Finding Motion Parameters from Spherical Motion Fields
(Or the Advantages of Having Eyes in the Back of Your Head)

R. C. Nelson and J. Aloimonos
Computer Vision Laboratory, University of Maryland, College Park, MD 20742, USA

Abstract. A theory is developed for determining the motion of an observer given the motion field over a full 360 degree image sphere. The method is based on the fact that for an observer translating without rotation, the projected circular motion field about any equator can be divided into disjoint semicircles of clockwise and counterclockwise flow, and on the observation that the effects of rotation decouple around the three equators defining the three principal axes of rotation. Since the effect of rotation is geometrical, the three rotational parameters can be determined independently by searching, in each case, for a rotational value for which the derotated equatorial motion field can be partitioned into 180 degree arcs of clockwise and counterclockwise flow. The direction of translation is also obtained from this analysis. This search is two dimensional in the motion parameters, and can be performed relatively efficiently. Because information is correlated over large distances, the method can be considered a pattern recognition rather than a numerical algorithm. The algorithm is shown to be robust and relatively insensitive to noise and to missing data. Both theoretical and empirical studies of the error sensitivity are presented. The theoretical analysis shows that for white noise of bounded magnitude $M$, the expected errors are at worst linearly proportional to $M$. Empirical tests demonstrate negligible error for perturbations of up to 20% in the input, and errors of less than 20% for perturbations of up to 200%.

I Introduction

The idea of obtaining information about the structure of the environment and observer motion from the two dimensional motion field is an old one. Many of the basic concepts were set forth in a non-mathematical fashion almost forty years ago by Gibson (1950). More recent work has provided a mathematical framework for many of these ideas (Koenderink and van Doorn 1975; Prazdny 1980; Longuet-Higgins and Prazdny 1984; Koenderink 1986). Development in robotics and computer vision have created an interest in obtaining exact quantitative solutions to the problem of determining shape and/or motion from image flow, and a considerable amount of theoretical work has been done in this connection. Solving the general problem under a planar perspective projection requires inverting a non-linear transformation and the mathematics involved are non-trivial. Several authors have obtained solutions to this general problem in various forms (Tsai and Huang 1984; Longuet-Higgins 1981). Additional work on extracting shape from motion under various constraints on the environment (e.g. assuming a representation in terms of planar or quadric surface patches) has been done by Kanatani (1984).

Most of these studies, however, have started with the assumption that detailed and accurate motion field information is available, and the solutions obtained tend to be unstable in the presence of even small amounts of noise. Tsai and Huang (1984) report 60% error for a 1% perturbation in the input in some instances using their method. The extraction from real image sequences of motion field information which is of sufficient accuracy to allow the use of these theoretical results has proven to be exceedingly difficult. Feature based methods require the solution of a correspondence problem, and differential methods must deal with the aperture problem using regularization or some other scheme. Neither problem has been satisfactorily resolved, and as a result, the mathematical developments have seen few practical applications (Ullman 1981).

A more qualitative approach was suggested by Prazdny (1981) who noted the obvious fact that the effects of rotational motion are equivalent to geometric
transformations of the motion field due to the translational component of the motion. Thus it is possible to undo the effects of rotation on the motion field without knowing anything about the three dimensional structure of the environment. Prazdny proposed determining the motion parameters by searching for a set of rotation parameters for which the derotated motion field satisfied the conditions for purely translational motion. Specifically, the flow was required to be radial with respect to a well defined focus of expansion or contraction. The proposal is based on the intuitively appealing notion that the curved projection paths generated by rotational motion are qualitatively different from the straight radial paths generated by translation, and that consequently, radial flow from a focus of expansion unambiguously indicates purely translational motion. The basic idea is slightly flawed by the fact that there are examples where a motion field due to motion with non-zero rotation is identical over the entire image plane to one produced by translation alone. This situation arises because for planar surfaces there are, in general, two possible interpretations for a given motion field (Tsai and Huang 1981). This would not in itself be a serious practical drawback, since real-world scenes generally contain non-coplanar points. However, the situation reflects the existence of a wide range of motions for which the motion field is approximately radial, and thus the method suffers from the same error sensitivity as other proposals.

Recently, it has been suggested that the general shape from motion problem may be essentially ill posed. Adiv (1985) argues that inherent near-ambiguities in the 3-D structure from motion problem may preclude full determination of the motion parameters in the presence of even slightly inaccurate data. The basic problem is that certain completely different observer motions produce motion fields which appear locally similar. The simplest example of this is the case where given a camera looking along the z axis, rotation about the x axis produces a flow which is similar to that produced by translation along the y axis. A close match can be extended over angular fields of view comparable to those of standard cameras. The extreme sensitivity to small errors in the motion field of some proposed solutions to the shape from motion problem is due to analogous effects.

It is notable, however, that although motion fields produced by different observer motions may match closely over restricted fields of view, they cannot do so globally, i.e., over the entire visual sphere. This is forbidden by the topological constraints on the possible motion fields. In the above example, if we project the global motion field onto a sphere and consider the flow around the circle corresponding to the intersection of the image sphere with the z–y plane, then a clockwise rotation about the x axis produces a constant counterclockwise image flow. A translation in the y direction, on the other hand, produces a flow which is clockwise over half the circle, and counterclockwise on the other half, no matter what the structure of the environment. The two flows cannot possibly match, even to first order, over the entire circle. The fact that the rotational contribution, relative to the translational flow, is opposite on opposite sides of a circle has been noted by Thompson in a proposal for “four-eyed vision” (Thompson and Kearney 1986). We show how the motion field over the full visual sphere can be used to determine the components of the observer motion unambiguously, even in the presence of noisy data.

II Determining Motion from Spherical Motion Fields

Let $S$ be the unit sphere with center point $c$. This will be called the image sphere. The spherical projection of a three-dimensional scene point $p$ is defined by the intersection of the ray from $c$ to $p$ with the image sphere. A spherical image is formed by the projection of scene points onto the image sphere. The spherical projection of a scene in the neighborhood of any point on the image sphere is locally similar to a perspective projection centered on that point, a consequence of the fact that the sphere is locally planar. Straight lines project to geodesics. The spherical projection has the advantages that all points on the image sphere are equivalent with respect to the center of projection, which is not true under perspective projection. This, combined with the fact that a spherical image contains information about all directions, makes it useful in applications where information must be correlated over wide angles.

Spherically projected motion fields were considered by Clocksin in an early paper on the determination of surface slope from image flow (Clocksin 1978), and have been used in the description of qualitative flow invariants by Koenderink (1986) among others, but in general, spherical projection has not been widely utilized. There are several reasons for this. One is the absence of a simple, uniform indexing system for points on the sphere analogous to $x$–$y$ indexing of pixels on the plane. A second reason is that physical imaging devices in common use produce an image that is well approximated by a perspective projection. A third reason is that there have been few applications which needed to correlate information over wide angles. Since our proposal is premised on the ability to analyze the image flow over a 360 degree field we use spherical projection as our basic representation.