$.vs20$

CHAPTER 1

Introduction

''One of the principal objects of theoretical research in any department of knowledge is to find the point of view from which the subject appears in its greatest simplicity''. -J.W. Gibbs

Humans effortlessly perceive the shape and motion of unfamiliar objects from their changing images. From a purely theoretical point of view this ability of the human visual system has intrigued perception psychologists for decades, and more recently computer vision scientists, for two reasons. First, valuable information is lost due to the projection of the three-dimensional scene onto the two-dimensional retina, and second, light as an intermediary of information transmission introduces certain ambiguities (e.g. the *aperture problem* discussed later) into the interpretation process. A primary goal of this research is to understand this phenomenon at the level of its *computational theory* (Marr, 1982). A computational study of this phenomenon is usually carried out in two stages. First is the *measurement of visual motion.* This involves computing the motion of image elements from the changing intensity pattern on the eye's retina. The second stage is the *interpretation* of this visual motion, i.e., to infer the three-dimensional shape and motion of objects given the visual motion. The first stage, the measurement of visual motion, has been intensively studied by a number of researchers (e.g.: Horn and Schunck, 1981; Hildreth, 1983; Waxman and Wohn, 1985). This book is concerned with the second stage, the interpretation of visual motion. It extends the previous work of Longuet-Higgins and Prazdny (1980) and Waxman and Ullman (1985) in this area in many ways. A general formulation of the problem and algorithms for the interpretation process are presented. An important goal of computer vision is to understand human vision from an information processing perspective. For this purpose, the task of vision can be divided into several stages, at least as a first approximation. Marr has suggested that the goal of the first stage of visual processing is to obtain descriptions of the physical properties of visible surfaces with respect to the viewer, properties such as distance, orientation, texture, and reflectance. This stage has been termed the *2 1/2 -D sketch* and the processes involved are called *early vision processes*. This early stage of processing is primarily bottom-up, relying on general knowledge about the world, but not on special high-level information about the scene to be analyzed. Computational studies and perceptual experiments (Marr and Poggio, 1977) suggest that early vision processes are generic ones that correspond to conceptually independent modules that can be studied, at least to a first order, in isolation. Examples of early vision processes are *edge detection* for finding sharp intensity changes, *stereopsis* for computing a depth-map from a stereo pair of images, *shape from shading*, *shape from texture*, *measurement of visual motion* and *interpretation of visual motion*. There is no proof yet that the paradigm for computational vision proposed by Marr and his collaborators is correct, but we adopt it in the belief that something similar should be true. In this framework we contend that a rigorous and thorough analysis of the individual visual modules is fundamental to understanding vision as an information processing task. For this reason, this study focuses on one module, the visual motion module. Existence of this module as an independent process in the human visual system is demonstrated in many perceptual studies (Wallach and O'Connell, 1953; Johansson, 1973, 1975; Ullman, 1979; see Figure 1). Another very important motivation for this research arises from its potential applications in machine vision systems. This work is directly relevant to autonomous land vehicle and aircraft navigation, robot manipulation of moving machine parts, and general machine vision systems. This thesis is a *computational study* of the problem of visual motion interpretation. Before proceeding further we define the two key terms *computational theory* and *computational approach*. These two terms were originally elucidated by Marr.

A formulation of a *computational theory* of an information processing task consists of four steps:

(i) identifying the input and the output entities,

(ii) specifying the relationship between the input and the output entities,

(iii) explicitly stating the conditions and assumptions under which the output entities are obtainable from the input entities using the relations specified in step two, and

(iv) proving that the output entities are indeed obtainable from the input entities under the conditions and assumptions stated in step three using the relations specified in step two.

A *computational approach* to an information processing task consists of two steps:

(i) constructing a representation for the input and the output entities and

(ii) developing an algorithm to transform input into output according to a computational theory of the task for the representation constructed in step one.

In addition to the above two steps, we would like to add a third requirement for any computational approach:

(iii) an implementation of steps one and two on some processing hardware, such as a computer system, and a demonstration of its correctness on a variety of cases.

Tsotsos (1987) has recently argued that a computational study of a task must also include a *complexity level analysis.*

For the problem of visual motion interpretation, the input is the visual motion field obtained from changing intensity patterns on the eye's retina and the output is a description of the changing scene. In this thesis a computational theory of this process is formulated and a computational approach is given in the precise sense defined above. In this approach the input representation is a set of *image parameters* and the output representation is a set of *scene parameters*. The image parameters describe the local visual motion in the image domain; these are the ''observables'' or ''knowns''. The scene parameters describe the local shape and motion of surfaces in the scene; these are the ''unknowns''. Therefore the goal of visual motion interpretation is to recover scene parameters from image parameters. Relations between the scene parameters and the image motion parameters are established and an algorithm is given to recover scene parameters from image motion parameters. The algorithm is implemented and tested on a computer system. The process of inferring the time-varying geometry of a scene from the corresponding visual motion is carried out locally, both in space and in time. The reason for this is that it is impractical to represent arbitrary shapes and motions of surfaces in the scene by a global parameterization scheme. Restriction to local analysis facilitates working in a smaller parameter space (but with less information, of course). Even this local analysis is assumed to be preceded by a detection of discontinuities in the visual motion corresponding to discontinuities in the geometry (distance, orientation, curvature, etc) and the motion (translation, rotation, acceleration, etc.) of surfaces in the scene. This can, in principle, be achieved by an appropriate modeling of the visual motion (e.g. requiring the motion field to be described by polynomials of fixed degree up to a preset tolerance). Having located such discontinuities in the motion field a local analysis is carried out in small image regions not containing these discontinuities to recover the structure and motion of the corresponding surfaces in the scene. A patching together of this local three-dimensional information is necessary to obtain a global description of the scene. The perception of motion from a monocular visual stimulus is investigated. A computational theory of the interpretation of visual motion caused by the projection of a moving surface is presented. The formulation of the theory is basically an extension of the earlier formulations of Longuet-Higgins and Prazdny (1980) and Waxman and Ullman (1985). Further, a computational approach, in the sense defined earlier, is given for the interpretation process. A major portion of this book is devoted to the study of visual motion resulting from rigid motion of objects. Here the problem is to determine the three-dimensional shape and rigid motion of surfaces from their image motion. Equations relating the local surface parameters (slopes and curvatures) and motion parameters (translation and rotation) to the spatial image motion parameters are derived. These equations are solved for planar surfaces and curved surfaces; in both cases the solution is derived in *closed form.* The solutions show that, in general, planar surfaces have two interpretations whereas for curved surfaces the interpretation is unique. Many theorems concerning the multiplicity of interpretations are proved. This formulation for the analysis of instantaneous image motion is then extended to the analysis of spatio-temporal image motion. In this formulation the equations relating the local orientation and motion of a surface and the first order spatio-temporal image motion derivatives are derived. For this case, again, the solution for the orientation and motion is derived in closed form. Further, an interesting case where a camera tracks a point on a moving surface is solved with the knowledge of the camera's tracking motion. An extension of this formulation to deal with non-uniform or accelerated motion is described. This extension is illustrated with a simple example. Finally the formulation for rigid motion is generalized to deal with non-rigid motion. This again is illustrated with a simple example. This general formulation leads to some new insights into the intrinsic nature of the image motion interpretation problem. It makes explicit the well known fact that the general problem of inverting the perspective projection transformation is *inherently ill-posed* (or underconstrained), and that additional assumptions about the physical world are necessary to solve the problem. It gives the minimum number of additional constraints (in the form of assumptions about the scene) necessary to solve the problem. For example, it exposes the fact that the *rigidity assumption*, the assumption that objects in the scene are rigid, is a powerful and sufficient constraint that results in a unique interpretation in most cases. The general formulation serves to address the two fundamental issues: *What information is contained in the image motion field? How can it be extracted?* We emphasize here that the approach taken in this thesis is based on the assumption that the physical phenomena affecting the visual motion, phenomena such as surface structure, translation, rotation, etc., vary ''smoothly'', both in space and in time. However this assumption is not a drawback as long as locations of discontinuities in the image motion field can be detected. The computational approach developed here has been implemented on a computer system and tested on a variety of cases to verify the approach. Many experimental results are included. An important conclusion of this thesis is that an independent perceptual module for visual motion processing is capable of providing a wealth of information under some natural assumptions. The study also discloses the limitations of such a module, limitations in the form of assumptions about the world that are necessary to extract scene information from visual motion. The computational theory of visual motion perception presented here provides a basis for further investigations. It motivates an inquiry into the *second order theory* of the visual motion processing module, i.e. the interaction of this module with other visual processing modules. It naturally raises certain questions: What types of interactions are necessary? What type of interactions are possible? How can the different modules cooperate to achieve their goals efficiently and robustly? An important outcome of this work is a computational approach which is potentially useful in machine vision systems. It provides a uniform representation and a unified algorithm applicable in a variety of situations. The approach is flexible in that *a priori* information can be easily incorporated in the form of additional constraints and it is general enough to be extensible to many situations not considered explicitly in this study. However, in order to appply the theory developed here to practical applications, visual motion needs to be measured very accurately. Accurate and robust measurement of visual motion has remained a difficult problem to this day. An overview of the computational stages in visual motion analysis is given in the next chapter. It defines some essential terminology and summarizes previous work in this area. Chapter 3 gives the formulation of the problem for rigid motion of surfaces. Image motion equations that relate the scene parameters and the image parameters are derived. Chapter 4 deals with solving the image motion equations and the multiplicity of interpretations. Chapter 5 extends the analysis to the temporal domain. Spatio-temporal parameters of image motion are related to the three-dimensional structure and motion parameters. A number of interesting cases are solved including a simple case of non-uniform or accelerated motion. Chapter 6 generalizes the formulation to the non-rigid motion case. Error sensitivity analysis of the computational approach is discussed in Chapter 7.

CHAPTER 2

Background

Early studies of the role of observer motion in the perception of the visual world have been reported by perception psychologists (Helmoltz, 1925; Gibson, 1950, 1966). However, it was not until the wide use of digital computers that a rigorous computational study of this phenomenon was undertaken. In recent years the computer vision community has addressed extensively the computational aspects of visual motion perception. This has led to significant advances in many respects. A variety of computational approaches have been developed to measure as well as interpret visual motion. In this chapter we give a brief overview of the different approaches. In this process we will define some necessary terminology, mention important issues, and summarize previous work. At the end we discuss the relation of our work to previous work. The visual stimulus is an intensity distribution on the eye's retina changing with time. This can be thought of as a function $I(x,y,t)$ where (x,y) specifies a location on the retina and t denotes time (see Figure 2). Generally the intensity function I is taken to be available in the form of a succession of images $I(x,y,t_0)$, $I(x,y,t_0+\delta t)$, $I(x,y,t_0+2\delta t)$, ... for small values of δt. Each such image is called an *image frame* and the succession of images is called an *image sequence*. The goal of early visual processing is to recover the geometry and the time-variation of the scene from I. This is an under-constrained problem because the imaging process results in the loss of significant information. The depth information along the direction of view is absent in the image, and in addition the information is ambiguous in that nothing can be said about changes in the scene which do not cause changes in the image. For example, consider a sphere whose surface has uniform reflectance everywhere. Suppose that the sphere is not translating and the illumination pattern is not changing with time. Then the intensity variation in the image of the sphere is only due to the changing surface orientation. In this case a rotation of the sphere about its center does not change the image and therefore the interpretation is ambiguous with respect to rotation. Another manifestation of this ambiguity is the ‘‘aperture problem’’ along iso-intensity contours discussed later. As an image provides only partial information about the scene, general assumptions about the scene are necessary to reconstruct a description of it from the image. Two assumptions that are usually applied are the *rigidity assumption* and the *smoothness assumption*. The rigidity assumption states that the shapes of objects in the world do not change with time. The smoothness assumption basically states that in a small space-time window the scene parameters and the image parameters are analytic and change slowly. In addition, it is assumed that a violation of this assumption at any point can be detected from the visual motion field. This requirement helps to demarcate boundaries that enclose regions within which the scene and image parameters are analytic and vary smoothly. Such boundaries have been termed *boundaries of analyticity* and the regions enclosed by them *regions of analyticity* by Waxman (1984b). Their recovery has been studied by Hildreth (1983), Adiv (1985a), Wohn (1984), Thompson, Mutch and Berzins (1985), Waxman and Wohn (1986), Wohn and Waxman (1986), and others. The smoothness assumption limits the number of local scene and image parameters (i.e. parameters that specify the scene and the image in a small space-time interval) to small values. These parameters may be coefficients of Taylor series expansions. Relations between these parameters and $I(x,y,t)$ are used to recover the scene. Previous approaches for this purpose can be broadly classified into two categories, *discrete* and *continuous;* a brief overview of each of these approaches is given below. In the discrete approach a set of distinct features are detected in the image and are tracked over time. Examples of such features are points, lines, edges, or regions. Note that point features corresponding to points in the scene trace out curves in the image space-time domain (x,y,t); edges corresponding to curves in the scene trace out surfaces, and regions corresponding to surface patches in the scene trace out solids. The time derivatives of the image ‘‘positions’’ of these features describe their image motion. The image motion of these features are used to determine the relative positions and motions of the corresponding features in the world-domain. If the image motions of these features are ‘‘large’’ compared to the temporal sampling rate, then the eye has to solve the *correspondence* problem, i.e. it has to establish which feature at one time instant corresponds to which feature at the next time instant. In the discrete approach, most of the research until recently has concentrated on determining position and motion by observing the features over a very short image sequence (from only two or three images). In particular, point features corresponding to motion of points in the scene have been studied extensively. In this case, under the assumption of smoothness of motion, the image motion of a set of points $(x_i, y_i) = (a_i(t), b_i(t))$ for $i = 1, 2, .., n$ can be parameterized by the Taylor series coefficients of a_i and b_i. A number of important results have been established for this case. Hay (1966) obtained the first known results for coplanar points moving rigidly. He showed that three successive image frames of four points uniquely determine their relative positions and motion. He also showed that two successive image frames of such points, in general, give rise to a two-fold ambiguity. Tsai and Huang (1980) arrived at similar conclusions independently. Tsai and Huang (1984) also obtained important results for points in an arbitrary configuration undergoing rigid body motion. They showed that, in general, just seven points in only two images (taken in a short time interval) are sufficient to determine their motion and relative positions. Longuet-Higgins (1981) showed that, given two views of eight points, their rigid motion and relative positions can, in general, be determined uniquely. Ullman (1979) showed that, for orthographic projection, three views of four points are sufficient to determine their positions up to a reflection. Necessary conditions for the recovery of relative positions and motions of points under orthographic projection have been derived recently by Aloimonos and Brown (1986). Also, Hoffman and Bennet (1985) have obtained many important results for orthographic projection. Recent work in this area has concentrated on obtaining robust solutions using more image frames and/or more points; e.g. Broida and Chellappa (1986a,b), Shariat (1986), Sethi and Jain (1987). Progress has also been made in stereo-motion fusion (Aloimonos and Rigoutsos, 1986; Aloimonos and Basu, 1986). For a review of this area see Aggarwal (1986). A valuable collection of recent papers on this and related topics in time-varying image analysis has been compiled by Martin and Aggarwal (1988). Line features corresponding to straight lines in the world domain have also been used (Liu and Huang, 1986; Mitiche, Seida, and Aggarwal, 1986; Spetsakis and Aloimonos, 1987). For a planar image the lines in the image domain can be specified by $a_i(t)x + b_i(t)y + c_i(t) = 0$ in a Cartesian coordinate system. For smooth motion, the Taylor coefficients of a_i, b_i and c_i describe the image motion. Liu and Huang (1986) have shown that, using six line correspondences over three image frames, general rigid body motion can be recovered. Mitiche, Seida, and Aggarwal (1986) have used the conservation of the angular configuration between lines as a rigidity constraint and shown that four line correspondences in three views are sufficient to determine the parameters of the lines and their rigid body motion. Spetsakis and Aloimonos (1987) present closed-form solution to this problem. Recently edge features in (x, t) or (y, t) space have been used to recover the geometry of the scene (Bolles and Baker, 1985; Marimont, 1986). The observer motion is assumed to be known and restricted to pure translation in a plane. A dense sampling of the $I(x, y, t)$ function is used, thus avoiding the correspondence problem. In comparison with other approaches, this approach is conceptually different in that the edge features are observed in a two-dimensional domain with one dimension being time and the other being space. In this case first the *image velocity field* or the *image flow* is determined. *Image flow* is a two-dimensional velocity field defined over the retina (or image plane in the case of a camera). This velocity field is the result of projecting onto the retina the instantaneous three-dimensional velocities of points on visible surfaces in the scene (see Figures 3). Image flow is also referred to as *optical flow* in the literature. An example of a scene and an image flow field generated by it are shown in Figure 4a and Figure 4b. Note that image flow is dependent only on the geometry of the scene and the motion of surfaces in the scene. In particular it does not depend on I. Image flow is determined by one of two methods: either by using partial derivatives of the intensity distribution $I(x, y, t)$ or by using features of the intensity distribution. We will discuss each of these methods briefly. Image flow can be estimated from I under certain assumptions. At any point on the retina, let (u, v) denote the image velocity vector, (I_x, I_y) denote the spatial intensity gradient vector, and I_t denote the partial derivative of intensity with respect to time. Then, assuming that the intensity of the image of a scene point remains unchanged with time, the following relation can be derived (Horn and Schunck, 1981):

$$(I_x, I_y) \cdot (u, v) = -I_t \ . \tag{2.1}$$

This is the well known *optical flow equation*. Let $\hat{\boldsymbol{n}}$ denote a unit vector along the intensity gradient (I_x, I_y). Then the optical flow equation can be written as

$$\hat{\boldsymbol{n}} \cdot (u, v) = \frac{-I_t}{\sqrt{I_x^2 + I_y^2}} \ . \tag{2.2}$$

In this expression all quantities except (u, v) are known (i.e. obtainable from I). Here the left hand side represents the magnitude of the component of image velocity along the spatial intensity gradient. Let the component of image velocity along the spatial intensity gradient be denoted by $\boldsymbol{v}_n$, i.e.

$$\boldsymbol{v}_n = (\hat{\boldsymbol{n}} \cdot (u,v)\,)\ \hat{\boldsymbol{n}}. \tag{2.3}$$

$\boldsymbol{v}_n$ at any point can also be thought of as the image velocity component normal to the iso-intensity contour through that point. Therefore, effectively, the optical flow equation implies that at any point only that component of image velocity normal to the iso-intensity contour through that point can be determined (see Figure 5). The other orthogonal component tangential to the iso-intensity contour cannot be determined. This indeterminacy of one of the image velocity components is called the *aperture problem* (see Figure 6). This is an inherent limitation and therefore heuristic assumptions are necessary to measure both components of image velocity. Usually a smoothness constraint in the form of the minimization of a functional is introduced (Horn and Schunck, 1981; Hildreth, 1983). Waxman and Wohn (1985) impose a polynomial model on (u,v) for this purpose. Additional related work on this topic are (Fennema and Thompson, 1979; Nagel, 1983a,b; Anandan, 1986). A Fourier domain method for image flow extraction is found in Fleet and Jepson (1985), Watson and Ahumada (1985), and Heeger (1987). Image flow can be estimated by observing spatial image features over time. Waxman and Wohn (1985) have suggested the use of image contours that correspond to surface markings in the scene. These contours should typically span a field of view of a few degrees. Suppose that one such contour is described by a set of parameters a_i for $i=1,2,...$ at some instant of time. Then the way these parameters change with time due to a deformation of the contour can be used to estimate image flow in a small neighborhood around the contour. Here also a smoothness constraint on the image flow (e.g. a polynomial model for the image flow) is necessary. In general the number of parameters that specify a contour is too large. In order to overcome this problem Kanatani (1985) has suggested using feature measures of these contours instead of the parameters that specify them. In principle, arbitrary image features can be used to estimate image flow. An instantaneous description of these features is specified by a set of parameters which satisfy a certain predicate and the change in these parameters with time may be used to estimate image flow. In the last two subsections we have dealt with computing image flow. Having done this, the next step is to find the scene motion and its geometry from this image flow. Here the relations between the scene parameters and the image flow parameters are derived and the resulting equations are solved. Deriving the equations happens to be easy, but solving them is non-trivial as the equations are non-linear. This thesis is primarily concerned with this step. Early studies of this stage dealt with restricted situations such as pure translation, pure rotation, etc. of the observer in a stationery environment (Gibson, Olum, and Rosenblatt, 1955; Lee, 1974; Clockskin, 1978; Bruss and Horn, 1983; Jain, 1983). Nakayama and Loomis (1974) showed how depth contours may be extracted from a representation of the retinal velocity field induced by motion of the observer. The first mathematical treatment of the general rigid body motion case was by Longuet-Higgins and Prazdny (1980). They showed that, at any instant of time, using up to second order spatial derivatives of the image flow, local orientation and rigid motion of a curved surface patch could, in principle, be determined. Waxman and Ullman (1985) used a kinematic approach to image flow analysis. Using a new representation for the image flow they showed that local geometry (slopes and curvatures) and motion could be recovered for both planar and curved surfaces. They also showed that there exist a class of situations for which there are multiple interpretations. The approach taken here is mainly based on these last two approaches. Recently, Longuet-Higgins (1984), Subbarao and Waxman (1985), and Kanatani (1985) have independently obtained closed-form solutions to planar surfaces in motion. Subbarao (1986b) used spatio-temporal derivatives of image flow and derived closed-form solutions for the structure and motion parameters. More recently, Waxman, Kamgar-Parsi, and Subbarao (1986), have obtained closed-form solutions for curved surfaces based on the solution of planes. A direct solution method and a complete analysis of multiplicity of interpretations for curved surfaces was given by Subbarao (1986b). Some related results are also reported in Negahdaripour (1986). Of the many generic approaches mentioned above, no single approach is adequate for all situations. Each approach is more suitable than others under appropriate circumstances. For example, if the scene has many easily identifiable feature points or lines, then the discrete approach is better. If there are no easily identifiable feature points, the surfaces in the scene are smooth, and have no texture, then the continuous approach based on intensity derivatives is better. If the surfaces in the scene are smooth and have dense texture patterns in the form of surface markings then the continuous approach based on features is better. However, we observe that image flow is fundamental to all approaches. In the discrete approach a very sparsely sampled image flow is used and in the continuous approach a dense image flow is used. Image flow is dependent solely on the geometry of the scene and the motion of surfaces in the scene; it is independent of illumination, reflectance, or the physics of light. This study concentrates on interpreting image flow. It is fundamentally a study of the time-varying projective geometry of surfaces moving through space (Waxman and Wohn, 1986). The problem is to invert the perspective projection transformation to recover the scene. Since the primary goal of this thesis is to explore what information can be extracted from a time-varying image, for the most part correct image flow is assumed to be available. This thesis extends previous work in many ways. Some important extensions are deriving analytic solutions for planar surfaces (Chapter 4), proof of multiplicity of interpretations for curved surfaces (Chapter 4), using both spatial and temporal derivatives of image flow for the interpretation process (Chapter 5), and generalizing the problem to arbitrary surface shapes and transformations under the smoothness assumption (Chapter 6). In addition to these extensions, many relevant previously known results have been obtained in the new approach (e.g. analytic solutions to rigid curved surfaces, Chapter 4). Therefore a major outcome of this work is a general theoretical framework and a unified computational approach.

CHAPTER 3

Formulation

In this chapter first we give an overview of the computational theory and the computational approach presented in this thesis. Then we describe the imaging geometry and formulate the problem for the rigid motion case. This part (i.e. the imaging geometry and the formulation for rigid motion) is basically similar to the formulation of Longuet-Higgins and Prazdny (1980) and Waxman and Ullman (1985). However the significance of this part lies in the details of the formulation. With minor modifications it can be extended to a more general case (e.g. non-rigid motion). For reasons discussed in Section 1.4, we restrict our formulation to a small field of view over a short period of time. Further, in this field of view and period of time, the following are assumed to be ‘‘smooth’’ and changing slowly (the smoothness assumption; see Section 2.2): (i) the shape of the visible surface patch, (ii) the transformation of the surface patch, and (iii) the visual motion field. The shape of the visible surface patch can be described by its distance to the observer, orientation with respect to the observer, curvatures, and higher order variations along the center of field of view. These are the *surface parameters* or the *structure parameters*. The transformation of the surface patch during a short period of time is described by its rigid body motion and shape deformation during that time. Measures of surface motion and deformation are the *transformation parameters*. The structure parameters and the transformation parameters together comprise the *scene parameters*. The scene parameters give a *local* description of the scene. The visual motion or image flow in the small space-time interval is described by the initial image velocity and its spatial and temporal derivatives along the center of field of view. These are the *image flow parameters*. Therefore, for the problem of visual motion interpretation, the input entities are the image flow parameters which are to be measured from the changing image; the output entities are the scene parameters which are to be obtained from the image flow parameters. In order to obtain the scene parameters from the image flow parameters we must first know how they are related. We will see in Chapter 6 that, in general, the relations between the scene parameters and the image flow parameters form an under-constrained system of equations. Therefore, in addition to the relations between the scene parameters and the image flow parameters we need to use additional constraints on the scene to obtain the scene parameters. One such constraint that is of practical importance is the *rigidity assumption* (see Section 2.2). We first formulate our theory for some interesting rigid motion cases. For these cases we derive the equations relating the scene parameters to the image flow parameters, describe a method of solving these equations, and state explicitly the conditions for the presence of unique and multiple solutions. For cases where an assumption other than the rigidity assumption is used, we give clear directions as to how one may proceed to formulate and solve the problem. This is illustrated with a simple example. These results constitute our computational theory. The computational approach presented here is based on the computational theory outlined above. The representation for the input and the output are Taylor coefficients of analytic functions. A stepwise computational algorithm is given for each of the cases considered here. The algorithm is quite straightforward and usually involves finding the roots of a low order polynomial (e.g. a cubic equation). This algorithm is implemented on a computer system and some interesting numerical examples where multiple solutions occur are included here. A first approximation to the human eye is a pin-hole camera. For the purpose of this study we assume a pin-hole camera with a planar projection screen (see Figure 7). A pin-hole camera with a curved projection screen corresponding to the eye’s retina would be closer to reality, but the screen geometry is entirely a matter of convenience and does not compromise the generality of the approach. Note that there is a one to one correspondence between an image on a curved screen such as the eye’s retina and an image on a planar screen. The camera model (adopted from Longuet-Higgins and Prazdny, 1980) is illustrated in Figure 8. (In this model, without loss of generality, the projection plane is taken to be *in front of* the pin-hole to avoid dealing with inverted images.) All the terms defined previously can be defined for this camera model by replacing ‘‘eye’’ by ‘‘camera’’ and ‘‘retina’’ by ‘‘image plane’’. Henceforth, all our references to key terms refer to this camera model unless stated otherwise. The choice of the planar screen geometry restricts our analysis to the field of view along the optical axis. However, the image flow in a field of view not along the optical axis can be analyzed by first projecting the image velocities onto a suitable plane perpendicular to the field of view (see Figure 9).

This projection process is quite straightforward (Kanatani, 1986). In the camera model, the origin of a Cartesian coordinate system $OXYZ$ forms the *center of projection* (corresponding to ''pin-hole'' or optical center) of the camera. The Z-axis is aligned with the *optical axis* and points in the direction of view (or line of sight). The *image plane* is at unit distance from the origin perpendicular to the optical axis. The image coordinate system oxy on the image plane has its origin at (0,0,1) and is aligned such that the x and y axes are, respectively, parallel to the X and Y axes. In this section and the next chapter we deal with the instantaneous image flow due to rigid motion of surfaces. Later we extend our formulation to progressively more complex cases. Let the relative motion of the camera with respect to a rigid surface along the optical axis be described by translational velocity (V_X, V_Y, V_Z) and rotational velocity $(\Omega_X, \Omega_Y, \Omega_Z)$ around the origin. Due to the relative motion of the camera with respect to the surface, a 2D image flow is created by the perspective image on the image plane. At any instant of time, a point P on the surface with space coordinates $(X,\ Y,\ Z)$ projects onto the image plane as a point p with image coordinates $(x,\ y)$ given by

$$x = X/Z \quad \text{and} \quad y = Y/Z\ . \tag{3.1a,b}$$

If the position of P is given by the position vector $\boldsymbol{R}(X,Y,Z)$ then its instantaneous velocity $(\dot{X},\ \dot{Y},\ \dot{Z})$ is given by $\boldsymbol{U} = -(\boldsymbol{V} + \boldsymbol{\Omega} \times \boldsymbol{R})$. Therefore we have

$$\dot{X} = -V_X - \Omega_Y\, Z + \Omega_Z\, Y\ , \tag{3.2a}$$

$$\dot{Y} = -V_Y - \Omega_Z\, X + \Omega_X\, Z\ , \tag{3.2b}$$

$$\dot{Z} = -V_Z - \Omega_X\, Y + \Omega_Y\, X\ . \tag{3.2c}$$

The instantaneous image velocity of point p can be obtained by differentiating equations (3.1a,b):

$$\dot{x} = \frac{\dot{X}}{Z} - \frac{X}{Z}\,\frac{\dot{Z}}{Z} \quad \text{and} \quad \dot{y} = \frac{\dot{Y}}{Z} - \frac{X}{Z}\,\frac{\dot{Z}}{Z} \tag{3.3a,b}$$

In the above two expressions we substitute for the appropriate quantities using relations (3.2a-c,3.1a,b) to obtain

$$\dot{x} = u = \left\{ x\,\frac{V_Z}{Z} - \frac{V_X}{Z} \right\} + [\,xy\ \Omega_X - (1 + x^2)\ \Omega_Y + y\ \Omega_Z\ ] \quad \text{and} \tag{3.4a}$$

$$\dot{y} = v = \left\{ y\,\frac{V_Z}{Z} - \frac{V_Y}{Z} \right\} + [\,(1 + y^2)\ \Omega_X - xy\ \Omega_Y - x\ \Omega_Z\ ]\ . \tag{3.4b}$$

These equations define the instantaneous image velocity field, assigning a unique two-dimensional velocity to every point $(x,\ y)$ on the surface's image. Note that the image velocity at a point $(x,\ y)$ (given by equations (3.4a,b)) in the image domain is due to the world velocity of a point $(xZ,\ yZ,\ Z)$ in the world domain. The value of Z is determined by the actual surface. Suppose that the visible surface is described by $Z = f\,(X,\ Y)$ in our camera-centered coordinate system; then, assuming that the surface is continuous and differentiable, a Taylor series expansion of f can be used to describe a small surface patch around the optical axis:

$$Z = Z_0 + Z_X\, X + Z_Y\, Y + \frac{1}{2}\, Z_{XX}\, X^2 + Z_{XY}\, X\, Y + \frac{1}{2}\, Z_{YY}\, Y^2 + O_3\,(X,\, Y) \tag{3.5}$$

for $Z_0 > 0$. In the above expression, Z_0 is the distance of the surface patch along the optical axis, Z_X and Z_Y are the slopes with respect to the X and Y axes, Z_{XX}, Z_{YY}, Z_{XY} are the curvatures, and the last term denotes higher order terms of the Taylor series with respect to X and Y. Using the method given in Appendix A the surface can be expressed in terms of the image coordinates of the image points:

$$Z(x,y) = Z_0 \left(1 - Z_X x - Z_Y y - \frac{1}{2} Z_{xx} x^2 - Z_{xy} xy - \frac{1}{2} Z_{yy} y^2 - O_3(x,y) \right)^{-1} \tag{3.6}$$

where

$$Z_{xx} = Z_0\, Z_{XX}\ ,\ \ Z_{xy} = Z_0\, Z_{XY}\ ,\ \ Z_{yy} = Z_0\, Z_{YY} \tag{3.7a-c}$$

and $O_3(x,\ y)$ denotes higher order terms. Note that the curvatures are scaled by the distance along the optical axis according to relations (3.7a-c) and therefore absolute curvatures are not recoverable. The slopes Z_X, Z_Y and the scaled curvatures Z_{xx}, Z_{yy}, Z_{xy} will be collectively referred to as the *structure parameters*. Substituting for Z from equation (3.6) into equations (3.4a,b) we obtain

$$u(x,y)=\left[x\frac{V_Z}{Z_0}-\frac{V_X}{Z_0}\right]\left[1-Z_Xx-Z_Yy-\frac{1}{2}Z_{xx}x^2-Z_{xy}xy-\frac{1}{2}Z_{yy}y^2-O_3(x,y)\right] \tag{3.8a}$$

$$+\left[xy\Omega_X-(1+x^2)\Omega_Y+y\Omega_Z\right],$$

$$v(x,y)=\left[y\frac{V_Z}{Z_0}-\frac{V_Y}{Z_0}\right]\left[1-Z_Xx-Z_Yy-\frac{1}{2}Z_{xx}x^2-Z_{xy}xy-\frac{1}{2}Z_{yy}y^2-O_3(x,y)\right] \tag{3.8b}$$

$$+\left[(1+y^2)\Omega_X-xy\Omega_Y-x\Omega_Z\right].$$

In the above equations, the distance Z_0 between the surface and the camera along the optical axis always appears in a ratio with the translational velocity $\boldsymbol{V}$ and therefore is not recoverable. Therefore, we adopt the following notation in presenting the image flow equations.

Translation parameters:

$$V_x = \frac{V_X}{Z_0}\ ,\quad V_y = \frac{V_Y}{Z_0},\quad V_z = \frac{V_Z}{Z_0}\ \text{ for } Z_0 > 0\ . \tag{3.9a-c}$$

The three components of rotation Ω_X, Ω_Y, Ω_Z and the three components of scaled translation V_x, V_y, V_z will be collectively referred to as the *motion parameters*. Now, since the surface is assumed to be smooth (i.e. continuous and differentiable) the instantaneous image velocity in a small neighborhood around the image origin may be expressed in the form of a Taylor series:

$$u(x,y)=u_0+u_xx+u_yy+\frac{1}{2}\,u_{xx}x^2+u_{xy}xy+\frac{1}{2}\,u_{yy}y^2+O_3(x,y)\text{ and} \tag{3.10a}$$

$$v(x,y)=v_0+v_xx+v_yy+\frac{1}{2}\,v_{xx}x^2+v_{xy}xy+\frac{1}{2}\,v_{yy}y^2+O_3(x,y) \tag{3.10b}$$

where the subscripts indicate the corresponding partial derivatives evaluated at the image origin and $O_3\,(x,\ y)$ denotes higher order terms of the Taylor series. In the above expression, the coefficients of the Taylor series expansion, u_0, v_0, u_x, . . . , etc., will henceforth be referred to as the *image flow parameters*. From the image velocity equations (3.8a,b), we can derive the following equations which relate the image flow parameters up to second order at the image origin (i.e. $x=y=0$) to the structure and motion parameters:

$$u_0 = -\,V_x - \Omega_Y\ , \qquad v_0 = -\,V_y + \Omega_X\ , \tag{3.11a,b}$$

$$u_x = V_z + V_x\, Z_X\ , \qquad v_y = V_z + V_y\, Z_Y\ , \tag{3.11c,d}$$

$$u_y = \Omega_Z + V_x\, Z_Y \qquad v_x = -\,\Omega_Z + V_y\, Z_X\ , \tag{3.11e,f}$$

$$u_{xx} = -\,2\,V_z\, Z_X + V_x\, Z_{xx} - 2\,\Omega_Y u_{xy} = -\,V_z\, Z_Y + V_x\, Z_{xy} + \Omega_X\ , \tag{3.11g,h}$$

$$u_{yy} = V_x\, Z_{yy}\ , \qquad v_{xx} = V_y\, Z_{xx}\ , \tag{3.11i,j}$$

$$v_{xy} = -\,V_z\, Z_X + V_y\, Z_{xy} - \Omega_Y \quad\text{and}\quad v_{yy}=-2V_zZ_Y+V_yZ_{yy}+2\Omega_X. \tag{3.11k,l}$$

These equations were originally derived by Longuet-Higgins and Prazdny (1980). The method we have described to derive the above equations can also be used to derive the equations relating the third and higher order image flow parameters to the structure and motion parameters. However we stop at second order as we have a sufficiently constrained system of equations (twelve equations in eleven unknowns). We will say more about this in the next section. For the rigid motion case, equations (3.11a-l) will be referred to as the *image flow equations.* The image flow equations (3.11a-l) form twelve non-linear algebraic equations in eleven unknowns (five structure parameters and six motion parameters). Since the system of equations is overdetermined it is found that, in general, the solution is unique, but since the equations are non-linear, in some exceptional situations multiple solutions do occur. In these equations we observe that none of the rotational components appear in the non-linear terms. All the non-linear terms are formed by the product of a structure parameter and a translation component. Further, *all* structure parameters appear only in products with the translation parameters. Therefore, it is translation through space that reveals surface structure. If there is no translation, all the structure parameters remain undetermined. Also, a curvature parameter always appears in product with a component of translation parallel to the image plane. Therefore, if there is no translation parallel to the image plane the curvatures remain undetermined by image flow parameters up to second order. There are also many situations where some of the image flow equations become dependent in which case multiple solutions occur. For example, if all the curvatures are zero (i.e. the surface is planar), or if the optical axis, direction of translation and surface normal all lie in a plane, then there are in general two solutions. A systematic analysis of all the different cases is given in the appendices. The solution in general cases and a summary of the nature of the solutions are given in the following chapters. Notice that measuring the image velocity derivatives up to second order (given by the coefficients of the Taylor series in equations (3.10a,b)) is adequate since we obtain a sufficient number of constraints (twelve) on the eleven unknowns. On the other hand spatial derivatives up to only first order are inadequate since the first six image flow equations give only six equations involving eight unknowns. Since our analysis is local, we need to solve for the structure and motion parameters separately for each small image neighborhood. However, interestingly, for planar surfaces eight image motion parameters specify the image motion field *globally* (e.g. see Subbarao and Waxman, 1985). An intuitive explanation of this is that the motion parameters are constant everywhere in both cases, but whereas the structure parameters for a plane constituting the two slopes are constant everywhere, for curved surfaces they change.

> *‘‘This analysis (of Longuet-Higgins and Prazdny, 1980) is another example of how computational theory can help empirical investigation. By solving the mathematics of the problem-- and this was surely long overdue-- Longuet-Higgins and Prazdny have provided a framework within which to enquire whether we humans actually do make use of optical flow, as Gibson suggested, and if we do, how.’’*
>
> -David Marr (1982)

CHAPTER 4

SOLVING THE IMAGE FLOW EQUATIONS

''....... I know how slippery they (the image flow equations) are!''

-H.C. Longuet-Higgins[1]

In this chapter we present closed-form solutions to the image flow equations for both planar and curved surfaces. For this purpose we assume that all the structure and motion parameters are finite. Detailed derivations of these solutions and the solutions in some degenerate cases are given in Appendices B,C, and F. In solving for the structure and motion parameters from the given image motion parameters, we use a new parameterization of the solution space; we use a trigonometric substitution which introduces two new variables r and θ which respectively correspond to the (signed) magnitude and direction of the translational component parallel to the image plane. The relation of the translation parameters (V_x, V_y, V_z) to r, θ is shown in Figure 10. This particular representation of the problem simplifies the task of solving the problem and proving many uniqueness results. For all rigid motion cases, we can first solve for r and θ by simultaneously solving a small set of equations (typically three) and from these we can compute the other unknowns. An important advantage of this method is that the set of relations used to compute the structure and motion parameters from r and θ are the same in the many different cases considered here. The only difference is in the expressions we use to solve for r and θ. Therefore, the computational approach given here forms a general method useful in many different cases. In solving the image flow equations we note that the first six equations (3.11a-f) relate the first-order spatial derivatives of image flow to the structure and motion parameters. These six equations do not contain the curvature terms and therefore form six equations in eight unknowns. The solution of these equations in terms of r and θ is derived in Appendix B (Theorem 5). Using the notation

$$s \equiv \sin\theta \ , \ \ c \equiv \cos\theta \ \text{ for } -\pi/2 < \theta \le \pi/2 \ , \tag{4.1a,b}$$

$$a_1 = u_y + v_x, \quad \text{and} \quad a_2 = u_x - v_y \tag{4.2a,b}$$

the solution is given by

$$V_x \equiv rc, \qquad V_y \equiv rs, \tag{4.3a,b}$$

$$V_z = u_x s^2 + v_y c^2 - a_1 cs, \qquad \Omega_Z = u_y s^2 - v_x c^2 + a_2 cs, \tag{4.3c,d}$$

$$Z_X = (a_1 s + a_2 c)/r, \qquad Z_Y = (a_1 c - a_2 s)/r, \tag{4.3e,f}$$

$$\Omega_X = v_0 + rs, \qquad \Omega_Y = -(u_0 + rc), \tag{4.3g,h}$$

The above relations give an explicit solution for the orientation and motion in terms of r and θ. Therefore once we have solved for r and θ we can solve for the two slopes and all the motion parameters. Notice that V_z and Ω_Z are given only in terms of θ. Therefore they can be determined from θ alone. This observation will be important later on. In order to solve for r and θ we need additional constraints. In the next two subsections we will consider solving for r and θ for planar and curved surfaces using the second order spatial derivatives of the image flow (i.e. equations 3.11g-l). The image flow equations are non-linear and therefore many degenerate cases are possible. Therefore we solve these equations systematically by considering all possibilities. However we shall not do so in this chapter. Here we present only the main ideas that underlie our approach; a thorough treatment is given in Appendix B. The solutions of the image flow equations in three degenerate cases are given in Appendix B. These cases are: (i) translation is zero, (ii) translation parallel to the image plane is zero, and (iii) the surface is planar and frontal. In the remainder of

[1] Personal communication.

this chapter we shall preclude the occurence of these cases. One case of special interest is when the surfaces in the field of view are planar or nearly planar. This is often true in small fields of view. But if a large surface is known to be planar we shall see later that it is not necessary to restrict our analysis to small fields of view; the image flow in the entire field of view spanned by that surface can be analyzed as a whole. The solution of the image flow equations for planar surfaces was originally obtained independently by Longuet-Higgins (1984), Kanatani (1985), and Subbarao and Waxman (1985), each by a different approach. In this section we rederive the results of Subbarao and Waxman (1985) in our new approach. For planar surfaces the equations involving the second order spatial derivatives of the image flow are obtained by setting all the curvature terms (i.e. Z_{xx}, Z_{yy} and Z_{xy}) to zero in equations (3.11g-l). This results in the following constraints:

$$v_{xy} = -V_z Z_X - \Omega_Y \quad , \quad u_{xy} = -V_z Z_Y + \Omega_X \, , \tag{4.4a,b}$$

$$u_{xx} = 2\, v_{xy} \;\; , \;\; v_{yy} = 2\, u_{xy} \;\; , \;\; u_{yy} = 0 \;\; , \;\; v_{xx} = 0 \, . \tag{4.4c-f}$$

We see that only the first two equations involve the structure and motion parameters and therefore are useful in solving for r and θ. The last four equations simply give constraints for the surface to be planar. Further it is easy to show that for planar surfaces all third and higher order spatial derivatives of image flow are identically zero. Therefore the image flow of a planar surface is specified *globally* by just eight parameters (two zeroth order, four first order, and two second order flow parameters); in contrast, for general curved surfaces piecewise quadric surface approximation is used and twelve parameters specify the image flow *locally*. Relations (4.4a,b) can be used to solve for r and θ. Details of the solution method are given in Appendix C. Equations (4.4a,b) give the following constraints on θ and r:

$$r^2 c - (v_{xy} - u_0)\, r - V_z\, (a_1 s + a_2 c) = 0 \, , \tag{4.5a}$$

$$r^2 s - (u_{xy} - v_0)\, r - V_z\, (a_1 c - a_2 s) = 0 \, . \tag{4.5b}$$

In these relations V_z can be expressed in terms of θ (using relation (4.3c)) yielding two equations in the two unknowns r and θ. Note that equations (4.5a,b) are quadratic in r. Equating the roots of these quadratic equations we obtain an equation in θ alone (see equation (C25) in Appendix C). Using this equation we can easily solve for θ by simply searching in the interval $(-\pi/2,\ \pi/2]$. From θ, r is obtained as the common root(s) of the quadratic equations (4.5a,b). In general, it is shown (Theorem 8, Appendix C) that there are two solutions for θ, and corresponding to each of these there is one solution for r (Theorem 4, Appendix D) given by

$$r = V_z\, \frac{(a_1 \cos 2\theta - a_2 \sin 2\theta)}{(v_{xy} - u_0) s - (u_{xy} - v_0) c} \, . \tag{4.6}$$

In the above method we solved for θ numerically by searching in the interval $(-\pi/2,\ \pi/2]$. For any given precision we can carry out this search in a constant amount of time (because the search space is finite). However there is another method by which we can obtain an analytic solution for θ. Here we first solve for the velocity V_z along the line of sight and then solve for θ. In Appendix C (Theorem 1) it is shown that V_z can be determined by solving a cubic equation. Also it is shown that all three roots of the cubic equation are real (Lemma 1, Appendix C) and that V_z is the middle root (Theorem 4, Appendix C). This root can be determined utilizing the analytic expressions for the roots of a cubic equation (Theorem 5, Appendix C). A numerical method may also be used for computational purposes. Using relation (4.3c) we can derive the following equation for θ in terms of V_z:

$$(u_x - V_z) \tan^2\theta - a_1 \tan\theta + (v_y - V_z) = 0. \tag{4.7}$$

This is a quadratic equation in $\tan\theta$ and therefore can be easily solved to obtain θ. Since the equation is quadratic, there are in general two solutions for θ. We have seen that the relations (4.3c,d) for the velocity of approach along the line of sight V_z and the rotational velocity around the line of sight Ω_Z are functions of θ only; they do not depend on r. Therefore relations (4.3c,d) can be used to establish upper and lower limits on V_z and Ω_Z. In particular, it is shown (Theorem 6, Appendix C) that the maximum and minimum values of V_z, Ω_Z are

$$V_z^{(max/min)} = \frac{u_x+v_y}{2} \pm \frac{\sqrt{a_1^2+a_2^2}}{2} \ , \qquad \textbf{(4.8a)}$$

$$\Omega_Z^{(max/min)} = \frac{u_y-v_x}{2} \pm \frac{\sqrt{(u_y+v_x)^2+(u_x-v_y)^2}}{2} . \qquad (4.8b)$$

Note that all terms on the right hand side of the above expressions are only first order flow parameters; no second or higher order parameters are involved. Further, the above limits hold irrespective of whether the surface is planar or curved. This is an interesting observation for the following reason. Biological vision systems have been found to be very quick in responding to approaching objects (which can potentially hurt the organism). This has been called the ‘‘looming effect’’ (Schiff, Caviness, and Gibson, 1962). In this context, an implication of the above result is that first order flow parameters are sufficient to determine the *maximum value of the velocity of approach*; second or higher order parameters, whose estimation requires higher quality images and more computation, are not necessary. Thus, an organism can, in principle, quickly evaluate the potential danger of an approaching object from only coarse information. This possibility, irrespective of the shape of the object, and even while the organism is rotating, is by itself interesting. (This observation may also be useful for obstacle avoidance in a robot navigation system.) In addition, the knowledge of an interval in which V_z lies may be useful in determining its actual value. For example, in the case of moving planar surfaces, V_z is obtained by solving a cubic equation; if an interval in which V_z lies is known in this case, then a simple binary search can be used to solve the cubic equation. Alternatively such information can be used to guess an initial solution for a numerical solution method. For the special case where an observer is moving in a static environment, the above result has an interesting consequence. (Examples of such a case are flying bees, birds, and helicopters.) In this case, by determining the bounds on the translational and angular velocities along some three mutually orthogonal viewing directions, bounds on the over all translational and rotational velocities of the observer can be determined from only first order image flow derivatives. In the image domain the bounds can be expressed in terms of the three first order *differential invariants* of image flow: *divergence*, *curl*, and *shear magnitude.* The *divergence* of the image flow corresponds to the *isotropic expansion (contraction)* of an image neighborhood (Figure 11a,b). The *curl* represents the rigid body *rotation* of the image neighborhood (Figure 12a,b). And the *shear magnitude* is a measure of the *shape deformation* of the image neighborhood (Figure 13a,b). These quantities are *invariant* with respect to the orientation of the image axes. Therefore they are called *differential invariants.* Analysis of image flow using such an invariant representation has been done by several researchers in the Computer Vision area (Koenderink and Van Doorn, 1975, 1976; Waxman and Ullman, 1985; Kanatani, 1986). It is shown in Appendix C that the bounds on V_z and Ω_Z can be expressed as

$$\text{Maximum/Minimum approach velocity}$$
$$= \frac{1}{2}\left(\text{Divergence} \pm \text{Shear magnitude}\right) , \qquad (4.8c)$$

$$\left(\begin{array}{l}\text{Maximum/Minimum angular velocity}\\ \text{around the viewing direction}\end{array}\right)$$
$$= \frac{1}{2}\left(-\text{Curl} \pm \text{Shear magnitude}\right) . \qquad (4.8d)$$

Relation (4.8c) is illustrated diagrammatically in Figure 14. The bounds can also be expressed in terms of invariant quantities in the scene domain. Denoting the translation parallel to the image plane by $\boldsymbol{V}_1$ (as in equation (4.15a)) and surface gradient vector by $\boldsymbol{P}$ (as in equation (4.16b)), the following relations are derived in Appendix C:

$$V_z^{(max/min)} = V_z + \frac{1}{2}\left(\boldsymbol{V}_1 \cdot \boldsymbol{P} \pm |\boldsymbol{V}_1||\boldsymbol{P}|\right) \quad \text{and} \qquad (4.8e)$$

$$\boldsymbol{k}\ \Omega_Z^{(max/min)} = \boldsymbol{k}\ \Omega_Z + \frac{1}{2}\left(\boldsymbol{V}_1 \times \boldsymbol{P} \pm \boldsymbol{k}\ |\boldsymbol{V}_1| |\boldsymbol{P}|\right), \qquad \textbf{(4.8f)}$$

where $\boldsymbol{k}$ is a unit vector along the $Z-$axis. There does not appear to be a straightforward interpretation of the above relations although they have a pleasing form. Koenderink and Van Doorn (1975) discuss the different types of image flows generated depending on the eigen values of the *velocity gradient tensor* (see Appendix C) of the image flow. Their work is directly relavent to the interpretation of the above equations. The number of interpretations of an image flow field is determined by the number of solutions to $(\theta\ ,\ r)$ obtained by solving equations (4.5a,b). It is shown (Theorem 4, Appendix D) that there are in general two solutions to equations (4.5a,b), and among these two solutions there is always one which corresponds to the correct physical interpretation. Suppose that the quantities corresponding to the correct interpretation are denoted by appending ''0'' to their subscripts. Then in Appendix D it is shown that, in terms of the actual physical interpretation, the two solutions to (θ, r) are

$$(\theta_0,\ r_0),\ (\theta'_0,\ -V_{z0}\ k_0) \qquad (4.9a,b)$$

where $\theta'_0,\ k_0$ are such that (see Figure 15)

$$Z_{X0} = k_0\ \cos\theta_0' \quad \text{and} \quad Z_{Y0} = k_0\ \sin\theta_0'\ . \qquad (4.10a,b)$$

Note that r_0 and θ_0 are the magnitude and direction respectively of the translation vector (V_{x0}, V_{y0}) parallel to the image plane whereas k_0 and θ'_0 correspond respectively to the magnitude and direction of the surface gradient vector (Z_{X0}, Z_{Y0}). Alternatively, k_0 and θ'_0 can be thought of as measures of, respectively, the *slant* and *tilt* of the surface. (Slant is the angle by which the surface dips away from the frontal plane, and tilt is the direction in which the dip takes place.) Therefore, the two solutions for (θ, r) are such that one corresponds to the translation vector parallel to the image plane and the other corresponds to the surface gradient vector scaled by the translation component along the line of sight. Thus, in general, an instantaneous image flow of a planar surface has two interpretations. Further, given one interpretation the other interpretation can be obtained using relations (4.10a,b). There are two exceptions to the two-fold ambiguity of interpretation. The first is when there is no translation along the line of sight, i.e. $V_z = 0$. In this case, from the solution (4.9b), we see that the spurious solution for r becomes zero. Since $r=0$ is ruled out due to previous consideration of degenerate cases, we get the unique solution (4.9a) for (θ, r). The second exception is when the direction of translation is parallel to the surface normal (A general case of the physical situation is shown in Figure 16.). In this case it is easy to show that $\theta_0 = \theta'_0$ and $r_0 = -V_{z0} k_0$. Therefore the two solutions (4.9a,b) merge into one, resulting in a unique interpretation.

We have seen that the instantaneous image flow of a planar surface has in general two interpretations. This ambiguity of interpretation may be resolved by using additional information. Here we consider two sources of information which may aid this process. First is the image flow of a second planar surface in the scene which is rigid and stationary with respect to the first one (e.g. two faces of a polyhedron). In this case, for the spurious solutions the orientations of the two surfaces are the same and the rotation parameters are, in general, different (Theorem 1, Appendix E). These solutions are not acceptable because different surfaces in the scene usually have different orientations and should have the the same rotation parameters as the surfaces are stationary with respect to each other. We call this the *spatial consistency* requirement. For resolving the ambiguity in this manner we need to first segment the image flow field into regions corresponding to different surfaces in the scene. Then the image flows of different surfaces are interpreted separately. The second source of information we consider for resolving the ambiguity is the change in the image flow after a short time interval. Here we require that the interpretation of the image flow at one time instant should be consistent with that at an instant a short time later. For example, we require that the orientation of the surface determined from the image flow at a given instant should be consistent with the orientation and rotation determined from the image flow a short time earlier. We call this the *temporal consistency* of solutions. It is shown in Appendix E (Theorem 3 and Figure 20) that this requirement is in general sufficient to single out the unique interpretation corresponding to the physical world. For curved surfaces the second spatial derivatives of the image flow depend on their curvatures. Therefore use of second order image flow parameters to solve for r and θ introduces the three curvatures as additional unknowns. Thus we need at least five more equations to solve for all the unknowns. However, we shall use all the six equations which relate the six second order flow parameters to the structure and motion parameters. Having previously considered some degenerate cases and the planar surface case, in the remaining part of this chapter we preclude the possibility of these cases. This in effect implies that we are dealing only with cases where the surface is curved (i.e. at least one of the curvatures is non-zero) and the translation parallel to the image plane is non-zero. For curved surfaces, a formulation and a solution method were originally proposed by Longuet-Higgins and Prazdny (1980). However, their method cannot be used if either there is no translation along the line of sight, or the direction of translation, the surface normal, and the line of sight all lie in a plane. Waxman and Ullman (1985) proposed a numerical solution method which could be used in all cases. Recently Waxman, Kamgar-Parsi, and Subbarao (1986) obtained a complete solution in closed form for this problem. However many important questions concerning the multiplicity of solutions still remained open. These questions were investigated in Subbarao (1986a) using a new approach. Further, closed-form solutions were also obtained in the new approach. This section presents the results in Subbarao (1986a). In previous approaches (Longuet-Higgins and Prazdny, 1980; Waxman and Ullman, 1985; Waxman, Kamgar-Parsi, and Subbarao, 1986) the image flow parameters were first transformed corresponding to a rotation of the image coordinate system (around the line of sight) and then the image flow equations were solved; after solving the equations in this rotated coordinate system, the solution for the structure and motion parameters was transformed back to the original coordinate system. In comparison, the solution method given here does not involve such an intermediate step; the image flow equations are solved directly in the original coordinate system. This saves some computation. Apart from this minor computational advantage, the solution method used here facilitates easy proofs of the conditions for the occurrence of unique and multiple solutions (see Appendix G). In Theorems 1,2 of Appendix F the solution is derived for a curved surface with non-zero translation parallel to the image plane. The solutions for the motion parameters and the slopes are given by relations (4.3a-h). The solution for the curvatures is

$$Z_{xx} = \frac{1}{r}\left[u_{xx}c + v_{xx}s - 2u_0 c - 2rc^2 + 2V_z Z_X c\right] \tag{4.11a}$$

$$Z_{yy} = \frac{1}{r}\left[u_{yy}c + v_{yy}s - 2v_0 s - 2rs^2 + 2V_z Z_Y s\right] \tag{4.11b}$$

$$Z_{xy} = \frac{1}{2r}\left[s(u_{yy}+2v_{xy}-u_{xx}) + c(v_{xx}+2u_{xy}-v_{yy})\right] \tag{4.11c}$$

(Note: the right hand sides of equations (4.11a,b) can be expressed in terms of only r and θ by substituting for Z_X, Z_Y and V_z from (4.3c,e,f).) In Theorem 2 of Appendix F the following three equations are derived for r and θ:

$$u_{yy}\tan^3\theta+(2u_{xy}-v_{yy})\tan^2\theta+(u_{xx}-2v_{xy})\tan\theta-v_{xx} = 0\ , \tag{4.12a}$$

$$2r^2cs+\left[v_{xx}c-(u_{xx}-2u_0)s\right]r-2V_z s(a_1 s+a_2 c)=0\ , \tag{4.12b}$$

$$2r^2cs+\left[u_{yy}s-(v_{yy}-2v_0)c\right]r-2V_z c(a_1 c-a_2 s)=0. \tag{4.12c}$$

(Equation (4.12a) is also derived in Longuet-Higgins and Prazdny (1980)). Equations (4.12a-c), after substituting for V_z from (4.3c), form three equations in the two unknowns: r, θ. Therefore, to solve the image flow equations, first we solve for θ by solving the cubic equation (4.12a). Then r is obtained as the common root of the two quadratic equations (4.12b,c). If there is no common root then the corresponding solution of θ is not valid. Thus there is one extra constraint and this in general results in a unique solution. Corresponding to each solution of r and θ obtained by solving equations (4.12a-c) we get one solution for the image flow equations (from 4.3a-h,4.11a-c) In general we can obtain an explicit solution for r by eliminating the term $2r^2cs$ from equations (4.12b,c). The solution is

$$r = \frac{2V_z\left[s(a_1 s+a_2 c)-c(a_1 c-a_2 s)\right]}{(v_{yy}-2v_0+v_{xx})c-(u_{xx}-2u_0+u_{yy})s}\ . \tag{4.13}$$

This solution should further satisfy either (4.12b) or (4.12c) in order for it to be acceptable. If either the denominator or the numerator on the right hand side is zero, then the above expression degenerates and cannot be used to solve for r. This is because we are solving for a case where the translation parallel to the image plane is finite and non-zero. In this case we have to go back to (4.12b,c) to obtain r. A complete computational algorithm for solving the image flow equations is given in Appendix F. The number of solutions for the image flow equations (3.11a-l) is equal to the number of solutions for r, θ obtained by solving equations (4.12a-c). Therefore conditions for the occurrence of multiple solutions can be obtained by an exhaustive analysis of these three equations. This analysis turns out to be a long and tedious exercise in algebra. This analysis is given in Appendix G. A summary of the analysis is given below. Since equations (4.12a-c) form an over-constrained system of equations (three equations in two unknowns) the solution is in general unique (Theorem 9, Appendix G), but multiple solutions are possible as the equations are non-linear. Equation (4.12a) is a cubic equation involving only θ and therefore is easily solved. This gives at most three solutions for θ (in the interval $-\pi/2 < \theta \le \pi/2$). We find (in Lemma 1 of Appendix G) that for ovoid surfaces (i.e. convex or bowl shaped surfaces) only one of the three roots of equation (4.12a) is real and therefore we obtain a unique solution for θ. For cylindrical surfaces up to two, and for saddle shaped surfaces up to three solutions for θ are possible (Lemma 1 of Appendix G). Note that, for a given θ we can immediately solve for V_Z and Ω_Z from relations 4.3c,d. Therefore, given θ, equations (4.12b) and (4.12c) reduce to simple quadratic equations in r and therefore can be easily solved. The common root(s) of these two equations is (are) the solution(s) for r. If there is no common root then the given solution for θ is not acceptable. If the roots of the two quadratic equations obtained from equations (4.12b,c) are identical (which is the case when the coefficients of the two equations are proportional) then we have two solutions for a given θ. This is found to be the case (Theorems 7 and 8 in Appendix G) when the direction of translation, the surface normal and the optical axis all lie in a common plane. However, in this case, the solution for r becomes unique when the two roots of the quadratic equation (4.12b or 4.12c) are equal. In one case (Theorem 8, Appendix G) the two roots become equal when the direction of translation is parallel to the surface normal. There are two special situations when equations (4.12b,c) reduce to a single linear equation in r (as compared to two quadratic equations in a general case) for a given θ : (i) a specular saddle surface (i.e. the tangent plane to the surface is frontal or parallel to the image plane and the surface itself is saddle shaped; see Figure 17) with mean scaled curvature equal to -1 (Theorem 5 of Appendix G), and (ii) a saddle or a cylindrical surface with mean scaled curvature equal to -1, no translation along the optical axis, and the slopes and curvatures satisfy a certain condition (G17) (Theorem 6 of Appendix G). In these situations all solutions for θ obtained by solving equation (4.12a) are acceptable (Theorem 3 of Appendix G) and each of these results in one solution for the image flow equations. In fact it is found that for case (i) above three solutions exist (Theorem 5 of Appendix G) and for case (ii) up to two solutions are possible (Theorem 6 in Appendix G). There is one case where there are two solutions for θ and for each of these there are two solutions for r, thus leading to a total of four solutions. This occurs when there is a rare coincidence (relations G6c,G23) of the translation vector, slopes, and curvatures (see Theorem 7, Appendix G). However the presence of four solutions is an artifact of using only up to second order image flow parameters. Use of higher order flow parameters should prevent any more than three solutions according to Maybank (1985). Maybank (1985) has also obtained some results concerning the multiplicity of interpretations. His formulation uses a polar projection (spherical image screen) camera model. The main result of his work is that *an ambiguous flow field has at most three values of angular velocity compatible with it.* This should be compared to our result above which gives explicit conditions for the occurrence of three solutions for the image flow equations. Maybank indicated an implicit method for constructing a case where three solutions exist, but did not give a numerical example. In Chapter 7 we give one example (Example 1) of this case which is constructed using our own algorithm in Appendix G. A more precise account of the nature of the solution in various cases is summarized in Section 4.4. For curved surfaces, occurrence of multiple solutions is rare. When they do occur, the interpretation is ambiguous only locally, both in space and in time. The extension of the ideas of spatial and temporal consistency discussed for planar surfaces is quite straightforward. One form of spatial consistency that can be imposed is by using third or higher order image flow parameters. Here we require that the solution for the structure and motion parameters obtained from second and lower order flow parameters be in agreement with the third (and higher) order flow parameters. For a fixed level of noise in the input image flow, use of higher order flow parameters is equivalent to extending the image flow analysis to a slightly larger image neighborhood; or, for a fixed size of image neighborhood, it is equivalent to using image flow with a lower noise level. This subsection summarizes the detailed results presented in Appendices B, D, and G. The most important results of these appendices are the explicit statement and proofs of all the conditions for the occurrence of up to four solutions, four being the maximum possible. Following is a summary of the nature of the solutions to the image flow equations (3.11a-l) in different cases. (Figure 16 may be helpful in visualizing some of these cases.) Here we list the different cases in a sequential order such that the occurrence of a particular case is detected by the absence of the previous cases and checking for the satisfaction of one or more constraints on the image flow parameters.

(i) If there is no translation, the structure parameters are undetermined and the solution is unique.

(ii) If either there is no translation parallel to the image plane or the surface is planar and frontal, then there can be up to three solutions for the image flow equations.

(iii) If the surface is planar then there are two solutions with the exception of the following cases:

(a) There is no translation along the optical axis.

(b) Translation is parallel to the surface normal.

(iv) There are three solutions if the surface is a specular saddle and the mean scaled curvature is -1 (i.e. $Z_{xy}^2 - Z_{xx}Z_{yy} > 0$, $Z_X = Z_Y = 0$ and $(Z_{xx}+Z_{yy})/2 = -1$.).

(v) There are two solutions if the surface is saddle (Figure 17) or cylindrical (i.e. $Z_{xy}^2 - Z_{xx}Z_{yy} \geq 0$), there is no translation along the optical axis and the mean scaled curvature is -1 except when the translation vector, surface normal and the optical axis all lie in a plane; in this case the solution is unique.

(vi) There can be up to four solutions if the translation vector, surface normal and the optical axis all lie in a plane.

(vii) In cases other than the ones listed above the solution is unique.

In this chapter the scene parameters and the image flow parameters are represented by the Taylor coefficients of appropriate analytic functions. The values of these Taylor coefficients are dependent on the orientation of the image coordinate system in the image plane. A rotation of the image coordinate system about the optical axis changes the values of these quantities. The transformation equations for the change in these quantities due to such rotations can be easily derived (e.g. see Waxman and Ullman, 1985; Kanatani, 1986). However, an image itself has no inherent coordinate system, and therefore our choice of a coordinate system is completely arbitrary. For this reason, we would like to use a canonical representation for the scene parameters and the image parameters where the quantities have conceptually ‘‘simple’’ physical interpretations. To clarify this point, let us consider the representation of the image flow parameters. Suppose that the image coordinate system is rotated counterclockwise about the optical axis by an angle θ and in this coordinate system the new values of the corresponding quantities are denoted by a prime. Then, for the image flow parameters, the transformation can be expressed in the form

$$\begin{bmatrix} u'_0 \\ v'_0 \\ u'_x \\ \cdot \\ \cdot \\ v'_{yy} \end{bmatrix} = \begin{bmatrix} & & \\ & T(\theta) & \\ & & \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \\ u_x \\ \cdot \\ \cdot \\ v_{yy} \end{bmatrix} \tag{4.13}$$

(Note: the fact that the transformation can be expressed in the above form is a property of this set of parameters.) In the above expression the new value of a flow parameter is given as a linear function of the initial flow parameters. Now consider the fact that our choice of Taylor coefficients to represent the image flow is arbitrary, and that we could have used any appropriate linear combinations of these Taylor coefficients to represent the image flow. In particular we would be interested in flow parameters $C_1, C_2, \ldots.$ such that the transformation matrix T is diagonal (i.e. all non-diagonal elements are zero). In fact, it is known that for our case of coordinate rotation there always exists one such set of parameters. For this set of parameters the transformation equation is of the form

$$\begin{bmatrix} C'_1 \\ C'_2 \\ \cdot \\ \cdot \\ C'_m \end{bmatrix} = \begin{bmatrix} e^{-in_1\theta} & & & & 0 \\ & e^{-in_2\theta} & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & e^{-in_m\theta} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \cdot \\ \cdot \\ C_m \end{bmatrix} . \tag{4.14}$$

where i is the imaginary unit, and n_k is an integer. In the above relation, all quantities are transformed independently of the others. There exists no mutual coupling. Thus, intuitively, each of these quantities describes a single independent property of the image flow. We call the quantities $C_1, C_2, \ldots$. the *fully reduced flow parameters*. An introduction to the application of group theoretical methods to derive canonical representations for image characteristics can be found in Kanatani (1986). The above discussion is mainly derived from this reference. Next we give the fully reduced parameterizations for the scene and the image flow representation (the parameterization for the image flow was derived by Kanatani, 1986). Also we give the image flow equations in this representation. Two of the advantages of such a representation are its conceptual simplicity and compactness.

$$\boldsymbol{V}_1 = V_x + iV_y\ ,\quad n=1 \tag{4.15a}$$

$$\boldsymbol{V}_2 = V_z\ ,\quad n=0 \tag{4.15b}$$

$$\boldsymbol{\Omega}_1 = \Omega_X + i\Omega_Y\ ,\quad n=1 \tag{4.15a}$$

$$\boldsymbol{\Omega}_2 = \Omega_Z\ ,\quad n=0 \tag{4.15b}$$

$$\boldsymbol{P} = Z_X + iZ_Y\ ,\quad n=1 \tag{4.16a}$$

$$\boldsymbol{K}_1 = \frac{Z_{xx}+Z_{yy}}{2}\ ,\quad n=0 \tag{4.16b}$$

$$\boldsymbol{K}_2 = \frac{Z_{xx}-Z_{yy}}{2} + iZ_{xy}\ ,\quad n=2 \tag{4.16c}$$

$$\boldsymbol{U} = u_0 + iv_0\ ,\quad n=1 \tag{4.17a}$$

$$\boldsymbol{T} = u_x + v_y\ ,\quad n=0 \tag{4.17b}$$

$$\boldsymbol{R} = v_x - u_y\ ,\quad n=0 \tag{4.17c}$$

$$\boldsymbol{S} = (u_x - v_y) + i(v_x + u_y)\ ,\quad n=2 \tag{4.17d}$$

$$\boldsymbol{H} = (0.5u_{xx} + v_{xy} - 0.5u_{yy}) + i(0.5v_{yy} + u_{xy} - 0.5v_{xx})\ ,\quad n=1 \tag{4.17e}$$

$$\boldsymbol{L} = (0.5u_{xx} - v_{xy} + 1.5u_{yy}) + i(0.5v_{yy} - u_{xy} + 1.5v_{xx})\ ,\quad n=1 \tag{4.17f}$$

$$\boldsymbol{M} = (0.5u_{xx} - v_{xy} - 0.5u_{yy}) + i(0.5v_{xx} + u_{xy} - 0.5v_{yy})\ ,\quad n=3\ . \tag{4.17g}$$

The image flow equations (3.11a-l) can be represented in terms of the above defined quantities as below (note: * denotes complex conjugate).

$$\boldsymbol{U} = -\boldsymbol{V}_1 + i\boldsymbol{\Omega}_1 \tag{4.18a}$$

$$\boldsymbol{T} = 2\boldsymbol{V}_2 + \text{Real}[\boldsymbol{V}_1\boldsymbol{P}^*] \tag{4.18b}$$

$$\boldsymbol{R} = -2\boldsymbol{\Omega}_2 + \text{Imaginary}[\boldsymbol{V}_1\boldsymbol{P}^*] \tag{4.18c}$$

$$\boldsymbol{S} = \boldsymbol{V}_1\boldsymbol{P} \tag{4.18d}$$

$$\boldsymbol{H} = -2\boldsymbol{V}_2\boldsymbol{P} + i2\boldsymbol{\Omega}_1 + \boldsymbol{V}^*{}_1\boldsymbol{K}_2 \tag{4.18e}$$

$$\boldsymbol{L} = 2\boldsymbol{V}_1\boldsymbol{K}_1 - \boldsymbol{V}^*{}_1\boldsymbol{K}_2 \tag{4.18f}$$

$$\boldsymbol{M} = \boldsymbol{V}_1\boldsymbol{K}_2 \tag{4.18g}$$

In the previous sections we first expressed the structure and motion parameters in terms of θ and r and then derived equations for θ and r. Similarly, here we first express all the unknowns in terms of $\boldsymbol{V}_1$ using relations (4.18a-d,f,g) and then obtain an equation in $\boldsymbol{V}_1$ using relation (4.18e). Following this method we get

$$\boldsymbol{\Omega}_1 = -i(\boldsymbol{U}+\boldsymbol{V}_1)\ , \tag{4.19a}$$

$$\boldsymbol{P} = \frac{\boldsymbol{S}}{\boldsymbol{V}_1}\ , \tag{4.19b}$$

$$\boldsymbol{V}_2 = \frac{\boldsymbol{T}-\text{Real}[\boldsymbol{V}_1\boldsymbol{P}^*]}{2}\ , \tag{4.19c}$$

$$\boldsymbol{\Omega}_2 = \frac{-\boldsymbol{R}+\text{Imaginary}[\boldsymbol{V}_1\boldsymbol{P}^*]}{2}\ , \tag{4.19d}$$

$$\boldsymbol{K}_2 = \frac{\boldsymbol{M}}{\boldsymbol{V}_1}\ , \tag{4.19e}$$

$$\boldsymbol{K}_1 = \frac{\boldsymbol{L}+\boldsymbol{V}^*{}_1\boldsymbol{K}_2}{2\boldsymbol{V}_1}\ , \tag{4.19f}$$

and finally, $\boldsymbol{V}_1$ is obtained by solving the equation

$$\boldsymbol{H} = \left(\text{Real}\left[\boldsymbol{S}^*\ e^{i2\arg(\boldsymbol{V}_1)}\right]-\boldsymbol{T}\right)\frac{\boldsymbol{S}}{\boldsymbol{V}_1} + 2(\boldsymbol{U}+\boldsymbol{V}_1) + \boldsymbol{M}\ e^{-i2\arg(\boldsymbol{V}_1)}\ . \tag{4.19g}$$

However, for computational purposes, the solution method given earlier is simpler.

CHAPTER 5

Using Temporal Information: Rigid Motion

''.... although the instantaneous velocity field contains sufficient information for the recovery of the 3-D shape, the reliable interpretation of local structure from motion requires the integration of information over a more extended time period.''

-Shimon Ullman (1984)

In the previous two chapters we have considered the interpretation of instantaneous image flow. Information about the spatial variation of image flow was used to determine the structure and motion of surfaces. In this chapter we extend the previous formulation to incorporate temporal information. For this purpose we assume that the surface structure and motion in the scene domain and image flow parameters in the image domain are locally smooth in the space-time domain. Use of temporal information is necessary for two reasons. First, in many cases we can trade noise-sensitive spatial information for relatively robust temporal information. For example, consider a small surface patch in motion which is approximately planar. In this case, in general, we can use the first order temporal derivatives of the motion field in place of the second-order spatial derivatives. The second and more important reason is that using temporal information makes it possible for us to deal with more complicated situations such as when the object in view is undergoing non-uniform (or accelerated) motion. Here we consider three important cases where the motion is rigid and uniform and one more case where the motion is non-uniform. Of the three rigid uniform motion cases, the first two relate to the case where the image motion is observed in a fixed image neighborhood and the third relates to the case where the camera tracks a point on the object in motion and the tracking motion of the camera is known. In all three cases we have solved for the local orientation and rigid motion of the surface patch around the line of sight using only the first-order spatial and temporal derivatives of the image velocity field. In comparison, in the previous chapter we used up to second-order spatial derivatives of image flow. Temporal information has scarcely been used in the past for image flow interpretation. Hoffman (1980) used time derivatives of image flow to recover shape and motion of surfaces for orthographic projection. (Our formulation is for the perspective projection case.) Recently Bandopadhyay and Aloimonos (1985) and Wohn and Wu (1986) have used temporal information to solve a restricted case of rigid motion. This case is the one we have considered in Section 5.3.1. A systematic way of incorporating temporal information in the interpretation of image flow was proposed in Subbarao (1986b). This chapter and most of the next chapter are based on this work. The essence of our approach is described in the next section and detailed solution methods are given in the following sections. The main extension that is necessary for the previous formulation is that we have to consider the structure and motion parameters to be functions of time. Each of these parameters is assumed to be changing ''smoothly'' with time so that their time variation in a small time interval can be parameterized by the first few Taylor coefficients of the corresponding function. For example a surface is represented by

$$Z = (Z_0+\dot{Z}_0 t+...) + (Z_X+\dot{Z}_X t+...)X + (Z_Y+\dot{Z}_Y t+...)Y + \cdots \tag{5.1}$$

In this case the scene parameters consists of the initial structure and motion parameters and the time derivatives of the motion parameters. The time derivatives of the structure parameters are determined by the motion parameters and their time derivatives. Therefore the time derivatives of the structure parameters are not a part of the scene parameters. In the image domain the image flow is assumed to be analytic in the space-time domain. The image flow is represented by

$$u(x,y,t) = u_0+u_x x+u_y y+u_t t+O_2\,(x,y,t) \quad \text{and} \tag{5.2a}$$

$$v(x,y,t) = v_0+v_x x+v_y y+v_t t+O_2\,(x,y,t) \tag{5.2b}$$

where the subscripts indicate the corresponding partial derivatives evaluated at the image origin and time $t=0$ and $O_2\,(x,\,y,\,t)$ indicates second and higher order terms of the Taylor series. The coefficients of this Taylor series are the spatio-temporal image flow parameters. Now our goal is to derive the relations between the scene parameters and the image flow parameters and use these relations to recover the scene parameters from the image flow parameters. We illustrate this process for a few simple cases and from this it will be clear how more complicated cases can be tackled by following similar steps. In relation (5.1), neglecting second and higher order terms in X, Y and t we have

$$Z = Z_0 + Z_X X + Z_Y Y + \dot{Z}_0 t \tag{5.3a}$$

Rearranging terms this can be rewritten as

$$Z \left(1 - \frac{X}{Z} Z_X - \frac{Y}{Z} Z_Y - \frac{t}{Z} \dot{Z}_0 \right) = Z_0 \ . \tag{5.3b}$$

Or, using relations (3.1), equation (5.3b) can be rewritten as

$$Z = Z_0 \ (1 - Z_X \ x - Z_Y \ y - \dot{Z}_0 \ (t/Z) \ )^{-1} \tag{5.3c}$$

Substitution of this into the image velocity equations (3.4a,b) gives

$$u = \left(x\frac{V_Z}{Z_0} - \frac{V_X}{Z_0} \right) (1 - Z_X x - Z_Y y - \dot{Z}_0 (t/Z)) + \left[xy\Omega_X - (1+x^2)\Omega_Y + y\Omega_Z \right] \tag{5.4a}$$

$$v = \left(y\frac{V_Z}{Z_0} - \frac{V_Y}{Z_0} \right) (1 - Z_X x - Z_Y y - \dot{Z}_0 (t/Z)) + \left[(1+y^2)\Omega_X - xy\Omega_Y - x\Omega_Z \right] \tag{5.4b}$$

In the above equations, as in the previous formulation, Z_0 always appears in ratio with the translational velocity $\boldsymbol{V}$ and therefore is not recoverable. Therefore, we adopt the same notation as before (see relations 3.9a-c) in presenting the image flow equations. Using relations (5.4a,b) we can derive the relations between the spatio-temporal image flow parameters and the scene parameters. We point out here that the relations between the spatial derivatives of image flow and the instantaneous scene parameters are the same as before (relations 3.11a-l). Therefore relations (3.11a-f) are the same in this case. The solution of relations (3.11a-f) are already given in terms of r and θ (relations 4.3a-h). Here we shall determine r and θ using the temporal derivatives of the image flow. The relations between these and the scene parameters are given for some simple but interesting cases. Also, for each case we give analytic solutions for r and θ. In deriving the equations relating r and θ to u_t and v_t we consider three different cases which are explained below. With the exception of the second case (in a limited sense which will be made clear later) we assume the motion to be uniform, i.e. all orders of the derivatives of $\boldsymbol{V}$ and $\boldsymbol{\Omega}$ with respect to time are zero. In each of these cases, we derive the equations relating u_t, v_t to the structure and motion parameters and use them to solve for θ and r. Assuming that the translational velocity $\boldsymbol{V}$ is uniform with respect to the camera's reference frame (i.e. $\boldsymbol{\Omega} = \boldsymbol{0}$ or $\boldsymbol{\Omega}$ is parallel to $\boldsymbol{V}$) we can derive the following from equations (5.4a,b):

$$u_t = V_x \ (\dot{Z}_0 / Z_0) \quad \text{and} \quad v_t = V_y \ (\dot{Z}_0 / Z_0) \ . \tag{5.5a,b}$$

Using relation (E10a) in Appendix E, we can express $\dot{Z}_0 / Z_0$ in terms of the scene parameters. Therefore, from relations (E10a,3.11a,b), relations (5.5a,b) can be expressed as

$$u_t = V_x \ p \quad \text{and} \quad v_t = V_y \ p \tag{5.6a,b}$$

where

$$p = - \ (u_0 \ Z_X + v_0 \ Z_Y + V_z) \ \ . \tag{5.6c}$$

Now we solve for θ and r using relations (5.6a-c). Taking the ratio of relations (5.6a,b) and using relations (4.3a,b) the solution for θ is obtained as

$$\tan\theta = \frac{v_t}{u_t} \ . \tag{5.7a}$$

We solve for r using relations (5.6a-c), (4.3a,b,e,f) to get

$$r = \frac{(u_t + u_0 a_2 + v_0 a_1)c + (v_t + u_0 a_1 - v_0 a_2)s}{-V_z}. \tag{5.7b}$$

Thus, given the image flow parameters u_0, v_0, u_x, . . . , etc. of equations (5.2a,b), we first solve for θ, V_z and Ω_Z from relations (5.7a,4.3c,d) and then solve for r from relation (5.7b). In this case, there are some situations in which the system of equations (3.11a-f,5.6a-c) becomes under-constrained and so cannot be solved completely. The first situation is when the distance of the surface along the line of sight (given by Z_0) remains constant. In this case p (given by $\dot{Z}_0 / Z_0$, see equation (E10a,E14)) is zero and therefore equations (5.6a,b) degenerate and in their place we get a single constraint $p = 0$. So we find that the system of equations cannot be solved. Another situation is when there is no translation along the line of sight, i.e. $V_z = 0$. In this case r is indeterminate. There are other degenerate cases such as when there is no translation parallel to the image plane $(V_x = V_y = 0)$, when the surface patch is a frontal $(Z_X = Z_Y = 0)$, etc., when the equations are partially solvable and some of the unknowns become undetermined. In a computational algorithm, the presence of such degenerate cases should be detected in the early stages and dealt with (e.g. see Appendix B). Bandopadhyay and Aloimonos (1985) had shown earlier that in this case, given the image velocities at *any* three non-collinear points where their temporal derivatives are non-zero, the motion parameters are uniquely determined. Wohn and Wu (1986) have also solved this case by a different approach. In the previous case we assumed that the translation and rotation are uniform with respect to the camera's reference frame. But in many real world situations, their magnitudes remain the same but their directions change due to the rotation of the camera's reference frame. For example, consider a ball in the air which is moving horizontally with respect to the ground and spinning along an axis not parallel to the direction of translation (see Figure 18). Suppose that during a short time interval the motion of the ball can be considered uniform (ignoring gravity) in the world reference frame. Then the relative translation of the ground as seen from a reference frame fixed with respect to the ball changes continuously in direction with time due to the rotation of the ball, although the magnitude remains the same. The solution method in this case is similar to that in the previous case except that the expressions for u_t and v_t are more complicated than before. In deriving expressions for u_t and v_t from relations (5.4a,b) we consider $\boldsymbol{V}$ and $\boldsymbol{\Omega}$ to be functions of time t. The rates of change of $\boldsymbol{V}$ and $\boldsymbol{\Omega}$ with time are given by

$$\dot{\boldsymbol{V}} = \boldsymbol{V} \times \boldsymbol{\Omega} \quad \text{and} \quad \dot{\boldsymbol{\Omega}} = \boldsymbol{\Omega} \times \boldsymbol{\Omega} = 0\,. \tag{5.8a,b}$$

Using the above relations, we can derive expressions for u_t and v_t from relations (5.4a,b) to be

$$u_t = V_z\, \Omega_Y - V_y\, \Omega_Z + V_x\, p \quad \text{and} \tag{5.9a}$$

$$v_t = V_x\, \Omega_Z - V_z\, \Omega_X + V_y\, p \tag{5.9b}$$

where p is, as before, given by relation (5.6c). Relations (5.9a,b) can be used to solve for θ and r. Using relations (4.3a,b,e,f), the right hand sides of equations (5.9a,b) can be expressed in terms of θ, r, V_z and Ω_Z. From the resulting equations we can solve for r to get

$$r = \frac{u_t + u_0\, V_z + c\, q}{-(\Omega_Z\, s + 2\, V_z\, c)} \tag{5.10a}$$

and

$$r = \frac{v_t + v_0\, V_z + s\, p}{\Omega_Z\, c - 2\, V_z\, s} \tag{5.10b}$$

where

$$q \equiv u_0\, (a_1\, s + a_2\, c) + v_0\, (a_1\, c - a_2\, s) \tag{5.10c}$$

Equating the right hand sides of the two equations (5.10a,b), substituting for V_z and Ω_Z in terms of θ using relations (4.3c,d), and simplifying, we can derive a fifth degree equation in $\tan\theta$. This derivation is given in Appendix H. θ is obtained by solving for the roots of the fifth degree polynomial. Therefore θ

may have up to five solutions, but requiring the solution to be consistent over time should give a unique solution in most cases. For example, if Z_X and Z_Y are the slope components of the surface patch at time $t = 0$, then these components after a small time dt should be (approximately) $Z_X + \dot{Z}_X\, dt$ and $Z_Y + \dot{Z}_Y\, dt$ where $\dot{Z}_X$ and $\dot{Z}_Y$ are given by relations (E10b,c). Having solved for θ we solve for V_z and Ω_Z from equations (4.3c,d). We then solve for r from either (5.10a) or (5.10b). In this case, there are two special situations which deserve mention. In both these cases, the orientation of the surface patch is indeterminate as there is no translation parallel to the image plane. These cases are summarized in Appendix H. One case of special interest is when the camera system deliberately tracks a point on the object's surface along the line of sight by rotating around the focus (about an axis perpendicular to the line of sight) and the voluntarily induced angular velocity and acceleration of the camera in order to track the point are known. If $\boldsymbol{\omega}$ and $\dot{\boldsymbol{\omega}}$ are respectively the angular velocity and acceleration of the camera, then the image velocity and acceleration of the point being tracked with respect to a ''stationary'' camera are given by $\boldsymbol{\omega} \times \hat{\boldsymbol{k}}\, f$ and $\dot{\boldsymbol{\omega}} \times \hat{\boldsymbol{k}}\, f$ where f is the focal length of the camera. In this case, due to the tracking motion of the camera, $\boldsymbol{V}$ and $\boldsymbol{\Omega}$ are changing with time in a complex manner. In this situation, the image velocity field in a small neighborhood around the image of the point being tracked over a short duration of time is given by

$$u\,(x,\,y,\,t) = (u_0 + \dot{u}\; t) + u_x\, x + u_y\, y + O_2\,(x,\,y,\,t) \quad \text{and} \tag{5.11a}$$

$$v\,(x,\,y,\,t) = (v_0 + \dot{v}\; t) + v_x\, x + v_y\, y + O_2\,(x,\,y,\,t) \tag{5.11b}$$

where $(\dot{u},\,\dot{v})$ is the acceleration of the image of the point being tracked at time $t = 0$. Notice that the above expressions are similar to relations (5.2a,b) except that $u_t,\, v_t$ are replaced by $\dot{u},\,\dot{v}$ respectively. The expressions for $\dot{u}$ and $\dot{v}$ are obtained from equations (5.4a,b) by considering x and y to be functions of time t (i.e. $x = X(t)\,/\,Z(t)$ and $y = Y(t)\,/\,Z(t)$) and differentiating and evaluating at the image origin and $t = 0$. Alternatively, they can be obtained by directly differentiating relations (3.1a,b) twice with respect to t and evaluating at the image origin. This has been derived in Appendix I to be

$$\dot{u} = v_0\;\Omega_Z + u_0\; V_z - V_x\; V_z \quad \text{and} \tag{5.12a}$$

$$\dot{v} = v_0\; V_z - u_0\;\Omega_z - V_y\; V_z\;. \tag{5.12b}$$

In Appendix I the solution for θ is derived to be

$$\tan\theta = \frac{v_0\; v_y + v_x\; u_0 - \dot{v}}{u_x\; u_0 + u_y\; v_0 - \dot{u}}\;. \tag{5.13}$$

Having solved for θ from the above equation, we solve for $V_z,\,\Omega_Z$ (using relations 4.3c,d). In terms of these quantities the solution for r is shown (in Appendix H) to be

$$r = u_0 c + v_0 s + \frac{(v_0\Omega_Z - \dot{u})c - (u_0\Omega_Z + \dot{v})s}{V_z}. \tag{5.14}$$

In this case, we find that when there is no translation along the line of sight (i.e. $V_z = 0$), r and θ are indeterminate. Although Ω_Z can be computed as $\dot{u}/v_0$ (or $-\dot{v}/u_0$) all other parameters of motion and structure remain undetermined. The problem of interpreting *instantaneous* image flow when a binocular camera tracks a feature point has been considered by Bandopadhyay, Chandra, and Ballard (1986). In the previous examples we have restricted the time dependence of the motion parameters. In general they can be arbitrary (but analytic) functions of time. We can in principle deal with these cases. Solving the general case involves using second and higher order image flow derivatives. Here we illustrate the method with a simple example which involves only first order image flow parameters. In this example, we restrict the situation in the following ways: no relative rotation between the camera and the surface patch, the surface is rigid and the translational acceleration is uniform. In this case, we can derive the following equations from equations (5.4a,b):

$$u_0 = -\,V_x\;, \qquad v_0 = -\,V_y\;, \tag{5.15a,b}$$

$$u_x = V_z + V_x\ Z_X\ , \quad v_y = V_z + V_y\ Z_Y\ , \tag{5.15c,d}$$

$$u_y = V_x\ Z_Y\ , \quad v_x = V_y\ Z_X\ , \tag{5.15e,f}$$

$$u_t = -\frac{\partial V_x}{\partial t} \quad \text{and} \quad v_t = -\frac{\partial V_y}{\partial t}\ . \tag{5.15g,h}$$

The term $\frac{dV_Z}{dt}$ corresponding to the acceleration along the line of sight does not appear in the above equations and therefore is not recoverable from the available information (knowing u_{xt} or v_{yt} would make it possible for us recover this term). Equations (5.15a-h) are overdetermined (eight equations in seven unknowns). Solving these equations is straightforward.

CHAPTER 6

The General Formulation

''The formulation is mathematically equivalent to more usual foundations. There are, therefore, no fundamentally new results. However, there is a pleasure in recognizing old things from a new point of view. Also, there are problems for which the new point of view offers a distinct advantage.'' -Richard P. Feynmann (1948)

Until now we have only considered rigid motion of objects. In this chapter we consider the general case of non-rigid motion. Restricted types of non-rigid motion problem has been addressed by some of researchers Koenderink and Van Doorn (1986), Chen (1985), and Ullman (1984). General non-rigid motion problem was recently formulated in Subbarao (1986b). A refined and extended version of this formulation will be presented here. The formulation for the non-rigid motion case is basically an extension of the rigid motion case. The primary difference is that here the instantaneous velocities of points on surfaces in the scene are considered to be functions of their positions in the scene. The formulation of a general non-rigid motion case has two stages: (i) the representation of non-rigid motion of surfaces, and (ii) relating the non-rigid motion parameters to the changing image flow in space and time. Here we describe the non-rigid motion of a small surface patch in terms of the deformation and motion of a small volume element embedding the surface patch (see Figure 19). This is an adequate representation because given the deformation parameters of the volume element the deformation of the embedded surface is computable (see Appendix J for more discussion of this). In fact we can recover from the image flow field only those deformation parameters which affect the embedded surface patch and in any case this is all that we want. For example, for a planar surface patch, the extension (or contraction) of a line segment normal to the planar surface is not recoverable from the image flow and we don't need it anyway because it has no effect on the surface patch. An alternative representation of surface deformation can be obtained by using a curvilinear coordinate system fixed in the surface. In this system, geometric points on the surface are labelled by two independent parameters and the partial derivatives of the velocities of material particles on the surface with respect to these parameters represent the surface deformation parameters (see Aris, 1962; Waxman, 1984a; McConnell, 1957). But the velocity gradient tensor representation we have used is simpler and may be more desirable because the deformation of a surface in the physical world is often due to deformation of the 3D object of which it forms a part. Let the instantaneous three-dimensional velocities of points in a small volume embedding a small surface patch along the line of sight be given by $\boldsymbol{U} = (\dot{X},\ \dot{Y},\ \dot{Z})$ where

$$\dot{X} = a_{10} + a_{11} X + a_{12} Y + a_{13} (Z - Z_0) + O_2(X, Y, Z) \tag{6.1a}$$

$$\dot{Y} = a_{20} + a_{21} X + a_{22} Y + a_{23} (Z - Z_0) + O_2(X, Y, Z) \tag{6.1b}$$

$$\dot{Z} = a_{30} + a_{31} X + a_{32} Y + a_{33} (Z - Z_0) + O_2(X, Y, Z)\ . \tag{6.1c}$$

In the above expressions the last terms denote the second and higher order terms with respect to $X,\ Y$ and Z. The 3×3 matrix defined by a_{ij} for $1 \le i,\ j \le 3$ is in fact the spatial *velocity gradient tensor* at the point $(0,\ 0,\ Z_0)$. An intuitive interpretation of this velocity gradient tensor and a_{ij} are given in Appendix J. Comparing the above expressions for a general non-rigid motion to relations (3.2a-c) for a rigid motion, we see that in the case of rigid motion

$$a_{11} = a_{22} = a_{33} = 0\ , \quad a_{23} = -a_{32} = \Omega_X\ , \tag{6.2a,b}$$

$$a_{31} = -a_{13} = \Omega_Y \quad \text{and} \quad a_{12} = -a_{21} = \Omega_Z\ . \tag{6.2c,d}$$

Therefore, non-zero values for the terms $a_{11},\ a_{22},\ a_{33},\ a_{12} + a_{21},\ a_{13} + a_{31}$, and $a_{23} + a_{32}$, imply a non-rigid motion. Substituting for Z from equation (5.3a) in equations (6.1a-c) and rearranging terms we obtain

$$\dot{X} = a_{10} + (a_{11} + a_{13} Z_X) X + (a_{12} + a_{13} Z_Y) Y + a_{13} Z_t t + O_2(X, Y, Z, t) \tag{6.3a}$$

$$\dot{Y} = a_{20} + (a_{21} + a_{23} Z_X) X + (a_{22} + a_{23} Z_Y) Y + a_{23} Z_t t + O_2(X, Y, Z, t) \tag{6.3b}$$

$$\dot{Z} = a_{30} + (a_{31} + a_{33} Z_X) X + (a_{32} + a_{33} Z_Y) Y + a_{33} Z_t t + O_2(X, Y, Z, t) \; . \tag{6.3c}$$

Now we wish to solve for a_{ij} and the local surface structure given the image flow field. Here, as there are more unknowns than before, we have to consider terms in the Taylor series expansion of the image velocity field beyond first order (see relations (5.2a,b)). The coefficients of this Taylor series are the new image flow parameters. The relations between these image flow parameters and the deformation, motion and local surface structure parameters are derived by a method similar to that in the rigid motion case (Chapter 3) except that in this case $\dot{X}$, $\dot{Y}$ and $\dot{Z}$ are taken to be as in relations (6.3a-c) instead of (3.2a-c). We illustrate this for a simple case where we need to consider only the first order image flow parameters. The general case is considered later. Consider a simple situation where a surface patch along the direction of view of a camera is translating uniformly and expanding (or contracting) along the X and Y directions. In this case we have

$$a_{12} = a_{13} = a_{21} = a_{23} = a_{31} = a_{32} = a_{33} = 0 \; , \tag{6.4a}$$

$$a_{10} = -V_X \; , \quad a_{20} = -V_Y \; , \quad a_{30} = -V_Z \; , \tag{6.4b-d}$$

$$a_{11} = -\frac{\partial V_X}{\partial X} \; , \quad \text{and} \quad a_{22} = -\frac{\partial V_Y}{\partial Y} \; . \tag{6.4e,f}$$

Now we can derive the following from equations (5.4a,b):

$$u_0 = -V_x \; , \quad v_0 = -V_y \; , \tag{6.5a,b}$$

$$u_x = V_z + V_x Z_X - \frac{\partial V_X}{\partial X} \; , \quad v_y = V_z + V_y Z_Y - \frac{\partial V_Y}{\partial Y} \; , \tag{6.5c,d}$$

$$u_y = V_x Z_Y \; , \quad v_x = V_y Z_X \; , \tag{6.5e,f}$$

$$u_t = V_x s' \quad \text{and} \quad v_t = V_y s' \tag{6.5g,h}$$

where

$$s' = V_x Z_X + V_y Z_Y - V_z \; . \tag{6.5i}$$

Equations (6.5) are eight equations in seven unknowns. The equations are overdetermined because of the restricted type of motion and deformation we have assumed. Solving these equations is straightforward. Combined non-rigid and non-uniform motions can be analyzed by considering the parameters a_{ij} in the previous case to be functions of time. This modification is similar to our extension in the previous chapter from uniform motion to non-uniform motion. In this case the velocity field in the scene domain takes the form:

$$\dot{X} = a_{10} + a_{11} X + a_{12} Y + a_{13} (Z - Z_0) + a_{14} t + \cdots \tag{6.6a}$$

$$\dot{Y} = a_{20} + a_{21} X + a_{22} Y + a_{23} (Z - Z_0) + a_{24} t + \cdots \tag{6.6b}$$

$$\dot{Z} = a_{30} + a_{31} X + a_{32} Y + a_{33} (Z - Z_0) + a_{34} t + \cdots \tag{6.6c}$$

Further, we can use finer local surface models (quadric, cubic, etc.) by considering longer Taylor series expansions of the surface expressed in the form $Z(X, Y, t)$ (see equation (5.1)). However, in this case we need to know how the general spatio-temporal derivatives of image flow are related to the scene parameters. We consider this next. In order to derive the equations relating the time and space-time derivatives of image flow to the scene parameters, it is necessary to know how the corresponding surface structure parameters are changing with time; i.e., if the surface is represented by

$$Z = b_0 + b_1 X + b_2 Y + b_3 X^2 + ... \qquad \textbf{(6.7)}$$

then we need to know how $b_0, b_1, b_2, ...$ are changing with time (compare the above equation with equation (3.5)). For this purpose, if the transformation of the surface in the scene is assumed to be ''smooth'' and analytic with time, then each of the surface structure parameters can be expressed in a Taylor series expansion as

$$b_i = b_{i\,0} + b_{i\,1} t + b_{i\,2} t^2 + \cdots \qquad \text{for } i = 0, 1, 2, ... \qquad (6.8)$$

(Compare the above with terms within parantheses in equation (5.1)). The transformation can be any combination of translation, rotation, deformation, and higher order variations of these quantities. The time dependence of these parameters is determined by the scene transformation parameters, i.e., a_{ij}. If a_{ij} are themselves changing with time then we can express them as functions of time as in the case of b_i (see equation (6.8)). The first order dependence of the structure parameters on time, denoted by $b_{i\,1}$, can be related to a_{ij} as follows. Differentiating equation (6.7) we obtain

$$\dot{Z} = \dot{b}_0 + \dot{b}_1 X + b_1 \dot{X} + \dot{b}_2 Y + b_2 \dot{Y} + \dot{b}_3 X^2 + ... \qquad (6.9)$$

In the above expression we substitute for $\dot{X}, \dot{Y}, \dot{Z}$ using equations (6.6a-c) respectively and then substitute for Z from equation (6.7). Simplifying the resulting expression we can obtain an expression of the form

$$C_0 + C_1 X + C_2 Y + C_3 X^2 + ... = 0. \qquad (6.10)$$

We can equate each of the coefficients C_i to zero in the above expression since the above equation should hold for every (X, Y) value. Using the set of equations $C_i = 0$ at time zero we can explicitly express $b_{i\,1}$ in terms of a_{ij} (see equations (E8-E10) as examples in Appendix E). In order to derive the second order dependence of the structure parameters we differentiate equation (6.9) and follow steps similar to the previous one. In general this method can be used to express all b_{ij} for $j > 0$ in terms of a_{ij}. Having obtained these relations, the equation of the surface as a function of time is given by equations (6.7) and (6.8). Using this representation of the surface, the equations relating the time and space-time derivatives to the scene parameters (i.e. $b_{i\,0}, a_{ij}$) can be derived. The method is similar to that of the rigid motion case in Chapter 3. The result is that we have a general method for obtaining the relation between the image flow parameters of any order and the scene parameters. Solving these equations constitutes the interpretation of image flow. As we generalize our method to incorporate more general motions (non-rigid, non-uniform, etc.) and finer local surface patch models (quadric, cubic, etc.) more unknown parameters are introduced. In these situations, the general principle is to consider sufficiently long Taylor series expansions of the image velocity field so that enough equations are available to solve for all the unknowns. Typically, each Taylor series coefficient of the image velocity field yields one equation (sometimes all the equations may not be independent as some of them may yield extra constraints such as a rigidity constraint, etc.). These Taylor series coefficients are to be extracted from the given image velocity field. Since this image velocity field is itself estimated from an image sequence, the longer the Taylor series, the higher the desired quality of the input image data (in terms of spatial and grey-level resolution). The problem of image flow interpretation in its general form is *inherently ill-posed* or under-constrained. In order to see this we first observe that, in general, each image flow parameter (which is assumed to be known) gives one image flow equation for the scene parameters (the unknowns). If we consider up to nth order Taylor coefficients in equations (5.2a,b), then it can be shown that we get $2\binom{n+3}{3}$ image flow equations where $\binom{n}{r}$ denotes the obvious binomial coefficient. In these equations, all scene parameters up to nth order will appear. Therefore the number of unknowns is obtained by summing the number of a_{ij} in equations (6.6a-c) and the number of b_i in equation (6.7) for $i > 0$ (b_0 has been taken to be the scaling factor which is indeterminate; see earlier discussion in Section 5.2 following equation (5.4b)). This sum can be shown to be $3\binom{n+4}{4} + \binom{n+2}{2} - 1$. Therefore, the number of equations increase as $O(n^3)$ whereas the number of unknowns increase as $O(n^4)$. Thus, for any given order of image flow parameters the number of equations is lower than the number of unknowns (see Table 1). In order to solve for the unknowns we will have to impose additional constraints on the scene parameters. For example, consider the case in section 5.4 where $n = 1$. The above formulas give eight image flow equations and seventeen unknowns. Now the rigidity assumption gives effectively six additional equations represented by equations (6.2a-d). The assumption that the motion is uniform (i.e. acceleration is zero with respect to an external reference frame) gives three additional equations (one for each component of $\dot{\mathbf{V}}$) as in equation (5.8a). (Note: $\dot{\mathbf{\Omega}}$ is a second order scene parameter and therefore equation (5.8b) does not give additional constraints.) Therefore we arrive at a situation where the number of equations exactly match the number of unknowns (seventeen each). (Only for $n=0$ we will have to consider an object centered coordinate system for our formulas to hold (in this paper we are using a camera centered coordinate system). In this case Ω_X and Ω_Y will not appear in equations that correspond to (3.11a,b)). In practice the assumptions of rigidity of motion, local planarity of surfaces, and constancy of motion with respect to time are useful. In general some model of the scene parameters is required for the interpretation process.

Order of Taylor coefficients:	0	1	2	3	4	..
Number of equations:	2	8	20	40	70	..
Number of unknowns:	3	17	50	114	224	..

Table 1. The numbers in the first row represent the maximum order of the Taylor coefficients considered for the scene parameters and the image flow parameters. We see that under any given column, the number of unknowns exceeds the number of equations.

‘‘The critical step in formulating the computational theory of stereopsis is the discovery of additional constraints on the process that are imposed naturally and that limit the result sufficiently to allow a unique solution. Finding such constraints is a true discovery-- the knowledge is of permanent value, it can be accumulated and built upon, and it is in a deep sense what makes this field of investigation into a science.’’ -David Marr (1977)

CHAPTER 7

Error Sensitivity and Numerical Examples

In the computational approach described here, the solution for the image flow equations is given by *explicit analytic expressions.* Therefore, a theoretical sensitivity analysis can be done by taking *error differentials.* If the uncertainty in the input image flow parameters are known, then approximate bounds on the maximum error in the 3D structure and motion parameters can be estimated. This analysis holds for *all* cases. In contrast, the sensitivity analyses of previous approaches are based on a few numerical examples; a general analysis was not possible as closed-form solutions were not available (Adiv, 1985a,b; Waxman and Ullman 1985; Waxman and Wohn 1985). The error in the output 3D structure and motion parameters depend on the uncertainty in the input flow parameters. The errors in the flow parameters in turn depend on two factors: the image quality (in terms of spatial resolution, gray level resolution, and the noise level in grey level and pixel registration) and the computational method employed in estimating the flow derivatives. If a tolerance is specified for the output structure and motion, then it is in principle possible to obtain an approximate idea of the required image quality. A sensitivity analysis based on error differentials gives only the worst case behavior. Therefore such an analysis is often inadequate in practical applications. A more accurate analysis is difficult unless a domain of application is specified. This difficulty arises from the non-linear nature of the problem. The maximum absolute error in the computation of an analytic function can be estimated using the total differential of the function (cf. Piskunov, 1974). Let $y = f(x_1, x_2, ..., x_n)$ be an analytic function and $\Delta x_1, \Delta x_2, ..., \Delta x_n$ be the errors in the corresponding arguments. Then, for sufficiently small absolute values of $\Delta x_1, \Delta x_2, ..., \Delta x_n$, the error Δy in y can be shown to satisfy the relation

$$|\Delta y| \le \left|\frac{\partial f}{\partial x_1}\right||\Delta x_1| + \left|\frac{\partial f}{\partial x_2}\right||\Delta x_2| + ... + \left|\frac{\partial f}{\partial x_n}\right||\Delta x_n| \tag{7.1}$$

Relation (7.1) can be used to estimate the maximum absolute errors in the scene parameters given the uncertainties in the image parameters. For example, consider the estimation of error in $\tan\theta$ by solving the cubic equation (4.12a). Let the cubic equation be represented by

$$x_1 + x_2 y + x_3 y^2 + x_4 y^3 = 0\ . \tag{7.2}$$

Then, using relation (7.1) we can obtain

$$|\Delta y| \le \frac{|\Delta x_1| + |\Delta x_2||y| + |\Delta x_3||y^2| + |\Delta x_4||y^3|}{|x_2 + 2x_3 y + 3x_4 y^2|}. \tag{7.3}$$

Therefore, if the uncertainties in the coefficients of the cubic equation are given, then the uncertainty in the roots can be estimated. Errors in the other unknowns can be estimated similarly. The actual error is usually much smaller than that given by relation (5.1). We give here the result of an experiment where the correct solution and the solution obtained from noisy input are given. This example is included to give an idea about the sensitivity of the approach. Bounds on the error are not estimated.

Example: The estimated image flow parameters for a curved surface in rigid motion using the *velocity functional method* (Waxman and Wohn, 1985) are given (from Wohn, 1984). The estimation was based on a noisy (5%) normal velocity field along a contour. The contour spanned approximately a ten degree field of view. The result of solving the image flow equations in this case is given below.

Noisy input image flow parameters:

u_0 :	-6.017523	u_x :	2.999429	u_y :	-0.088125
u_{xx} :	3.009457	u_{xy} :	0.023188	u_{yy} :	3.052289
v_0 :	-3.965109	v_x :	0.087267	v_y :	2.997560

v_{xx} : 2.020104 v_{xy} : -0.015687 v_{yy} : 2.046979

Solution of image flow equations:

(V_x, V_y, V_z) : (5.991399 3.984022 2.998529)
(O_x, O_y, O_z) : (0.018913 0.026124 -0.087696)
(Z_X, Z_Y) : (-0.000008 -0.000005)
(Z_{xx}, Z_{yy}, Z_{xy}) : (0.506775 0.001294 0.509323)

Original scene parameters:

Z_0 : 1.000000
(V_x, V_y, V_z) : (6.000000 4.000000 3.000000)
(O_x, O_y, O_z) : (0.034966 0.017453 -0.087266)
(Z_X, Z_Y) : (0.000000 0.000000)
(Z_{xx}, Z_{yy}, Z_{xy}) : (0.500000 0.00000 0.500000)

Here we see that that the computed solution compares well with the original values.

In general it has been observed that the image flow parameters recovered by Wohn (1984) are very robust for planar surfaces, but are error sensitive for curved surfaces, perhaps due to the inadequate spatial resolution of the images. Consequently the 3D interpretation is reliable for planar surfaces but is unsatisfactory for curved surfaces. The solution method described here was implemented on a Symbolics 3600 computer. Random values of motion and structure parameters were generated and the image flow parameters were computed using relations (3.11a-l). These image flow parameters were given as input to a program to solve the image flow equations. The program was successfully run on hundreds of test examples. Many of these examples were specifically designed (using the results in Appendix B and G) to produce the cases of multiple interpretations. Some non-trivial examples of these cases are given here. The validity of each of these examples is easily verified by computing the image motion parameters for the different solutions using relations (3.11a-l) and comparing them to the input image motion pameters. (All values are rounded to the sixth decimal place.)

Example 1 : For a curved surface with non-zero translation parallel to the image plane, there are three solutions if the surface is a specular saddle and the mean curvature is -1.

Input image motion parameters:

u_0 : 9.560000 v_0 : 13.570000 u_x : -9.140000 v_x : 8.960000
u_y : -8.960000 v_y : -9.140000 u_{xx} : 14.563000 v_{xx} : -3.402000
u_{xy} : -5.825180 v_{xy} : -40.404280 u_{yy} : 4.557000 v_{yy} : 30.542000

The set of solutions for (θ, r):

{ (-0.035108 , -50.740273) , (1.381851 , -10.399291) , (1.329556 , -7.785441) }.

Solution 1:
(V_x, V_y, V_z) : (-50.709006 , 1.781027 , -9.140000)
(O_X, O_Y, O_Z) : (15.351027 , 41.149006 , -8.960000)
(Z_X, Z_Y) : (0.000000 , 0.000000)
(Z_{xx}, Z_{yy}, Z_{xy}) : (-1.910134 , -0.089866 , 0.417602)

Solution 2:
(V_x, V_y, V_z) : (-1.953224 , -10.214214 , -9.140000)
(O_X, O_Y, O_Z) : (3.355786 , -7.606776 , -8.960000)
(Z_X, Z_Y) : (0.000000 , 0.000000)
(Z_{xx}, Z_{yy}, Z_{xy}) : (0.333065 , -2.333065 , 4.700416)

Solution 3:
(V_x, V_y, V_z) : (-1.860000 , -7.560000 , -9.140000)
(O_X, O_Y, O_Z) : (6.010007 , -7.700000 , -8.960000)
(Z_X, Z_Y) : (0.000000 , 0.000000)
(Z_{xx}, Z_{yy}, Z_{xy}) : (0.450000 , -2.450000 , 6.363006)

Example 2 : For a curved surface with non-zero translation, there are two solutions if it is saddle or cylindrical, there is no translation along the line of sight, its mean scaled curvature is -1 and the slopes and curvatures are related by relation (G17).

Input image motion parameters:

u_0 : -13.090000 v_0 : 10.130000 u_x : -2.025000 v_x : 6.870000
u_y : -0.618000 v_y : -3.050400 u_{xx} : 2.625994 v_{xx} : -12.576974
u_{xy} : -23.874837 v_{xy} : 10.120568 u_{yy} : -28.805994 v_{yy} : 32.836974

The set of solutions for (θ, r):

{ (-1.187512 , 31.733480) , (-0.545848 , 4.738576) }

Solution 1 :
(V_x, V_y, V_z) : (11.867317 , -29.430945 , 0.000000)
(O_X, O_Y, O_Z) : (-19.300945 , 1.222683 , -1.848000)
(Z_X, Z_Y) : (-0.170637 , 0.103646)
(Z_{xx}, Z_{yy}, Z_{xy}) : (0.427338 , -2.427338 , -0.385419)

Solution 2:
(V_x, V_y, V_z) : (4.050000 , -2.460000 , 0.000000)
(O_X, O_Y, O_Z) : (7.670000 , 9.040000 , -5.640000)
(Z_X, Z_Y) : (-0.500000 , 1.240000)
(Z_{xx}, Z_{yy}, Z_{xy}) : (5.112591 , -7.112591 , -7.788849)

Example 3 : There are four solutions when the vector given by (slope along the X-axis, slope along the Y-axis, 1+mean scaled curvature) is parallel to the translation vector.

Input image motion parameters :

u_0 : -8.300000 v_0 : 2.530000 u_x : 8.886919 v_x : -4.112594
u_y : -1.332594 v_y : 3.122606 u_{xx} : -30.694250 v_{xx} : 3.251107
u_{xy} : -12.390270 v_{xy} : -4.858356 u_{yy} : 14.094250 v_{yy} : 1.808893

The set of solutions for (θ, r):

{ (-0.378468 , 2.733441) , (-0.378468 , -5.917901)
(1.192328 , -15.826524) , (1.192328 , -4.995007) }.

Solution 1:
(V_x, V_y, V_z) : (2.540000 , -1.010000 , 2.040000)
(O_X, O_Y, O_Z) : (1.520000 , 5.760000 , 1.390000)
(Z_X, Z_Y) : (2.695638 , -1.071887)
(Z_{xx}, Z_{yy}, Z_{xy}) : (-3.218917 , 5.548918 , -6.337371)

Solution 2:
(V_x, V_y, V_z) : (-5.499101 , 2.186651 , 2.040000)
(O_X, O_Y, O_Z) : (4.716651 , 13.799101 , 1.390000)
(Z_X, Z_Y) : (-1.245098 , 0.495098)
(Z_{xx}, Z_{yy}, Z_{xy}) : (1.486797 , -2.563010 , 2.927191)

Solution 3:
(V_x, V_y, V_z) : (-5.847863 , -14.706508 , 9.969525)
(O_X, O_Y, O_Z) : (-12.176508 , 14.147863 , 1.390000)
(Z_X, Z_Y) : (0.185128 , 0.465571)
(Z_{xx}, Z_{yy}, Z_{xy}) : (-0.221066 , -2.410154 , -0.757158)

Solution 4:
(V_x, V_y, V_z) : (-1.845643 , -4.641518 , 9.969525)
(O_X, O_Y, O_Z) : (-2.111518 , 10.145643 , 1.390000)
(Z_X, Z_Y) : (0.586574 , 1.475146)
(Z_{xx}, Z_{yy}, Z_{xy}) : (-0.700440 , -7.636498 , -2.399032)

Example 4 : Given the first order spatial and temporal image flow derivatives for a rigid motion case where the angular velocity and the magnitude of translation are constant with time, but the direction of translation changes due to angular velocity (see Section 5.4), up to five interpretations are possible.

For this case about fifty random numerical examples were generated. For these examples, it was found that, most often the number of possible interpretations was three (about three out of four); occasionally (about one out of five) there were five possible interpretations, and in a few cases (about one out of twenty) the interpretation was unique. Below we give one case where there are five possible interpretations:

Input image motion parameters:

u_0 : -9.150000 v_0 : -8.970000 u_x : 54.466200 v_x : 21.655800
u_y : 0.062400 v_y : -1.488400 u_t : 304.508958 v_t : 303.101922

The set of solutions for (θ, r):

{ (1.129612 , 0.762626) , (0.6196659 , 4.598889) , (0.545963 , 9.899050)
(0.235251 -31.567504) , (-1.014546 , 6.088551) }

Solution 1:
(V_x, V_y, V_z) : (3.214788 , -5.170647 , 48.605154)
(O_X, O_Y, O_Z) : (-14.140647 , 5.935212 , -31.082672)
(Z_X, Z_Y) : (1.823151 , 9.688064)

Solution 2:
(V_x, V_y, V_z) : (-30.698008 , -7.357963 , -3.371209)
(O_X, O_Y, O_Z) : (-16.327963 , 39.848008 , -7.792830)
(Z_X, Z_Y) : (-1.884077 , -0.255887)

Solution 3:
(V_x, V_y, V_z) : (8.460000 , 5.140000 , 3.960000)
(O_X, O_Y, O_Z) : (-3.830000 , 0.690000 , 9.030000)
(Z_X, Z_Y) : (5.970000 , -1.060000)

Solution 4:
(V_x, V_y, V_z) : (3.743829 , 2.670866 , 7.116296)
(O_X, O_Y, O_Z) : (-6.299134 , 5.406171 , 12.123849)
(Z_X, Z_Y) : (12.647452 , -3.221688)

Solution 5:
(V_x, V_y, V_z) : (0.325650 , 0.689602 , 37.877619)
(O_X, O_Y, O_Z) : (-8.280398 , 8.824350 , 17.707709)
(Z_X, Z_Y) : (57.081497 , -54.184914)

''The purpose of computing is insight, not numbers.''

-Richard W. Hamming

CHAPTER 8

Conclusions

We started with the problem of determining the shape and rigid motion of a surface given the visual motion. We first solved this problem for planar and curved surfaces from instantaneous visual motion, and then solved it for constant space motion using visual motion in a small space-time interval. Finally, we generalized our formulation to arbitrary structure and transformation of surfaces. We found that the general problem is inherently under-constrained and that additional constraints on the scene are necessary to solve the problem. From a practical standpoint, perhaps the most important part of this research is that related to the rigid motion of surfaces. The assumption of rigid motion or ‘‘nearly rigid motion’’ holds in most cases in the world. The existence of a closed-form solution in this case suggests that a visual system can solve this problem quickly, and using minimal processing hardware. The many results concerning the multiplicity of interpretations suggests that cooperation between processors processing ‘‘adjacent regions’’ of visual motion can expedite obtaining a unique interpretation. Also, the computational approach developed for this case is potentially useful in building machine vision systems. It provides a theoretical basis for further investigations into designing robust methods for interpreting visual motion. From a theoretical viewpoint, however, the generalized formulation is important. It reveals the inherent illposedness of the problem. It makes explicit the solvability of the problem for a given set of constraints on the scene. This gives us clues as to what implicit constraints the human visual system might be using in order to solve the problem. More importantly, it provides an overall theoretical and computational framework from which to view the visual motion problem. The problem of visual motion perception is not yet solved completely. However, we now have clear directions for future work. We know how to relate the scene parameters and the image parameters in a general case; and we know the nature of these equations. In the future several questions need to be addressed. We do not yet have a general solution method for non-rigid motion. Interpretation of binocular visual motion needs to be investigated thoroughly. General methods for the *measurement* of image flow in the *spatio-temporal domain* have not been developed until now. Possible interactions of the visual motion module with other modules of vision are yet to be explored. New ‘‘multi-resolution analysis’’ methods may reduce computation in the measurement as well as the interpretation of image flow.

APPENDIX A

Expressing a surface in terms of image coordinates

Here we give a method of deriving the function that maps image points (x, y) on the image plane to points on the surface in the scene along the optical axis. Since the image at a point (x, y) on the image plane corresponds to the point (xZ, yZ, Z) in the scene, our goal is to express Z in terms of (x, y) and the surface structure parameters. In Longuet-Higgins and Prazdny (1980) and Waxman and Ullman (1985) Z was so expressed only up to second order terms of (x, y). Below we give a systematic method which can be used to express Z up to any desired order of terms in (x, y). Assuming that the surface is smooth and is given by $Z = f(X, Y)$ we can expand $f(X, Y)$ in a Taylor series:

$$Z = a_0 + a_1 X + a_2 Y + a_3 X^2 + a_4 XY + a_5 Y^2 + a_6 X^3 + \ldots.. \tag{A1}$$

Using equations (3.1a,b) and equation (A1) we can obtain an implicit expression for Z in terms of the image coordinates x, y :

$$Z = a_0 + Z\,(a_1 x + a_2 y + \underline{Z}\,(a_3 x^2 + a_4 xy + a_5 y^2 + Z\,(a_6 x^3 + \ldots..)\,)\,) \tag{A2}$$

Now we systematically substitute for the appropriate Zs on the right hand side to eliminate second and higher order terms in Z on the right hand side of equation (A2). Substituting the entire right hand side of equation (A2) for the Z underlined in equation (A2) we get

$$Z = a_0 + Z\,(a_1 x + a_2 y + (a_0 + Z\,(a_1 x + a_2 y + \ldots.)) \tag{A3}$$

$$(a_3 x^2 + a_4 xy + a_5 y^2 + Z\,(a_6 x^3 + \ldots..)))$$

Rearranging terms in equation (A3) we have

$$Z = a_0 + Z(a_1 x + a_2 y + a_0 a_3 x^2 + a_0 a_4 xy + a_0 a_5 y^2 + a_0 \underline{Z}(a_6 x^3 + \ldots..) \tag{A4}$$

$$+\underline{Z}(a_1 x + a_2 y + \ldots.)(a_3 x^2 + a_4 xy + a_5 y^2 + Z(a_6 x^3 + \ldots..)))$$

We again substitute for the Zs underlined in equation (A4) the entire expression on the right hand side of equation (A4):

$$Z = a_0 + Z(a_1 x + a_2 y + a_0 a_3 x^2 + a_0 a_4 xy + a_0 a_5 y^2 + (a_0^2 a_6 + a_0 a_1 a_3) x^3 + \ldots). \tag{A5}$$

Continuing this recursive substitution procedure, Z can be expressed explicitly in terms of the image coordinates x, y to any required order of terms. Using $O_3(x, y)$ to denote third and higher order terms equation (A5) can be written as

$$Z = a_0 + Z\,(a_1 x + a_2 y + a_0 a_3 x^2 + a_0 a_4 xy + a_0 a_5 y^2 + O_3(x, y))\,. \tag{A6}$$

Rearranging terms in equation (A6) we get

$$a_0 = Z\,(1 - a_1 x - a_2 y - a_0 a_3 x^2 - a_0 a_4 xy - a_0 a_5 y^2 - O_3(x, y))\,. \tag{A7}$$

Equation (A7) can be used to obtain an expression for Z which is explicit up to second order terms:

$$Z = a_0\,[1 - a_1 x - a_2 y - a_0 a_3 x^2 - a_0 a_4 xy - a_0 a_5 y^2 - O_3(x, y)]^{-1}\,. \tag{A8}$$

APPENDIX B

Some degenerate cases and conditions

In this appendix we mainly consider some degenerate cases. Conditions for the presence of these cases are given in terms of the image flow parameters and in each case the solution of the image flow equations is given. This enables us to analyze a general case by precluding the occurrence of these degenerate cases. This strategy of analysis is almost a necessity due to the non-linear nature of the image flow equations. At the end we also give conditions on the image flow parameters for a surface in motion to be planar and curved. The intepretation of image flow in these two cases is considered in later appendices. Below we systematically consider the different cases in a sequential order. The solution is presented in a sequence of theorems and lemmas. Typically, a theorem or a lemma gives the solution when a specified condition is true. The condition part usually specifies that the condition for none of the preceding theorems or lemmas is true but a certain condition specific to this (theorem/lemma) holds. This style of presentation suggests an implementation algorithm based on testing for conditions on the image flow parameters. Whenever the proof of a theorem or a lemma is obvious or simple we have not included the proof here. *Theorem 1 :* **The condition**

$$u_x=v_y=u_y+v_x=u_{xx}-2u_0=u_0-v_{xy}=v_{yy}-2v_0=v_0-u_{xy}=u_{yy}=v_{xx}=0. \quad \textbf{(B1)}$$

is true *if and only if* **there is no translation, i.e.** $V_x=V_y=V_z=0$**.**

Lemma 1: **Under the condition stated in the above theorem the solution of the image flow equations (3.11a-l) is**

$$Z_X,\ Z_Y,\ Z_{xx},\ Z_{yy} \text{ and } Z_{xy} \text{ are indeterminate}, \quad \textbf{(B2a)}$$

$$(V_x,\ V_y,\ V_z) = (0,\ 0,\ 0), \quad \text{and} \quad \textbf{(B2b)}$$

$$(\Omega_X,\ \Omega_Y,\ \Omega_Z) = (v_0,\ -u_0,\ u_y)\ . \quad \textbf{(B2c)}$$

Theorem 2 : **The condition under Theorem 1 is false and**

$$(u_x-v_y=u_y+v_x=u_{xx}-2v_{xy}=v_{yy}-2u_{xy}=u_{yy}=v_{xx}=0) \quad \textbf{(B3)}$$

if and only if **one of the following is true:**
(i) there is no translation parallel to the image plane, i.e.

$$V_x = V_y = 0 \quad \textbf{(B4a)}$$

(ii) the surface is frontal, planar, and $V_x=0$**, i.e.**

$$Z_X = Z_Y = Z_{xx} = Z_{yy} = Z_{xy} = V_x=0\ . \quad \textbf{(B4b)}$$

(iii) the surface is frontal, planar, and $V_y=0$**, i.e.**

$$Z_X = Z_Y = Z_{xx} = Z_{yy} = Z_{xy} = V_y=0\ . \quad \textbf{(B4c)}$$

Lemma 2a : **Under the conditions of the above theorem there can be up to three solutions for the image flow equations (3.11a-l), one solution each for the three cases (B4a),(B4b), and (B4c). The solutions are**

$$(i)\ \ Z_{xx},\ Z_{yy} \text{ and } Z_{xy} \text{ are indeterminate}, \quad \textbf{(B5a)}$$

$$(V_x,\ V_y,\ V_z) = (0,\ 0,\ u_x), \quad \textbf{(B5b)}$$

$$(\mathbf{\Omega}_X,\ \mathbf{\Omega}_Y,\ \mathbf{\Omega}_Z) = (v_0,\ -u_0,\ u_y)\ , \qquad \textbf{(B5c)}$$

$$(Z_X,\ Z_Y) = \left((u_0 - v_{xy})/u_x,\ (v_0 - u_{xy})/u_x \right) \qquad \textbf{(B5d)}$$

$$(ii)\ \ Z_X = Z_Y = Z_{xx} = Z_{yy} = Z_{xy} = 0\ , \qquad \textbf{(B6a)}$$

$$(\ V_x,\ V_y,\ V_z) = (0,\ u_{xy} - v_0,\ u_x)\ , \qquad \textbf{(B6b)}$$

$$(\mathbf{\Omega}_X,\ \mathbf{\Omega}_Y,\ \mathbf{\Omega}_Z) = (u_{xy},\ -v_{xy},\ u_y)\ , \qquad \textbf{(B6c)}$$

and

$$(ii)\ \ Z_X = Z_Y = Z_{xx} = Z_{yy} = Z_{xy} = 0\ , \qquad \textbf{(B7a)}$$

$$(\ V_x,\ V_y,\ V_z) = (v_{xy} - u_0,\ 0,\ u_x)\ , \qquad \textbf{(B7b)}$$

$$(\mathbf{\Omega}_X,\ \mathbf{\Omega}_Y,\ \mathbf{\Omega}_Z) = (u_{xy},\ -v_{xy},\ u_y)\ . \qquad \textbf{(B7c)}$$

Theorem 3 : **Conditions under Theorem 1 and Theorem 2 are both false and**

$$(u_{yy} = v_{xx} = u_{xx} - 2v_{xy} = v_{yy} - 2u_{xy} = 0) \qquad \textbf{(B8)}$$

if and only if **the surface is planar, and there is a finite translation parallel to the image plane, i.e.**

$$Z_{xx} = Z_{yy} = Z_{xy} = 0\ , \qquad \textbf{(B9a)}$$

$$V_x \neq 0 \quad \text{OR} \quad V_y \neq 0\ . \qquad \textbf{(B9b)}$$

This case of planar surfaces is treated in the next appendix. *Theorem 4 :* **If conditions under Theorem 1, Theorem 2, and Theorem 3 are all false, then the surface is curved (i.e. at least one of the curvatures is non-zero) and translation parallel to the image plane is non-zero; i.e.**

$$(\ Z_{xx} \neq 0 \text{ OR } Z_{yy} \neq 0 \text{ OR } Z_{xy} \neq 0\ )\ \ \text{AND} \qquad \textbf{(B10a)}$$

$$V_x \neq 0 \quad \text{OR} \quad V_y \neq 0\ . \qquad \textbf{(B10b)}$$

This case of curved surfaces is considered later (Appendices F and G). The condition part of the above theorem (i.e. conditions under Theorem 1, Theorem 2, and Theorem 3 are all false) will be referred to as the *curved-surface- condition. Theorem 5:* **Suppose that translation parallel to the image plane is not zero and let r and θ be such that**

$$V_x \equiv r\ \cos\theta \ \text{ and }\ V_y \equiv r\ \sin\theta \text{ for } -\pi/2 < \theta \leq \pi/2\ . \qquad \textbf{(B11a,b)}$$

Then, using the notation

$$s \equiv \sin\theta \ \text{ and }\ c \equiv \cos\theta\ , \qquad \textbf{(B12a,b)}$$

the motion and orientation are given by relations (4.3a-h).

Proof : **Relations (4.3a,b,g,h) are easily obtained from relations (3.11a,b) and (B11a,b). From relations (4.2a,b), (3.11c-f), and (B11a,b,B12a,b) we can get**

$$a_1 = rcZ_Y + rsZ_X \ \text{ and }\ a_2 = rcZ_X - rsZ_Y\ . \qquad \textbf{(B13a,b)}$$

Solving for Z_X and Z_Y from above equations, we get relations (4.3e,f). Now, from relations (3.11c), (B11a), and (4.3e) we can get

$$V_z = u_x - a_1 cs - a_2 c^2\ . \qquad \textbf{(B14a)}$$

Or, using relation (4.2b) and the identity $s^2 + c^2 = 1$,

$$V_z = u_x(s^2 + c^2) - a_1 cs - a_2 c^2\ . \qquad \textbf{(B14b)}$$

Relation (4.3c) can be obtained from the above relation. The derivation of relation (4.3d) is similar to

that of relation (4.3c) •

APPENDIX C

Solving the image flow equations: planar surfaces

Throughout this section we assume that the surface in motion is planar, non-frontal, and has a non-zero translation parallel to the image plane. Condition for the presence of this case in terms of the image flow parameters is given in Theorem 3 of Appendix B. We give here a method to solve the image flow equations by first solving for the translation along the line of sight.

Theorem 1 : **The translation along the line of sight V_z is a root of the cubic equation**

$$\alpha^3 + C_1 \alpha^2 + C_2 \alpha + C_3 = 0 \qquad \textbf{(C1)}$$

where, using the notation

$$A = u_0 - v_{xy} \ , \ B = v_0 - u_{xy} \ , \qquad \textbf{(C2a,b)}$$

$$C_1 = -(u_x + v_y) \ , \qquad \textbf{(C3a)}$$

$$C_2 = -(\frac{1}{4}A^2 + \frac{1}{4}B^2 + a_1^2 - u_x v_y) \ , \text{ and} \qquad \textbf{(C3b)}$$

$$C_3 = \frac{1}{4}A^2 v_y + \frac{1}{4}B^2 u_x - \frac{1}{2} a_1 AB \ . \qquad \textbf{(C3c)}$$

Proof : **We eliminate all parameters except V_z from equations (3.11a-f,4.4a,b). From equations (3.11a,b) we have**

$$\Omega_Y = -u_0 - V_x \quad \text{and} \quad \Omega_X = v_0 + V_y. \qquad \textbf{(C4a,b)}$$

Substituting for Ω_X and Ω_Y from above in equations (4.4a,b) and rearranging terms, we find

$$V_x = V_z Z_X - A \quad \text{and} \quad V_y = V_z Z_Y - B \ , \qquad \textbf{(C5a,b)}$$

where A and B are as in (C2a,b).

Using the above relations, we eliminate V_x and V_y from (3.11c-f) to get

$$u_x = V_z + (V_z Z_X - A) Z_X \qquad \textbf{(C6a)}$$

$$v_y = V_z + (V_z Z_Y - B) Z_Y \qquad \textbf{(C6b)}$$

$$a_1 = (V_z Z_Y - B) Z_X + (V_z Z_X - A) Z_Y \ . \qquad \textbf{(C6c)}$$

We have here three equations in three unknowns. Equations (C6a,b) are quadratic in Z_X and Z_Y, respectively, with solutions given in terms of V_z as

$$Z_{X+},\ Z_{X-} = \frac{A \pm \sqrt{A^2 - 4V_z(V_z - u_x)}}{2V_z} \qquad \textbf{(C7a)}$$

$$Z_{Y+},\ Z_{Y-} = \frac{B \pm \sqrt{B^2 - 4V_z(V_z - v_y)}}{2V_z} \ . \qquad \textbf{(C7b)}$$

Substituting for Z_X and Z_Y from (C7a,b) into (C6c) and simplifying yields

$$2\,a_1 V_z + AB = \pm\sqrt{[\,A^2 - 4V_z(V_z - u_x)\,]\ \ [\,B^2 - 4V_z(V_z - v_y)\,]}\ . \tag{C8}$$

Squaring expression (C8) and simplifying gives

$$V_z\ (\ V_z^3 + C_1\ V_z^2 + C_2\ V_z + C_3\ ) = 0\ . \tag{C9}$$

Here we have a quartic equation in V_z. However it can be shown that the coefficient C_3 is equal to zero *if and only if* **the translation along the line of sight is zero. Therefore the quartic equation reduces to the cubic equation (C1) •**

Longuet-Higgins (1984) obtained the same cubic equation by a different approach.

Lemma 1: **All three roots of the cubic equation are real and are given by**

$$\alpha_0,\ \alpha_+,\ \alpha_- = V_z,\ \frac{1}{2}\left\{ \mathbf{v}\cdot\boldsymbol{\lambda}\ \pm\ \sqrt{(\mathbf{v}\cdot\mathbf{v})\,(\boldsymbol{\lambda}\cdot\boldsymbol{\lambda})} \right\} \tag{C10}$$

where $\mathbf{v}$ is the scaled translational velocity vector with its Z-component reversed, i.e. $\mathbf{v} = (V_x,\ V_y,\ -V_z)$, and $\boldsymbol{\lambda}$ is a vector normal to the planar surface given by $\boldsymbol{\lambda} = (\ Z_X,\ Z_Y,\ -1)$.

Proof : **By expressing the coefficients $C_1,\ C_2,\ C_3$ in terms of $V_x,\ V_y,\ V_z,\ Z_X$ and Z_Y this lemma is easily verified •**

Below we state a theorem which helps in identifying the required solution α_0 for V_z from the spurious solutions α_+ and α_-. The theorem effectively implies that if the spurious solutions are not equal to α_0 then they result in complex valued solution for either slope or both slopes. In contrast α_0 always results in real valued solutions for the slopes. This fact is proved by considering the signs of the radicands in equations (C7a,b).

Theorem 2 :

$$A^2 - 4\,\alpha_0\ (\alpha_0 - u_x) \ge 0 \quad \text{and} \quad B^2 - 4\,\alpha_0\ (\alpha_0 - v_y) \ge 0\ , \tag{C11a,b}$$

$$A^2 - 4\,\alpha_+\ (\alpha_+ - u_x) \le 0 \quad \text{or} \quad B^2 - 4\,\alpha_+\ (\alpha_+ - v_y) \le 0\ , \tag{C11c,d}$$

$$A^2 - 4\,\alpha_-\ (\alpha_- - u_x) \le 0 \quad \text{or} \quad B^2 - 4\,\alpha_-\ (\alpha_- - v_y) \le 0\ . \tag{C11e,f}$$

Proof : **This can be easily proved by expressing all quantities in terms of the structure and motion parameters •**

Theorem 3 : **If for a non-middle root of equation (C1) (i.e. α_+ or α_-) both radicands (C7a,b) vanish, then it is equal to α_0 •**

We now state another theorem which implies that the correct solution α_0 for V_z is the middle root of the cubic equation. This fact helps us to select α_0 from among the three roots without actually evaluating the radicands (C7a,b) for each of the roots.

Theorem 4 :

$$\alpha_- \le \alpha_0 \le \alpha_+\ . \tag{C12}$$

Proof : **This is proved using the relations (C11a-f). Let**

$$A_+\ ,\ A_-\ = (1/2)\ (u_x \pm \sqrt{u_x^2 + A^2}\ ) \tag{C13a}$$

and

$$B_+\ ,\ B_-\ = (1/2)\ (v_y \pm \sqrt{v_y^2 + B^2}\ )\ . \tag{C13b}$$

Then from relation (C11a)

$$A^2 - 4\,\alpha_0\,(\alpha_0 - u_x) \geq 0$$

$$\rightarrow\ \ \alpha_0^2 - u_x\,\alpha_0 - \frac{1}{4}A^2 \leq 0$$

$$\rightarrow\ \ (\alpha_0 - A_-)\,(\alpha_0 - A_+) \leq 0$$

$$\rightarrow\ \ A_- \ \leq\ \alpha_0\ \leq\ A_+. \qquad \textbf{(C14a)}$$

Similarly, from relation (C11b) we can show that

$$B_- \ \leq\ \alpha_0\ \leq\ B_+\,. \qquad \textbf{(C14b)}$$

Hence

$$\max\,(A_-,\ B_-)\ \leq \alpha_0 \leq\ \min\,(A_+,\ B_+). \qquad \textbf{(C14c)}$$

Similarly, from relations (C11c-f) we can show that

$$\alpha_+ \geq \max\,(A_+,\ B_+) \quad \text{and} \quad \alpha_- \leq \min\,(A_-,\ B_-)\,. \qquad \textbf{(C15a,b)}$$

Relations (C14c) and (C15a,b) imply that α_0 is the middle root of the cubic equation (C1) (i.e. $\alpha_- \leq \alpha_0 \leq \alpha_+$) •

The following theorem gives an analytic solution for α_0.

Theorem 5 : **Let**

$$p = \frac{C_2}{3} - \frac{C_1^2}{9} \quad \text{and} \quad q = \frac{C_1^3}{27} - \frac{C_1\,C_2}{6} + \frac{C_3}{2}. \qquad \textbf{(C16a,b)}$$

Then, if $p = 0$,

$$\alpha_0 = \sqrt[3]{-2q} - C_1/3\ ; \qquad \textbf{(C17a)}$$

if $q = 0$,

$$\alpha_0 = -\,C_1/3\ ; \qquad \textbf{(C17b)}$$

otherwise,

$$\alpha_0 = 2\,d\,\cos\left(\frac{\pi+\psi}{3}\right) - \frac{C_1}{3} \qquad \textbf{(C17c)}$$

where

$$d = \text{sign}(q)\sqrt{|\,p\,|} \quad \text{and} \quad \psi = \cos^{-1}\left(\frac{q}{d^3}\right). \qquad \textbf{(C18a,b)}$$

Proof : **This is a direct consequence of the closed-form solutions for a cubic equation; see Rektorys (1969) •** *Theorem 6 :* **The first order flow parameters determine lower and upper bounds on the velocity of approach V_z of a surface along the line of sight. The bounds are**

$$V_z^{(max/min)} = \frac{u_x+v_y}{2} \pm \frac{\sqrt{a_1^2+a_2^2}}{2} \qquad \textbf{(C19)}$$

where $a_1,\ a_2$ are as in relations (4.2a,b).

Proof : **By simple trigonometric manipulation, expression (4.3c) for V_z can be reexpressed as**

$$V_z = \frac{u_x+v_y}{2} - \frac{a_1}{2}\sin 2\theta - \frac{a_2}{2}\cos 2\theta \ . \qquad \textbf{(C20)}$$

Differentiating the right hand side above and equating the resulting expression to zero we can show that the θs corresponding to the extrema of V_z are given by

$$\tan 2\theta = \frac{a_1}{a_2} \ . \qquad \textbf{(C21)}$$

From the above expression we have

$$\sin 2\theta = \frac{a_1}{\sqrt{a_1^2+a_2^2}} \ \text{ and } \ \cos 2\theta = \frac{a_2}{\sqrt{a_1^2+a_2^2}} \ \text{ for } 0 \le 2\theta \le 2\pi \ . \qquad \textbf{(C22a,b)}$$

Substituting for sin2θ and cos2θ from the above expressions in expression (C20) we can get relation (C19) • A similar theorem can be stated for Ω_Z, the rotational velocity about the line of sight. An interpretation of the bounds is discussed at the end of this appendix. Having solved for $V_z = \alpha_0$, we can obtain the remaining unknowns by back substituting its value in reverse order in the steps leading to the derivation of the cubic equation (C1). Assuming $V_z \neq 0$ the solution method is as follows. Substitute the value of V_z into equation (C8) and determine the sign of the right hand side which satisfies this equation. If this is ''+'', then the two solutions for the slopes are (Z_{X+}, Z_{Y+}) and (Z_{X-}, Z_{Y-}). If it is ''–'', the two solutions for the slopes are (Z_{X+}, Z_{Y-}) and (Z_{X-}, Z_{Y+}). If the sign is indeterminate, as when the right hand of equation (C8) is zero, it can be arbitrarily chosen as either sign gives the same solution. At this point we can solve for V_x and V_y using relations (C5a,b), and Ω_X and Ω_Y using relations (C4a,b). Finally, Ω_Z is found from relation (3.11e/f). The above method cannot be used if $V_z=0$. In this case we need to use an alternative method. Assuming $V_z=0$ the solution for the other unknowns can be easily obtained (see Waxman and Ullman, 1985). However we would like to use a uniform method in all cases. Therefore, below we give a method where θ and r are determined first (possibly from a knowledge of V_z), and then the other unknowns from relations (4.3a-h). Also this method conforms to the other problems considered later on (e.g. curved surfaces, and planar surfaces using spatio-temporal derivatives). *Theorem 7 :* **For planar surfaces the constraints on r and θ are**

$$r^2 c - (v_{xy} - u_0)\, r - V_z\, (a_1 s + a_2 c) = 0 \ , \qquad \textbf{(C23a)}$$

$$r^2 s - (u_{xy} - v_0)\, r - V_z\, (a_1 c - a_2 s) = 0 \qquad \textbf{(C23b)}$$

where V_z is expressed in terms of θ as in relation (4.3c).

Proof : **From relations (4.3c,e-h) and (4.4a,b) we can derive**

$$v_{xy} - u_0 = rc - V_z\, (a_1 s + a_2 c)/r \ , \qquad \textbf{(C24a)}$$

$$u_{xy} - v_0 = rs - V_z\, (a_1 c - a_2 s)/r \ . \qquad \textbf{(C24b)}$$

Since the translation parallel to the image plane is assumed to be non-zero, we have $r \neq 0$ and therefore relations (C23a,b) are easily obtained from relations (C24a,b) • Considering relations (C23a,b) to be quadratic equations in r and equating the roots of these equations, we can obtain

$$Bc - As = \pm s\sqrt{A^2 + 4cV_z(a_1 s + a_2 c)} \pm c\sqrt{B^2 + 4sV_z(a_1 c - a_2 s)} \qquad \textbf{(C25)}$$

where the two $\pm$ signs are used to denote four combinations of signs. In the above relation we can substitute for V_z in terms of θ from relation (4.3c) and get an equation involving only θ. The resulting equation can be solved to obtain θ by a simple numerical method (say by searching for θ in the interval $(-\pi/2, \pi/2])$. Having solved for θ we can obtain V_z from relation (4.3c) and then solve for r from relations (4.5a,b). However it is easier to solve for θ in terms of V_z which, as we have seen, can be obtained by solving a cubic equation.

Theorem 8 : **tanθ is given by the roots of the quadratic equation**

$$(u_x - V_z)\tan^2\theta - a_1 \tan\theta + (v_y - V_z) = 0. \tag{C26}$$

Proof : **This follows from relation (4.3c) • First we solve for V_z as the middle root of the cubic equation (C1). Then we solve the quadratic equation (C26) to get θ. There are in general two solutions for θ and for each of them we solve for r by solving the quadratic equations (C23a,b). The common root(s) of these equations are the solutions for r. From r and θ the other unknowns are obtained from relations (4.3a-h). In order to interpret the bounds on V_z and Ω_Z we make the following observation. To a first order, the image velocity field in a small field of view around the direction of view can be described by**

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \tag{C27}$$

The above expression represents an *affine transformation.* **In this expression, the vector $[u_0, v_0]^T$ gives the** *pure translation* **of the image region at the image origin; the 2×2 tensor on the right hand side is the** *velocity gradient tensor.* **This tensor can be expressed uniquely as the sum of a** *symmetric tensor* **and an** *anti-symmetric tensor* **as below:**

$$\begin{bmatrix} u_x & u_y \\ v_x & v_y \end{bmatrix} = \begin{bmatrix} u_x & (v_x+u_y)/2 \\ (v_x+u_y)/2 & v_y \end{bmatrix} + \begin{bmatrix} 0 & (u_y-v_x)/2 \\ -(u_y-v_x)/2 & 0 \end{bmatrix}. \tag{C28}$$

In Fluid Mechanics literature (e.g.: Aris, 1962), the symmetric tensor of a velocity gradient tensor is called the *deformation* **or** *rate of strain* **tensor and the anti-symmetric tensor is called the** *spin tensor.* **These tensors have nice physical interpretations. We will borrow these well known ideas from Fluid Mechanics to interpret our results. The independent parameter $u_y - v_x$ of the spin tensor is called the** *spin* **or** *vorticity.* **It is also the** *negative curl* **of the image velocity field at the image origin, i.e.**

$$-\text{curl} = u_y - v_x\ . \tag{C29}$$

This can be easily verified from relation (C27). It gives the *rigid body rotation* **of the image neighborhood at the image origin (Figure 12b). By setting all terms except the curl term to zero, i.e.**

$$u_0 = v_0 = (v_x + u_y) = u_x = v_y = 0, \tag{C30}$$

we can obtain the image flow field corresponding to this term. The term results in a purely *rotational flow field* **(see Figure 12a). The deformation tensor gives the deformation of the image neighborhood at the image origin. We can interpret this tensor in terms of its eigen values. The two eigen values of this tensor are in fact $V_z^{\max}$, $V_z^{\min}$, given by relation (4.8a). The sum of the eigen values (which is also the trace of the original tensor) is the** *divergence* **of the image velocity field at the image origin, i.e.**

$$\text{divergence} = u_x + v_y\ . \tag{C31}$$

This can be easily verified from relation (C27). This quantity gives the *isotropic expansion* **or contraction of the image neighborhood at the image origin (Figure 11b). The image flow corresponding to the divergence term is obtained by setting other terms to zero, i.e.**

$$u_0 = v_0 = u_y = v_x = (u_x - v_y) = 0\ . \tag{C32}$$

The result is a *purely divergent flow* **(see Figure 11a). The difference of the two eignen values of the deformation tensor is the** *magnitude of pure shear* **of the image neighborhood at the image origin, i.e.,**

$$\text{Shear magnitude} = \sqrt{(u_y + v_x)^2 + (u_x - v_y)^2}\ . \tag{C33}$$

The image neighborhood undergoes a contraction along one direction and an expansion orthogonal to it under constant area (see Figure 13b). The directions of contraction and expansion are aligned with the two eigen vectors of the deformation tensor. The image flow corresponding to a pure shear transformation is obtained by setting all but the shear terms to zero, i.e.,

$$u_0 = v_0 = u_x + v_y = u_y - v_x = 0\ . \tag{C34}$$

An example of a pure shear flow is shown in Figure 13a. In summary, a small circular image element

at the image origin translates rigidly with velocity $[u_0,\ v_0]^T$, rotates as a rigid area with spin $u_y - v_x$, dilates according to the sum of the eigen values of the deformation tensor, and undergoes a stretch and compression at constant area according to the difference of the eigen values of the deformation tensor (along mutually orthogonal axes aligned with the eigen vectors) (Koenderink and Van Doorn, 1975, 1976; Waxman and Wohn, 1986). In view of our above discussion and equations (C29,C31,C33), and (4.8a,b), relations (4.8c,d) can be derived. The quantities: divergence, curl, and shear magnitude are all *invariant* with respect to the orientation of the image axes. Their values are unaffected by a rotation of the image coordinate system. This can be easily shown by considering how the image flow derivatives u_x, u_y, v_x, v_y are transformed by a rotation of the image coordinate system (e.g. see Kanatani, 1986). Hence they are called *differential invariants* of image flow. Let us now interpret what the bounds mean in the scene domain. For this sake we introduce two vectors, $\boldsymbol{V}_1$ as in relation (4.15a) which is the direction of translation parallel to the image plane, and $\boldsymbol{P}$ as in relation (4.16a) which is the gradient of the object's surface with respect to the image plane. Now, from equations (3.11c,d,C31,4.15a,4.16a) we can show that

$$\text{divergence} = 2V_z + \boldsymbol{V}_1 \cdot \boldsymbol{P}\ . \qquad \textbf{(C35a)}$$

Let $\boldsymbol{k}$ be a unit vector along the Z axis. Then, from equations (3.11e,f,C29,4.15a,4.16a) we can show that

$$-\text{curl}\ \boldsymbol{k} = 2\Omega_Z \boldsymbol{k} + \boldsymbol{V}_1 \times \boldsymbol{P}\ . \qquad \textbf{(C35b)}$$

Also, from equations (3.11c-f,C33,4.15a,4.16a) we can show that

$$\text{Shear magnitude} = |\boldsymbol{V}_1||\boldsymbol{P}|\,. \qquad \textbf{(C35c)}$$

The above relations (C35a-c) show how the differential invariants of image flow are related to the three-dimensional motion and surface orientation. Some of the terms in these relations are in agreement with our intuition, for example the appearance of V_z in divergence and Ω_Z in curl. Now, from equations (4.8c,d,C35a-c) we can derive relations (4.8e,f). See the main text for some comments on these relations.

APPENDIX D

Nature of solutions for planar surfaces

Theorem 1 : **The interpretation of the image flow of a planar surface is unique when the translation along the line of sight is zero.**

Proof : V_z **is the translation component along the line of sight. Setting** V_z **to zero in relations (C24a,b) we see that the resulting equations uniquely determine** r **and** θ**. Each solution of** (θ, r) **gives one solution to the structure and motion parameters according to relations (4.3a-h). Hence the theorem •**

Theorem 2 : **There are two solutions to** θ **which correspond respectively to the direction of translation parallel to the image plane and the surface tilt (** *tilt* **is the angle between the x-axis and the projection of surface normal on the image plane).**

Proof : **By expressing the coefficients of the quadratic equation (C26) in terms of the structure and motion parameters (using relations (4.2a,3.11c-f)) its roots can be shown to be**

$$\tan\theta = \frac{V_y}{V_x} \; , \; \frac{Z_Y}{Z_X} \qquad \textbf{(D1)}$$

The first solution corresponds to the direction of translation parallel to the image plane and the second one corresponds to the surface tilt •

Theorem 3 : **If the direction of translation is parallel to the surface normal then the solution for the structure and motion parameters is unique.**

Proof : **First we observe that** (V_x, V_y, V_z) **is the translation vector and** $(Z_X, Z_Y, -1)$ **is the surface normal. If these vectors are parallel then their corresponding components are proportional. Therefore we can show that**

$$\frac{V_y}{V_x} = \frac{Z_Y}{Z_X} \; , \;\; V_x = -V_z \, Z_X \; , \quad \text{and} \quad V_y = -V_z \, Z_Y \; . \qquad \textbf{(D2a,b,c)}$$

From relations (D1,D2a) it follows that the two solutions for θ **are equal. In addition we can show these two results: (i) In equations (C23a,b) the corresponding coefficients of** r **become proportional, and therefore the two quadratic equations in** r **become linearly dependent; consequently both equations have the same pair of roots. (ii) The two roots of both (C23a,b) are equal. Therefore we get a unique solution for** r**. Thus, in effect, we have a unique solution for** (θ, r) **and therefore a unique solution for the structure and motion parameters •**

Theorem 4 : **If neither the translation along the line of sight is zero nor the direction of translation is parallel to the surface normal then there are two solutions for the structure and motion parameters.**

Proof :

Case 1. **The translation vector, surface normal, and optical axis lie in a plane.**

In this case it is easy to see that the two solutions for θ **given in relation (D1) are equal. However in this case it can be shown that equations (C23a,b) are equivalent and give two distinct solutions for** r**. Therefore, effectively** (θ, r) **has two solutions.**

Case 2 : **This corresponds to the situation where Case 1 is false. Under this assumption, multiplying equation (C23a) by s and equation (C23b) by c we get**

$$r^2 cs - (v_{xy} - u_0)\, r - V_z\, (a_1 s + a_2 c) = 0\ , \tag{D3a}$$

$$r^2 cs - (u_{xy} - v_0)\, r - V_z\, (a_1 c - a_2 s) = 0\ . \tag{D3b}$$

Subtracting equation (D3a) from (D3b) and simplifying we get

$$r = V_z\, \frac{(a_1 \cos 2\theta - a_2 \sin 2\theta)}{(v_{xy} - u_0)s - (u_{xy} - v_0)c}\ . \tag{D4}$$

Under the conditions stated in the theorem it can be shown that the right hand side of this equation is non-zero and finite. Therefore the above equation gives a unique solution for r for each θ. In this case θ has two distinct solutions. Therefore we get two solutions for (θ, r).

In each of the two cases we get two solutions for the structure and motion parameters • The image flow equations always have one solution which corresponds to the actual shape and motion of the surface in the physical world. Therefore when there are multiple solutions it is natural to try to relate the spurious solutions to the actual shape and motion of the surface. In order to do this we will adopt the following notation: we will denote the parameters corresponding to the actual or the ''correct'' solution by appending a ''0'' to the subscripts of the respective parameters, and we will denote the spurious solutions by appending distinct integers to these subscripts. For example, r_0 denotes the actual translation parallel to the image plane whereas r_1, r_2, ..., etc. denote the spurious solutions for the translation parallel to the image plane. In addition, we will define the quantities k, θ' such that

$$Z_X = k \cos\theta' \quad \text{and} \quad Z_Y = k \sin\theta'\ . \tag{D5a,b}$$

Note that θ' corresponds to surface tilt. We also use the abbreviations

$$c' = \cos\theta' \quad \text{and} \quad s' = \sin\theta'\ . \tag{D6a,b}$$

With this notation, the two solutions of θ are

$$\tan\theta_0 = \frac{V_y}{V_x} \quad \text{and} \quad \tan\theta_1 = \frac{Z_Y}{Z_X} = \tan\theta'_0\ . \tag{D7a,b}$$

Theorem 5 : **The solutions of (θ, r) are**

$$(\theta_0,\ r_0),\ (\theta'_0,\ -V_{z0}\ k_0)\ . \tag{D8a,b}$$

Proof : **The common root(s) of equations (C23a,b) are the solutions of r. Expressing the roots of these equations in terms of the structure and motion parameters we can show that, for θ = θ_0 the common root is r_0, and for θ = θ_1 it is $-V_{z0}\ k_0$ •**

Waxman and Ullman (1985) obtained a result equivalent to the above theorem.

APPENDIX E

Resolving the ambiguity of interpretation: planar surfaces

We have seen that an instantaneous image flow of a planar surface has in general two interpretations. In this section we will show that a unique interpretation is obtained in two cases: (i) the instantaneous image flow of two planar patches moving rigidly is given, and (ii) for a single planar surface the image flows at two time instants separated by a ''short'' time interval are given. Waxman and Ullman (1985) first obtained the relation between the two solutions for a planar surface. Denoting the quantities corresponding to the ''spurious'' solution by a hat on top, the relations are

$$\hat{Z}_X = -V_x / V_z \;, \tag{E1a}$$

$$\hat{Z}_Y = -V_y / V_z \;, \tag{E1b}$$

$$\hat{V}_x = -Z_X V_z \;, \tag{E1c}$$

$$\hat{V}_y = -Z_Y V_z \;, \tag{E1d}$$

$$\hat{V}_z = V_z \;, \tag{E1e}$$

$$\hat{\mathbf{\Omega}}_X = \mathbf{\Omega}_X - V_y - V_z Z_Y \;, \tag{E1f}$$

$$\hat{\mathbf{\Omega}}_Y = \mathbf{\Omega}_Y + V_x + V_z Z_X \;, \tag{E1g}$$

$$\hat{\mathbf{\Omega}}_Z = \mathbf{\Omega}_Z + V_x Z_Y - V_y Z_X \;. \tag{E1h}$$

Given either of the two solutions for a planar surface we can obtain the other using these relations. Therefore the pairs of solutions that satisfy relations (E1a-h) have been termed *dual solutions* **by Waxman and Ullman (1985). These relations can be easily verified from relations (4.3a-h) and the fact that the solutions for (θ, r) are (θ_0, r_0) and $(\theta'_0, -V_{z0} k_0)$ (see Theorem 5, Appendix D).**

Theorem 1 : **If two planar patches are moving as a rigid body and neither of them passes through the origin then the interpretation of their instantaneous image flow is unique.**

Proof : **The fact that the two planar patches are moving rigidly implies that there is no relative motion between the two; their motion parameters are the same. Further, since neither of them passes through the origin, we have $Z_0 \neq 0$. This is important because the translation is scaled by Z_0. Suppose that solving the flow field for the structure and motion of the two surfaces yields the dual solution pairs $(\boldsymbol{S}^{(1)}, \boldsymbol{SD}^{(1)})$ and $(\boldsymbol{S}^{(2)}, \boldsymbol{SD}^{(2)})$ where**

$$\boldsymbol{S}^{(1)} = (V_x^{(1)},\, V_y^{(1)},\, V_z^{(1)},\, \mathbf{\Omega}_X{}^{(1)},\, \mathbf{\Omega}_Y{}^{(1)},\, \mathbf{\Omega}_Z{}^{(1)},\, Z_X{}^{(1)},\, Z_Y{}^{(1)}) \tag{E2a}$$

$$\boldsymbol{S}^{(2)} = (V_x^{(2)},\, V_y^{(2)},\, V_z^{(2)},\, \mathbf{\Omega}_X{}^{(2)},\, \mathbf{\Omega}_Y{}^{(2)},\, \mathbf{\Omega}_Z{}^{(2)},\, Z_X{}^{(2)},\, Z_Y{}^{(2)}) \;, \tag{E2b}$$

and $\boldsymbol{SD}^{(1)}$ and $\boldsymbol{SD}^{(2)}$ are related to $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$ by relations (E1a-h). Then we shall show that the solutions corresponding to the actual interpretation for the structure and motion of the two surfaces, say $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$, are uniquely determined.

Suppose that the two surfaces are not parallel (i.e. they have different orientations); then for $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$ we have

$$(Z_X{}^{(1)},\, Z_Y{}^{(1)}) \;\neq\; (Z_X{}^{(2)},\, Z_Y^{(2)}) \;. \tag{E3a}$$

But for the slopes of $SD^{(1)}$ and $SD^{(2)}$, from relations (3.9a-c) and (E1a,b), we always find that

$$(\hat{T}_X{}^{(1)}\,,\ \hat{T}_Y{}^{(1)}\ )\ =\ (\hat{T}_X{}^{(2)}\,,\ \hat{T}_Y^{(2)}\ ) \qquad \textbf{(E3b)}$$

since

$$(\ -V_x^{(1)}\ /\ V_z^{(1)}\ ,\ -V_y^{(1)}\ /\ V_z^{(1)}\ )\ =\ (\ -V_x^{(2)}\ /\ V_z^{(2)}\ ,\ -V_y^{(2)}\ /\ V_z^{(2)}\ )\ , \qquad \textbf{(E3c)}$$

which reflects the rigid body motion of the planar surface pair. Furthermore, $S^{(1)}$ and $S^{(2)}$ have the same rotation parameters, whereas, in general, this is not true with $SD^{(1)}$ and $SD^{(2)}$. Therefore, $SD^{(1)}$ and $SD^{(2)}$ can be ruled out as they are inconsistent.

If the two planar surfaces are in fact parallel but distinct (i.e. $Z_0^{(1)} \neq Z_0^{(2)}$), then $S^{(1)}$ and $S^{(2)}$ still have the same rotation parameters whereas $SD^{(1)}$ and $SD^{(2)}$ do not. Thus, again we can rule out $SD^{(1)}$ and $SD^{(2)}$ as unacceptable.

Therefore, for two (or more) planar surfaces moving as a rigid body (e.g. faces of a polyhedron), a unique 3-D interpretation may be derived from an instantaneous image flow field. Of course, the flow field itself must first be segmented into regions corresponding to individual planar surfaces (Waxman, 1984b; Wohn 1984) •

Now we give a result which will be used in proving the second uniqueness result. At time instant t, let a planar surface in motion be described by

$$Z = Z_0 + Z_X\,X + Z_Y\,Y \qquad \textbf{(E4)}$$

in the coordinate system shown in Figure 8, where Z_0 is the Z intercept of the plane and Z_X, Z_Y are the X and Y slopes respectively. Also, let $\boldsymbol{V}$ and $\boldsymbol{\Omega}$ be the relative translational and rotational velocities in space of an observer with respect to the planar surface. This *relative motion* **in space is** *assumed constant* **over short time intervals (over three image frames). We represent the relative motion of the observer and the structure of the plane at time t in matrix form:**

$$\boldsymbol{S} = \begin{bmatrix} V_x & V_y & V_z \\ \Omega_X & \Omega_Y & \Omega_Z \\ Z_X & Z_Y & Z_0 \end{bmatrix}. \qquad \textbf{(E5)}$$

In this matrix, the first row represents the translational velocity components scaled by distance Z_0, the second row the rotational velocity components and the third row the slopes of the plane and the distance scale factor. The motion and structure parameters, represented by S above, change with time due to the relative motion between the plane and the observer's reference frame. In the following theorem we derive expressions for these parameters after a small time interval dt has elapsed.

Theorem 2 : **If the structure of a planar surface and the instantaneous motion of an observer relative to the planar surface at time t is S given by (E5), and at time $t+dt$ is SE given by**

$$\boldsymbol{SE} = \begin{bmatrix} V'_x & V'_y & V'_z \\ \Omega'_X & \Omega'_Y & \Omega'_Z \\ T'_X & T'_Y & Z'_0 \end{bmatrix}, \qquad \textbf{(E6)}$$

then S and SE are related by

$$\begin{bmatrix} V'_x \\ V'_y \\ V'_z \end{bmatrix} = \begin{bmatrix} V_x \\ V_y \\ V_z \end{bmatrix} + \begin{bmatrix} -p & \Omega_Z & -\Omega_Y \\ -\Omega_Z & -p & \Omega_X \\ \Omega_Y & -\Omega_X & -p \end{bmatrix} \begin{bmatrix} V_x \\ V_y \\ V_z \end{bmatrix} dt\ , \qquad \textbf{(E7a)}$$

where p is as defined in (E14),

$$(\mathbf{\Omega}'_X,\ \mathbf{\Omega}'_Y,\ \mathbf{\Omega}'_Z) = (\mathbf{\Omega}_X,\ \mathbf{\Omega}_Y,\ \mathbf{\Omega}_Z), \quad \text{and} \tag{E7b}$$

$$\begin{bmatrix} T'_X \\ T'_Y \\ Z'_0 \end{bmatrix} = \begin{bmatrix} Z_X \\ Z_Y \\ Z_0 \end{bmatrix} + \begin{bmatrix} \mathbf{\Omega}_Y Z_X - \mathbf{\Omega}_X Z_Y & \mathbf{\Omega}_Z & \mathbf{\Omega}_Y \\ -\mathbf{\Omega}_Z & \mathbf{\Omega}_Y Z_X - \mathbf{\Omega}_X Z_Y & -\mathbf{\Omega}_X \\ Z_0(\mathbf{\Omega}_Y + V_x) & -Z_0(\mathbf{\Omega}_X - V_y) & -Z_0 V_z \end{bmatrix} \begin{bmatrix} Z_X \\ Z_Y \\ 1 \end{bmatrix} dt.$$

Proof : **To compute the change in the structure parameters during the time interval dt, we first take the time derivative on either side of equation (E4) to get**

$$\dot{Z} = \dot{Z}_0 + \dot{Z}_X X + Z_X \dot{X} + \dot{Z}_Y Y + Z_Y \dot{Y} \tag{E8}$$

Substituting for $(\dot{X}, \dot{Y}, \dot{Z})$ in the above expression from relations (3.2a-c), and for Z from relation (E4), and rearranging terms, we get

$$\left\{ \frac{dZ_0}{dt} - Z_0\ (\ (\mathbf{\Omega}_Y + V_X/Z_0)\ Z_X - (\mathbf{\Omega}_X - V_Y/Z_0)\ Z_Y - V_Z/Z_0\ ) \right\} +$$

$$\left\{ \frac{dZ_X}{dt} - (\ Z_X\ (\mathbf{\Omega}_Y\ Z_X - \mathbf{\Omega}_X\ Z_Y) + (\mathbf{\Omega}_Y + \mathbf{\Omega}_Z\ Z_Y)\ ) \right\}\ X +$$

$$\left\{ \frac{dZ_Y}{dt} - (\ Z_Y\ (\mathbf{\Omega}_Y\ Z_X - \mathbf{\Omega}_X\ Z_Y) - (\mathbf{\Omega}_X + \mathbf{\Omega}_Z\ Z_X)\ ) \right\}\ Y \ = \ 0\ . \tag{E9}$$

In the above expression, since X, Y are independent parameters of points on the plane in motion, we can equate each of the three terms above to zero separately to get the exact differentials for the slopes and distance:

$$dZ_0 = Z_0 \left[\ (\mathbf{\Omega}_Y + V_x)\ Z_X - (\mathbf{\Omega}_X - V_y)\ Z_Y - V_z \right] dt\ , \tag{E10a}$$

$$dZ_X = \left[Z_X\ (\mathbf{\Omega}_Y\ Z_X - \mathbf{\Omega}_X\ Z_Y) + (\mathbf{\Omega}_Y + \mathbf{\Omega}_Z\ Z_Y) \right] dt\ , \tag{E10b}$$

$$dZ_Y = \left[Z_Y\ (\mathbf{\Omega}_Y\ Z_X - \mathbf{\Omega}_X\ Z_Y) - (\mathbf{\Omega}_X + \mathbf{\Omega}_Z\ Z_X) \right] dt\ . \tag{E10c}$$

Using the above relations, we can get expressions for the new structure parameters at time $t+dt$ as

$$T'_X = Z_X + dZ_X\ ,\ \ T'_Y = Z_Y + dZ_Y \ \text{ and } \ Z'_0 = Z_0 + dZ_0\ . \tag{E11a-c}$$

The new translational velocity $\boldsymbol{V}'$ at time $t+dt$ is found (in the absence of accelerations) from

$$\boldsymbol{V}' = \boldsymbol{V} + \boldsymbol{V} \times \mathbf{\Omega}\ dt\ . \tag{E12}$$

Dividing $\boldsymbol{V}'$ by Z'_0 and simplifying, we get the new scaled velocity components

$$V'_x = V_x + (V_y\ \mathbf{\Omega}_Z - V_z\ \mathbf{\Omega}_Y - V_x\ p)\ dt + O\,(dt^2)\ , \tag{E13a}$$

$$V'_y = V_y + (V_z\ \mathbf{\Omega}_X - V_x\ \mathbf{\Omega}_Z - V_y\ p)\ dt + O\,(dt^2)\ , \tag{E13b}$$

$$V'_z = V_z + (V_x\ \mathbf{\Omega}_Y - V_y\ \mathbf{\Omega}_X - V_z\ p)\ dt + O\,(dt^2)\ , \tag{E13c}$$

where

$$p \equiv (\mathbf{\Omega}_Y + V_x)\ Z_X - (\mathbf{\Omega}_X - V_y)\ Z_Y - V_z\ . \tag{E14}$$

The rotational velocity remains unchanged since

$$\mathbf{\Omega}' = \mathbf{\Omega} + \mathbf{\Omega} \times \mathbf{\Omega}\ dt \ = \ \mathbf{\Omega}\ . \tag{E15}$$

Relations (E10-E15) prove the theorem • *Theorem 3 :* **Given the image flow of a planar surface at two time instants separated by a "short" time interval, its structure and motion are uniquely determined.**

Proof : **The scheme of our proof is given in Figure 20. An image flow field F at time t generated by a planar surface in motion has two solutions, say S given by (E5) and SD which is related to S by relations (E1a-h); i.e.**

$$\boldsymbol{SD} = \begin{bmatrix} -Z_X V_z & -Z_Y V_z & V_z \\ \Omega_X - V_y - V_z Z_Y & \Omega_Y + V_x + V_z Z_X & \Omega_Z + V_x Z_Y - V_y Z_X \\ -V_x / V_z & -V_y / V_z & Z_0 \end{bmatrix} . \qquad \textbf{(E16)}$$

For these two solutions, assuming that the *motion in space is constant* **over a short time, the slope parameters change with time due to rotation and the scaled translational velocity components change with time due to the combined effect of translation and rotation. As a result, after a short time δt, S and SD evolve into two new solutions, say SE and SDE, respectively. SD and SE are given in terms of S by relations (E7a-c) and (E16) respectively. SDE can be obtained by substituting the elements of SD for the corresponding elements of S in expression (E7) for SE. The proof of uniqueness of interpretation is that, at time t, although S and SD are related by (E1a-h) corresponding to the two interpretations of the planar surface, at time $t+dt$, their time evolved solutions, SE and SDE respectively, are not related by the relations (E1a-h). A simple way to verify that SE and SDE do not satisfy the relations (E1a-h) is to note that their Z-components of translational velocity are not the same (i.e. relation (E1e) is not satisfied). It can be shown easily that the values of this velocity component for SE and SDE are related by**

$$V_{z\,(\boldsymbol{SDE})} = V_{z\,(\boldsymbol{SE})} + \left(V_x^2 - V_z^2\, Z_X^2 + V_y^2 - V_z^2\, Z_Y^2 \right) dt \, . \qquad \textbf{(E17)}$$

The two velocity components are the same (i.e. the term with dt above vanishes) when the direction of translational velocity is parallel to the plane's normal, in which case the interpretation is unique according to our discussion in Section 4.2.4.

Thus, if F' is the instantaneous flow field at time $t+\delta t$, then either SE or SDE will be one of its two solutions. If SE is a solution of the image flow field F' then the actual motion and orientation of the planar surface at time t are given by S •

APPENDIX F

Solving the image flow equations: curved surfaces

Throughout this section we assume that the surface in motion is curved (i.e. at least one of the curvatures is non-zero), and translation parallel to the image plane is not zero. Conditions for the presence of this case in terms of the image flow parameters is given in Theorem 4 of Appendix B. We give here a method to solve the image flow equations by first solving for θ, then for r and then for the other unknowns. *Theorem 1:* **Suppose that the** *curved-surface-condition* **(see Appendix B, Theorem 4) is true. Then the solutions for the curvatures are given in terms of θ and r by relations (4.11a-c).**

Proof : **In equation (3.11g) we substitute for V_x and Ω_Y from relations (4.3a,h), multiply the resulting equation by c and rearrange terms to get**

$$rc^2\, Z_{xx} = u_{xx}c - 2u_0 c - 2rc^2 + 2V_z Z_X c. \tag{F1a}$$

In equation (3.11j) we substitute for V_y from (4.3b) and multiply the resulting equation by s to get

$$rs^2 Z_{xx} = v_{xx}s\ . \tag{F1b}$$

Adding equations (F1a,b) and simplifying we can get equation (4.11a). Derivation of equation (4.11b) is similar to that of (4.11a) above. In this case we multiply (3.11l) by s, (3.11i) by c, add the resulting equations, do appropriate substitutions, and simplify to get (4.11b). From equations (3.11g,k,4.3a,b) we get

$$u_{xx}-2v_{xy} = rcZ_{xx} - 2rsZ_{xy} \tag{F2a}$$

From equations (3.11j,4.3b) we have

$$Z_{xx}s = v_{xx}/r. \tag{F2b}$$

Multiplying (F2a) by s and using (F2b) we get

$$s(u_{xx}-2v_{xy}) = cv_{xx} - 2rs^2 Z_{xy}. \tag{F2c}$$

Now we use equations (3.11l,h,4.3a,b) to get

$$v_{yy}-2u_{xy} = rsZ_{yy} - 2rcZ_{xy}. \tag{F3a}$$

From equations (3.11i,4.3a) we have

$$Z_{yy}c = u_{yy}/r. \tag{F3b}$$

Multiplying (F3a) by c and using (F3b) we get

$$c(v_{xx}-2u_{xy}) = su_{yy} - 2rc^2 Z_{xy}. \tag{F3c}$$

Adding (F2c,F3c) and simplifying we get (4.11c) • *Theorem 2 :* **When the curved-surface-condition is true, the parameters r and θ are related to the image flow parameters by the relations (4.12a-c).**

Proof : **In equation (F2c) we substitute for Z_{xy} from (4.11c) and simplify to get**

$$s(u_{xx}-2v_{xy}) = v_{xx}c + s^3(u_{xx}-2v_{xy}) - s^3 u_{yy} + s^2 c(v_{yy}-2u_{xy}) - s^2 cv_{xx}. \tag{F4a}$$

Using the identity $s^2+c^2 = 1$, this can be rewritten as

$$s(u_{xx}-2v_{xy}) = v_{xx}c + (1-c^2)s(u_{xx}-2v_{xy}) - s^3 u_{yy} \tag{F4b}$$

$$+ s^2 c(v_{yy}-2u_{xy}) - (1-c^2)cv_{xx}.$$

Simplifying the above equation and dividing by c^3 we get (4.12a). In equation (3.11g) we substitute

for V_x and Ω_Y from relations (4.3a,h), multiply the resulting equation by s and rearrange terms to get

$$rcs\ Z_{xx} = u_{xx}s - 2u_0 s - 2rcs + 2V_z Z_X s. \qquad \textbf{(F5a)}$$

In equation (3.11j) we substitute for V_y from (4.3b) and multiply the resulting equation by c to get

$$rcsZ_{xx} = v_{xx}c\ . \qquad \textbf{(F5b)}$$

Subtract (F5b) from (F5a), substitute for Z_X from (4.3e), and simplify to obtain (4.12b). Derivation of equation (4.12c) is similar to that of (4.12b) above. In this case we multiply (3.11l) by c, (3.11i) by s, subtract the resulting equations, do appropriate substitutions, and simplify to get (4.12c) • First check if the input image flow corresponds to a degenerate case as discussed in Appendix B, and if so, obtain the solution according to Appendix B. Otherwise first solve for θ by solving the cubic equation (4.12a) and then solve for r by simultaneously solving equations (4.12b,c). Now, for each set of solutions obtained for θ, r, solve for the other unknowns from relations (4.3a-h,4.11a-c). If the input image flow parameters are in error due to noise, then there may be no solution for θ and r which satisfies all three equations (4.12a-c). In this case we can solve for θ and r by using only two of these three equations. If this gives multiple solutions then we select the one which is closest to satisfying the third equation. In this solution method the solution we obtain could be complex valued. For example the solution for r obtained by solving a quadratic equation may have a small imaginary part. In this case we may either ignore the imaginary part or take the magnitude to be the solution. In general it will be necessary to use some heuristics to deal with noisy input data.

A note on solving polynomial equations in tanθ: **while solving a cubic, quadric, or a linear equation in tanθ, the coefficient of the highest power of tanθ may be zero. In this case we can take one of the solutions for θ to be $\pi/2$ and then proceed to solve the next lower order equation in tanθ. This gives correct results because we have assumed $-\pi/2 < \theta \le \pi/2$.**

APPENDIX G

Conditions for multiple interpretations: curved surfaces

In this section we give conditions for the presence of multiple interpretations for the structure and motion of a moving curved surface. Throughout this section we assume that we are dealing with a curved surface with non-zero translation parallel to the image plane. The condition for this case in terms of the image flow parameters is given in Theorem 4 of Appendix B. The number of solutions to the image flow equations is equal to the number of solutions for r and θ obtained by solving equations (4.12a-c). Therefore we analyze equations (4.12a-c) exhaustively to derive the conditions for multiple interpretations. The following two lemmas are related to the nature of the solution for θ as determined by equation (4.12a). The results of these lemmas were known earlier (Waxman, Kamgar-Parsi, and Subbarao, 1986).

Theorem 1 : **The roots of the cubic equation (4.12a) are**

$$\tan\theta = \frac{V_y}{V_x}\ ,\ \frac{1}{Z_{yy}}\,(-Z_{xy} \pm \sqrt{Z_{xy}^2 - Z_{xx}Z_{yy}}\,)\ . \qquad \textbf{(G1)}$$

Proof : **The coefficients of the cubic equation (4.12a) can be respectively expressed in terms of the structure and motion parameters as**

$$V_x Z_{yy}\ ,\ \ 2V_x Z_{xy} - V_y Z_{yy}\ ,\ \ V_x Z_{xx} - 2V_y Z_{xy}\ ,\ \ -V_y Z_{xx}\ . \qquad \textbf{(G2)}$$

Therefore, equation (4.12a) can be factored as

$$(V_x \tan\theta - V_y)\,(Z_{yy} \tan^2\theta + 2Z_{xy} \tan\theta + Z_{xx}) = 0\ . \qquad \textbf{(G3)}$$

The roots of the above equation are given by (G1) •

Lemma 1 : **There is one real solution for θ if the surface is ovoid (i.e. bowl shaped), there are two real solutions if the surface is cylindrical and there are three real solutions if the surface is a saddle.**

Proof : **The shape of a surface can be inferred from the sign of the** *Gaussian curvature* **(cf. O'Neill, 1966). The Gaussian curvature has the same sign as the expression**

$$Z_{xx}Z_{yy} - Z^2{}_{xy}\ . \qquad \textbf{(G4)}$$

If the sign is positive then the surface is ovoid, if it is zero then it is a cylinder (or a plane) and if it is negative then it is a saddle. From this and the expressions for the solution of θ given by (G1) the proof is clear • Although equation (4.12a) may give up to three solutions for θ, in most cases only one of them corresponding to the correct physical interpretation satisfies equations (4.12b,c) for some real value of r. In the following we consider cases where θ has two or more solutions. Theorem 2 to Lemma 3d correspond to this case. *Theorem 2:* **For a given set of image flow parameters, the curved-surface-condition is true, and**

$$(v_{yy} - 2v_0 + v_{xx} = 0)\ \text{ and} \qquad \textbf{(G5a)}$$

$$(u_{xx} - 2u_0 + u_{yy} = 0) \qquad \textbf{(G5b)}$$

if and only if **one of the following is true :**

$$(i)\quad (Z_X = Z_Y = 0)\ \text{ and }\ (Z_{xx} + Z_{yy} + 2 = 0) \qquad \textbf{(G6a)}$$

$$(ii)\quad (V_z = 0)\ \text{ and }\ (Z_{xx} + Z_{yy} + 2 = 0) \qquad \textbf{(G6b)}$$

(iii) $(V_z \neq 0)$ and $(Z_{xx}+Z_{yy}+2 \neq 0)$ and **(G6c)**

$$\frac{V_x}{Z_X} = \frac{V_y}{Z_Y} = \frac{V_z}{1+(Z_{xx}+Z_{yy})/2}$$

(Note: Case (i) implies that the surface is frontal (or specular) and the mean scaled curvature is -1; Case (ii) implies that there is no translation along the optical axis and the mean scaled curvature is -1; Case (iii) implies that the direction of translation is parallel to the vector: $(Z_X, Z_Y, 1+(Z_{xx}+Z_{yy})/2$.)

Proof : **From equations (3.11a,b,g,i,j,l) we have**

$$v_{yy}-2v_0+v_{xx} = V_y(Z_{xx}+Z_{yy}+2)-2V_zZ_Y \quad \text{and} \qquad \textbf{(G7a)}$$

$$u_{xx}-2u_0+u_{yy} = V_x(Z_{xx}+Z_{yy}+2)-2V_zZ_X \ . \qquad \textbf{(G7b)}$$

Therefore, conditions (G5a,b) hold if and only if

$$V_y(Z_{xx}+Z_{yy}+2) = 2V_zZ_Y \quad \text{and} \qquad \textbf{(G8a)}$$

$$V_x(Z_{xx}+Z_{yy}+2) = 2V_zZ_X \ . \qquad \textbf{(G8b)}$$

Now recall that $V_x \neq 0$ or $V_y \neq 0$ since $r \neq 0$. Now consider the logical expression

$$(G8a) \text{ and } (G8b) \text{ and } \left[(Z_{xx}+Z_{yy}+2=0) \text{ or } (Z_{xx}+Z_{yy}+2 \neq 0)\right] \text{ and}$$

$$(V_x \neq 0 \text{ or } V_y \neq 0) \text{ and } (V_z = 0 \text{ or } V_z \neq 0). \qquad \textbf{(G9)}$$

Expanding the above logical expression and simplifying, we get the disjunction of the three clauses (G6a-c). •

Theorem 3 : **Under the conditions of Theorem 2 when either (G6a) or (G6b) is true equations (4.12b,c) are independent of θ, i.e. they cannot be used to solve for θ (nor do they give rise to any constraint on θ).**

Proof : **For the cases (G6a,b) we get from equations (4.12b,c) the following constraint on θ :**

$$(v_{yy}-2v_0+v_{xx}) \cos\theta = (u_{xx}-2u_0+u_{yy}) \sin\theta \ . \qquad \textbf{(G10)}$$

The above constraint is identically true for all values of θ under the conditions (G5a,b). • If conditions of Theorem 2 are true then we can determine which one of the three conditions (G6a-c) is actually true in that order by further testing the image flow parameters.

Theorem 4 : **Under the conditions of Theorem 2,**

(i) case (G6a) is true if and only if $a_1 = a_2 = 0$,

(ii) case (G6a) is false and case (G6b) is true if and only if

$$\tan\theta = \frac{a_1 \pm \sqrt{a_1^2 - 4u_xv_y}}{2u_x} \ . \qquad \textbf{(G11)}$$

(iii) cases (G6a,b) are false and (G6c) is true if and only if

$$\tan 2\theta = a_1/a_2 \ . \qquad \textbf{(G12)}$$

Proof : **Case (i) is obvious from equations (4.3e,f). Case (ii) is easily proved by equating the expression (4.3c) for V_z to zero. To prove the last case, we solve for θ from equations (4.12b,c) as follows. Eliminating the term $2r^2cs$ from the expressions (4.12b,c) and rearranging terms, we get the following constraint on θ :**

$$r\left[(v_{yy}-2v_0+v_{xx})c-(u_{xx}-2u_0+u_{yy})s\right] \tag{G14}$$
$$= 2V_z\left[s(a_1s+a_2c)-c(a_1c-a_2s)\right] \ .$$

Under the conditions (G5a,b), the left hand side of equation (G14) is identically zero. Therefore equating the right hand side to zero we get equation (G12).• *Theorem 5 :* **Under the conditions of Theorem 2, equations (3.11a-l) can have three solutions** *if and only if* **case (G6a) is true (i.e. the surface is frontal and the mean scaled curvature is -1) and the surface is a saddle, i.e.**

$$Z^2{}_{xy}-Z_{xx}Z_{yy}>0 \ . \tag{G15}$$

Proof : **In this case, all the values of θ computed by solving equation (4.12a) are valid since there is no other extra constraint on θ (from Theorem 4). Under condition (G15) there are three solutions for θ given by relations (G1). For each value of θ this case results in a unique value for r obtained by solving (4.12b) or (4.12c) given by**

$$r = -\frac{1}{2}\left\{\frac{v_{xx}}{s}+\frac{u_{yy}}{c}\right\} . \tag{G16}$$

Therefore, there are three solutions in this case. • *Theorem 6 :* **Under the conditions of Theorem 2 suppose that case (G6a) is false. Then equations (3.11a-l) can have up to two solutions if case (G6b) is true (i.e. translation along the optical axis is zero and the mean scaled curvature is -1) and the two solutions of equation (G11) are also the roots of the the cubic equation (4.12a).**

Proof : **In this case equations (4.12a) and (G11) are the only constraints on θ. For each solution common to these equations we get a unique value for r obtained by solving equations (4.12b,c) given by relation (G16). Therefore there can be up to two solutions for the image flow equations (3.11a-l). •**

Lemma 2 : **Under the conditions stated in Theorem 6 the structure parameters of the surface satisfy (G6b and) the following constraint:**

$$\frac{Z_Y}{Z_X} = \frac{1}{Z_{yy}}\left(-Z_{xy}\pm\sqrt{Z_{xy}^2-Z_{xx}Z_{yy}}\right) . \tag{G17}$$

Proof : **Substituting for terms on the right hand side of equation (G11) in terms of the structure and motion parameters (from relations (3.11c-f) and (4.2a)) we can show that the two solutions for θ are**

$$\tan\theta = \frac{V_y}{V_x} \ , \ \frac{Z_Y}{Z_X} \ . \tag{G18}$$

Since these are also the roots of the cubic equation (4.12a), from Theorem 1 we conclude that the structure parameters are related as in (G17). • As can be seen from the expressions for the two roots in relation (G18), the solution for θ becomes unique when $V_xZ_Y = V_yZ_X$, i.e. the translation vector $(V_x,\ V_y,\ V_z)$, the surface normal $(Z_X,\ Z_Y,\ -1)$ and the optical axis (the Z-axis) all lie in a plane. *Theorem 7 :* **Under conditions of Theorem 2 if the cases (G6a,b) are false and the two solutions of equation (G12) for θ are also the roots of the cubic equation (4.12a) then up to four solutions are possible for the image flow equations (3.11a-l).**

Proof : **Here the common solutions of the equations (4.12a) and (G12) are the valid solutions for θ. Using the trigonometric identity**

$$\tan 2\theta = \frac{2\tan\theta}{1-\tan^2\theta} \tag{G19}$$

we can show that the solutions of equation (G12) are

$$\tan\theta = \frac{-a_2 \pm \sqrt{a_1^2 + a_2^2}}{a_1} . \quad \textbf{(G20)}$$

Expressing the terms on the right hand side of equation (G20) in terms of the structure and motion parameters (using relations (3.11c-f) and (4.2a,b)), the two roots for θ in this case can be shown to be

$$\tan\theta = \frac{V_y}{V_x} , - \frac{V_x}{V_y} \quad \left(OR \ \frac{Z_Y}{Z_X} , - \frac{Z_X}{Z_Y} \right) . \quad \textbf{(G21)}$$

Note that the two roots are such that their absolute difference is equal to $\pi/2$ radians. For each of the above two roots we may solve for r from either (4.12b) or (4.12c). In this case equations (4.12b) and (4.12c) are identical and can be written as

$$r^2 + \frac{1}{2}\left(\frac{v_{xx}}{s} + \frac{u_{yy}}{c} \right) r - V_z \frac{a_1}{2cs} = 0 . \quad \textbf{(G22)}$$

Equation (G22) gives up to two solutions for r for each θ. Therefore up to four solutions are possible in this case. •

Lemma 3a : **Under the conditions stated in Theorem 7 the structure parameters satisfy (G6c and) the following constraint:**

$$-\frac{Z_X}{Z_Y} = \frac{1}{Z_{yy}} \left(-Z_{xy} \pm \sqrt{Z_{xy}^2 - Z_{xx} Z_{yy}} \right) . \quad \textbf{(G23)}$$

Proof : **Since the two roots for θ given by (G21) are also the roots of the cubic equation (4.12a), from Theorem 1 we see that relation (G23) should be true of the structure parameters. • As pointed out in Appendix D (see the discussion at the begining of the section titled ‘‘Relations between the two interpretations’’ in Appendix D), the image flow equations always have one solution which corresponds to the actual shape and motion of the surface in the physical world. Here also we use the notation described in Appendix D to give the relation between the correct and the spurious solutions. Also, for the case considered in Theorem 7 and other cases (considered later) where the translation vector, the surface normal, and the optical axis all lie in a plane, we will define a quantity k such that**

$$Z_X = kc \quad \text{and} \quad Z_Y = ks . \quad \textbf{(G24)}$$

Lemma 3b : **Under the conditions stated in Theorem 7, the spurious solutions are related to the actual shape and motion parameters by**

$$\theta_1 = \theta_0 \pm \pi/2 , \quad r_1 = -V_{z0} k_0 , \quad \textbf{(G25a,b)}$$

and r_2 and r_3 are the roots of the quadratic equation

$$r^2 + \frac{1}{2}\left(\frac{s_0 Z_{xx0}}{\pm c_0} + \frac{c_0 Z_{yy0}}{\mp s_0} \right) r + (V_{z0} r_0 k_0) = 0 . \quad \textbf{(G25c)}$$

Proof : **Relation (G25a) is easily derived from the two solutions for θ given by (G21) (note that $\tan\theta_0 = V_y/V_x$). In order to derive (G25b), in equation (G22) we substitute $s \leftarrow s_0$ and $c \leftarrow c_0$, and substitute for all the image motion parameters in terms of r_0 and θ_0 (using relations (4.3a,b),(3.11i,j), (3.11e,f) and (G24)). Then using relation (G6c) we can derive**

$$r^2 + (V_{z0} k_0 - r_0) r - (V_{z0} r_0 k_0) = 0 . \quad \textbf{(G26)}$$

The two roots of the above quadratic equation are r_0 and $-V_{z0} k_0$. Hence the relation (G25b). To derive relation (G25c) we first substitute $s \leftarrow s_1$, $c \leftarrow c_1$ in relation (G22). Then, as before, we substitute for all the image motion parameters and θ_1 in terms of r_0 and θ_0 to get relation (G25c). •

Lemma 3c : **Under the conditions stated in Theorem 7, for $\theta=\theta_0$ the solution for r is unique when the direction of translation is parallel to the surface normal.**

Proof : **In this case the solutions for r are the roots of the quadratic equation (G26) which are r_0 and $-V_{z0}k_0$. Using relations (4.3a,b) and (G24) we can easily show that these two roots are equal when the translation vector (V_x, V_y, V_z) is parallel to the surface normal vector $(Z_X, Z_Y, -1)$. •**

Lemma 3d : **Under the conditions stated in Theorem 7, the spurious solution for θ given by relation (G25a) is a valid solution if and only if the roots of the quadratic equation (G25c) are real. • Using relations (G6c), (G23) and the condition that the roots of the quadratic equation (G25c) be real we have constructed an algorithm to generate numerical examples which result in up to four solutions for the image flow equations. One such example is given in Chapter 7 (Example 3). In Section G2.4 we have seen one case where multiple interpretations arise because of multiple solutions for both θ and r. Next we consider another case of multiple interpretations characterized by a unique solution for θ but two solutions for r.** *Theorem 8 :* **Suppose that the curved-surface-condition is true and (G5a,b) are false. Then a solution for θ (obtained by solving equation (4.12a)) gives two solutions for r (and consequently for the image flow equations (3.11a-l))** *if and only if* **the following conditions are true of θ :**

$$\tan\theta = \frac{v_{yy}-2v_0+v_{xx}}{u_{xx}-2u_0+u_{yy}} \quad \text{and} \quad \tan 2\theta = \frac{a_1}{a_2} \tag{G27a,b}$$

Proof : **We will first prove the** *only if* **part. For a given θ we solve for r by solving the two quadratic equations (4.12b,c) (each of which may yield up to two roots) and take the root(s) common to both as the solution. If there is no common root or the roots are complex then the given θ is not a valid solution. If there is one common root then the solution for r is unique. If both roots are (real and) common then the coefficients of the two quadratic equations have to be proportional (i.e. the two quadratic equations become linearly dependent). Equating the ratios of the corresponding coefficients of the two equations (4.12b,c) we have**

$$\frac{2cs}{2cs} = \frac{v_{xx}c-(u_{xx}-2u_0)s}{u_{yy}s-(v_{yy}-2v_0)c} = \frac{s(a_1s+a_2c)}{c(a_1c-a_2s)} = 1\ . \tag{G28}$$

Therefore, equating the numerator and the denominator of the second term in equation (G28) we get condition (G27a) and equating the numerator and the denominator of the third term we get condition (G27b). The *if* **part can be similarly proved. • Note that the equality of the numerator and the denominator of the third term in equation (G28) together with (4.3a,b) and (4.3e,f) implies**

$$\tan\theta \;=\; V_y/V_x \;=\; Z_Y/Z_X\ . \tag{G29}$$

Equation (G29) implies that the translation vector $(V_x,\ V_y,\ V_z)$, the surface normal $(Z_X,\ Z_Y,\ -1)$ and the optical axis (the Z-axis) all lie in a common plane. Also note that we can check that there is no θ for which there are two solutions by first solving for θ from equation (G27a) and checking for the validity of (G27b) (or, the right hand sides of equations (G27a) and (G27b) can be directly related by using the relation (G19) between tanθ and tan2θ). In this case, the condition for the solution for r to be unique can be derived as follows: since equations (4.12b) and (4.12c) are linearly dependent, we add them to get the following equation:

$$r^2+\frac{1}{4}\left(\frac{v_{xx}-v_{yy}+2v_0}{s}+\frac{u_{yy}-u_{xx}+2u_0}{c}\right)r-V_z\frac{a_1}{2cs} = 0\ . \tag{G30}$$

In terms of the actual structure and motion parameters, the above equation can be shown to reduce to equation (G26) whose roots are r_0 and $-V_{z0}k_0$. When these two roots are equal we can show that the translation vector is parallel to the surface normal (see Lemma 3c). *Theorem 9 :* **Suppose that the curved-surface-condition is true, (G5a,b) are false, and there is no θ which satisfies (G27a,b) (i.e. the translation vector and the surface normal do not lie in a plane) then there exists a unique solution for the image flow equations (3.11a-l).**

Proof : **To prove this, we simply observe that there is always one solution for θ computed by solving equation (4.12a) which corresponds to** $\tan\theta = V_y/V_x$ **and this θ and the corresponding** r **computed from, say, (4.12b) always satisfies the constraint equation (4.12c). Multiple solutions are ruled out due to the previous theorems. •**

APPENDIX H

Solving for θ when the direction of V changes

Equating the right hand sides of the two equations (5.10a,b), substituting for V_z and Ω_Z in terms of θ using relations (4.3c,d), and simplifying, we get the following equation for θ:

$$(b_1 + b_2 \cos^2\theta + b_3 \sin^2\theta + b_4 \cos\theta \sin\theta) \quad \textbf{(H1)}$$

$$(b_5 \cos^3\theta + b_6 \cos^2\theta \sin\theta + b_7 \cos\theta \sin^2\theta + b_8 \sin^3\theta) +$$

$$(c_1 + c_2 \cos^2\theta + c_3 \sin^2\theta + c_4 \cos\theta \sin\theta)$$

$$(c_5 \cos^3\theta + c_6 \cos^2\theta \sin\theta + c_7 \cos\theta \sin^2\theta + c_8 \sin^3\theta) = 0$$

where b_i and c_i are constants given by

$$b_1 = u_t\ ,\quad b_2 = u_0\, u_x + a_1\, v_0\ ,\quad b_3 = u_0\, u_x\ ,\quad b_4 = -\, a_2\, v_0 \quad \textbf{(H2a-d)}$$

$$b_5 = -\, v_x\ ,\quad b_6 = a_2 - 2\, v_y\ ,\quad b_7 = 2\, a_1 + u_y\ ,\quad b_8 = -\, 2\, u_x \quad \textbf{(H2e-h)}$$

$$c_1 = v_t\ ,\quad c_2 = v_0\, v_y\ ,\quad c_3 = v_0\, v_y + a_1\, u_0\ ,\quad c_4 = a_2\, u_0 \quad \textbf{(H3a-d)}$$

$$c_5 = 2\, v_y\ ,\quad c_6 = -\, v_x - 2\, a_1\ ,\quad c_7 = a_2 + 2\, u_x \text{ and } \quad c_8 = u_y\ . \quad \textbf{(H3e-h)}$$

Now, multiplying b_1, c_1 in relation (H1) by $\cos^2\theta + \sin^2\theta$ and simplying, equation (H1) can be further reduced to

$$d_1 \tan^5\theta + d_2 \tan^4\theta + d_3 \tan^3\theta + d_4 \tan^2\theta + d_5 \tan\theta + d_6 = 0 \quad \textbf{(H4)}$$

where

$$d_1 = (b_1 + b_3) b_8 + (c_1 + c_3) c_8\ , \quad \textbf{(H5a)}$$

$$d_2 = b_4 b_8 + (b_1 + b_3) b_7 + c_4 c_8 + (c_1 + c_3) c_7\ , \quad \textbf{(H5b)}$$

$$d_3 = (b_1 + b_2) b_8 + b_4 b_7 + (b_1 + b_3) b_6 + (c_1 + c_2) c_8 + c_4 c_7 + (c_1 + c_3) c_6\ , \quad \textbf{(H5c)}$$

$$d_4 = (b_1 + b_2) b_7 + b_4 b_6 + (b_1 + b_3) b_5 + (c_1 + c_2) c_7 + c_4 c_6 + (c_1 + c_3) c_5\ , \quad \textbf{(H5d)}$$

$$d_5 = (b_1 + b_2) b_6 + b_4 b_5 + (c_1 + c_2) c_6 + c_4 c_5\ , \text{ and} \quad \textbf{(H5e)}$$

$$d_6 = (b_1 + b_2) b_5 + (c_1 + c_2) c_5\ . \quad \textbf{(H5f)}$$

In this case, there are two special situations which deserve mention. In both these cases, the orientation of the surface patch is indeterminate as there is no translation parallel to the image plane. For brevity, the two situations are summarized below:

$$[\,(u_t = -\, u_0\, u_x) \text{ and } (v_t = -\, v_0\, v_y) \text{ and } (u_x = v_y) \text{ and } (u_y = -\, v_x) \text{ and} \quad \textbf{(H6a)}$$

$$(u_x \neq 0 \text{ or } u_y \neq 0)\,] \rightarrow [\,(V_x = V_y = 0) \text{ and } (Z_X, Z_Y \text{ are indeterminate})\,]$$

$$[\,(u_t = -\, u_0\, u_x) \text{ and } (v_t = -\, v_0\, v_y) \text{ and } (u_x = v_y) \text{ and } (u_y = -\, v_x) \text{ and} \quad \textbf{(H6b)}$$

$(u_x = 0)$ and $(u_y = 0)\,] \rightarrow [\,((V_x = V_y = 0)$ and $(Z_X, Z_Y$ are indeterminate$))$

or $((Z_X = Z_Y = 0)$ and $(V_x, V_y$ are indeterminate$))\,]$

APPENDIX I

Solving for r and θ when the camera tracks a point

Differentiating equation (3.1a) twice with respect to time t yields

$$\ddot{x} = \frac{\ddot{X}}{Z} - 2\,\frac{\dot{X}}{Z}\,\frac{\dot{Z}}{Z} + \frac{X}{Z}\left(2\,\frac{\dot{Z}^2}{Z^2} - \frac{\ddot{Z}}{Z}\right) . \tag{I1}$$

In the above expression, $\dot{X}$, $\dot{Y}$ and $\dot{Z}$ are given by relations (3.2a-c) and $\ddot{X}$, $\ddot{Y}$ and $\ddot{Z}$ are easily derived from these. For example,

$$\ddot{X} = -\,\boldsymbol{\Omega}_Y\,\dot{Z} + \boldsymbol{\Omega}_Z\,\dot{Y} . \tag{I2}$$

From these, we express $\ddot{x}$ in terms of only $\boldsymbol{V}$, $\boldsymbol{\Omega}$, X, Y and Z and evaluate it at the image origin, i.e. $(X, Y, Z) = (0, 0, Z_0)$ where $Z_0 > 0$. Denoting $\ddot{x}$ evaluated at the image origin by $\dot{u}$ we can derive

$$\dot{u} = \boldsymbol{\Omega}_Z\,(\boldsymbol{\Omega}_X - V_y) - V_z\,(\boldsymbol{\Omega}_Y + V_x) - V_x\,V_z . \tag{I3}$$

Using relations (3.11a,b) the above equation can be reexpressed as

$$\dot{u} = v_0\,\boldsymbol{\Omega}_Z + u_0\,V_z - V_x\,V_z \tag{I4}$$

or

$$V_x = (u_0\,V_z + v_0\,\boldsymbol{\Omega}_Z - \dot{u})\,/\,V_z . \tag{I5a}$$

Similarly, starting from equation (3.1b) and following steps similar to those above, we can derive

$$V_y = (v_0\,V_z - u_0\,\boldsymbol{\Omega}_Z - \dot{v})\,/\,V_z . \tag{I5b}$$

In equations (I5a,b) we substitute for all unknowns in terms of r and θ from (4.3a-d) and eliminate r and solve for θ to get relation (5.13). Relation (5.14) which gives the solution for r is easily obtained from relations (I5a,b) and (4.3a,b).

APPENDIX J

Surface deformation parameters

We have chosen to describe the deformation of a small surface patch in 3D space in terms of the deformation of a small volume element embedding the surface patch (Figure 19). To a first approximation, the deformation parameters of a small volume element are given by the components of its velocity gradient tensor. The physical interpretation of the velocity gradient tensor shows that an arbitrary time variation of a small surface patch can be expressed as the combined effect of a pure translation, a pure rotation, a pure accleration and a deformation (see the last part of this Appendix). Also, the velocity gradient tensor representation gives explicit conditions for rigid motion, pure translation, etc.. Consider a Cartesian coordinate system with axes x_1, x_2 and x_3. The gradient tensor of a velocity vector $\boldsymbol{v} = (v_1, v_2, v_3)$ can be written as the sum of symmetric and antisymmetric parts,

$$\frac{\partial v_i}{\partial x_j} = \frac{1}{2}\left\{\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right\} + \frac{1}{2}\left\{\frac{\partial v_i}{\partial x_j} - \frac{\partial v_j}{\partial x_i}\right\} \tag{J1a}$$

$$= e_{ij} + \omega_{ij} \quad i, j = 1, 2, 3 . \tag{J1b}$$

It can be shown that the three independent parameters of the antisymmetric tensor ω_{ij} correspond to the components of a rigid body rotation, and, if the motion is a rigid one (composed of a translation plus a rotation), all the components of the symmetric tensor e_{ij} will vanish. For this reason the tensor e_{ij} is called the *deformation* or *rate of strain* tensor and its vanishing is necessary and sufficient for the motion to be without deformation, that is, rigid. A component e_{ii} of this tensor gives the rate of longitudinal strain of an element parallel to the x_i axis. A component e_{ij}, $i \neq j$, represents one-half the rate of decrease of the angle between two segments originally parallel to the x_i and x_j axes respectively. For a more detailed treatment of these topics, see Aris (1962). From our discussions above, the interpretation of the motion and deformation parameters a_{ij} in equations (6.1a-c) with respect to $(X, Y, Z, t) = (0, 0, Z_0, 0)$ can be summarized as follows:

$$(a_{10}, a_{20}, a_{30}) : \text{rigid body translation} \tag{J2a}$$

$$\frac{1}{2}(a_{23} - a_{32}, a_{31} - a_{13}, a_{12} - a_{21}) : \text{rigid body rotation} \tag{J2b}$$

$$(\dot{a}_{10}, \dot{a}_{20}, \dot{a}_{30}) : \text{rigid body acceleration} \tag{J2c}$$

$$(a_{11}, a_{22}, a_{33}) : \text{measures stretching} \tag{J2d}$$

$$\frac{1}{2}(a_{12} + a_{21}, a_{23} + a_{32}, a_{31} + a_{13}) : \text{measures shear .} \tag{J2e}$$

It is interesting to note that an arbitrarily time-varying surface patch can be described, to a first approximation, in terms of a rigid translation plus a rigid rotation plus a rigid acceleration plus a deformation.

BIBLIOGRAPHY

Adiv, G. 1985a (July). Determining 3-D motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-7, pp. 384-401.**

Adiv, G. 1985b (December). Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field. *Proceedings of DARPA Image Understanding Workshop*, **pp. 399-412.**

Aggarwal, J.K. 1986 (May). Motion and time-varying imagery - an overview. *Proceedings of IEEE Workshop on Motion: Representation and Analysis*, **pp. 1-6.**

Aloimonos, J., and C.M. Brown. 1986. Perception of structure from motion: I : Optic flow vs. discrete displacements, II: Lower bound results. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, **pp. 510-517.**

Aloimonos, J., and A. Basu. 1986. Shape and 3-D motion from contour without point to point correspondences: General principles. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, **pp. 518-527.**

Aloimonos, J., and I. Rigoutsos. 1986. Determining the 3-D motion of a rigid planar patch without correspondence under perspective projection. *Proceedings of IEEE Workshop on Motion: Representation and Analysis*, **pp. 167-174.**

Anandan, P. 1986. Measuring Visual Motion From Image Sequences. Ph.D. dissertation, COINS Department, University of Massachusetts, Amherst.

Aris, R. 1962. *Vectors, Tensors, and the Basic Equations of Fluid Mechanics*. **Englewood Cliffs: Prentice-Hall.**

Bandopadhyay, A., and J. Aloimonos. 1985 (March). Perception of rigid motion from spatio-temporal derivatives of optical flow. TR 157, Department of Computer Science, University of Rochester.

Bandopadhyay, A., B. Chandra, and D.H. Ballard. 1986. Active navigation: tracking an environmental point considered beneficial. *Proceedings of IEEE Workshop on Motion: Representation and Analysis*, **pp. 23-29.**

Bolles, R.C., and H.H. Baker. 1985 (October). Epipolar-plane image analysis: a technique for analyzing motion sequences. *Proceedings of Third IEEE Workshop on Computer Vision: Representation and Control***, pp. 168-178.**

Broida, T.J., and R. Chellappa. 1986a (January). Estimation of object motion parameters from noisy images. *IEEE Transactions on Pattern Analysis and Machine Intelligence***, PAMI-8, pp. 90-99.**

Broida, T.J., and R. Chellappa. 1986b (May). Kinematics and structure of a rigid object from a sequence of noisy images. *Proceedings of IEEE Workshop on Motion: Representation and Analysis***, pp. 95-100.**

Bruss, A. R. and B. K. P. Horn. 1983. Passive navigation. *Computer Vision, Graphics and Image Processing,* **Vo. 21, No. 1, pp. 3-20.**

Chen, S. 1985 (October). Structure from motion without the rigidity assumption. *Proceedings of Third IEEE Workshop on Computer Vision: Representation and Control***, pp. 105-112.**

Clocksin, W. 1980. Perception of surface slant and edge labels from optical flow: A computational approach. *Perception,* **9, pp. 253-269.**

Fennema, C. L. and W. R. Thompson. 1979. Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing,* **Vo. 9, No. 4, pp. 301-315.**

Feynman, R.P. 1948. Space-time approach to non-relativistic quantum mechanics. *Reviews of Modern Physics,* **20, p. 267.**

Fleet, D. J. and A. D. Jepson. 1985. Velocity extraction without form interpretation. *Proceedings of the Third Workshop on Computer Vision: Representation and Control,* **pp. 179-185, October, Bellaire, Michigan.**

Gibson, J.J. 1950. *The Perception of the Visual World***. Boston: Houghton Mifflin.**

Gibson, J.J. 1966. *The Senses Considered as Perceptual Systems***. Boston: Houghton Mifflin.**

Gibson, J.J., P. Olum, and F. Rosenblatt. 1955. Parallax and perspective during aircraft landings. *American Journal of Psychology***, 68, pp. 372-385.**

Hay, J.C. 1966. Optical motions and space perception: an extension of Gibson's analysis. *Psychological Review,* **73, pp. 550-565.**

Heeger, D. J. 1987. Optical flow from spatio-temporal filters. *Proceedings of the First International Conference on Computer Vision,* **London, pp. 181-190.**

Helmholtz, H. von. 1925. *Treatise on Physiological Optics*. **Ed. J.P.C. Southall, New York: Dover.**

Hildreth, E.C. 1983 (August). The Measurement of Visual Motion. Ph.D. Thesis, M.I.T.

Hoffman, D.D. 1980 (December). Inferring shape from motion fields. *A.I. Memo No. 592,* **M.I.T.**

Hoffman, D.D., and B.M. Bennet. 1985. Inferring the relative three-dimensional positions of two moving points. *Journal of the Optical Society of America* **A 2, pp. 350-353.**

Horn, B.K.P., and B.G. Schunck. 1981. Determining optical flow. *Artificial Intelligence,* **17, pp. 185-204.**

Jain, R. 1983. Direct computation of the focus of expansion. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **Vol. 5, No. 1, pp. 58-64.**

Johansson, G. 1973. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics,* **14, pp. 201-211.**

Johansson, G. 1975. Visual motion perception. *Scientific American,* **232, pp. 76-88.**

Kanatani, K. 1985. Structure from motion without correspondence: general principle. *Proceedings of 9th International Joint Conference on Artificial Intelligence***, pp. 886-888.**

Kanatani, K. 1986. *Group Theoretical Methods in Image Understanding***, CAR-TR-214, Center for Automation Research, University of Maryland.**

Koenderink, J.J., and A.J. van Doorn. 1975. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Optica Acta,* **22, pp. 773-791.**

Koenderink, J.J., and A.J. van Doorn. 1976. Local structure of movement parallax of the plane. *Journal of the Optical Society of America***, 66, pp. 773-791.**

Koenderink, J.J., and A.J. van Doorn. 1986. Depth and shape from differential perspective in the presence of bending deformations. *Journal of the Optical Society of America***, A 3, pp. 242-249.**

Lee, D.N. 1974. Visual information during locomotion. *Perception: Essays in Honor of James J. Gibson***, I.D.G. MacLed and O.Pick, eds. Ithaca, N.Y.: Cornell University Press.**

Liu, Y., and T.S. Huang. 1986 (May). Estimation of rigid body motion using straight line correspondences. *Proceedings of IEEE Workshop on Motion: Representation and Analysis*, **pp. 47-52.**

Longuet-Higgins, H.C., and K. Prazdny. 1980. The interpretation of a moving retinal image. *Proceedings of the Royal Society of London*, **B 208, pp. 385-397.**

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

Longuet-Higgins, H.C. 1984. The visual ambiguity of a moving plane. *Proceedings of the Royal Society of London*, **B 223, pp. 165-175.**

Marimont, D.H. 1986 (May). Projective duality and the analysis of image sequences. *Proceedings of IEEE Workshop on Motion: Representation and Analysis*, **pp. 7-14.**

Marr, D. 1977. Artificial intelligence - a personal view. *Artificial Intelligence,* **9, pp. 37-48.**

Marr, D. 1982. *Vision.* **San Francisco: Freeman.**

Marr, D., and T. Poggio. 1977. From understanding computation to understanding neural circuitry. *Neurosciences Research Progress Bulletin*, **15, pp. 470-488.**

Martin, W. N., and J. K. Aggarwal (Editors). 1988. *Motion Understanding: Robot and Human Vision,* **Kluwer Academic Publishers, Norwell, Massachusetts.**

Maybank, S.J. 1985. The angular velocity associated with the optical flowfield arising from motion through a rigid environment. *Proceedings of the Royal Society of London*, **A 401, pp. 317-326.**

McConnell, A.J. 1957. *Applications of Tensor Analysis*. **New York: Dover.**

Mitiche, A., S. Seida, and J.K. Aggarwal. 1986 (May). Line-based computation of structure and motion using angular invariance. *Proceedings of IEEE Workshop on Motion: Representation and Analysis*, **pp. 175-180.**

Nagel, H. H. 1983a. Displacement vectors derived from second-order intensity variations in image sequences. *Computer Vision, Graphics and Image Processing,* **Vo. 21, No. 1, pp. 85-117.**

Nagel, H. H. 1983b. Constraints for the estimation of displacement vector fields from image sequences. *Proceedings of the International Joint Conference on Artificial Intelligence,* **Karlsruhe, West Germany, pp. 945-951.**

Nakayama, K., and J.M. Loomis. 1974. Optical velocity patterns, velocity sensitive neurons and space perception. *Perception*, **3, pp. 63-80.**

Negahdaripour, S. 1986. Direct Passive Navigation. Ph.D. thesis. Dept. of Mechanical Engineering, M.I.T., Cambridge, Massachusetts.

O'Neill, B. 1966. *Elementary Differential Geometry*. **New York: Academic Press, p. 204.**

Piskunov, N. 1974. *Differential and Integral Calculus, Vol. I*, **Moscow: Mir Publishers, p. 264.**

Rektorys, K. (Editor). 1969. *Survey of Applicable Mathematics.* **Cambridge: MIT Press, pp. 77-79.**

Schiff, W., J.A. Caviness, and J.J. Gibson. 1962. Persistent fear responses in rhesus monkeys in response to the optical stimulus of ''looming''. *Science*, **136, pp. 982-983. Sethi, I. K. and R. Jain. 1987. Finding trajectories of feature points in a monocular image sequence.** *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **PAMI-9, No. 1, pp. 56-73.**

Shariat, H. 1986. The Motion Problem: How To Use More Than Two Frames. Ph.D. dissertation, University of Southern California, Dept. of Electrical Engineering, Los Angeles, California.

Sinha, S.S. and A.M. Waxman. 1984 (July). An image flow simulator. CAR-TR-71, Center for Automation Research, University of Maryland.

Spetsakis, M. E. and J. Aloimonos. 1987. Closed-form solution to the structure from motion problem from line correspondences. *Proceedings of the Sixth National Conference on Artificial Intelligence,* **pp.738-743, Seattle, Washington.**

Subbarao, M. 1986a (April). Interpretation of image motion fields: rigid curved surfaces in motion. CAR-TR-199, Center for Automation Research, University of Maryland.

Subbarao, M. 1986b (May). Interpretation of image motion fields: a spatio-temporal approach. *Proceedings of IEEE Workshop on Motion: Representation and Analysis*, **pp. 157-166.**

Subbarao, M., and A.M. Waxman. 1985 (October). On the uniqueness of image flow solutions for planar surfaces in motion. *Proceedings of Third IEEE Workshop on Computer Vision: Representation and Control*, **pp. 129-140.**

Thompson, W. B., K. M. Mutch, and V. A. Berzins. 1985. Dynamic occlusion analysis in optical flow fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **Vol. PAMI-7, No. 4, pp. 374-383.**

Tsai, R.Y., and T.S. Huang. 1981 (November). Estimating three-dimensional motion parameters of a rigid planar patch. Technical Report R-922, Coordinated Science Laboratory, University of Illinois.

Tsai, R.Y., and T.S. Huang. 1984 (January). Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence***, PAMI-6, pp. 13-26.**

Tsotsos, J. K. 1987. A 'Complexity Level' analysis of vision. *''Proceedings***of London, England, pp. 346-355.**

Ullman, S. 1979. *The Interpretation of Visual Motion***. Cambridge: MIT Press.**

Ullman, S. 1984. Maximizing the rigidity: the incremental recovery of 3-D structure from rigid and rubbery motion. *Perception***, 13, pp. 255-274.**

Wallach, H., and D. N. O'Connell. 1953. The kinetic depth effect. *Journal of Experimental Psychology***, 45, pp. 205-217.**

Watson, A. B. and A. J. Ahumada. 1985. Model of human visual motion sensing. *Journal of the Optical Society of America A,* **2 (2), pp. 322-342.**

Waxman, A.M. 1984a. Dynamics of a couple-stress fluid membrane. *Studies in Applied Mathematics***, 70, pp. 63-86.**

Waxman, A.M. 1984b (April). An image flow paradigm. *Proceedings of IEEE Workshop on Computer Vision: Representation and Control,* **pp. 49-57.**

Waxman, A.M., and J.H. Duncan. 1986. Binocular image flows. *Proceedings of IEEE Workshop on Motion: Representation and Analysis***, pp. 31-38.**

Waxman, A.M., B. Kamgar-Parsi, and M. Subbarao. 1986 (February). Closed-form solutions to image flow equations for 3-D structure and motion. CAR-TR-190, Center for Automation Research, University of Maryland.

Waxman, A.M., and S. Ullman. 1985. Surface structure and three-dimensional motion from image flow kinematics. *International Journal of Robotics Research* **4 (3), pp. 72-94.**

Waxman, A.M. and K. Wohn. 1985. Contour evolution, neighborhood deformation and global image flow: planar surfaces in motion. *International Journal of Robotics Research***, 4 (3), pp. 95-108.**

Waxman, A.M. and K. Wohn. 1986. Image flow theory: a framework for 3-D inference from time-varying imagery. To appear in *Advances in Computer Vision***, ed. C.M. Brown, Hillside, NJ: Erlbaum.**

Wohn, K. 1984. A contour-based approach to image flow. Ph.D. Thesis, Department of Computer Science, University of Maryland.

Wohn, K. and A.M. Waxman. 1985 (July). Contour evolution, neighborhood deformation and local image flow: Curved surfaces in motion. CAR-TR-134, Center for Automation Research, University of Maryland.

Wohn, K. and J. Wu. 1986 (January). Recovery of 3-D motion from first-order deformation. Robotics Laboratory Technical Report, Harvard University.

$#

Figure 1. Ullman's (1979) two cylinder experiment which demonstrates the perception of three-dimensional shape and motion from visual motion. Positions of about 100 ponts lying on the surfaces of two coaxial cylinders were stored in a computer's memory. The orthographic projections of these points on a frontal plane were computed and displayed on a CRT screen. (The outlines of the cylinders were not presented in the actual display.) Although the density of dots in the projected image increased at the edges of each cylinder, in the combined image of the two cylinders the dot pattern was complex and was ineffective in revealing the two cylinders. However when a changing projection of the dots was presented corresponding to a rotation of the two cylinders, the shape and motion of the cylinders were immediately percieved. Since each instantaneous view contained no shape information, visual motion alone was sufficient to recover both shape and motion of the cylinders.

Figure 2. The domain of visual intensity stimulus function *I*.

Figure 3. *Image flow* **(also referred to as** *optical flow)* **can be defined as the velocity field obtained by projecting the instantaneous three-dimensional velocities of points on visible surfaces in the scene on to the image screen. In the above diagram,** *P* **is a point on a visible surface and** *p* **is its image. If** ***U*** **is the instantaneous velocity of** *P* **in the world, then its projection** ***v*** **on the image plane is the velocity of** *p***. Thus image flow describes the motion of images of geometric points in the scene.**

Figure 4. (a) A scene consisting of a sphere in front of a cone with background at infinity. (b) The instantaneous image flow associated with a general rigid body motion of the camera. (These pictures were generated using the image flow software developed by Sinha and Waxman, 1984.)

Figure 5. The optical flow equation constrains at each point only the component of image velocity at that point normal to the iso-intensity contour passing through that point. Therefore the observed motion of brightness patterns determine only the normal displacement of iso-intensity contours. Their tangential displacement remains undetermined.

Figure 6. Aperture problem. The motion of a brightness pattern in a small observation window (aperture) cannot be determined uniquely. C: **An iso-intensity contour in the first image frame.** C'**: the iso-intensity contour in the second image frame which corresponds to** C **in the first frame.** $\boldsymbol{v}_n$: **normal displacement of intensity pattern.** $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\boldsymbol{v}_n$ **: possible displacements of the brightness pattern.**

Figure 7. A pin-hole camera.

Figure 8. Camera model and coordinate systems (after Longuet-Higgins and Prazdny, 1980). $OXYZ$ **is a Cartesian coordinate system fixed to the camera.** O **is the focus and the positive** Z **axis is the direction of view. The image plane is located at unit distance from** O **along the direction of view.** oxy **is the image coordinate system with its origin at (0,0,1). A point** P **at** (X,Y,Z) **in the world produces an image** p **at** (x,y) **on the image plane where** $x=X/Z$ **and** $y=Y/Z$**. The instantaneous relative rigid body motion of the camera with respect to a surface** S **is given by the translation vector** (V_X,V_Y,V_Z) **and the rotation vector** $(\Omega_X,\Omega_Y,\Omega_Z)$**. If** $\boldsymbol{U}$ **is the instantaneous relative three-dimensional velocity of the point** P **(on surface** S**) then the image velocity of point** p **is** $\boldsymbol{v}$ **where** $\boldsymbol{v}$ **is the projection of** $\boldsymbol{U}$ **on the image plane.**

Figure 9. Analysis of image flow in a field of view not along the optical axis. The observed original image sequence in the field of view is first projected onto a plane perpendicular to the field of view. The image flow parameters are computed from this projected image sequence. The scene parameters (in a coordinate system rotated with respect to the original axes) are then obtained by solving the image flow equations.

Figure 10. The relation of translation parameters to *r* and θ.

Figure 11. (a) A *divergent* image flow field. (b) The effect of divergent image flow on an image neighborhood.

Figure 12. (a) A *rotational* image flow field. (b) The effect of rotational image flow on an image neighborhood.

Figure 13. (a) A *shear* image flow field. (b) The effect of shear image flow on an image neighborhood.

Figure 14. The relation of the bounds on velocity of approach V_z to the differential invariants of image flow. A similar diagram can be drawn for angular velocity Ω_Z where 'divergence' is replaced by '-curl'.

Figure 15. The surface gradient vector, k_0 and θ_0.

Figure 16. The 3D geometry of a general case of motion.

Figure 17. A specular saddle surface. The tangent plane to the surface is frontal (i.e. perpendicular to the direction of view or parallel to the image plane) and the surface itself is saddle shaped.

Figure 18. An object undergoing uniform translation with respect to an external reference frame and spinning uniformly about an axis fixed with respect to itself. The individual components of translation with respect to a coordinate system fixed to the object change continuously due to rotation, but the magnitude of translation remains the change.

Figure 19. A surface element embedded in a volume element. Analytic deformation of the surface element can be described in terms of an analytic deformation of the embedding volume element.

Figure 20. A scheme for resolving the two-fold ambiguity of interpretation for planar surfaces.

TABLE OF CONTENTS