2]_\times \mathbf{M}$, where $\mathbf{M}^{-\mathrm{T}}$ is any matrix that maps the epipolar lines from one image to the other, and $\mathbf{e}_2$ is the epipole in the second image. In the special case of circular motion, it follows that

$$\mathbf{F} = [\mathbf{e}_2]_\times \mathbf{W}. \tag{3}$$

Note that $\mathbf{F}$ has 6 dof: 2 to fix $\mathbf{e}_2$, and 4 to determine $\mathbf{W}$. From (2), if the camera intrinsic parameters are known, 2 parameters are enough to define $\mathbf{W}$ and thus $\mathbf{F}$ will have only 4 dof.

An alternative parameterization for the fundamental matrix in the case of circular motion [14,5,9] is given by

$$\mathbf{F} = [\mathbf{v}_x]_\times + \kappa \tan\frac{\theta}{2}(\mathbf{l}_{\mathrm{s}}\mathbf{l}_{\mathrm{h}}^{\mathrm{T}} + \mathbf{l}_{\mathrm{h}}\mathbf{l}_{\mathrm{s}}^{\mathrm{T}}), \tag{4}$$

where θ is the angle of rotation between the cameras. The constant κ can be determined from the camera intrinsic parameters [9] if $\mathbf{l}_{\mathrm{s}}$, $\mathbf{v}_x$ and $\mathbf{l}_{\mathrm{h}}$ are properly normalized. θ is the only parameter which depends on the particular pair of cameras being considered, while the other 4 terms are common to all pairs of images in the sequence. When the camera intrinsic parameters are known, 2 parameters are enough to fix $\mathbf{l}_{\mathrm{s}}$ and $\mathbf{v}_x$. Since $\mathbf{v}_x$ must lie on $\mathbf{l}_{\mathrm{h}}$, only 1 further parameter is needed to fix $\mathbf{l}_{\mathrm{h}}$. As a result, the fundamental matrix has only 4 dof.

3 Algorithms

Before the 3D model can be reconstructed from the silhouettes of the head, the motion of the camera has to be estimated. By using the parameterization shown in (4), the $\binom{N}{2} = N(N-1)/2$ fundamental matrices relating all possible pairs of cameras in a sequence of N images, taken by a rotating camera with known intrinsic parameters, can be defined with the 3 parameters which fix $\mathbf{l}_{\mathrm{s}}$, $\mathbf{v}_x$ and $\mathbf{l}_{\mathrm{h}}$, together with the $N-1$ angles of rotation between adjacent cameras. By enforcing the epipolar constraint on the corresponding epipolar tangent points, these $N+2$ motion parameters can be estimated by minimizing the reprojection errors of corresponding epipolar tangents (see Fig. 3).

Since a silhouette has at least two epipolar tangencies (one at its top and another at its bottom), there will be totally $2\binom{N}{2} = N(N-1)$ measurements from all pairs of images. Due to the dependence between the associated fundamental matrices, however, these $N(N-1)$ measurements only provide $2N$ (or 2 when $N=2$) independent constraints on the $N+2$ parameters. As a result, a solution will be possible if $N \geq 3$.
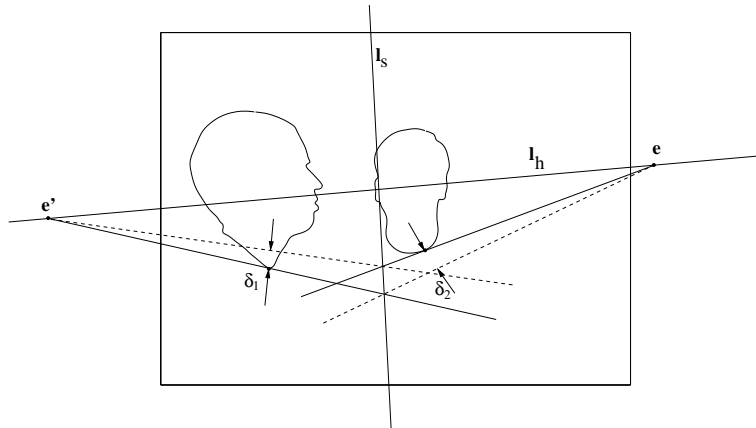

Figure 3. The parameters of the fundamental matrix associated with each pair of images in the sequence can be estimated from the reprojection errors of epipolar tangents. The solid lines are tangents to the silhouettes passing through the epipoles, and the dashed lines are the epipolar lines corresponding to the tangent points

The minimization of the reprojection errors will generate a consistent set of fundamental matrices, which, together with the camera intrinsic parameters, can be decomposed into a set of camera matrices describing a circular motion compatible with the image sequence. The algorithm for motion estimation is summarized in Algorithm 1. Having the motion of the camera estimated, the 3D model can then be reconstructed from the silhouettes using the simple triangulation technique introduced in [15].

4 Experiments and Results

In order to evaluate the performance of the algorithm described in Section 3, 2 human head sequences, each with 10 images, were acquired using the setup shown in Fig. 4. The camera is mounted to the extensible rotating arm of the tripod, whose height can be adjusted according to the height of the subject. Each image in the sequence was taken after rotating the arm of the tripod roughly by $20°$, with the subject standing close to the tripod. The intrinsic parameters of the camera are obtained from an offline calibration process. The silhouettes of the heads are tracked using cubic B-spline snakes [2] (see Fig. 5 and 6).

Algorithm 1 Estimation of the motion parameters from silhouettes

    track the silhouettes of the head using cubic B-spline snakes;
    initialize $\mathbf{l}_s$, $\mathbf{l}_h$ and the $N-1$ angles between the N cameras;
    while not converged **do**
      for each image pair **do**
        form fundamental matrix;
        locate epipolar tangents;
        compute reprojection errors;
      end for
      update parameters to minimize the sum of reprojection errors;
    end while

Figure 4. Experimental setup used to acquire image sequences around human heads. The camera is mounted to the rotating arm of the tripod with the subject standing close to the tripod. Although the camera motion is constrained to be circular, the camera orientation and rotation angle are unknown
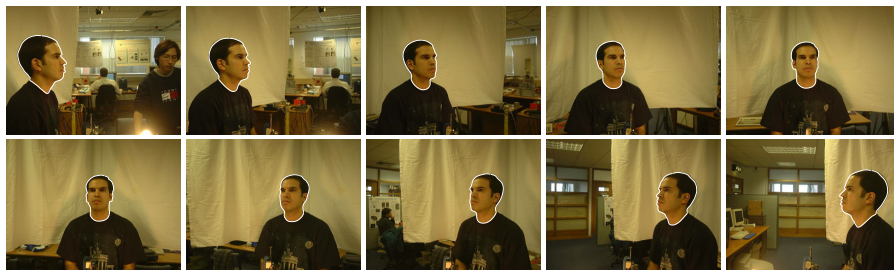

Figure 5. Image sequence (I) used in the experiment, with the silhouettes of the head tracked using cubic B-spline snakes
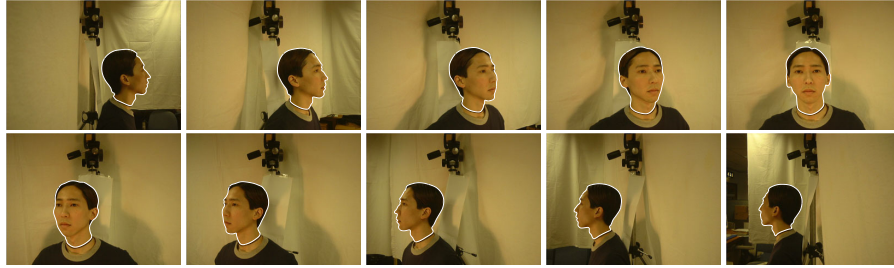
Figure 6. Image sequence (II) used in the experiment, with the silhouettes of the head tracked using cubic B-spline snakes

The initial guess for the horizon and the image of the rotation axis was picked by observation, and the angles of rotation were initialized as $10°$ respectively. The sum of the reprojection errors was minimized using the Levenberg-Marquardt algorithm [7].The reconstructed 3D head models can be found in Fig. 7 and 8. The shapes of the ears, lips, noise and eyebrows demonstrate the quality of the 3D models recovered.

5 Conclusions

In this paper we have presented a simple and practical system for building 3D models of human heads from image sequences. No prior model is assumed, and in fact the system can be applied to a variety of objects. The only constraint on the camera is that it must perform circular motion, though the exact camera orientations and positions are unknown. Besides, the camera is not required to perform a full rotation and there is no need for using a dense image sequence. The silhouettes of the head are the only information used for both motion estimation and reconstruction, circumventing the lack, instability and occlusion of landmarks on faces. The silhouettes also provide a natural and compact way of parameterizing the head model, concentrating contours around regions of high curvature. The experimental results show the accuracy of the acquired model.

Acknowledgements

References

1. R. Cipolla, K. Åström, and P.J. Giblin. Motion from the frontier of curved surfaces. In *Proc. 5th Int. Conf. on Computer Vision*, pages 269–275, 1995.
2. R. Cipolla and A. Blake. Surface shape from the deformation of apparent contours. *Int. Journal of Computer Vision*, 9(2):83–112, 1992.
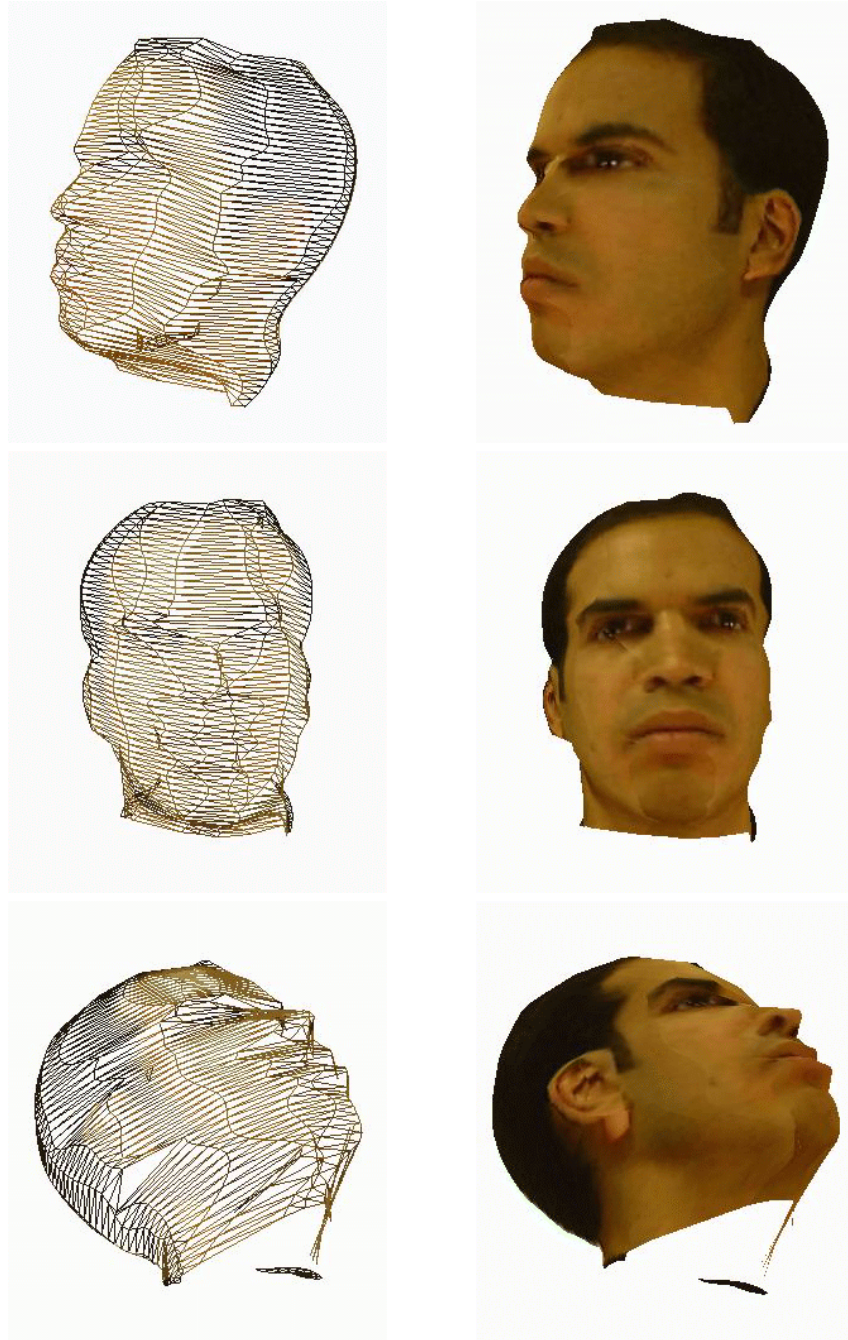
Figure 7. Different views of the VRML model from the model building process using the 10 images in Fig. 5
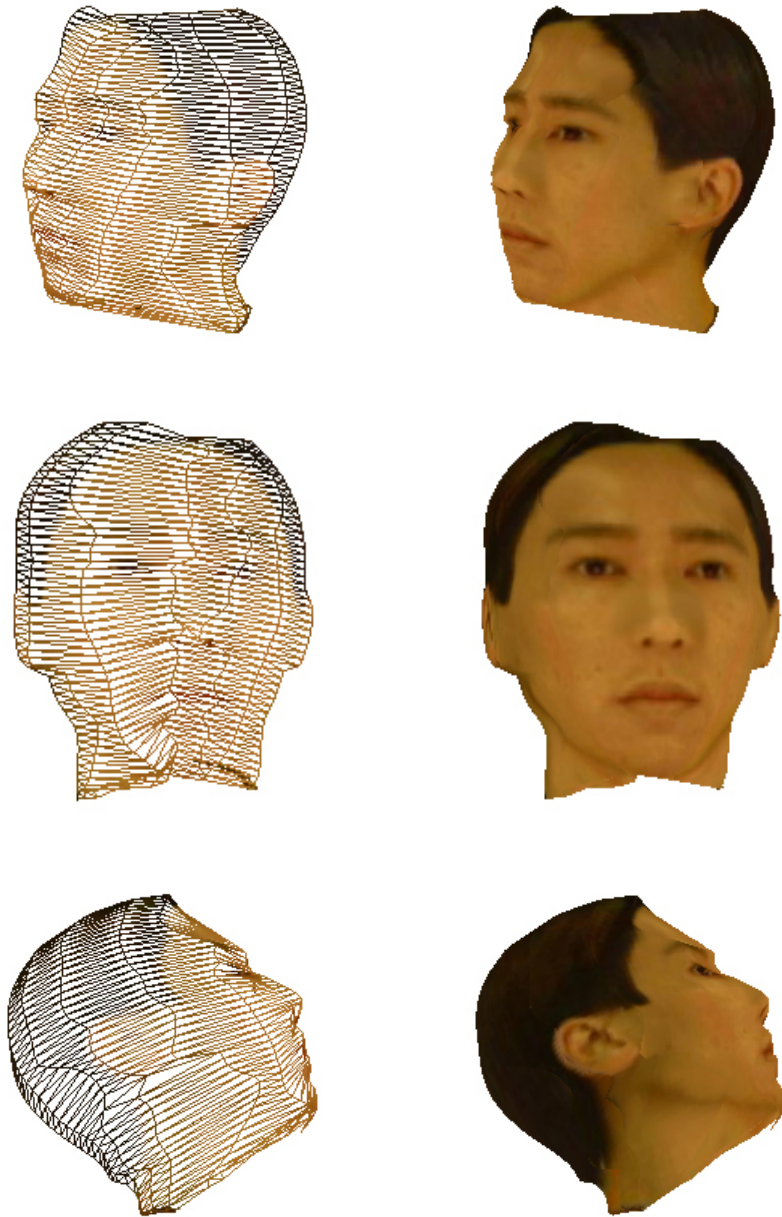
Figure 8. Different views of the VRML model obtained from the model building process using the 10 images in Fig. 6

3. R. Cipolla and P. J. Giblin. *Visual Motion of Curves and Surfaces*. Cambridge University Press, Cambridge, 1999.

4. O. Faugeras. Stratification of three-dimensional vision: Projective, affine and metric representations. *J. Opt. Soc. America A*, 12(3):465–484, 1995.

5. A. W. Fitzgibbon, G. Cross, and A. Zisserman. Automatic 3D model construction for turntable sequences. In *3D Structure from Multiple Images of Large-Scale Environments, European Workshop SMILE'98*, Lecture Notes in Computer Science 1506, pages 155–170, 1998.

6. H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

7. D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, USA, second edition, 1984.

8. Q.-T. Luong and O. Faugeras. The fundamental matrix: Theory, algorithm, and stability analysis. *Int. Journal of Computer Vision*, 17(1):43–75, 1996.

9. P. R. S. Mendonça, K.-Y. K. Wong, and R. Cipolla. Recovery of circular motion from profiles of surfaces. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, Lecture Notes in Computer Science 1883, pages 151–167, Corfu, Greece, Sep 1999. Springer–Verlag.

10. P. R. S. Mendonça, K.-Y. K. Wong, and R. Cipolla. Camera pose estimation and reconstruction from image profiles under circular motion. In D. Vernon, editor, *Proc. 6th European Conf. on Computer Vision*, volume II, pages 864–877, Dublin, Ireland, Jun 2000. Springer–Verlag.

11. J. Porrill and S. B. Pollard. Curve matching and stereo calibration. *Image and Vision Computing*, 9(1):45–50, 1991.

12. R. Vaillant and O. D. Faugeras. Using extremal boundaries for 3D object modelling. *IEEE Trans. on Pattern Analysis and Machine Intell.*, 14(2):157–173, 1992.

13. T. Vetter. Synthesis of novel views from a single face image. *Int. Journal of Computer Vision*, 28(2):103–116, June 1998.

14. T. Vieville and D. Lingrand. Using singular displacements for uncalibrated monocular visual systems. In *Proc. 4th European Conf. on Computer Vision*, volume II, pages 207–216, 1996.

15. K.-Y. K. Wong, P. R. S. Mendonça, and R. Cipolla. Reconstruction and motion estimation from apparent contours under circular motion. In T. Pridmore and D. Elliman, editors, *Proc. British Machine Vision Conference*, volume 1, pages 83–92, Nottingham, UK, Sep 1999. British Machine Vision Association.

16. K.-Y. K. Wong, P. R. S. Mendonça, and R. Cipolla. Camera calibration from symmetry. In R. Cipolla and R. Martin, editors, *The Mathematics of Surfaces IX*, pages 214–226, Cambridge, UK, Sep 2000. Institute of Mathematics and its Applications, Springer–Verlag.

17. Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *Int. Journal of Computer Vision*, 27(2):161–195, 1998.