The Epipolar Parametrization*

Richard S. Weiss
Dept. of Computer Science
University of Massachusetts, Amherst, MA 01003-4610

Abstract

The epipolar parametrization arises naturally in the reconstruction of surfaces from profiles with known camera motion. This is a special case of a local parametric representation which is a mesh. Local parametric representations can be combined into a global one by computing the transformation in parameter space on the overlap of the patches. This paper also discusses the applicability of this type of model to problems in grasp configuration determination and pose determination.

1 Introduction

Different reconstruction algorithms may produce different types of information about a surface. It is important to have a representation that bridges the gap between the modeling and data collection on the one hand and the application on the other. This paper focusses mostly on the model acquisition but also touches on the requirements for some tasks such as pose determination and grasp configuration planning. An assumption made here is that modeling objects will be automated or semi-automated and will directly use images. The modeling process can involve stereo on viewpoint independent features (texture or edges), laser range sensors, or generalized stereo on profiles (reconstruction from profiles with known camera motion).

There are two types of mathematical models for surfaces[1]. One is the *implicit* representation given by an equation $F(X, Y, Z) = 0$, e.g. superquadrics and algebraic surfaces. F is also called an *inside-outside function.* This is convenient as a global representation if the function F has a nice form, e.g. low degree polynomial or superquadric, and fits the surface well. However, if this is not the case, the surface will need to be decomposed into parts such that each part can be represented both simply and with desired accuracy. The second type is a *parametric* surface representation, which is a map from $D \subseteq R^2 \rightarrow R^3$, i.e. $(X(u,v), Y(u,v), Z(u,v))$, e.g. generalized cylinders and splines. Superquadrics also have a simple representation as parametric surface patches. The set D together with the map is called a chart or coordinate *patch* because it can be thought of as mapping a u, v-Cartesian coordinate system onto the surface. Parametric representations are more limited in the sense that a surface which might be represented globally by a single simple function, e.g. $X^2 + Y^2 + Z^2 = 1$, may require several parametric patches. Note that any parametric patch representation can be converted to an implicit representation, and in the case of algebraic surfaces this can be done in closed form.

Since 3D data for a surface often comes in the form of discrete points or discrete points that are connected in some way, the concepts of implicit and parametric representations need to take this into account. Sampling a surface as a set of 3D points which are connected by arcs produces a graph on the surface of the object. Triangulations and meshes are special cases of graphs. One can think of a triangulation as being a polyhedral representation of the surface. Every triangle becomes a map from a subset of the plane to the surface. For this paper, a mesh is defined as a graph of degree four, so that there are two families of polygonal curves that are transversal. One can also decompose a graph on the object surface into planar sub-

*The author would like to acknowledge the support of ARPA and TACOM under contract DAAE07-91-C-R035 and NSF under grants IRI-920892 and IRI-9116297.

[1] I am ignoring CSG and volumetric models

graphs that overlap, and if one maps each of these subgraphs to a region in the plane, then this would correspond to choosing a parametrization for each patch. Note that in this case the nodes in the different subgraphs correspond when there is overlap. In general, if one starts with a set of graphs that cover overlapping patches on the surface, one may have to subdivide each graph to get nodes that match.

I would postulate that a graph representation is suitable for any of the standard techniques for collecting 3D data. The simplex angle image, which can be used for recognition is a derivative of a degree three graph representation that satisfies certain regularity conditions [12]. In addition, sensor fusion is easily done with a graph representation because one needs to model uncertainty at the resolution of the data, otherwise fitting data to an implicit (inside-outside) function makes certain assumptions about the shape of the object and may weight data points inappropriately. For example, when fitting a superellipsoid to data there are a number of issues. First, using the evaluation of the inside-outside function is different from using geometric distance. In general, one can approximate geometric distance based on the gradient of the inside-outside function [18, 21], but even then different points will have different covariances which need to be taken into account. However, the most important problem related to representation is that the family of functions itself affects the result, e.g. finding the closest superellipsoid to the data may be significantly different from finding the closest polynomial surface of fixed degree. Note that if one actually has some additional constraints on the surface, e.g. one knows that it is a cylinder or surface of revolution, then such models are entirely appropriate and the covariance of each measurement should be used in fitting such a parametrized family of models to the data. Of course, fitting an implicit function model can be done from a graph representation, but it would not be necessary to go through that intermediate step, and the representation of uncertainty would be on the parameter level rather than the data level.

The important issues for representation and modeling are:

- Representing accuracy of the model, which is determined by sensor error and sampling density.
- Determining what parts of the surface have not been sampled, i.e. the boundaries of observation.
- Facility of constructing the model from image data.

These issues are addressed in section 3 in the case of the epipolar representation. The epipolar representation or epipolar parametrization arises naturally in the reconstruction of surfaces from a sequence of profiles where the motion of the observer is known or can be recovered from features visible in the image sequence. This leads to a set of local parametric models on patches that must be combined in order to represent global information. It may happen that only part of the surface is observable, and even then it may require several trajectories by the camera to reconstruct that part. Kutulakos and Dyer [15] have developed strategies for moving the camera or the object so that a maximal subset of the surface is covered by such a set of patches.

2 Epipolar parametrization

Given a smooth surface M and a curve $c(t)$ of camera centers, we have, for each t, a *critical set* or *contour generator* Σ_t on M consisting of those points $\mathbf{r}$ where the 'visual ray' or viewline from $c(t)$ to $\mathbf{r}$ is tangent to M (see Fig 1). The *epipolar plane* is the plane spanned by this ray and the tangent to the curve $\mathbf{c}$ of centers. In practice the epipolar plane is computed from a visual ray to one camera center and the baseline connecting the two centers [5, p.170].

On the surface M an *epipolar curve* through $\mathbf{r}$ is defined as one whose tangent is along the visual ray as shown in Fig. 1. As $\mathbf{c}$ moves with time, the visual ray slips along the epipolar curve. Note that this is almost never the intersection of the epipolar plane with the surface, unless the camera moves in a straight line. In general, the critical sets and epipolar curves make a coordinate grid on M: a local parametrization $r(t, u)$ can be found in which the critical sets are given by t =constant and the epipolar curves by u = constant. In practice one has a discrete set of views and a finite sampling of each profile, which is the image of the critical set in a viewplane or viewsphere. This can result in a mesh where the nodes are the intersections of the two families of curves. It is the 'epipolar parametrization' of M which is used in [1, 2] to reconstruct

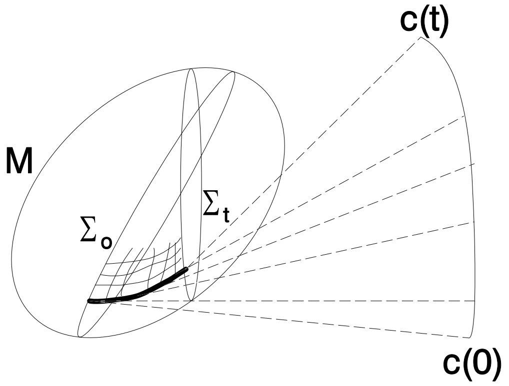

Figure 1: A surface M and segment of camera path from $\mathbf{c}(0)$ to $\mathbf{c}(t)$. Also shown are the two corresponding critical sets Σ_0, Σ_t, a segment of epipolar curve (drawn heavily), viewlines (dashed) tangent to the epipolar curve, and a local coordinate grid of critical sets and epipolar curves.

M from its profiles, although b-splines were used to represent critical sets. In [20] it is shown that reconstruction from the epipolar parametrization is readily transformable into an optimal estimation problem. The epipolar parametrization also has another very interesting property: the viewlines associated with points $r(t,u)$ and $r(t+\delta t, u)$ of M will (being lines in space) generally not intersect. However, for the epipolar parametrization, the point at which these lines come *closest* to one another is (as $\delta t \to 0$) on the surface M. So this parametrization is the best one to choose if one uses the pseudo-intersection of viewlines as an approximation to the surface. In addition, the best trajectories for minimizing these distances are planar trajectories. This leads to strategies that use a sequence of planar trajectories each of which may only produce a model of a small region on the surface. The boundaries of these patches, which depend on the camera trajectories, have been characterized in [10].

There are basically two types of situation which occur at the boundaries of an epipolar patch. The first type is called the *frontier*, where the epipolar plane becomes tangent to M. It is precisely at such points that the epipolar curve becomes singular, and the epipolar parametrization breaks down. To examine the situation at the frontier one can use the 'spatio-temporal surface' as in [9, 10]. The second type of situation occurs in the case of occlusion. Occlusion may happen in two ways. First, the surface normal may turn away from the camera. This event is typically a cusp and is characterized by the fact that the epipolar curve and critical set become tangent. In the image, the profile is seen as ending at a point, and the locus of points on the surface which project to such end points is called the line of cusps. Another type of occlusion occurs at T-junctions. In this case, part of the critical set is obscured by another part of the surface, even though the normal is pointing toward the camera. These points are characterized as the distal points of contact of bitangent viewing rays. Taken together, the frontier and the *natural boundary,* which is the set of points of occlusion, form the boundary of each epipolar patch. A more complete description will appear in [10], and some of the results have appeared in [9].

To summarize those results, the (local) epipolar parametrization of M has a boundary when any of the following occur:

(i) The critical sets form an envelope on M (frontier points).

(ii) The critical set and epipolar curve on M are smooth and tangent to one another. The profile has a singularity which is a cusp or has higher order.

(iii) The viewing ray is tangent to the surface at a point closer to the camera. (T-junction points).

(iv) The critical set on M is singular (having an isolated point or a crossing).

3 The Epipolar parametrization as a representation

One of the issues is how to represent the accuracy of the model in such a way that incorporates both accuracy of the depth estimates and the sampling density of measurements. As noted above, there are criteria in the process of reconstruction from profiles for detecting the boundaries of the reconstruction process. Therefore, it is possible to keep track of parts of the surface that have not been observed. As for the accuracy of depth, the approach of Szeliski and Weiss [20] uses linear smoothing and explicitly models the covariance of the position and curvature at each point in the mesh.

It is clear that local parametric surface patches can be easily produced from reconstruction from profiles. In addition, information about curvature is recovered. One of the important issues is how the parts are combined into a global representation. It is important to do this because combining observation data from different views improves accuracy and is necessary for a consistent model, e.g. the profile segments produced from different patches should agree so that it won't matter which patch is used. This is potentially a difficult problem. If feature points are available on the surface, then they can be used to identify corresponding points in different views. Otherwise, one might try to find the rotation and translation that brings two patches into alignment on maximal subsets. In general this might be difficult, but for the epipolar parametrization there is additonal geometric information available, for example at the frontier, where epipolar curves meet in cusps.

4 Applications: representation of errors

A critical function of vision in navigation, recognition, or enhanced reality systems is to determine the position and orientation of an object relative to the viewer. The pose (i.e. position and orientation) of an object has six parameters: three rotational and three translational. Most pose algorithms take as input an initial "guess" of the approximate object pose and correspondences between a set of projected model features and 2D image features. Chen, Stockman, *et al.* [3] computed pose from profiles using curvature information. One would like to know how closely the profile of the generated view matches the actual view. A graph representation which also stores curvature information is suitable for this algorithm, and accuracy information in the form of a covariance matrix at each node can be used to predict the accuracy of the profiles. A global figure of merit, which could be used to evaluate a model is the maximum over all views of the average distance between the predicted profile and the actual profile.

For grasping, there are a couple of ways to use geometric models. One is finding stable grasp configurations. A stable grasp configuration is given by the null space or force closure equation:

$$\sum_{i=0}^{d} w_i = 0 \tag{1}$$

where d is the number of contacts and w_i is the wrench due to contact i. This expression can be minimized and evaluated at any set of d points for which the normals and friction cones are given. The most time-consuming aspect of this process is finding zeros of this expression. An iterative search is possible and requires a direction for moving the contacts so that the magnitude of (1) is decreased. A local parametric representation such as the epipolar representation, which captures curvature at each node satisfies these requirements.

Another way in which a geometric model can be used is in planning the motion for achieving the grasp. This task requires a prediction of where the manipulator will make contact with the surface if it approaches along a specified ray. From this standpoint, the question that must be answered is where given ray will intersect the surface. This leads again to a representation of accuracy of the surface that is not just a value at each point but a matrix representing the error ellipsoid.

5 Conclusion

This paper shows how a reconstruction algorithm and graph representation fit together to represent the information needed to solve some specific applications. The graph representation is a discrete form of the epipolar parametrization that results from the reconstruction from profiles. It meets the informational requirements of pose determination and stable grasp formation. Issues of speed have not been addressed and may require auxiliary representations. The issues of modeling accuracy and determining the boundary between observed and unobserved parts of the surface are also addressed.
I

References

[1] A.Blake and R.Cipolla, 'Robust estimation of surface curvature from deformations of apparent contours', *Image and Vision Computing* 9 (1991), 107–112.

[2] R. Cipolla and A. Blake, 'Surface shape from the deformation of apparent contours', *Internat. J. of Computer Vision* 9 (1992), 83-112.

[3] J.-L. Chen, G.C. Stockman, and K. Rao, "Recovering and Tracking Pose of curved 3D Objects from 2D IMages," cvpr93 pp. 233-239.

[4] A. Blake and M. Taylor, "Planning planar grasps of smooth contours," *IEEE Proc. Int. Conf. on Robotics and Automation.* IEEE Computer Society Press: Los Alamitos. 1993, pp. II-834-839.

[5] O.D. Faugeras, *Three Dimensional Computer Vision*, MIT Press, 1993.

[6] O.D. Faugeras. 'On the motion of 3-D curves and its relation to optic flow,' *Proc 1st ECCV,* pp 107–117, Springer Verlag, April 1990.

[7] P.J. Giblin, J.E. Rycroft and F.E. Pollick 'Recovery of an unknown axis of rotation from the profiles of a rotating surface', *J.Optical Soc. America* 1994

[8] P.J.Giblin and R.S.Weiss, 'Reconstruction of surface from profiles', *Proc. First Internat Conf. on Computer Vision*, London, 1987, Computer Society of the IEEE, pp. 136–144.

[9] P.J.Giblin and R.S.Weiss, 'Epipolar fields on surfaces', Proceedings of ECCV, Stockholm 1994, Springer Lecture Notes on Computer Science, Ed. J.-O.Eklundh, 800 (1994) 14-23.

[10] P.J.Giblin and R.S.Weiss, 'Epipolar curves on surfaces', to appear in *Image and Vision Computing.*

[11] R. Grupen and R. Weiss. "Force Domain Models for Multifingered Grasp Control." In *Proc. Int. Conf. on Robotics and Automation*, IEEE Computer Society Press: Los Alamitos. 1991.

[12] K. Higuchi, H. Delingette, M. Hebert, K. Ikeuchi, "Merging Multiple views using a spherical representation." *Proc. Second CAD-Based Vision Workshop.* IEEE Computer Society Press: Los Alamitos. 1994, pp. 124-131.

[13] T. Joshi, N. Ahuja, J. Ponce, *Structure and Motion Estimation from Dynamic Silhouttes,* University of Illinois Tech. Rpt. UIUC-BI-AI-RCV-94-01, 1994.

[14] K.N.Kutulakos and C.R.Dyer, 'Occluding contour detection using affine invariants and purposive viewpoint control', *IEEE Conf. on Computer Vision and Pattern Recognition*, Seattle, Washington (1994) 323-330.

[15] K.N.Kutulakos and C.R.Dyer, 'Global surface reconstruction by purposive control of observation motion', *IEEE Conf. on Computer Vision and Pattern Recognition*, Seattle, Washington (1994), 331-338.

[16] J.R.Steenstrom and C.I.Connolly, 'Constructing object models from multiple images', *Int. J. of Computer Vision*, 9 (1992), 185-212.

[17] J. Ponce, S. Sullivan, J.-D. Boissonat, and J.-P. Merlet, "On characterizing and computing three- and four-finger force closure grasps of polyhedral objects," *IEEE Proc. Int. Conf. on Robotics and Automation.* IEEE Computer Society Press: Los Alamitos. 1993, pp. II-821-827.

[18] S. Sullivan, L. Sandford, and J. Ponce, "On Using Geometric Distance Fits to Estimate 3D Object Shape, Pose, and Deformation from Range, CT, and Video Images," *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* New York, June 1993, pp. 110-115.

[19] R.Szeliski, 'Rapid octree construction from image sequences,' *CVGIP: Image Understanding* 58 (1993), 23-32.

[20] R. Szeliski and R. Weiss, 'Robust shape recovery from occluding contours using a linear smoother,' *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* New York, June 1993, pp.666-7.

[21] G.Taubin, R. Bolle, and D.B. Cooper, "Representing and comparing shapes using shape polynomials," *Proc. IEEE Conference on Computer Vision and Pattern Recognition,* 1989. pp. 510-516.

[22] R. Vaillant and O.D. Faugeras, 'Using extremal boundaries for 3-D object modeling,' *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 14, No. 2, 1992, pp. 157–173.