Ellipsoid Reconstruction from Three Perspective Views*

S.D.MA and L.LI
National Pattern Recognition Laboratory
Institute of Automation
Beijing 100080
Email Address: masd@prlsun2.ia.ac.cn

Abstract

We present a method to reconstruct an ellipsoid from its occluding contours observed in three images. We derive a linear relationship between an ellipsoid and its perspective projection. From this relationship, we show that, if cameras are calibrated, an ellipsoid can be reconstructed from its three views and the solution is, in general, unique; if the cameras are weakly calibrated, then the reconstruction is also unique but up to projectivity. Our method has been successfully tested on synthetic data and on real image data.

1 Introduction

One of the important tasks for a computer vision system is to reconstruct 3D objects in the scene from their images. Many methods proposed are edge based in which we first detect the edge points on image contours. Image contours are the projection of the surface bounding contours. There are two distint types of contours that bound a surface, which are called extremal and discontinuity. A discontinuity contour marks the abrupt termination of a smooth surface (e.g., when it intersects another surface), while at an extremal contour, the surface normal turns away smoothly from the viewer.

It is well-known that extremal contours are a rich source of information. (To distinguish between the extremal contour on an object in 3D space and its image, the extremal contour will be referred to as the rim, and its image will be called the occluding contour in this paper). Any rim on an object is a special curve, defined by the fact that the optical ray is tangent to the surface of the object at each rim point(see figure 1). The rim is viewpoint dependent, which means that rims observed from different viewpoints do not correspond to the same curve on the object. The rim and its image have been studied by many researchers. They presented the methods to reconstruct the surface from its occluding contours. The standard technique of stereo vision fails because the occluding contours in two images do not correspond to the same rim on the surface in 3D space. Kriegman and Ponce [7] present a method to recover the pose of a 3D object from a single image. Assuming the object model is available in a CAD-based vision system, they used the occluding contours to recover the pose of an object. The equations they should solve are, in general, highly non-linear. Giblin and Weiss [3] present an algorithm for computing a depth map for a smooth surface from a sequence of images by modeling the object as the envelop of its tangents. The algorithm requires that the viewing directions be coplanar. Vaillant and Faugeras [15] present an algorithm to reconstruct the local shape of a surface along the rim from three views. They parameterize the local surface patches with respect to the Euler angles of the normal to the surface. Using this algorithm, they can compute the depth and the second fundamental forms of the surface along the rim. They also discuss the important issue of identifying occluding contours in the image using their differential features.

In this paper we deal with a special case in which we assume that the 3D object to be reconstructed is an ellipsoid. This work is related to our previous work [9] and to the work of Karl[6], who showed that an ellipsoid can be reconstructed from its *three orthogonal projections* and the relationship between the ellipsoid and its orthogonal projections is linear. In our previous work[9], we presented a method to recover the quadric surface from its perspective views but the computation is non-linear. In this paper, we derive the linear relationship between the representation of an ellipsoid and its perspective views. It is shown that, if cameras are calibrated, an ellipsoid can be reconstructed from its occluding contours observed from at least three views and the computation is linear. Projective reconstruction is a recent topic of investigation. It has been shown that without camera calibration, or more exactly, with cameras that are weakly calibrated [2], [11], [4], the 3D scene can be reconstructed up to a projective transformation. Since the reconstruction of an ellipsoid needs three views, the weak calibration in this paper is in the sense of Hartley [5] and Shashua [12]. It was shown in their papers that, as in the role played by fundamental matrix in two-view case, a $3 \times 3 \times 3$ tensor can be estimated from point or line correspondences in three views. A similar relation was obtained by Aloimonos [13]. The estimation of this trifocal tensor can be considered as the weak calibra-

*This work is supported by the Chinese National Science Foundation and the Chinese National High Technology Program

1015-4651/96 $5.00 © 1996 IEEE
Proceedings of ICPR '96

tion of three cameras. From this trifocal tensor, the projective matrices of three cameras can be computed and then the 3D scene can be reconstructed up to a projectivity.

Our paper is organized as follows: In the next section, we describe the perspective transformation between an object in 3D space and its image. It is shown that the projective matrices can be assumed to be known by camera calibration or weak calibration respectively for metric space reconstruction or projective reconstruction. In section 3, we provide the linear relationship between an ellipsoid and its perspective views. The reconstruction algorithm is provided in section 4. We show that an ellipsoid can be recovered from at least three views by solving a system of linear equations. We provide four geometric invariants which can be used to establish the correspondence of the occluding contours in the images. Experiments using simulated data and real image data are presented in section 5. The conclusion is given in the last section.

2 Perspective Transformation and Camera Calibration

The perspective transformation between any 3D point and its projection in the i^{th} image (see figure 1) can be represented as

$$\mathbf{u}_i \simeq \mathbf{M}_i \mathbf{x} \qquad i = 1 \sim n \tag{1}$$

where $\simeq$ indicates the equality up to a scale factor, $\mathbf{x} = (x, y, z, 1)^t$ are the homogeneous coordinates of any point in object space, $\mathbf{u}_i = (u_i, v_i, 1)^t$ are the coordinates of the projection in i^{th} image, and $\mathbf{M}_i$ is a 3×4 matrix, which is called the camera matrix or projective matrix. Our method presented in this paper can be used for metric reconstruction as well as for projective reconstruction. In both cases, we assume the projective matrices $\mathbf{M}_i$ $(i = 1 \sim n)$ are available from either weak (projective) or strong (Euclidian metric) camera calibration.

1. For the reconstruction in metric space, it is well-known that the projective matrices $\mathbf{M}_i$ $(i = 1 \sim n)$ can be obtained by camera calibration [14] [1] [16]. By certain definition of the coordinate system, the projective matrix of the first camera $\mathbf{M}_1$ is equal to $(\mathbf{I}|\mathbf{0})$, where $\mathbf{I}$ is a 3×3 indentity matrix and $\mathbf{0}$ is a 3D null vector.
2. In the case of projective reconstruction, it has been shown that without camera calibration, or more exactly, with weak camera calibration [2], [11], [4], the 3D scene can be reconstructed up to a projectivity. Since the reconstruction of an ellipsoid needs three views, the weak calibration in this paper is in the sense of Hartley [5] and Shashua[12]. It was shown in their papers that, as in the role played by fundamental matrix in the two-view case, a $3 \times 3 \times 3$ tensor can be estimated from point or line correspondences in three views. The estimation of this trifocal tensor can be considered as the weak calibration of the three cameras. From this trifocal tensor, the projective matrices $\mathbf{M}_i, (i = 1 \sim 3)$ of three cameras can be computed. Then we can use these projective matrices to reconstruct the scene up to projectivity. In this case, the image or the object coordinate system can be respectively any 2D or 3D projective coordinate system. Since we reconstruct the 3D scene only up to projectivity, we can always assume the first projective matrix is equal to $(\mathbf{I}|\mathbf{0})$ as pointed out by Hartley [5].

We assume in the following sections that all the projective matrices are available by the camera calibration or weak calibration, and for simplicity, the projective matrix of the first camera is equal to $(\mathbf{I}|\mathbf{0})$. The method described in the following sections is valid both for metric and projective reconstruction, provided we keep in mind that the projective matrices are given respectively by strong or weak camera calibration.

3 Linear Relationship Between an Ellipsoid and its Occluding Contour

3.1 Geometric Constraints

In this subsection, we give the geometric constraint between an ellipsoid and its image in the first camera, assuming the projective matrix is equal to $(\mathbf{I}|\mathbf{0})$.

An ellipsoid in the scene can be represented as

$$\mathbf{x}^t \mathbf{A} \mathbf{x} = 0 \tag{2}$$

where $\mathbf{x}$ is any point on the surface of the ellipsoid and $\mathbf{A}$ is a 4×4 symmetric matrix.

For later convenience, we denote

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{33} & \mathbf{A}_4 \\ \mathbf{A}_4^t & a_{44} \end{pmatrix} \tag{3}$$

where $\mathbf{A}_{33}$ is the 3×3 upper left submatrix of $\mathbf{A}$ and $\mathbf{A}_4 = (a_{41}, a_{42}, a_{43})^t$ is a three dimensional vector.

From the fact that any line joining the optical center of the camera and any point on the rim of an ellipsoid is tangent to the ellipsoid(see figure 1), it can be proved [9] that the extremal contour of an ellipsoid is a planar conic, and in turn its perspective projection (called the occluding contour) in the image is also a conic. We represent this occluding contour by

$$\mathbf{u}^t \mathbf{Q} \mathbf{u} = 0 \tag{4}$$

where $\mathbf{u} = (u, v, 1)^t$ is any point on the occluding contour in the image of the first camera, and $\mathbf{Q}$ is a 3×3 symmetric matrix. We have proved, in our previous paper [8], the following equation relating the representation of the ellipsoid in 3D space and its occluding contour.

$$kQ = A_{33} - \frac{1}{a_{44}} A_4 A_4^t \quad (5)$$

where A_{33} and A_4 are defined by the matrix A (see Eq. (3)), and Q is a known matrix estimated from the image. k is an unknown scale factor. Since A is a symmetric 4×4 matrix and since A and tA (t is any non-zero scale factor) represent the same ellipsoid, A is defined by 9 parameters. Considering the unknown scale factor k, we have 10 unknown parameters in Eq. (5).

Eq. (5) is a basic constraint relating the surface A and its image contour. It is interesting to notice that the occluding contour contains not only the information of its corresponding rim but also the global information of the ellipsoid.

It can be seen from Eq. (5) that

1. The matrices in both sides of Eq. (5) are 3×3 symmetric, and thus it contain at most 6 independent equations. Since we have 10 unknowns, it is not possible to recover the ellipsoid from a single image.
2. All the equations contained in Eq. 5 are non-linear.

In the next subsection, we will show that the non-linear relation can become linear.

3.2 Linear Relationship and Multiple Views

It is proved that Eq. (5) can be rewritten as

$$kQ^{-1} = (A^{-1})_{33} \quad (6)$$

where $(A^{-1})_{33}$ indicates the upper left 3×3 submatrix of A^{-1}.

For the reason of limited space of this paper, we neglect the proof of above relation, which can be found in [10]. The equivalence of Eq. (5) and Eq. (6) is very interesting, because it can be seen that although the relationship between Q and A is non-linear, but between Q^{-1} and A^{-1} are linearly related!

We must generalize Eq. (6) to the case of multiple views. As mentioned in section (3.1), Eq. (6) *holds only for the first camera* when assuming its projective matrix M_1 is equal to $(I|0)$. In general we have the following Proposition.

Proposition 1 If an ellipsoid in 3D space is represented by $x^t A x = 0$, if M is the projective matrix, and if the occluding contour of the ellipsoid in the image is represented by Q then we have

$$kQ^{-1} = MA^{-1}M^t \quad (7)$$

For the reason of limited space of this paper, we neglect the proof of above proposition, which can be found in [10].

It is interesting to notice that, given Q and M, by considering the entries of A^{-1} as variables, Eq. (7) is linear. Since the matrices on both sides are 3×3 symmetric. Eq. (7) consists of six linear equations of the variables k and 9 entries of A^{-1} (One of its entries can be assumed to be one). We have a total of 10 variables. In the next section, we will prove that three views are necessary in order to solve for A.

4 Ellipsoid Reconstruction from Multiple Views

4.1 Multiple views constraints

Suppose we have n cameras. Then Eq. (7) can be rewritten as

$$k_i Q_i^{-1} = M_i B M_i^{-1} \qquad i = 1 \sim n \quad (8)$$

where $B = A^{-1}$, Q_i ($i = 1 \sim n$) is the occluding contour estimated from i^{th} image, and M_i ($i = 1 \sim n$) is the projective matrix of the i^{th} camera. Our purpose is to solve for B given Q_i and M_i ($i = 1 \sim n$). We denote

$$A^{-1} = B = \begin{pmatrix} B_{33} & b_4 \\ b_4^t & 1 \end{pmatrix} \quad (9)$$

where B_{33} is the 3×3 upper left submatrix and B_4 is a 3D vector. Note we can assume the last element of B is equal to 1 because B can only be determined up to a scale factor.

By denoting

$$M_i = (\bar{M}_i | m_i) \qquad i = 1 \sim n \quad (10)$$

$$\bar{Q}_i = \bar{M}_i^{-1} Q_i^{-1} (\bar{M}_i^t)^{-1} \quad (11)$$

$$\bar{m}_i = \bar{M}_i^{-1} m_i \quad (12)$$

where $\bar{M}_i$ is the left 3×3 submatrix of M_i, and by substituting Eq. (9) into Eq. (8) we obtain

$$k_i \bar{Q}_i = B_{33} + B_4 \bar{m}_i^t + \bar{m}_i B_4^t + \bar{m}_i \bar{m}_i^t \quad (13)$$

4.2 Two views are not enough to recover an ellipsoid

We assume the first camera matrix M_1 is equal to $(I|0)$. Then from Eqs. (11) (12), we have $\bar{Q}_1 = Q_1^{-1}$ and $\bar{m}_1 = m_1 = 0$. So when $i = 1$, Eq. (13) becomes

$$k_1 \bar{Q}_1 = B_{33} \quad (14)$$

From Eq. (14) and an equation obtained from Eq. (13) for a second camera (i = 2), we have 2×6 linear equations and 11 unknowns (k_1, k_2 and 9 entries of B). But we will show that these equations are not all independent.

By subtracting Eq. (14) from Eq. (13) (i=2) to eliminate $\mathbf{B}_{33}$, we obtain

$$k_2\bar{\mathbf{Q}}_2 - k_1\bar{\mathbf{Q}}_1 = \mathbf{B}_4\bar{\mathbf{m}}_2^t + \bar{\mathbf{m}}_2\mathbf{B}_4^t + \bar{\mathbf{m}}_2\bar{\mathbf{m}}_2^t \qquad (15)$$

Or equivalently

$$k_2\bar{\mathbf{Q}}_2 - k_1\bar{\mathbf{Q}}_1 - (\mathbf{B}_4 + \frac{1}{2}\bar{\mathbf{m}}_2)\bar{\mathbf{m}}_2^t - \bar{\mathbf{m}}_2(\mathbf{B}_4 + \frac{1}{2}\bar{\mathbf{m}}_2)^t = 0 \qquad (16)$$

We see that Eq. (16) consists of six *homogeneous* linear equations with 5 unknowns (k_1, k_2 and three elements of $\mathbf{B}_4$). Since these unknown variables are not all equal to zero (k_1 and k_2 are not equal to zero from Eq. (5)), the number of independent equations included in Eq. (16) is 4 (more details can be found in [10]), or in other words, the number of solutions of k_i and $\mathbf{B}_4$ is infinite.

4.3 Reconstruction from three views

If we have three images, the reconstruction algorithm can be described as follows:

1. Given three occluding contours $\mathbf{Q}_i$ respectively in three images and three projective matrices $\mathbf{M}_i$, we obtain three matrix equations from Eq. (13) by taking the index i = 1, 2, 3.

2. As described in the previous subsection, we can eliminate $\mathbf{B}_{33}$ to obtain Eq. (15) and the equation

$$k_3\bar{\mathbf{Q}}_3 - k_1\bar{\mathbf{Q}}_1 = \mathbf{B}_4\bar{\mathbf{m}}_3^t + \bar{\mathbf{m}}_3\mathbf{B}_4^t + \bar{\mathbf{m}}_3\bar{\mathbf{m}}_2^t \qquad (17)$$

   Then we have 8 linear equations included in Eqs. (15) (17). We use the linear least-square method to solve for $\mathbf{B}_4, k_1, k_2, k_3$ from these linear equations. It can be easily shown theoretically and experimentally that the solution is unique if the three cameras are located in different positions.

3. By substituting $\mathbf{B}_4$ and k_2 into Eq. (13) (taking index i=2) we can solve for $\mathbf{B}_{33}$

4. Using Eq. (9) we obtain $\mathbf{A}$, which is the representation of the ellipsoid.

4.4 Establishing the correspondence

As we mentioned in the above section, Eq. (15) (or Eq. (17)) contains six linear equations but only 4 are independent, which means there exists intrinsic relationship between the parameters in these equation. It is shown that, by eliminating k_i ($i = 1, 2, 3$) and $\mathbf{B}_4$ from Eqs. (16) (17), we can obtain (more details can be found in [10]) 4 quantities which remain invariant *if the occluding contours in different images correspond to the same ellipsoid*:

$$\lambda_1 = \frac{\triangle_{12}\bar{\mathbf{Q}}_i}{\triangle_{13}\bar{\mathbf{Q}}_i}, \qquad \lambda_2 = \frac{\triangle_{13}\bar{\mathbf{Q}}_i}{\triangle_{23}\bar{\mathbf{Q}}_i} \qquad i = 1, 2 \qquad (18)$$

$$\lambda_3 = \frac{\triangle'_{12}\bar{\mathbf{Q}}_i}{\triangle'_{13}\bar{\mathbf{Q}}_i}, \qquad \lambda_4 = \frac{\triangle'_{12}\bar{\mathbf{Q}}_i}{\triangle'_{23}\bar{\mathbf{Q}}_i} \qquad i = 1, 3 \qquad (19)$$

where $\triangle_{ij}\bar{\mathbf{Q}}_k$ and $\triangle'_{ij}\bar{\mathbf{Q}}_k$ are defined by the projective matrices. Thus we can use these invariants to establish the correspondence of the occluding contours.

5 Experiments

We have successfully verified our method using both synthetic and real image data. In simulation, we generate images of two ellipsoids using computer graphics techniques. The ellipsoids are located at a distance of about 1.5m from the cameras. We generated 3 images to simulate three cameras in 3 different viewpoints. The three cameras are positioned such that the baseline lengths of each pair of cameras are all identical. It is shown that the parameters of both the sizes and the poses of the reconstructed ellipsoid are very close to the actual data. For the reason of the limited space, we can not show these results in this paper.In the real image experiment, we take three images of two spheres at three viewpoints by moving a CCD camera. The distance between the camera and the spheres is about 1m. Since we can not know the actual position and orientation of the spheres with respect to the cameras, we show in table 1 only the actual size of the spheres and that of the sphere reconstructed. Table 2 and 3 shows the invariants estimated respectively from the first-second and first-third image pair.

	Ball 1	Ball 2
Actual Radius	110	90
Computed Radius	107.35	90.8

Table 1. Reconstruction results of the two balls

Invariants	First Image		Second Image	
	Ball 1	Ball 2	Ball 1	Ball 2
λ_1	2.27	15.05	2.33	15.03
λ_2	0.0019	0.00039	0.0019	0.00039

Table 2. Invariants λ_1 and λ_2 for the two balls

Invariants	First Image		Third Image	
	Ball 1	Ball 2	Ball 1	Ball 2
λ_3	7.08	13.93	7.20	13.92
λ_4	0.00058	0.00042	0.00058	0.00042

Table 3. Invariantsλ_3 and λ_4 for the two balls

6 Conclusion

We have presented a method to reconstruct an ellipsoid from its three perspective views. All the computations in the algorithm are linear. The occluding contour of an ellipsoid is a conic. It is well known that a conic in the image can be extracted from at least 5 points. Thus even when it is partially occluded, a conic can be extracted from the image. As a result, using our method we can globally reconstruct the ellipsoid even when its surface is partially occluded. Establishing the correspondence of the primitives in different images is a serious fundamental problem for the point or line based reconstruction techniques from multiple views. We have shown that conics contain more information than points and lines and their correspondence is implicitly determined by their parameters.

References

[1] O.D.Faugeras and G.Toscani, "The Calibration Problem for Stereo", Proc. IEEE Conference on Computer Vision and pattern Recognition, pp15-20, 1986.

[2] O.D.Faugeras, "Waht Can Be Seen in Three-dimensions With an Uncalibrated Stereo Rig?", in Proc. of ECCV, pp563-578, Italy, June 1992.

[3] P. Giblin and R. Weiss, "Reconstruction of Surfaces from Profiles", Proc. ICCV, London, U.K, pp136-144, 1987.

[4] R. Hartley, "Projective Reconstruction and Invariants from Multiple Images", IEEE Trans. on PAMI, PAMT-16, No 10, pp 1036-1040, 1994

[5] R. Hartley, "A Linear Method for Reconstruction from Lines and Points", Proc. of ICCV'95, pp882-887, Boston, USA, June, 1995.

[6] W.C. Karl, G.C. Verghese and A.S. Willsky, "Reconstructing Ellipsoids from projections", CVGIP: Graphical Models and Image Processing, Vol.56, No.2, March, pp124-139, 1994.

[7] D.J. Kriegman and J. Ponce, "On Recognition and Positioning Curved 3D objects From Image Contours,", IEEE Trans. PAMI-12(12), pp1127-1137, 1990.

[8] S.D. Ma, "Conics-Based Stereo, Motion Estimation, and Pose Determination", Int. Journal of Computer Vision, Vol. 10:1, pp7-25, 1993.

[9] S.D.MA and X.Chen, "Quadric Surface Reconstruction From Its Occluding Contours", Proc. of ICPR'94, Isreal, 1994.

[10] S.D.MA and L.LI, "Ellipsoid Reconstruction From its Perspective Views", Technical Report, National Pattern Recognition Lab., Institute of Automation, Chinese Academy of Sciences, Beijing, China.

[11] R.Mohr and E.Arbogast, "It can be done without camera calibration", RR 805-1 IMAG 106 LIFIA, France, fev. 1990.

[12] A. Shashua and M. Werman, "Trilinearity of Three Perspective Views and Its Associated Tensors", Proc. of ICCV'95, Boston, USA, pp 920-925, June, 1995.

[13] M.E.Spetsakis and J.Y.Aloimonos, "A Unified Theory of Structure from Motion", Proc. of DARPA IU Workshop, 1990.

[14] R.Y.Tsai, "An Efficient and Accurate Camera Calibration Technique For 3D Machine Vision", Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp364-374, 1986

[15] R.Vaillant and O. Faugeras, "Using Extremal Boundaries for 3D Object Modeling", IEEE Trans. PAMI-14(2), pp157-173, 1992.

[16] G.Q.Wei and S.D.MA, "Implicit and Explicit Camera Calibration: Theory and Experiments", IEEE Trans. on PAMI, PAMI-16, No.5, May, 1994 1994, Isreal.

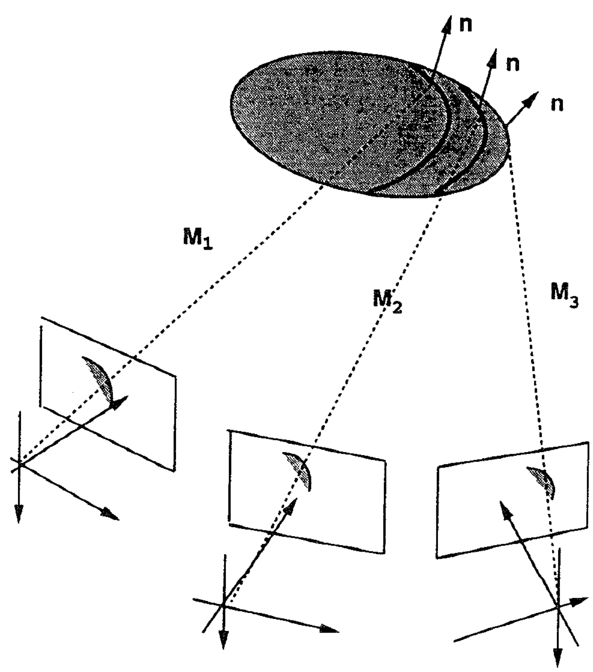

Fig 1 Rims on an ellipsoid and its images