Reconstruction from Uncalibrated Affine Silhouettes

Paul McIlroy
http://mi.eng.cam.ac.uk/~pmm33
Tom Drummond
http://mi.eng.cam.ac.uk/~twd20

Machine Intelligence Laboratory
Department of Engineering
University of Cambridge
Cambridge, UK

In this paper we address the problem of model building in environments we do not control. An example problem is the reconstruction of aircraft geometry from images obtained from the ground. We show that using six or more unconstrained scaled orthographic silhouette views it is possible to recover the geometry of the aircraft and the motion between the camera views using only the two outer epipolar tangency constraints in each image.

The silhouette outline is comprised of a subset of the set of *apparent contours* [2] which are the projections of the *contour generators* which divide the visible part from the occluded part of the object surface. The contour generators are functions of both the surface geometry and viewpoint. Contour generators corresponding to two distinct viewpoints of the same surface are, in general, different space curves. These space curves intersect at *frontier points*, which provide a pair of point correspondences between the two viewpoints [1, 3].

The epipolar lines are parallel for an affine camera model so only the relative rotation between the viewpoints is required to determine the angle of the epipolar lines. The two epipolar tangent points at which the epipolar lines at this angle first touch the silhouette outline are the projections of the two outer frontier points. These point correspondences provide an epipolar tangency constraint on the camera motion between the viewpoint pair.

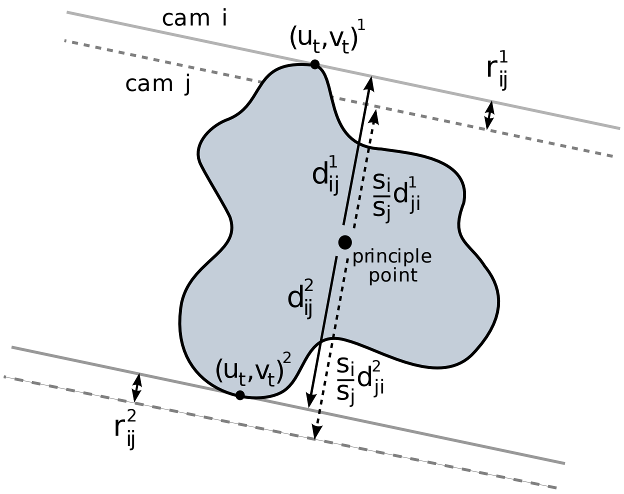


Figure 1: Parallel epipolar lines and residual errors for an affine camera

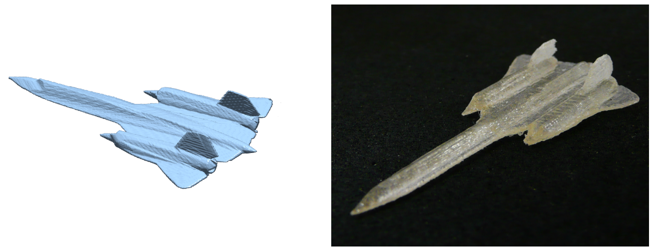
Each scaled orthographic camera has six parameters to be determined, three for the rotation relative to the world, two for the image coordinates of the principal point and one for the scale. Thus there are a total of $6N$ parameters in the N-camera system. The world reference frame may be chosen arbitrarily such that the first camera's rotation matrix is the identity, the principal point is in the centre of the image and it has unit scale. This constrains the world origin to lie on the optical axis of the first camera. By fixing one of the principal point coordinates in a second camera we fully specify the 3D position of the world origin. We have arbitrarily chosen seven of the $6N$ camera parameters in the system to define the world coordinate frame, leaving $(6N-7)$ free parameters to be solved using the constraints from the outer two epipolar tangents.

There are $\frac{1}{2}N(N-1)$ camera pairs in the system with each camera pair providing two unique constraints from the outer epipolar tangents, so for a unique solution the number of cameras, N, must satisfy

$$(6N-7) \geq N(N-1). \tag{1}$$

It is therefore possible to recover motion using the epipolar tangency constraint from a minimum of six silhouette views.

We use the following parameterization for the scaled orthographic camera model. $\mathbf{R}$ is a 3×3 rotation matrix, (u_0, v_0) is the principal point



(a) (b)

Figure 2: Reconstructed model: (a) rendered; (b) 3D printer output

and s is a scale parameter. The mapping between a point $\mathbf{X}$ in world coordinates and the same point $\mathbf{X}_c$ expressed in camera coordinates is given by

$$\mathbf{X}_c = \mathbf{R}_c\mathbf{X}. \tag{2}$$

The scaled orthographic projection of $\mathbf{X}_c$ may be expressed as

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} s & 0 & 0 \\ 0 & s & 0 \end{bmatrix} \mathbf{X}_c + \begin{bmatrix} u_0 \\ v_0 \end{bmatrix}. \tag{3}$$

The camera motion is obtained by an optimization process which minimises a cost function based on the epipolar tangents,

$$E = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sum_{k=1}^{2} (r_{ij}^k)^2, \tag{4}$$

where r_{ij}^k is the residual error between the k^{th} epipolar tangent point in camera i and the epipolar line from camera j as illustrated in Figure 1. The residual error may be expressed in terms of d_{ij}, the perpendicular distance from principal point to epipolar tangent line in camera i , and d_{ji}, the corresponding distance in camera j, such that

$$r_{ij} = d_{ij} - \frac{s_i}{s_j} d_{ji}, \tag{5}$$

where s_i and s_j are the scale parameters of cameras i and j respectively. The perpendicular distance from the principal point (u_0, v_0) to the epipolar tangent line through tangent point (u_t, v_t) is given by

$$d_{ij} = \frac{1}{|\sin\rho_{ij}|} \begin{bmatrix} 0 & u_t - u_0 & 0 \\ -(v_t - v_0) & 0 & 0 \end{bmatrix} \mathbf{R}_i \mathbf{R}_j^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \tag{6}$$

where ρ_{ij} is the angle between the two cameras' optical axes and

$$|\sin\rho_{ij}| = \left\| \mathbf{R}_i^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \times \mathbf{R}_j^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\|. \tag{7}$$

This camera parameterization permits a closed form solution for the partial derivatives required in the optimization.

In an experiment designed to investigate the accuracy of the recovered camera configuration, a model aircraft was rotated through 90° with images captured at increments of 10°. The rotation angle between cameras was determined for the resulting camera configuration and the rms error for each interval was 0.27°. An error of this magnitude is expected in the manual positioning of the object. The rms error for the same process repeated with synthetic data was 0.02°. Figure 2 shows the final reconstructed model and the output from a 3D printer.

[1] K. Åström, R. Cipolla, and P. Giblin. Generalised epipolar constraints. *International Journal of Computer Vision*, 33(1):51–72, 1999.

[2] R. Cipolla and A. Blake. Surface shape from the deformation of apparent contours. *International Journal of Computer Vision*, 9(2):83–112, 1992.

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