PROCESSING DYNAMIC IMAGE SEQUENCES FROM A MOVING SENSOR

Daryl T. Lawton

COINS Technical Report 84-05

February 1984

This work was supported in part by the Office of Naval Research under grant number N00014-75-C-0459 and the Advanced Research Projects Agency under grant number N00014-82-K-0464.

Processing Dynamic Image Sequences from a Moving Sensor

A Dissertation Presented

By

DARYL TALIESEN LAWTON

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 1984

Department of Computer and Information Science

This research was supported in part by:

The Office of Naval Research
Grant Number N00014-75-C-0459

and

The Advanced Research Projects Agency
Grant Number N00014-82-K-0464

Processing Dynamic Images
from a Moving Sensor

A Dissertation Presented

By

DARYL TALIESEN LAWTON

Approved as to style and content by:

Edward M. Riseman, Chairperson of Committee

Allen R. Hanson, Member

Nico S. Spinelli, Member

Robert L. Huguenin, Outside Member

Edward M. Riseman, Chairman
Department of Computer and Information Science

For my Genetic Buddies,

Sam, Sarah, Fritzie, and Dennis

ACKNOWLEDGMENTS

I have many people to thank, in ways more special than are possible here. I owe a tremendous amount to my principle advisors, Ed Riseman and Al Hanson (or Ednal to many of us) for several types of support and even more types of patience. Ed in particular lovingly badgered me through some periods of intense laziness and anguish. Without him, this thesis, and much else besides, would not have been completed. I would also like to thank Nico Spinelli for many enjoyable and insightful discussions and to Bob Huguenin for his cheerful enthusiasm and also putting up with being referred to, somewhat ambiguously, as my outside member.

I am very happy and proud to be a part of the UMASS VISIONS and MOTIONS groups for shared experiences, software, and a wide range of generally excessive behavior. I would particularly like to thank Joachim Rieger, Terry Weymouth, Frank Glaser, Gilad Adiv, George Reynolds, P. Anandan, Janet Turnbull, Martha Steenstrup, Ken Overton, Bert Shaw, Tom Williams, John Prager, Charles Kohl, Ralf Kohler, Steve Levitan, Chip Weems, Steve Epstein, Kate Greenspan, and the little Tex Master, Lenny Wesley.

Dr. Nelson Corby of the Machine Intelligence Laboratory of General Electric in Schenectady, New York made possible some of the industrial image sequences I have been working with.

ABSTRACT

Processing Dynamic Image Sequences
from a Moving Sensor

February, 1984

Daryl T. Lawton

B.S., University of California at Santa Cruz

M.S., Ph.D., University of Massachusetts at Amherst

Directed by: Professor Edward M. Riseman

A fundamental problem in motion processing research has been the discrepancy between the precision and reliability with which image displacements can be determined and the sensitivity of inference procedures to noise and resolution errors. There are also indications that these inference procedures are inherently unstable and, in some cases, ambiguous. The approach of this thesis has been to deal with restricted cases of motion for which the inference of the motion parameters, image displacements, and environmental depth, can be combined into a single, uniform, and mutually constraining computation. These restricted cases of motion are sufficient for a wide range of real-world tasks, especially since other associated sensing devices can be used to ascertain the other parameters of motion. We then apply the procedure developed for translational motion to local portions of image sequences to process general sensor motion as if it were composed of independent local environmental translations. The resulting representation can considerably simplify the processing of less restricted and general motion.

The procedure for processing translational motion robustly combines the determination of image displacements with the extraction of the direction of sensor motion. We present several experiments showing its behavior in a variety of situations. We also consider various extensions to this procedure for such things as developing it as a hierarchical computation; processing translational blur patterns; dealing with multiple independently moving objects; and using the translational procedure in the control of an autonomous vehicle.

Results are presented for two other restricted cases of motion: pure sensor rotation and motion constrained to a known plane. The results are similar to the translational case except that certain simple cases of planar motion are found to be inherently ambiguous.

We then process less restricted and general sensor motion by applying the procedure for translational motion processing to local areas of images. This results in a low level description of motion called the Environmental Direction of Motion Field (or EDMF) which associates a direction of environmental motion with extracted image features. This representation can greatly simplify the recovery of sensor motion parameters. We also develop the constraints associated with object rigidity in determining the inference of sensor motion parameters, and then show how these constraints are simplified by information in the EDMF.

We conclude with a summary of the major results of the thesis and mention future work, chiefly in the areas of architectures for real time motion processing, and applications to more challenging and specific domains.

TABLE OF CONTENTS

. . . . . .

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

The importance of processing dynamic information is obvious. Change is a basic and pervasive aspect of reality. Artificial perceptual systems which cannot deal with such dynamic information will be severely limited. They would not be able to determine basic causal and structural relations in the environment. They would not be able to move about and directly explore the world. These fundamental concerns, coupled with recent advances in sensor technology and attainable computing power, have made image motion processing an active area of research.

The work in dynamic image processing can be roughly divided into two types of techniques: those for determining the changes in a sequences of images and those for inferring environmental information from these transformations. Much basic work has been done on determining the displacements of distinguishable image points over time and inferring sensor motion and environmental depth from these displacements. A fundamental problem that has emerged in all this work is the discrepancy between the precision and reliability with which image displacements can be determined and the sensitivity of the inference procedures to noise and resolution errors. For example, some of the inference procedures require high order derivatives to be extracted from the determined image displacements. Additionally, there are indications that the problem itself is inherently unstable and, in some cases, ambiguous. This has lead to an interesting state of affairs: formulations which are often elegant but do not work in motion processing of real world situations, and therefore have limited practical application.

The approach of this thesis has been to deal with restricted cases of motion for which the inference of the motion parameters, image displacements, and, to some extent, environmental depth, can be combined into a single, uniform, and mutually constraining computation. These restricted cases of motion are sufficient for a wide range of real-world tasks, especially since other associated sensing devices can be used to ascertain the other parameters of motion. Finally, we apply the procedure developed for translational motion to local portions of image sequences to process general sensor motion as if it were composed of independent local environmental translations. The resulting representation can considerably simplify the processing of less restricted and general motion. A brief outline of the thesis follows.

Thesis Outline

Chapters two and three present background information on motion processing. In chapter two we review the general problems and previous work in image motion processing. In chapter three we review the basic structural relations between image displacements and sensor motion.

In chapter four we present a procedure for processing image sequences produced by translational motion of a sensor relative to a stationary environment. The procedure robustly combines the determination of image displacements with the extraction of the direction of sensor motion. Several experiments are performed to show the behavior of the procedure in different situations. As a part of the implementation we develop a simple feature extraction process.

In chapter five we consider various extensions to the translational procedure. These include developing the procedure as a hierarchical computation to increase

its speed; processing the blur patterns produced by prolonged exposures during translational motion; dealing with multiple independently moving objects; and using the translational procedure in the control of an autonomous vehicle by using devices to stabilize the sensor or directly determine the other parameters of motion.

In chapter six we consider two other restricted cases of motion: pure sensor rotation and motion constrained to a known plane. The results are very similar to the translational case except that certain simple cases of planar motion are found to be inherently ambiguous.

In chapter seven we process less restricted and general sensor motion by applying the procedure for translational motion processing to local areas of images. This results in a low level description of motion called the Environmental Direction of Motion Field (or EDMF) which associates a direction of environmental motion with extracted image features. This representation can greatly simplify the recovery of sensor motion parameters. We consider different ways of computing the EDMF and how sensor motion can be determined from it. We present a simple computation for the case of motion constrained to an unknown plane. We also develop the constraints associated with object rigidity in determining the inference of sensor motion parameters, and then show how these constraints are simplified by information in the EDMF.

In chapter eight we summarize the major results of the thesis and mention future work, chiefly in the areas of architectures for real time motion processing, and application to more challenging and specific domains.

CHAPTER II

THE NATURE OF MOTION PROCESSING

Introduction

A general outline of motion processing is shown in figure 1. This figure indicates a basic control loop in which the changes in a sequence of images are determined and represented, a model is inferred from these transformations, and the model is used to predict and constrain the processing of further and ongoing image transformations.

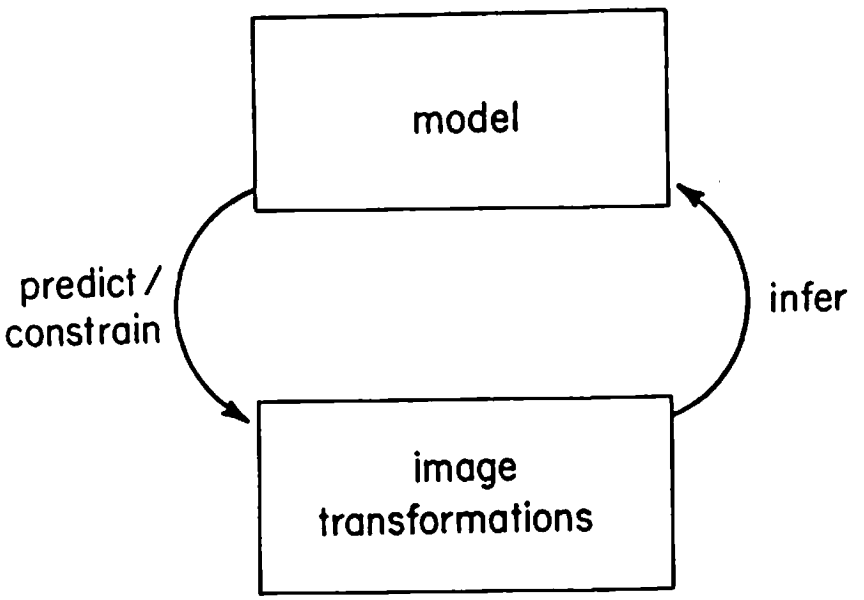


Figure 1. The General Structure of Motion Processing

Each of these elements — the image transformations, the inference of the model, the model itself, and the predictions — typically correspond to several different processes and representations which can vary significantly with application. In this representation, the beginning of the processing is ambiguous because of the circular nature of the organization. This is an aspect of what we will refer to as the start-up problem, and is concerned with whether it is possible to determine image transformations without an initial model. Generally, there is always an initial model which is either based upon domain specific information about the type of image transformations that can be expected to occur, or implicit in the procedures for determining image transformations by basing them upon general environmental properties such as continuity of motion and environmental surfaces.

One implication of the start-up problem is that motion processing always involves assumptions about the environment in which it is used. In many applications, these assumptions are quite specific and task dependent, as in target tracking. In others, the assumptions are more abstract and the resulting procedures have more general application, as in the case of constrained types of continuous motion, constrained types of environmental objects, or image transformations. A general area of research in motion processing has been concerned with the analysis of image sequences produced by rigid body motions in the environment. This problem lends itself to a theoretical development which does not become overly complex, yet also reflects a very common occurrence in the real world. A particular image transformation which this analysis can utilize is also well known — optic flow. This may be thought of as an almost classical problem in image processing: the inference of environmental information from the optic flow field generated by rigid body motions. In much of what follows, the static environment is viewed as a single rigid body and relative motion is induced by sensor motion.

Optic Flow

Optic flow is the vector field representing the changes in the positions of the images of environmental points over time. It was introduced by the psychologist J.J. Gibson [Gibs50, Gibs66, Gibs79] based, to some extent, on his experiences as a bomber pilot during the Second World War. Gibson was struck with how different patterns and extents of image displacements could specify critical environmental information for the control of behavior, such as heading, immediacy of collisions, and environmental layout. Gibson's analysis has proven to be extremely suggestive and stimulating, but incomplete, in two critical aspects. He assumed the optic flow field was a given and did not deal with the computational difficulties in determining it. He also did not explicitly (at least initially and never completely) analyze how environmental information was extracted from the flow field. Both of these problems have come to form the basis of much research by psychologists, psychophysicists, and researchers in computer vision. It is this latter work, concerning the computation of optic flow and the formation of environmental inferences from optic flow, upon which we will focus.

There is some ambiguity in the definition of optic flow in the literature (even with respect to the phrase itself, since optical flow or even optic flows are used). Some refer to the flow field as being entirely independent of images, and instead view it as a representation of the changes in environmental directions over time. To others it is a basic description of image motion determined from image intensity changes and not necessarily related to environmental motions. Both of these perspectives have validity and the sense to which we are referring should be clear from the context of whether we are dealing with computing optic flow or forming environmental inferences from a flow field. A further source of ambiguity is that some people refer to the optic flow field as a continuous vector field in which the

vectors are instantaneous velocity vectors, while others refer to it as a field of discrete displacement vectors. Throughout this thesis, we refer to it as a set of discrete displacement vectors.

Computing Optic Flow

Computing optic flow involves the determination of the displacements of image points over a sequence of images. There are several problems in this computation involving the effects of image resolution, the types of dramatic changes in image structure that can occur during motion (such as occlusion), and the now well-known stimulus matching or correspondence problem. To begin with, the notion of an environmental point corresponding to a distinguishable image point is an abstraction which is difficult to realize computationally. An image point is actually a small image area which can correspond to an appreciable surface area in the environment. One aspect of this observation is that actual flow fields do not have an arbitrarily high level of precision. The flow vector at a point may actually summarise the composite activities of an area in the environment. Another implication is the emergence or disappearance of detail as environmental surfaces are approached or receded from. In such situations, features which are meaningful and trackable at one environmental distance may no longer be meaningful at another distance. This provides motivation for the hierarchical procedures for flow field computation that we discuss below. It also reflects an important assumption applied throughout motion processing: during motion the image structures will change sufficiently slowly to allow the changes to be determined, but not so dramatically that correspondence becomes unrecognisable at successive instants. Often this is not a valid assumption and reflects another basic problem with computing optic flow. Highly significant information can be obtained from particular situations at which the optic flow field

becomes non-existent or singular, and thus difficult to compute. These situations are related to image events such things as occlusion, the motion of specularities, and the presense of smooth extremal boundaries. Another source of confusing changes are the wide range of general noise effects in image formation.

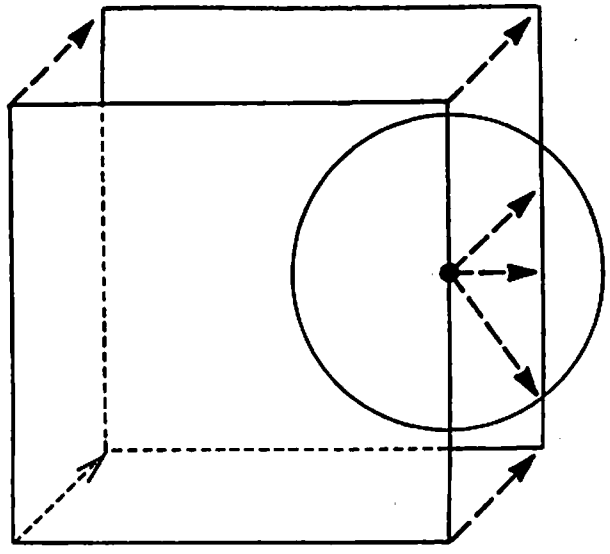


Figure 2. The Stimulus Matching Problem

The stimulus matching or correspondence [Burt76, Huan81, Thom81, Ullm81] problem refers to the ambiguity in determining image displacements, and is particularly problematic with nondistinctive portions of image structures or homogeneous image areas. The difficulties are simply exemplified by the situation illustrated in figure 2 which shows a square undergoing a diagonal displacement. The information obtainable at a portion of one of the edges only constrains the locally observed

edge motion to a wide range of potential displacements. The general form of the stimulus matching problem involves the manner in which local determination of displacements can result in a globally coherent interpretation of the changes in an image sequence.

Techniques developed to date for computing optic flow can be grouped into matching techniques and differential techniques. Both of these techniques have to deal with the problems just described and are distinguished by the different assumptions under which they operate. Both can be expressed hierarchically (though it is more typical for matching procedures). This allows the procedures to be expressed uniformly across different image resolutions, and a flow field to be determined by utilizing required consistencies between image displacements in images at different resolutions.

Matching Techniques

Matching techniques can be roughly distinguished by the types of image structures upon which they operate and the criteria by which matches of image structures in successive images are determined. Image structures can be ordered by the extent and the locality of processing required in their extraction and the complexity of the structural relations in their description. In general, the more abstract the image structure, the more stable it becomes over a sequence of images because the ambiguity in determining matches is reduced. For example, if a complete semantic analysis of each image has been performed in a sequence taken from a sensor moving relative to a house, it is easier to match at the level of extracted houses in the successive images than a less abstract and more local feature level, such as a vertical edge. There are fewer things to match and they cover an area of the image

significantly larger than their potential displacements.

Examples of image structures that have been (or could be) used in motion analysis, organized in terms of increasing abstraction are distinctive raw image subareas [Agga81b, Barn80, Dres81, Hann74, Levi73, Mora81, Quam71], parameterized tokens describing local image subareas [Hara82, Hara83, Lee82, Prag79], edges [Agga81a, Burr77, Mart79], regions [Medi83, Nage77, Nage78, Radi81, Roac79], structural descriptions of edges and regions [Brad83, Jaco80], instantiated environmental surfaces [Will80], and various high level semantic interpretations [Badl75, Tsot80].

Procedures for determining optic flow have generally been restricted to matching features whose extraction involves very little processing and are based on local image structures and computations. This is a consequence of optic flow being viewed as a very primitive description of image motion from which much information that is useful for higher level processes will be derived. From this perspective, flow processing should not be dependent on the processes to which its results will contribute. Also, when more abstract descriptions are used, although the determinations of matches becomes more viable, the determination of specific image displacement becomes less exact. This reflects a general problem that has been largely ignored by researchers in motion (with some important exceptions, notably Tsotos [Tsot80]): the mechanisms by which matches at different semantic levels of image descriptors can be combined into a coherent interpretation of an image sequence. Here, the matches between lower level image structures could be constrained by the matches determined at higher levels of surface or semantic description. The same question is involved in prediction of feature displacements from a model in which the model may consist of relatively distinct, multilevel information, and is used to constrain the interpretation and displacements of low level, local processes and features.

In general, most matching procedures that have been developed do not explicitly deal with the dramatic-change and resolution problems. Due to the assumption that most image structures will change slowly over time, if dramatic changes do occur, they will be reflected by a break-down in the matching processes. The basic approach to the stimulus matching problem has been to characterize global properties of the displacement field in a manner which directs the evaluation of image displacements. This is done in different ways. Matching structures at a more abstract or symbolic level typically involves matching strings or graph-like structures. There are solutions to this type of problem using dynamic programming or heuristic search techniques to minimize some global distortion measure reflecting the extent of graph similarity [Barr72, Chen82, Hara78, Shap82]. In another form of match processing typically applied to less abstract features, a global property such as smoothness or continuity of the displacement field is used to form a local constraint on the flow field computation. This constraint leads to a local, iterative, relaxation type procedure in which a given feature displacement must be consistent, under the criteria of smoothness, with the displacements of its spatially neighboring features [Barn80, Prag79]. Updating rules take the form of setting a feature's estimate of its correct displacement to the average of its neighbors.

Generalized Hough transform approaches to matching [Agga81b, Ball81, O'Rou81, Davi83] somewhat reverse the relation between local computations and global field properties when compared to the relaxation-based matching approaches just described. In the generalized Hough approaches, the properties of a displacement field are parameterized and represented in an n-dimensional histogram to which the local image measurements contribute. For example, the global structure of the flow field can be restricted to being a particular type of transformation, such as an affine transformation in the plane. Each local process for determining an image displacement evaluates the consistency of its potential displacements with

the values of the parameters describing each affine transformation (up to some level of parametric resolution). Globally, the parameter value most consistent with all of the potential image displacements will have the most favorable evaluation (or response in the histogram). Once a global interpretation has been determined, it can then be refined with increased resolution in the parameter space about the coarse solution.

Differential Techniques

Differential techniques are based on direct measurements of intensity changes perpendicular to an image gradient in order to determine one component of the optic flow at a point. These measurements are expressed as a function of the temporal changes in image intensity and the image gradient at a point. The other component is then determined by using an additional constraint derived from assumptions concerning the global structure of the flow field. These generally involve smoothness of the flow field or the type of transformations that can describe the displacement field. In a manner similar to the matching techniques, these constraints can be developed computationally as local, iterative processes in which global consistency is achieved via propagation similar to solutions of diffusion equations [Horn80, Glaz81, Glaz83a, Terz83]. In a few applications [Fenn79, Thom81], the local measurements can also be integrated by their independent contributions to a global histogram which expresses the parameter values of particular types of image transformations. Differential techniques can also be used to roughly constrain the motion of boundaries [Marr79] without trying to derive the optic flow. These constraints can be used to get rough qualitative motion information along closed contours, such as expansion, image motion in a rough direction, or the occurrence of rotation.

The key attributes of differential techniques is that they are based on very local, simple computations that may be performed at a low level of processing. They are also based on some unrealistic assumptions that show up when these techniques are uniformly applied to actual image sequences. These assumptions concern smoothness and often linearity in the image intensity gradients, limited extents of motion, and the constancy of image brightness over time. The smoothness assumption breaks down at surface occlusion boundaries, or wherever dramatic image changes occur such as at reflectance boundaries. Differential techniques also tend to produce dense fields, whose value is not clear, especially since the interpolation is performed in a manner that may adversely affect the inference of motion parameters. Researchers are focusing on some of these problems: Schunk [Schu83] has tried to characterize the effects of occlusion so that the computation of image displacements are selectively shut off in such areas. Nagel [Nage83], Hildreth [Hild82], and Kearney [Kear82] are working with more complex image gradients and integrating the components of information to the degree they provide unambiguous displacement information at boundaries.

Hierarchical Processing

A basic paradigm in computer vision is the use of hierarchical representations and processes [Burt82, Hans80, Rose83, Tani80, Uhr78]. This allows different resolutions and scales of image events to be handled uniformly. Additionally, the consistent agreement among hierarchically organized processes is a basic control strategy for a wide range of high and low level interpretation tasks. Hierarchical processing can produce significant computational reductions, wherein results from processing performed rapidly at lower resolutions of image information are used to

direct and constrain more detailed and extensive processing of higher resolution image information. Given the increase in computational requirements over static image processing, hierarchical mechanisms are extremely important in real-time motion processing.

The use of hierarchical processing in motion typically involves representing an image at different filtered spatial frequencies and using the processing at lower spatial frequencies to constrain the processing at higher spatial frequencies [Burt82, Glaz83b, Grim81, Luca81, Wong78]. The matches determined for the larger spatial structures in an image are used to initialize the computation for the displacements of the smaller structures. In hierarchically organised processing, the resolution problem is handled implicitly by representing an image sequence at multiple resolutions simultaneously. The stimulus matching problem is dealt with by taking advantage of the fact that matches have a tendency to be less ambiguous at lower spatial frequencies because there are fewer gross image structures and they are large relative to their potential displacements. However, the problems of dramatic change associated with flow field computation affects hierarchical processing because image structures may appear and disappear at different levels of resolution and errors produced at a lower image resolutions can propagate to the higher resolution images. Some filtering schemes [Burt83, Glaz83b] have been proposed to deal with this inherent problem by detecting the occurrence of a failure in the matching procedure and shutting off the initialisation of image displacements in the higher resolution images.

Inference of Environmental Information

Work in the inference of environmental information from flow fields has gen-

erally been restricted to the case of rigid body motion or linked systems of rigid bodies [Webb81]. There is very little general understanding in the interpretation of non-rigid environmental motions. Often, such work is task dependent as in the interpretation of image sequences of moving cloud formations and beating hearts [Tsot80].

The problem of inferring environmental information from a flow field produced by rigid body motion is often termed the shape-from-motion problem (i.e., how to determine the shape of objects or environmental depth from a flow field or a sequence of flow fields); or, somewhat confusingly, the motion-from-motion problem (i.e., how to determine the parameters of object or sensor motion from a flow field or sequence of flow fields). Theoretically, these problems are equivalent, though there are practical difficulties in inferring one from the other.

There have been significant milestones in formulating solutions to these problems in motion processing research. One set of results has dealt with the minimal conditions that are necessary for determining object shape and sensor motion in terms of the number of flow vectors across an image sequence [Fang83b, Lawt80, Meir80, Roac80, Ullm79, Webb81, Yen83]. In this work, researchers derive various sets of simultaneous nonlinear equations whose solution would constitute the appropriate inference. Since these equations cannot be solved directly, various optimization procedures are required. In another set of formulations developed primarily by Nagel [Nage81] and Prazdny [Praz81], the inference of sensor motion parameters is expressed as a search through the rotational subspace of the total set of rigid body motion parameters. Prazdny's development is rather geometrical and Nagel's is more algebraic, but they are basically similar. In 1981, Tsai and Huang [Tsai82], simultaneously with Longuet-Higgins [Long81], developed a closed form solution which could be solved by direct means.

Given these developments over the past several years, it is somewhat alarming that none of the techniques have been successfully applied to flow fields computed from anything like realistic image sequences. In fact, only in the recent work of Huang and Fang [Fang83a, Fang83b] and Jerian and Jain [Jeri83] has there even been an explicit evaluation of a procedure on such images. This work has shown the particular difficulties familiar to motion researchers: extreme sensitivity to noise and resolution, dependence upon the type and extent of motion, and general instability.

A possible exception to these difficulties may be a procedure recently developed by Rieger and Lawton [Rieg83, Lawt83]. The technique is restricted to recovering the parameters of sensor motion relative to a stationary environment and is based upon the fact that the decomposition of a flow field into its rotational and translational components can be directly obtained at image positions where a significant depth variation occurs in the environment [Long80], such as at some occlusion boundaries. This results in a very simple analysis which does not involve solving unstable equations. The basic practical difficulty associated with this technique is that it is dependent on the analysis of a flow field at occlusion boundaries where the flow field tends to be most errorful. Dealing with this effect requires a computation which may reduce the precision of the inference of the sensor motion parameters.

There are many reasons, not all of which are fully understood, why the inference of motion parameters and environmental depth has been difficult. Some of the formulations involve image measurements, such as higher order derivatives of an instantaneous vector velocity field which are difficult to obtain and are also quite noise sensitive when applied to discrete image sequences [Praz80, Long80]. There are also many cases of motion which are inherently ambiguous. One of these is discussed in chapter VI of this thesis and concerns a rather typical case of terrestrial motion in which the rotational and translational field components are nearly impos-

sible to separate. In recent work concerning the interpretation of images containing multiple independently moving objects, Adiv [Adiv84] appears to have found cases in which independently moving objects with different parameters of motion, can, when considered together, result in a globally consistent, but incorrect, interpretation. Another problem affecting shape from motion formulations is the baseline effect which is common to stereo. The baseline effect expresses that the resolution and accuracy of depth inferences are a decreasing function of the distance between the sensor locations at which images are formed. For motion, where the sensor displacements are generally small between successive instants, the environmental inference would tend to be poor, but could be compensated by the availability of more and more images over time.

There has been almost no stability analysis of the systems of equations for inference from optic flow. Along these lines, recent work by colleagues and myself [Stee83] has given empirical indications of the instabilities in the inference procedures under certain conditions. We have been exploring the use of a highly parallel array architecture for inferring motion parameters from flow fields. This processing amounts to sampling and evaluating 200,000 points in the five dimensional space of determinable rigid body motion parameters at near video rates. This roughly shows the appearance of the error surface these system of equations may describe. What this work indicates is that the space is very bumpy and jagged, full of local optima, that would make solutions difficult, especially in the presence of noise.

There have been several responses to these difficulties. One approach has been to utilize optimization procedures which are based on global evaluation of the expressions for the inference of motion parameters from flow fields instead of local, iterative optimization procedures. Examples of these approaches are the work with generalized Hough transforms [Adiv84, Ball81] and the procedure involving highly

parallel architectures mentioned above [Stee83]. Some researchers are beginning to perform an explicit analysis of the stability of the different solutions [Shaw83], while others are trying to develop qualitative inference techniques which are hoped to be more robust [Thom83], and still others are beginning to investigate the inference of motion and shape from image transformations other than optic flow, such as the analysis of contour shape changes [Davi82]. Currently, much of this work is preliminary.

Another response to these inadequacies has been to deal with restricted cases of motion. Here too, the work has been limited in application to realistic image sequences with principle results having been achieved by Williams [Will80] and Dreschler and Nagel [Dres81]. These restricted cases of motion can be of significant practical use, since in many cases some of the parameters of motion can be determined by other sensing devices. Additionally, general motion can be locally interpreted, temporally and spatially, as consisting of certain restricted types of motion.

In the research presented in this thesis, we will develop procedures for various cases of restricted motion, and show how to use the procedures for translational motion to locally interpret more general motion. In this regard, it is useful to summarize related work in vanishing point extraction and translational motion processing. The determination of the vanishing point in a static image is closely related to determining the direction of translation. In perspective projection, parallel lines in the environment map onto lines radiating from the vanishing point in the image. For translational motion, the environmental motion paths correspond to the parallel lines in the perspective case. Techniques for extraction of a vanishing point have been explored by Kender [Kend79], Nakatani [Naka80], and in a more general framework by Ballard [Ball81]. The use of the Hough transform in this work is sim-

ilar to the global sampling of the error measure developed in chapter IV. It would be interesting if the determination of edges could be combined with the determination of the vanishing point, in a manner similar to the concurrent determination of image displacements and the translational axis in the work presented in chapter IV.

Williams [Will80] was the first to develop algorithms for interpreting natural complex images produced by an optic sensor translating relative to environmental objects. This work consisted of two processes: one for inferring the direction of translation given environmental depth information and the other for inferring depth given the direction of motion. These processes used an error measure describing the consistency of depth information and the inferences of feature motion along image displacement paths. His work indicated that these two processes, for inferring depth and the direction of motion, could be combined.

The primary weakness of Williams' work was the necessary restriction to planar surfaces at one demonstrated orientation. Additionally, in the case of unknown environmental depth and translation, the processing is quite complex — involving segmentation, resegmentation, and coordinating the processes for inferring depth and for inferring the direction of translation. The method we develop in chapter IV requires no restrictions on the orientation of surfaces or shape of environmental objects, and involves only a simple procedure for evaluating an error measure. It also indicates that the direction of sensor motion should be determined prior to, or concurrently with, environmental depth.

CHAPTER III

DISPLACEMENT FIELD STRUCTURE

Introduction

In this chapter we review the relations between sensor motion relative to rigid body objects and the structure of the corresponding field of image displacements. Basic results from kinematics [Whit44] and geometry [Coxe61] allow arbitrary rigid body motions of the camera to be decomposed into a rotation about its focal point followed by a translation. This permits image motions to be described as consisting of two components: a rotational and a translational field. The rotational field contains information concerning sensor orientation relative to the environment, while the translational component contains information concerning environmental depth and the relative displacements of the sensor and environmental objects. This decomposition forms the basis of procedures for recovering camera motion parameters from displacement fields [Nage81, Praz81].

Describing Rigid Body Motion

In this section we review some basic terminology for describing image and environmental motion, the particular coordinate systems employed, and how rigid body motions are described in terms of sensor motion.

Terminology

It is necessary to have terms for describing the motion of features in an image sequence and the corresponding motion of environmental points. We define an Image Displacement Vector to be a two-dimensional vector describing the displacement of an image feature from one image to the next. An Image Displacement Field is the set of image feature displacement vectors for successive images. An Image Displacement Sequence indicates the positions of an image feature over several successive images. Though we are dealing with discrete image sequences, it is often possible to describe the continuous curve along which an image feature point is moving. This curve is called the Image Displacement Path.

Corresponding to image motions we use a set of terms for describing environmental motions. An Environmental Displacement Field is the set of three-dimensional vectors indicating the positions of environmental points at successive instants. An Environmental Displacement Sequence indicates the position of an environmental point over several successive instants. An Environmental Displacement Path describes the three-dimensional curve that an environmental point is moving along for a particular motion.

Coordinate Systems

We utilize two coordinate systems in this exposition: a fixed system based on the environment and another based on the sensor. The fixed environmental coordinate system is a Cartesian coordinate system. The sensor coordinate system (or camera model) is referred to throughout this thesis and consists of a planar retina embedded

in a three-dimensional Cartesian coordinate system (X, Y, Z), with the origin at the focal point and the optical axis aligned with the positive $Z-$axis (figure 3). The X and Y axes correspond to the gravitationally intuitive horizontal and vertical directions, respectively. The image plane is parallel to the XY plane and located at a distance of one focal length along the Z axis.

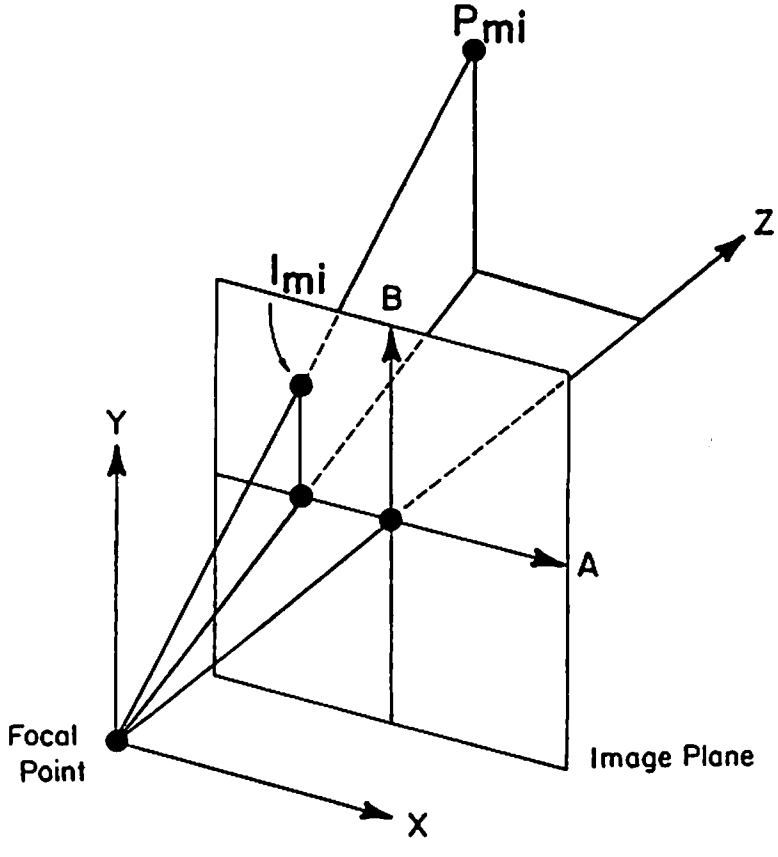


Figure 3. Camera Model.

Positions in the image plane are described using a 2-D coordinate system with the axes A and B aligned with the X and Y axes of the camera coordinate system, respectively. The origin of the image plane coordinate system is determined

by the intersection of the image plane and the $Z-$axis. The vector P_{mi} refers to the position of an environmental point in the sensor coordinate system and the vector I_{mi} refers to the position of the intersection of the ray of projection for P_{mi} with the image plane. The first index of these vectors is used to specify a particular image from a sequence of images. The second index specifies a particular environmental point. Setting the focal length to one, the relations between P_{mi}, z_{mi}, and positions in the image plane determined by perspective projection are:

$$P_{mi} = (x_{mi}, y_{mi}, z_{mi})$$

$$I_{mi} = (a_{mi}, b_{mi}, 1) \tag{1}$$

$$I_{mi} = (\frac{x_{mi}}{z_{mi}}, \frac{y_{mi}}{z_{mi}}, 1)$$

$$P_{mi} = z_{mi} I_{mi}$$

The position and the orientation of the sensor relative to the environmental coordinate system at time t is described by the vector $P(t)$ and the matrix $O(t)$, where $P(t)$ is the position of the origin of the sensor coordinate system at time t, and $O(t)$ describes the orientation of the sensor coordinate system by its direction cosines. The matrix $O(t)$ is obtained by translating the sensor coordinate system to the origin of the environmental coordinate system and determining the angles between the axes of the two coordinate systems. Denoting the coordinate axes of the camera coordinate system as (X_c, Y_c, Z_c) and those of the environmental coordinate system as (X, Y, Z) yields:

$$O(t) = \begin{pmatrix} \cos(X, X_c) & \cos(X, Y_c) & \cos(X, Z_c) \\ \cos(Y, X_c) & \cos(Y, Y_c) & \cos(Y, Z_c) \\ \cos(Z, X_c) & \cos(Z, Y_c) & \cos(Z, Z_c) \end{pmatrix} \quad (2)$$

Decomposing Rigid Body Motion

There are some basic results in kinematics which allow arbitrary rigid body motions to be expressed as consisting of a rotation about an axis positioned at an arbitrary point followed by a translation. These are stated as

> A rotation about any axis is equivalent to a rotation through the same angle about any axis parallel to it, together with a simple translation in a direction perpendicular to the axis. The converse is also true, the rotation of a rigid body about any axis, preceded or followed by a translation in a direction perpendicular to the axis, are together equivalent to a rotation of the body about a parallel axis [Whit44].

Thus, the orientation of a body will change the same for parallel axes of rotation with the same extent of rotation, regardless of where they are positioned. This implies that the axis of rotation can be positioned anywhere so long as it is followed by the appropriate translation. Thus, we can canonically describe sensor motion as an initial rotation about an axis positioned at the origin of the sensor coordinate system (bringing the sensor into the same orientation at successive instants) followed by a translation (bringing the sensor in coincidence at the successive instants). This will also decompose an image displacement field into a field produced solely by the

rotation of the sensor and a field produced solely by the translation of the sensor. Each of these fields contains different information.

More specifically, given the sensor at successive positions and orientations $(P(t), O(t))$ and $(P(t+1), O(t+1))$, its motion is described as an initial rotation about the origin of the sensor coordinate system described by the matrix R such that $O(t+1) = O(t) * R$, followed by a translation T with respect to the environmental coordinate system such that $P(t+1) = P(t) \times T$. Thus,

$$O(t)^{-1} \times O(t+1) = R \tag{3}$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ P_x(t) & P_y(t) & P_z(t) & 1 \end{pmatrix} = T$$

Properties of Pure Rotational Displacement Fields

Let us consider rotational fields that are produced by rotation about an axis containing the origin of the sensor coordinate system. The basic property of such fields is that the image displacements are totally a function of image position and can yield no information concerning environmental depth. That is, given the position of an image point at time t and the sensor rotation R, its position at time $t+1$

is determined.

To describe the general structure of rotational flow fields, consider the image displacement path generated by a particular image point under sensor rotation. In figure 4a we see an axis of rotation positioned at the origin of the coordinate system and a ray of projection determined by some image point I_{mi}. The effect of the rotation will be that the ray of projection will generate the surface of a cone. The image displacement path for the rotation of this image point will then be determined by the intersection of this cone with the image surface, i.e. a conic section.

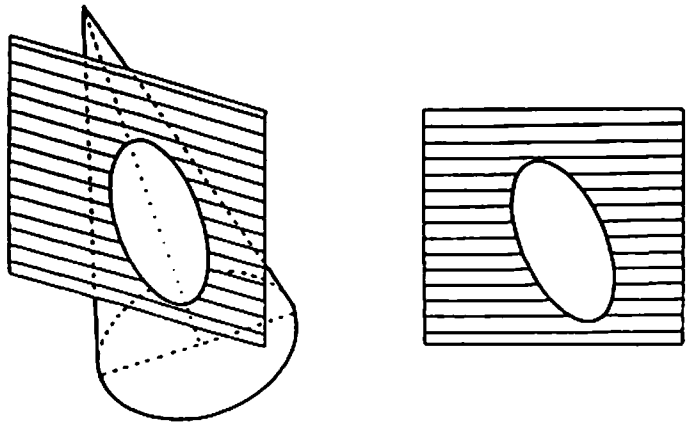


Figure 4a. Rotational Displacement Paths. The figure on the left shows the intersection of an image plane with the cone determined by the axis of rotation positioned at the focal point and a given image position vector. The figure on the right shows the resulting conic image displacement path.

One should note that for points along the same ray of projection, the image displacements under a given rotation will all be the same. Thus, there is no basis upon which to infer environmental depth under rotational motion because the angles between rays of projection remains fixed.

Now let us consider sensor rotation analytically with the axis of rotation represented as a unit vector $R = (R_x, R_y, R_z)$. For any environmental point $P = (x, y, z)$, we can describe the cone generated by the rotation to be:

$$c = \cos(\theta) = \frac{P \cdot R}{|P|} \tag{4}$$

where θ is the angle between R and P. To determine the image displacement paths, we expand this equation with z set to 1 (corresponding to the location of the image plane):

$$c = \frac{xR_x + yR_y + R_z}{\sqrt{x^2 + y^2 + 1}} \tag{5}$$

By squaring both sides and reorganising terms, this equation may be expressed as an implicit function in the general form of a conic:

$$\begin{aligned} F(x,y) = {} & x^2(R_x^2 - c^2) + y^2(R_y^2 - c^2) + 2x(R_xR_z) \\ & +2y(R_yR_z) + 2xy(R_xR_y) + (R_z^2 - c^2) = 0 \end{aligned} \tag{6}$$

The partial derivatives of this equation yield the tangents to the image displacement path:

$$\frac{\partial F(x,y)}{\partial x} = 2x(R_x^2 - c^2) + 2(R_x R_z) + 2y(R_x R_y) \tag{7}$$

$$\frac{\partial F(x,y)}{\partial y} = 2y(R_y^2 - c^2) + 2y(R_y R_z) + 2x(R_x R_y)$$

Note that for the rotational axis aligned with the Z axis, $R = (0,0,1)$ substitution into equation 6 yields

$$x^2 + y^2 = \frac{1}{c^2} - 1 \tag{8}$$

This describes a family of circles in the image plane centered at $(0,0,1)$ and indexed by the particular values of c in the range 0 to 1 (figure 4b). For the rotational axis $R = (0,1,0)$ substitution into equation 6 yields

$$y^2 \frac{1-c^2}{c^2} - x^2 = 1 \tag{9}$$

This describes a family of hyperbolas indexed by values of c in the range 0 to 1 (figure 4c).

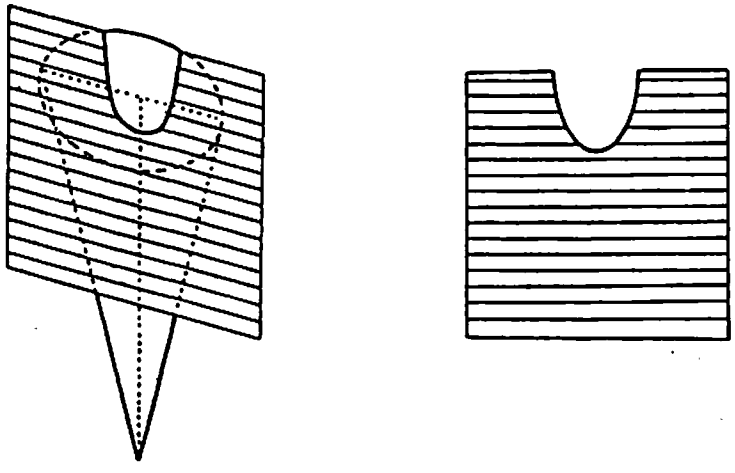


Figure 4b. Displacement Paths for Rotations about the (0,0,1) axis.

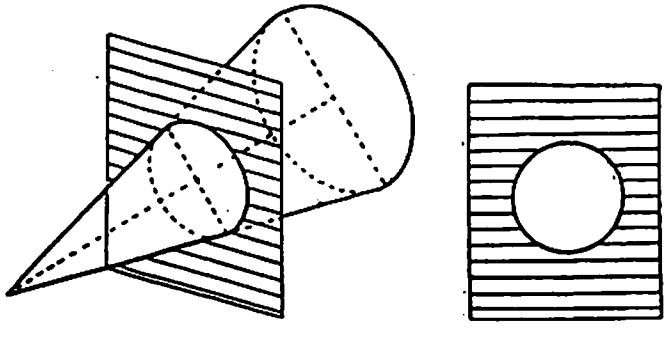


Figure 4c. Displacement Paths for Rotations about the (0,1,0) axis.

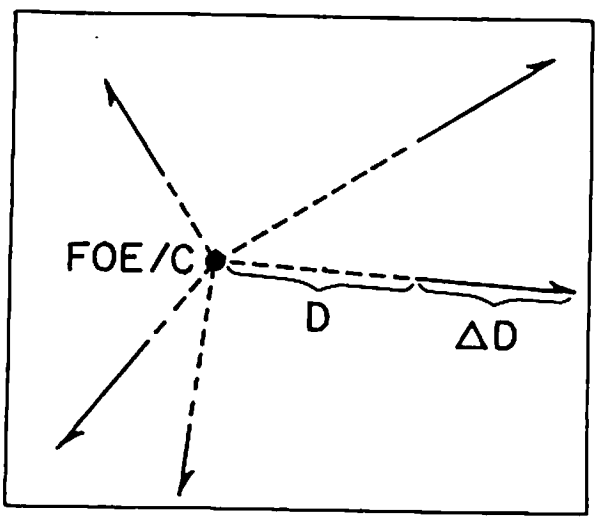
Translational Field Properties

For purely translational motion the sensor orientation is fixed relative to the environmental coordinate system and the motion is described by an axis of translation. The image displacement paths are determined by the intersection of the translational axis with the image plane. If the translational axis intersects the image plane on the positive half of the axis, the point of intersection is called a Focus of Expansion (FOE) and the image motion is along straight lines radiating from it. This corresponds to sensor motion towards visible environmental points. If the translational axis intersects the image plane on the negative half of the axis, the point is called a Focus of Contraction (FOC) and the image displacement paths are along straight lines converging towards the FOC. This corresponds to camera motion away from visible environmental points. The intersections of axes parallel to the image plane are points at infinity and thus may be considered to be either an FOE or FOC in opposite directions. This ambiguity is one reason we refer to the directions of motion determined by the translational axes themselves instead of the intersections with the image plane.

Given the direction of translation and the image displacements of a set of environmental points, the relative depths of these points can be computed by solving the inverse perspective transform [Roge76]. Relative depth can also be simply inferred from the position of a feature and the extent of its displacement relative to an FOE or an FOC. This relation is expressed as

$$\frac{D}{\Delta D} = \frac{Z}{\Delta Z} \tag{10}$$

where Z is the value of the Z component of an environmental point at time $t+1$, ΔZ is the extent of environmental displacement along the Z axis from time t to time $t+1$, D is the distance of the corresponding image point from the FOE or FOC at time t, and ΔD is the displacement of the image point from time t to time $t+1$. Thus, the Z value of an environmental point can be recovered from image measurements in units of ΔZ, or what has been termed Time-Until-Contact by Lee [Lee76, Lee80] (figure 5a and 5b). To the degree that the sensor displacement can be accurately monitored, absolute depth of surface points can be computed.



$$\frac{Z}{\Delta Z} = \frac{D}{\Delta D}$$

Figure 5a. Relation between relative environmental depth and the extent of image displacement with respect to the FOE/C.

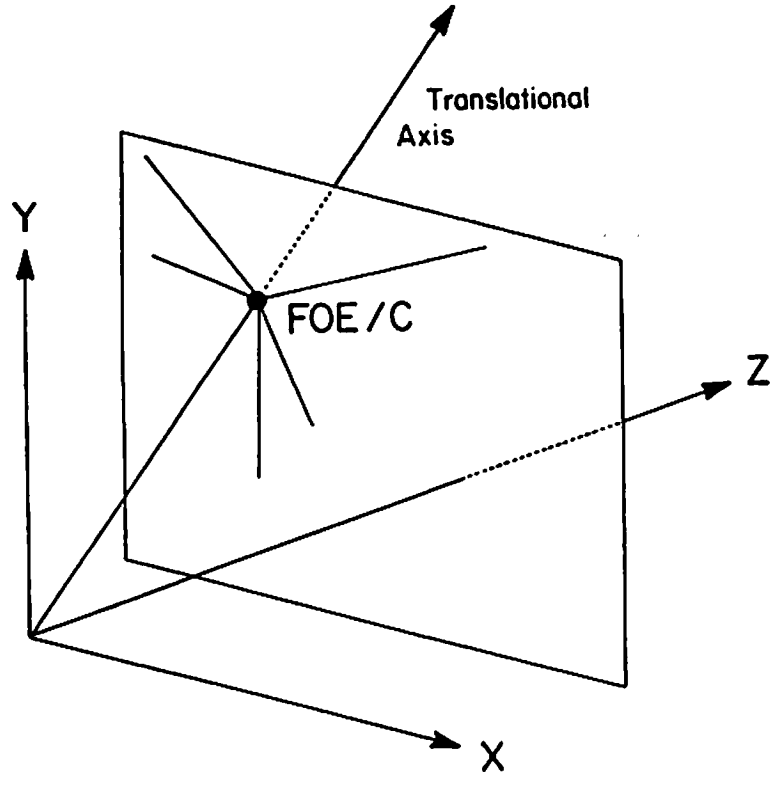


Figure 5b. The FOE/C is determined by the intersection of the image plane with the translational axis.

Composite Field Properties

The effects of composite image motions produced by sensor rotation and translation can be analyzed as follows for an image feature I_{mi} which undergoes a displacement D to position I_{ni} at time n (figure 6a). The motion can be described as an initial displacement R to a position J_{mi} due solely to the rotation of the sensor, which is followed by a displacement T from J_{mi} to I_{ni} along the translational displacement path determined by the straight line containing image points J_{mi} and the FOE determined by the translational parameters.

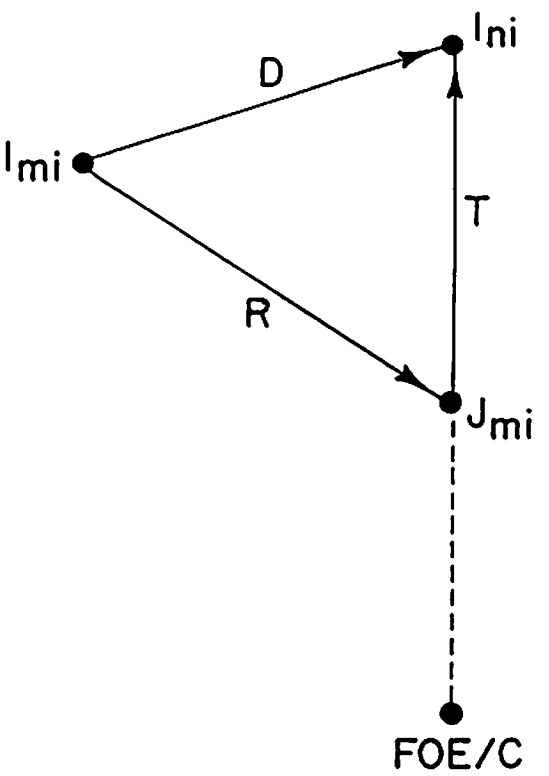


Figure 6a. Composite Field Structure.

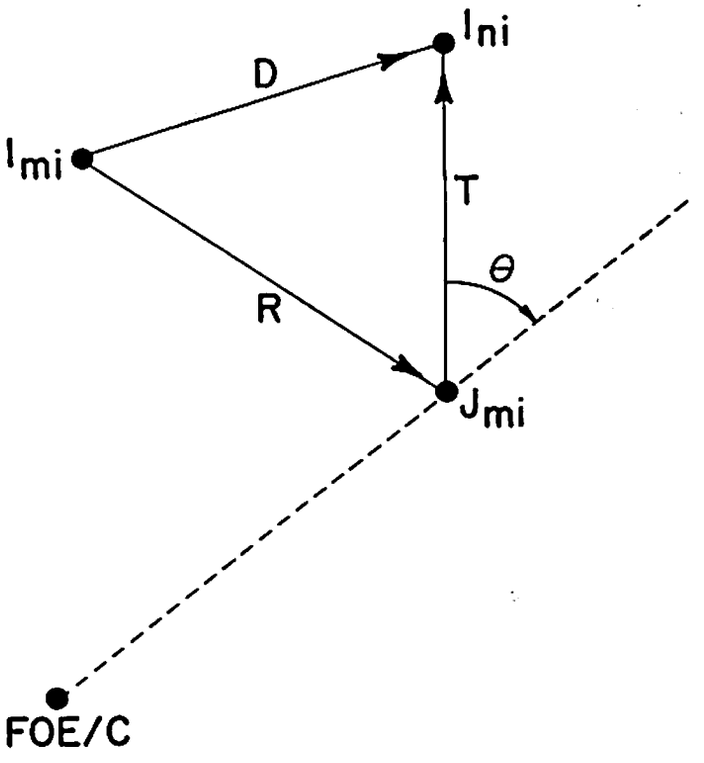


Figure 6b. Error Measure from Composite Field Structure

These structural properties will be used to develop measures describing the consistency of a given image displacement with hypothesized sensor rotation and translation parameters (figure 6b). As above, for an image point I_{mi}, the rotational parameters induce an image displacement to some position J_{mi}. This point and the FOE corresponding to a particular translational axis, determine an expected translational displacement path. The angle between this displacement path and the

vector $I_{ni} - J_{mi}$ corresponds to the discrepancy between the image displacement and the hypothesized values of the sensor motion parameters. We will utilize this measure to evaluate motion parameters with respect to determined displacement fields in chapters VI and VII. This local consistency measure has also been used in generalized Hough transforms so that each image displacement vector can scale its vote against a particular set of motion parameters corresponding to the extent of this determined angle [Stee83].

CHAPTER IV

PROCESSING TRANSLATIONAL MOTION

Introduction

In this chapter we present a procedure for processing image sequences produced by translational motion. The computation robustly combines the determination of the translational motion parameters, image displacements, and environmental depths of visible surfaces. The procedure consists of two basic steps: Feature Extraction and Search. The feature extraction process finds small image areas which may correspond to distinguishing, and therefore trackable, parts of environmental objects. The direction of translational motion is then found by a search across hypothesized FOE/C positions to determine a set of image displacement paths for the extracted features which minimizes an error measure of total feature mismatch along these displacement paths, and also yields consistent displacements for the features.

The feature extraction process finds distinctive points which are positioned at points of high curvature along contours determined by simple processes such as thresholding, zero-crossing extraction and local contrast measurements. Particular forms of the feature extraction process can lead to effective and very rapid computation on proposed image processing architectures.

The search process minimizes an error measure defined over a unit sphere, with each point on the sphere corresponding to a different direction of sensor translation. A given direction of translation constrains the motion of extracted image features

to straight lines which radiate from or converge onto a single point in the image plane. Thus, the error measure associates a point on the unit sphere, corresponding to a particular translational axis, with a number describing the degree of total feature mismatch along the displacement paths determined by the translational axis. Experiments have shown this error measure to be smooth and with a distinct minimum in a large neighborhood about the correct translational axis. This allows simple search methods to be effective.

We present several experiments showing the results of applying the procedure in various situations. The experiments indicate that it is robust and applicable to a wide range of real world image sequences. In the next chapter, we review particular extensions for implementing the procedure in a hierarchical computational framework, dealing with independently translating objects, translational blur-streaks, and implications for autonomous navigation.

Extraction of Interesting Points

The feature extraction process is used to determine small areas (referred to as image points or features) in an image that are distinct from their respective neighboring areas. This distinctiveness limits the potential matches of these image areas in suceeding images and suggests the possibility that these points may be trackable over time. These image features may also reflect a correspondence to actual and significant features in the environment, such as points of high curvature on object boundaries, texture elements, surface markings, etc. However, there are some features, termed false features, which may be selected but which result from noise, occlusion, and light source effects and have behavior which is currently difficult to interpret. Features can be represented either as arrays of numbers extracted as a subimage directly from an image, or as parameterized tokens describing local image properties. We refer to features exclusively as small arrays of data values centered at some point in an image at some time t.

Following Moravec [Mora77, Mora81], the method of feature extraction used here is based upon finding image areas which are significantly different than their neighboring areas. Using correlation measures bounded between 1 (for perfect correlation) and 0, the distinctiveness of a feature is 1 minus the best correlation value obtained when the feature is correlated with its immediately neighboring areas (excluding correlation with itself). Good features can then be selected by finding the local maxima in the values of the distinctiveness measure over an image. There are several metrics available for similarity of two $n \times n$ arrays $A_{i,j}$ and $B_{i,j}$. We have utilized the following measures:

Normalized Correlation

$$\frac{\sum_i \sum_j A_{i,j} B_{i,j}}{\sqrt{\sum_i \sum_j A_{i,j} A_{i,j}} \times \sqrt{\sum_i \sum_j B_{i,j} B_{i,j}}} \tag{11}$$

Moravec Correlation [Mora77]

$$\frac{\sum_i \sum_j A_{i,j} B_{i,j}}{(\sum_i \sum_j A_{i,j} A_{i,j} + \sum_i \sum_j B_{i,j} B_{i,j})/2.0} \tag{12}$$

Normalized Absolute Value Difference

$$1.0 - \frac{\sum_i \sum_j |A_{i,j} - B_{i,j}|}{\sum_i \sum_j A_{i,j} + \sum_i \sum_j B_{i,j}} \tag{13}$$

All of these measures have a value of 1 for a perfect match. Of these, the first choice is the most conventional, the second is a good approximation to the first and more efficient, and the third is the quickest to evaluate.

We further constrain the neighborhoods over which the features are selected to contours determined by other processes, such as zero-crossing extraction and thresholding, which are sensitive to edges. This yields interesting points which are locally distinctive and exhibit high curvature along extracted contours containing the point.

Feature Extraction Using Zero-Crossings

The use of zero-crossings to determine significant image contours at different levels of resolution has been proposed and extensively studied by Marr et. al. [Hild80, Marr80]. In this processing an image is convolved with Gaussian-Laplacian masks ($\nabla^2 G$) of different positive widths and thresholded at zero to determine zero-crossing contours. These contours are significant since they correspond to the points of greatest change in the convolved image. The distinctiveness measure can be applied to points along these contours in the convolved image, with the local maxima determining the position of potential features. This generally has the effect of finding points of high curvature along the zero-crossing contour, although points apparently corresponding to local occlusion vertices and weak maxima will also be extracted.

Many weak features which are local maxima of distinctiveness can be removed by suppressing those which are at points of low curvature along the zero-crossing contours (a cheaper method for dealing with this is presented in the discussion of this chapter). For features which are local distinctiveness maxima, we approximate the curvature along the contour by the inner product of the normalized vectors describing the relative positions of the nearest local maxima along the contour (figure 7). These values are then thresholded between 1.0 (corresponding to high curvature) and -1.0 (corresponding to low curvature) to reflect feature strength.

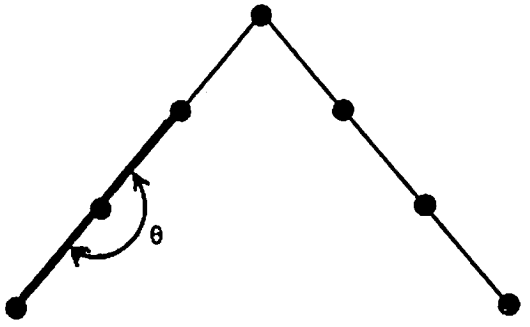


Figure 7. Computation of curvature for low curvature suppression of extracted features.

The images in figure 8a and figure 8b were taken from a gyroscopically stabilized movie camera held by a passenger in a car traveling down a country road in Massachusetts [Will80]. They are 128x128 pixel images with 6 bits of resolution in intensity and will be referred to as the roadsign images. Figure 8c shows the zero-crossings extracted from the initial roadsign image using a $\nabla^2 G$ mask with a positive width of 5 pixels. The distinctiveness values were computed using features which were 5x5 pixel arrays extracted from the convolved image and centered on pixels which were adjacent to the zero-crossing contour and of positive value. These features were correlated, using Moravec's norm, with their 8 immediately neighboring features. Figure 8d shows the local maxima in the distinctiveness measure positioned with respect to the zero-crossing contour. Figure 8e shows the results of suppressing low-curvature points using a threshold set to -0.8 radians (143 degrees).

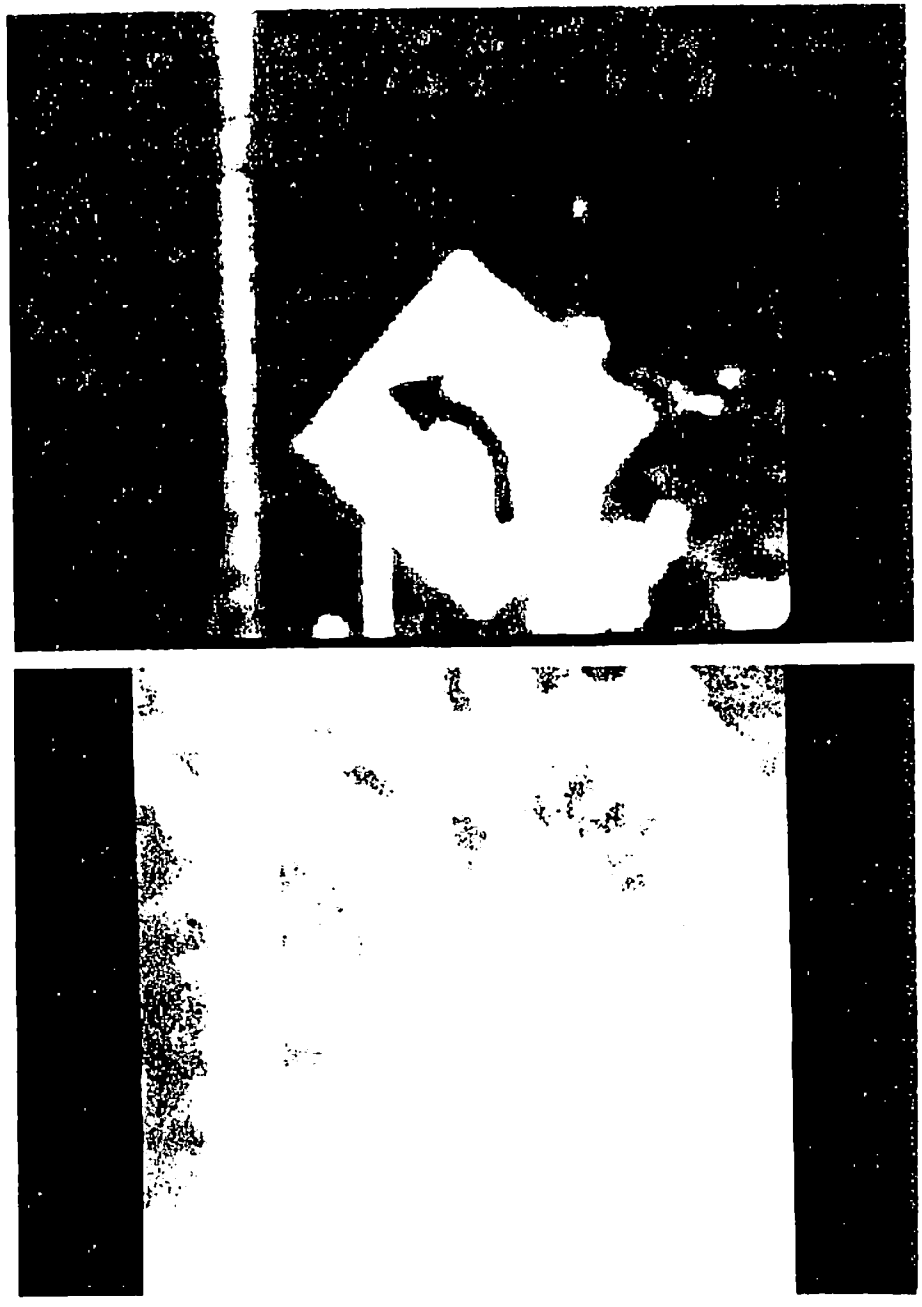


Figure 8a. Roadsign Image 1. The upper image has the intensity values normalized across the entire image. The lower image uses a restricted range of intensity values to show the dark, low contrast tree texture.

Figure 8b. Roadsign Image 2. The upper image has the intensity values normalized across the entire image. The lower image uses a restricted range of intensity values to show the dark, low contrast tree texture.

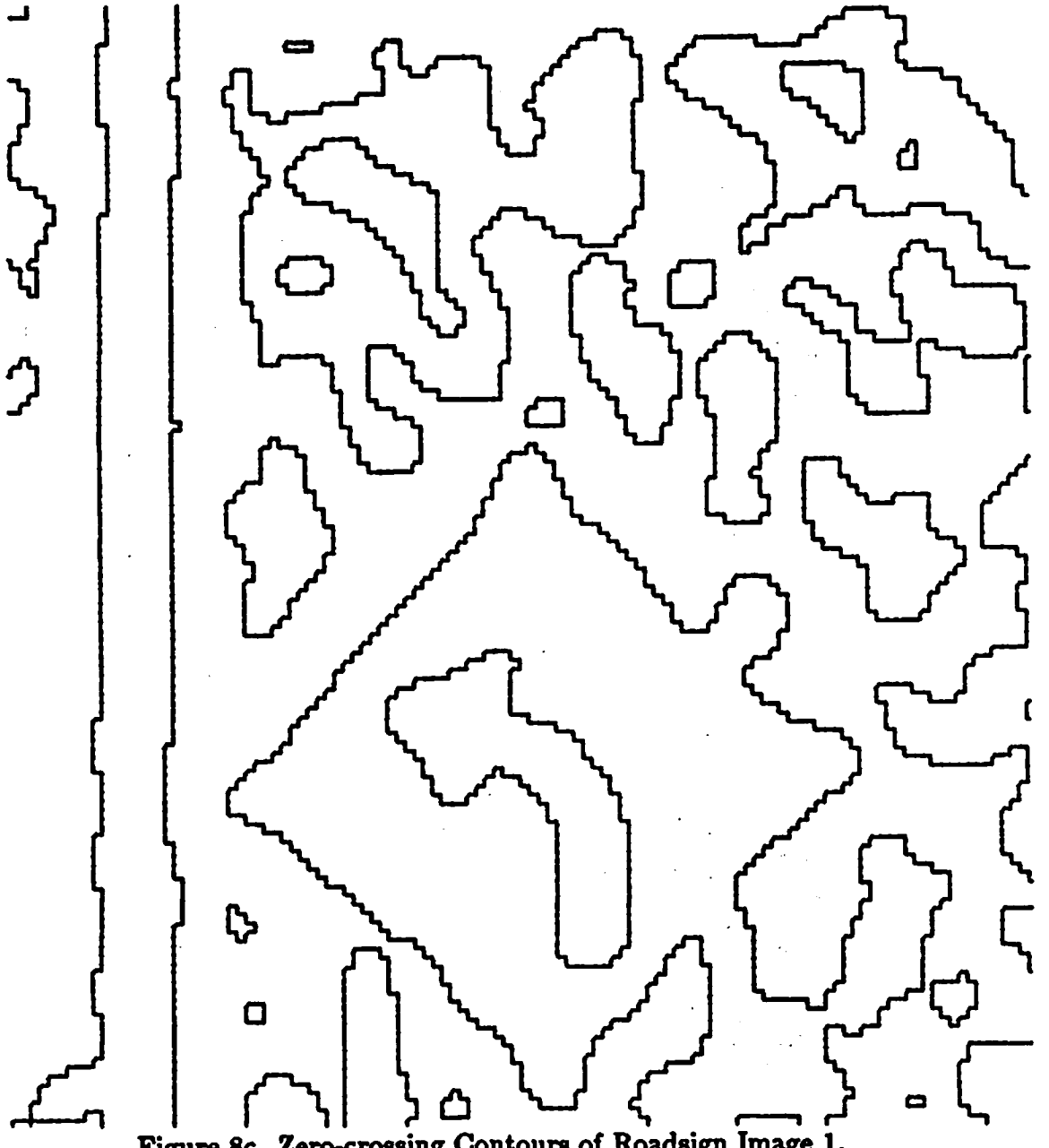


Figure 8c. Zero-crossing Contours of Roadsign Image 1.

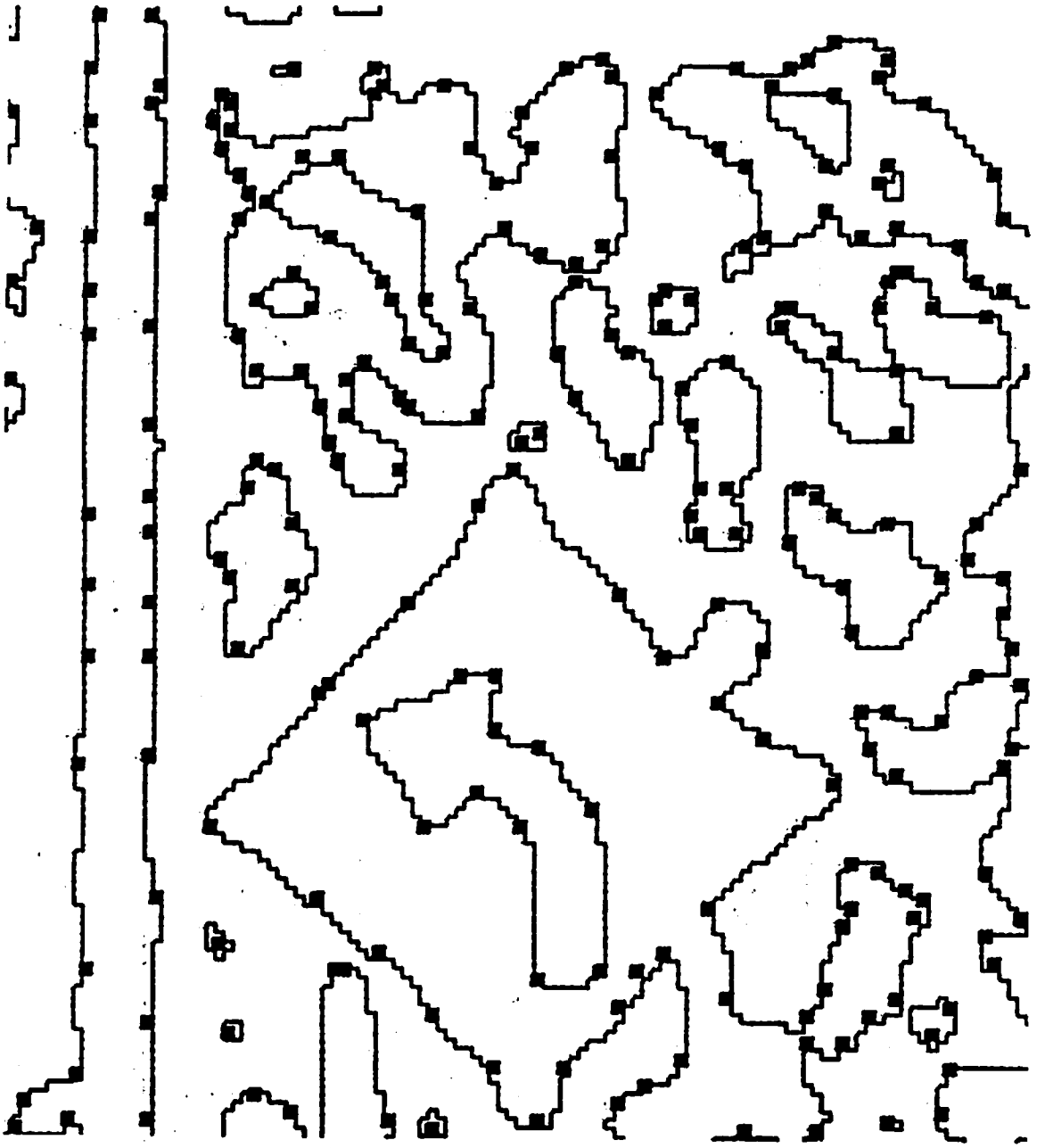


Figure 8d. Local Maxima of Interest Measure.

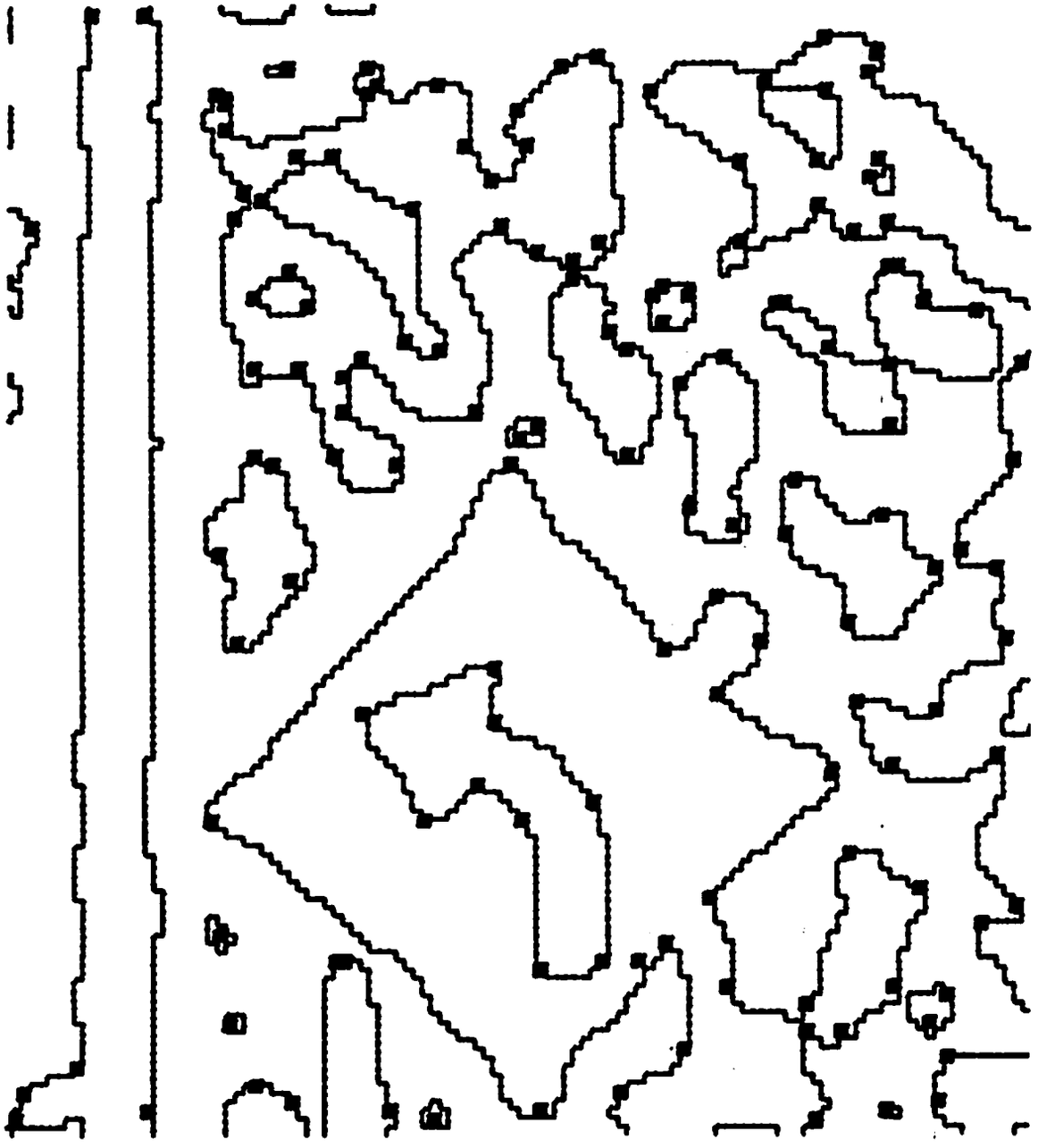


Figure 8e. High Curvature Points along Zero-Crossing Contour

Use of features based on zero-crossings requires specification of the sizes of the convolution masks that are employed, and a decision whether to position extracted feature points with respect to the unprocessed image or the convolved images. It is usually beneficial to use masks of various widths for sensitivity to features at different levels of resolution. In this case, the translational processing described below can be applied independently to the different pairs of images formed by convolving the original successive images with the different masks. Alternatively, as was done above, features can be extracted from the original, unfiltered image at the positions where features were determined in the convolved images, though experience with large masks has shown that this approach can position features significant distances from their apparent position in the original image.

Feature Extraction Using Threshold Contours

Another simple operation to determine image contours is thresholding. The values of the threshold can be determined in a variety of ways: using fixed increments, finding peaks and valleys in the image intensity histogram, or using techniques sensitive to image contrast across the contours produced by a particular threshold [Kohl81,Wesk75].

The images in figure 9a and 9b were produced from a solid state camera held by a robot manipulator translating toward some industrial parts lying on a table. The images are 128x128 pixel images with 6 bits of intensity resolution. These will be referred to as the industrial images. Analysis of the image intensity histogram, using the procedures described in [Kohl81], determined a clear break in the histogram at an intensity level of 10 in the image. This corresponded to separation of the dark background and the brighter objects in the scene. Figure 9c shows the extracted contour and figure 9d the local maxima in the distinctiveness measure of image features centered on pixels adjacent to the contour and of intensity value greater than or equal to ten. Figure 9e shows the extracted feature points after low curvature suppression using a threshold set to -0.8 radians (corresponding to an angle of 143 degrees).

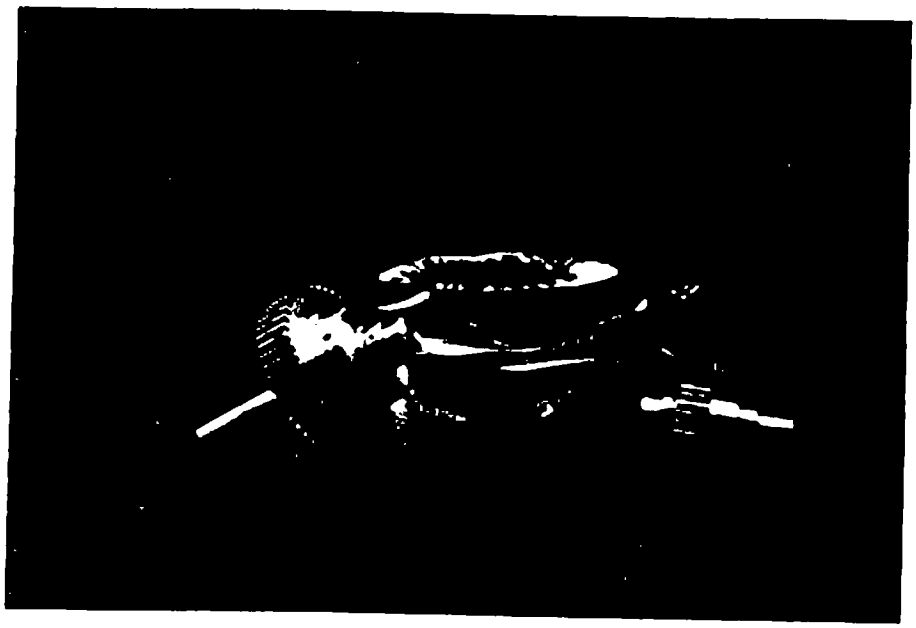


Figure 9a. Industrial Image 1.

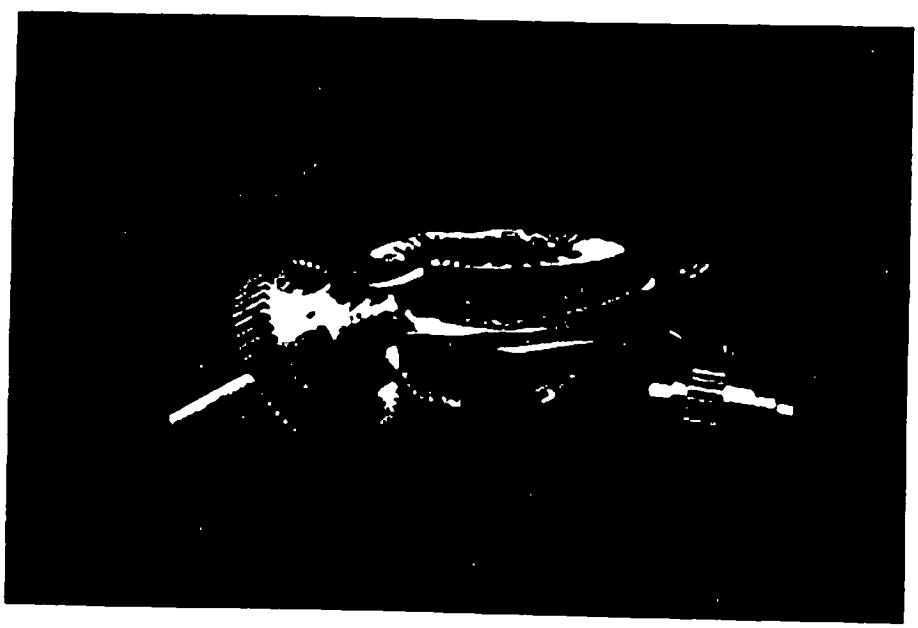


Figure 9b. Industrial Image 2.

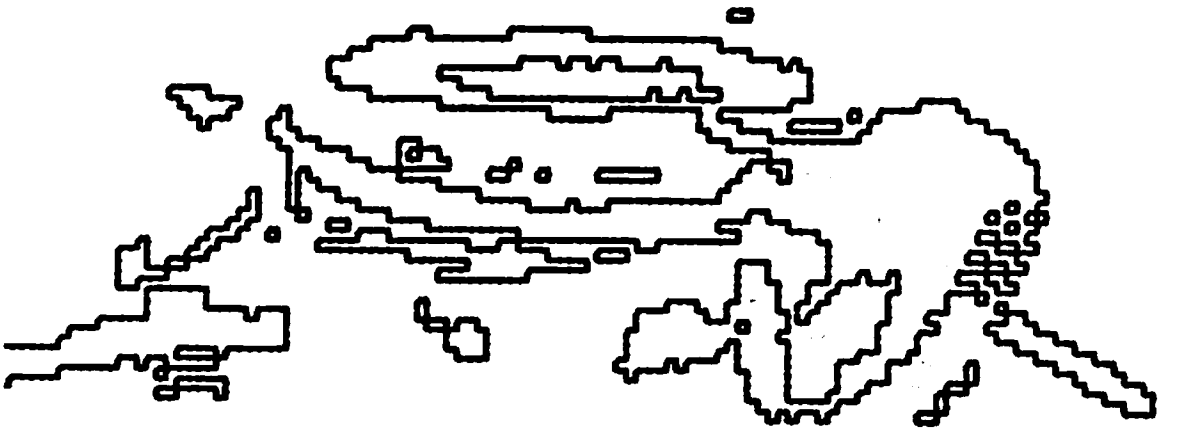


Figure 9c. Threshold Contour of Industrial Image 1.

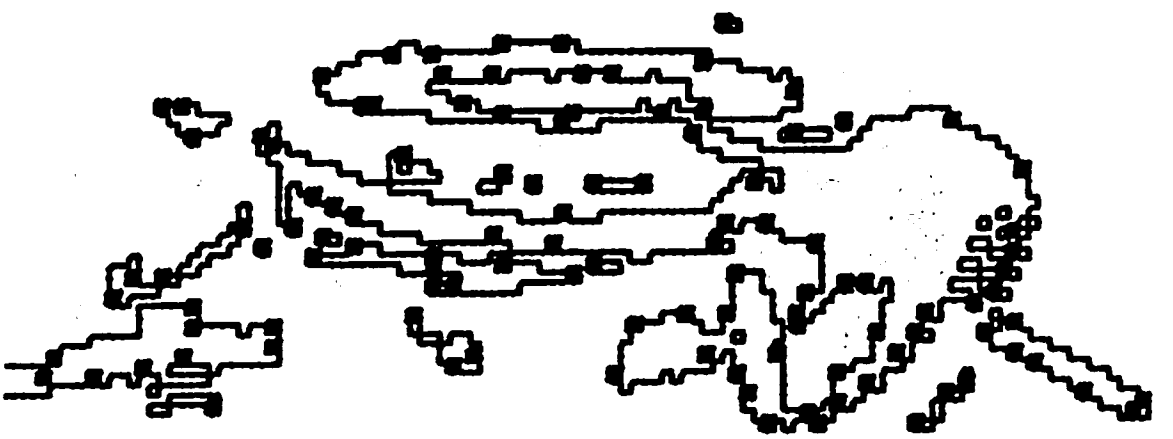


Figure 9d. Local Maxima of Interest Operator.

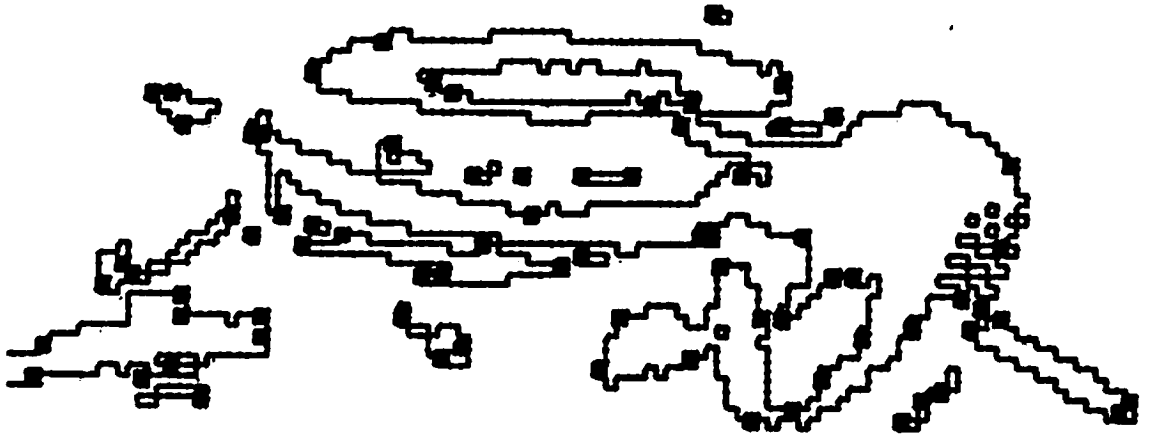


Figure 9e. High Curvature Points along Threshold Contour.

Determining the Axis of Translation

The procedure for determining the translational axis minimizes an error measure which describes the extent of feature mismatch along the image displacement paths determined by an hypothesized translational axis. Note that the image displacements are determined simultaneously with the direction of motion. For example, figure 10 shows an FOE determined by a potential translational axis and the corresponding image displacement paths for some extracted features. Also shown is the match profile for correlation of a particular feature along a segment of its displace-

ment path in the succeeding image. The adequacy of a potential translational axis for describing the motion between successive images is measured by summing the error associated with the best match for each of the features along their respective image displacement paths.

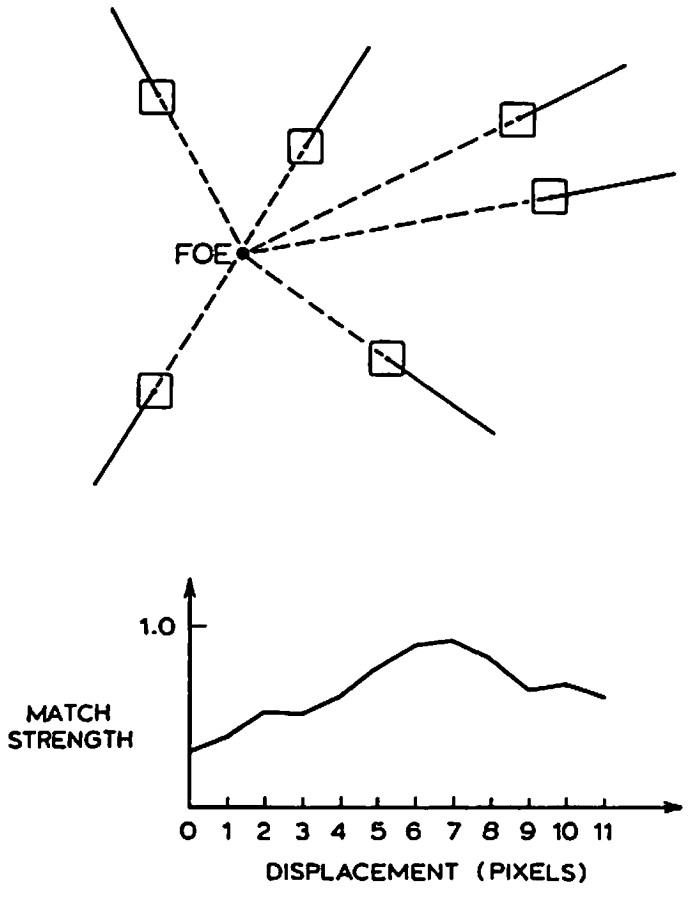


Figure 10. Translational Displacement Paths for a hypothesized FOE and a match function on one feature.

The set of all possible translational axes describes a unit sphere called the translational direction sphere. For reasons discussed below, the search procedures are defined with respect to this sphere, rather than the image plane itself. The error measure associates a point on the direction of translation sphere with a number describing the quality of feature matches along the image displacement paths determined by the corresponding hypothesized translational axis. This error value is computed by first finding the best match for each feature along a segment of its image displacement path using one of the normalised match metrics above. Each of these values is then subtracted from one, and all the resulting values are added together to form an error measure. Thus, for a set of N features in an initial image, a hypothesized translational axis, and use of one of the match metrics above, the error measure E is

$$E = \sum_{i=1}^{N} [1.0 - bestmatch(i)] \tag{14}$$

where $bestmatch(i)$ is the best match value associated with feature i along the appropriate image displacement path.

The error measure utilizes the different correlation norms described above and different interpolation processes for determining positions along an image displacement path. The choices among these generally involve a trade-off between the speed of evaluating the error measure and the precision with which the translational axis can be determined.

The interpolation process approximates feature values along the image displacement path from one image onto another. Depending on the accuracy required, positions along the image displacement path can be approximated roughly by setting the coordinates of the feature's position to the nearest integer value, or more accurately by performing a bilinear subpixel interpolation of the feature at each of a set of selected positions along the image displacement path. The basic trade-off is between speed and accuracy, with subpixel interpolation being more expensive.

The error measure was computed in two forms in the experiments below: a fast evaluation form and a precise evaluation form. The fast form uses the absolute value norm and the nearest integer approximation to determine feature position along the image displacement paths. The fast form is useful for evaluating image sequences with several extracted features to determine the rough position of the global minimum. However, the fast form may not be adequate for fine determination of the translational axis because of the nearest integer approximation for feature position.

The precise form of evaluation uses the Moravec norm and bilinear interpolation. It has been found to vary smoothly with respect to small changes in the position of a translational axis.

Utility of the Direction of Translation Sphere

There are significant advantages in defining the error measure with respect to a unit sphere instead of the potential positions of FOEs and FOCs in the image plane. The sphere is a bounded surface which makes uniform global sampling of the error measure feasible. In contrast, when the image plane is used directly, the resolution in the position of the translational axis varies. For example, the FOEs determined by translational axes separated by very small angles will be separated by larger and larger distances in the image plane as FOEs are placed further from the visible image. The effect of using the image plane on the error measure is a loss of resolution with large flat areas surrounding FOEs that are distant from the visible portions of the image.

Finally, special criteria must be used to distinguish between FOEs and FOCs if the error measure is defined relative to the image plane. Roughly parallel image displacements could correspond to an FOE off to one side or an FOC off to the opposite side of the image plane. On the direction of translation sphere, the corresponding translational axes would be close, while on the plane they are widely separated at plus and minus infinity.

Search Organization

The search process used here consists of two phases: An initial global sampling of the error measure to determine its rough shape and then a local search to determine

a minimum. The local search begins at the position where the minimum value was determined by the global sampling. The local search is a gradient descent procedure using a diminishing step-size. That is, it begins with an initial fixed step size and determines a local minimum using it. The step-size is then reduced and the procedure repeated until the step-size is at the desired resolution for the determination of the translational axis. In the experiments below the initial step-size was set to 0.1 and then reduced successively to 0.025 and 0.005 radians.

As will be seen in the following experiments, the error measure is smooth, with a single minimum in a large neighborhood around the correct translational axis. Thus, the global sampling can be quite sparse or the initial step size of the local search quite large.

Experiments

The following experiments were performed using the roadsign and industrial image sequences. They represent a wide range of situations. The first experiment involves determining the translational axis from the industrial image sequence using the features indicated in figure 9e. In this sequence the translational axis intersects the image plane in a visible portion of the image. The second experiment involves processing the industrial image sequence using a smaller number of features. In the third experiment the roadsign image sequence is processed using the features extracted at the positions indicated in figure 8e. Here, the intersection of the translational axis and the image plane is not in the visible portion of the image.

The fourth experiment involves processing the roadsign image sequence, but using the features extracted prior to low-curvature suppression. This has the effect of introducing weak and spurious features into the error measure computation. The fifth experiment involves processing the roadsign images using features extracted from a small area of the initial image.

In all of the experiments, the maximal displacement along an image displacement path was set to 10 pixels. Displacements were in increments of 1 pixel along the image displacement paths. Features were 7x7 pixel arrays centered at the positions indicated in the figures.

We use a 2-D, polar coordinate system to describe the points on the direction of translation sphere over which the error measure is evaluated. The axes of translation are unit vectors based at the origin of the camera coordinate system and are described by two angles (ϕ_1, ϕ_2) (figure 11). For an axis of translation, V, based at the origin, ϕ_1 is the angle between the $(0,1,0)$ vector and the edge determined by the intersection of the YZ plane and the plane determined by the X axis and V. ϕ_1 thus specifies one of the pencil of planes containing the X axis. ϕ_2 is then used to express V as a vector in the specified plane. ϕ_2 is the angle between $(-1,0,0)$ and V. Note that for all angles a and b, $(a,0) = (b,0)$ and $(a,\pi) = (b,\pi)$ which corresponds to points lying along the X axis.

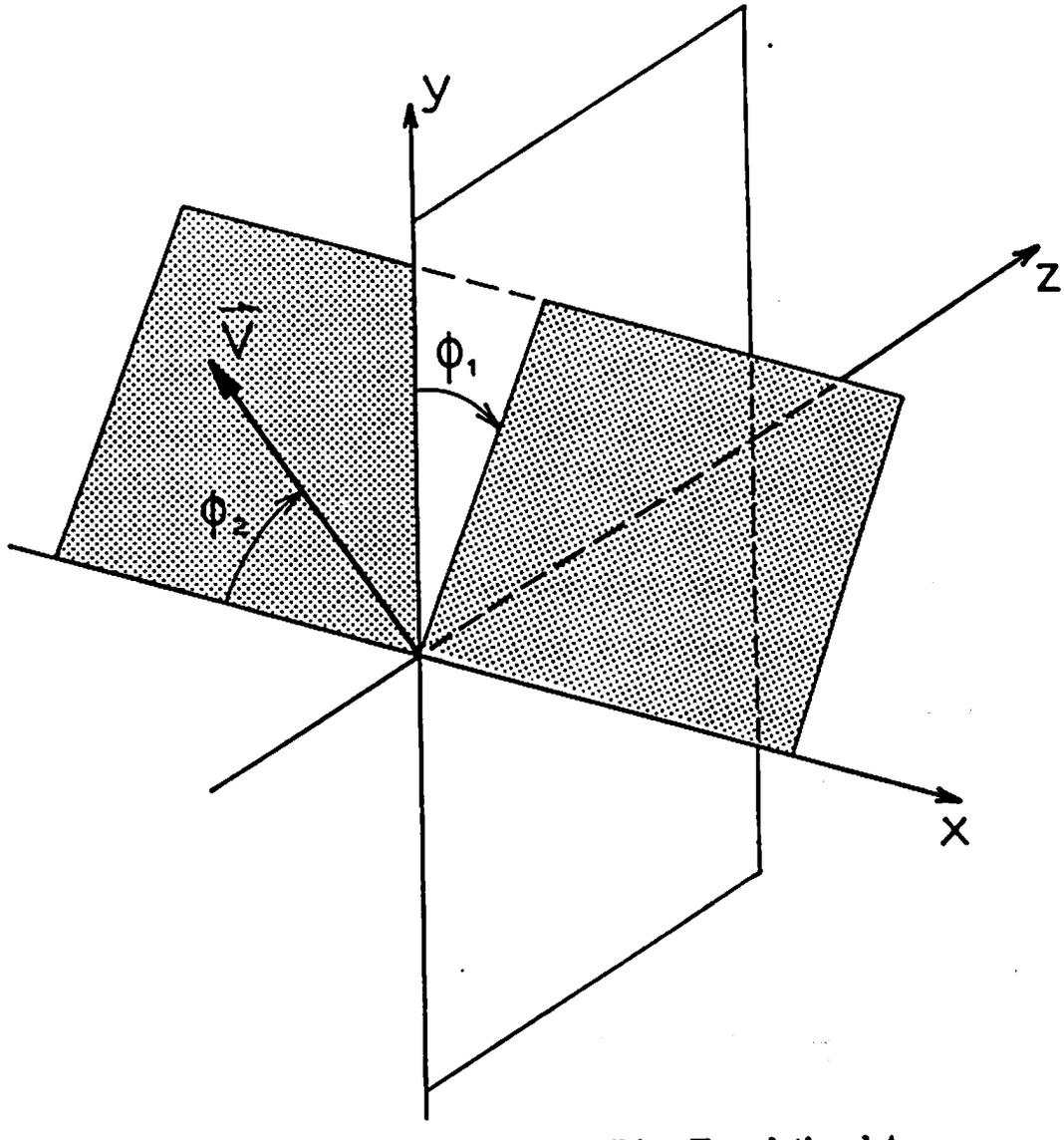


Figure 11. Coordinate System for Describing Translational Axes

For each experiment, the results of processing are contained in 3 tables. The first two (tables a and b) indicate the values of the error measure during the global sampling of points using a fixed angular increment (equal to $\frac{\pi}{10}$ or 18 degrees) in

(ϕ_1, ϕ_2) coordinates on the direction of translation sphere. The first of these tables corresponds to translational axes which intersect the image plane at FOEs. The second basically corresponds to those which intersect the image plane at FOCs. Each of these tables is also presented as an intensity plot and a contour plot. In the intensity plot, error is proportional to intensity so darker areas imply lower values of error. In the contour plots, the positions of local minima are marked with a " - " and the local maxima are marked with a " + ". Certain distortions appear in these figures because they result from mapping the unit sphere onto planes. Thus values near the right and left hand sides of the figures are actually closer to each other on the unit sphere than those points nearer the center. Additionally, the positions on the extreme left-hand side of the figures actually correspond to the same point on the direction of translation sphere which flattens the error surface plots at these positions.

The third table (table c) shows the minimal value determined by the global sampling process that is used to initiate the local search, and the successive values of the error measure determined during the local search. In this table, the position of the translational axis is referred to in terms of (X, Y, Z) camera coordinates, in addition to (ϕ_1, ϕ_2) coordinates, so that translational axes computed under different situations can be compared.

Industrial Images

The procedure was applied to the industrial images using the features extracted at the positions shown in figure 9e. Tables 1a and 1b show the global sampling of the error measure using the fast form of evaluation. Note the minima at $(\phi_1, \phi_2) = (5\frac{\pi}{10}, 4\frac{\pi}{10}) = (1.571, 1.257)$ radians. Table 1c shows the successive values of the local search using the precise form of evaluation. The determined translational axis is (-0.139, -0.099, 0.985). The image displacements determined for these features are shown in figure 12.

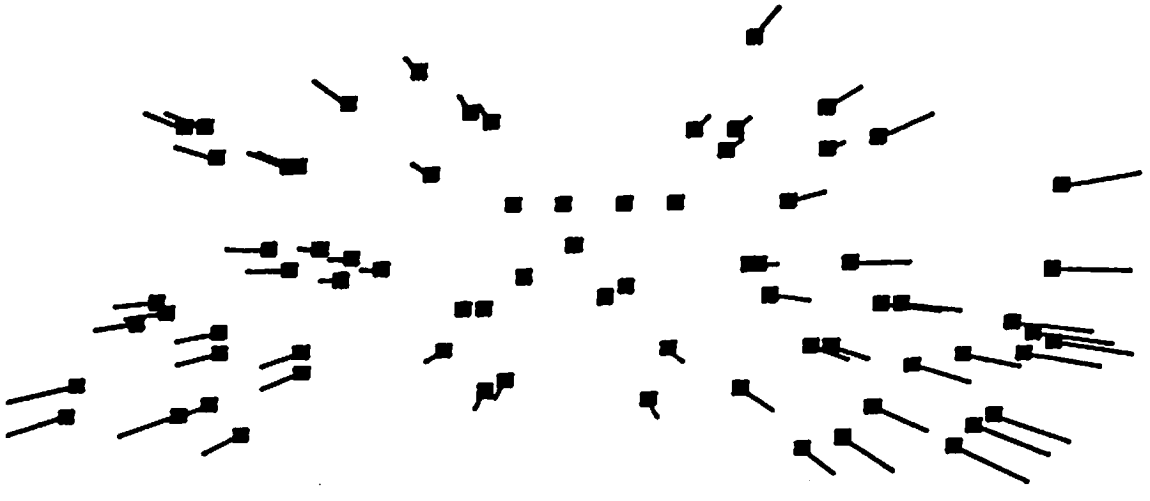


Figure 12. Industrial Image Displacements.

	0	1	2	3	4	5	6	7	8	9
0	22.00	23.73	24.16	25.15	26.03	26.02	26.52	25.30	24.17	23.18
1		20.76	24.90	26.09	26.18	27.17	27.27	28.46	27.28	24.21
2		20.29	23.22	25.60	25.03	26.88	26.21	27.13	25.33	23.65
3		20.31	21.60	24.71	25.45	25.61	25.67	24.95	24.75	23.90
4		20.17	19.97	21.17	23.74	22.65	23.33	24.07	24.20	24.25
5		21.45	20.31	19.52	14.45	15.76	20.53	24.22	24.48	24.82
6		21.04	20.66	20.78	18.12	17.38	19.94	22.71	23.86	24.56
7		21.25	22.51	21.86	23.55	24.75	23.10	21.89	23.01	24.22
8		22.19	22.49	24.23	25.33	26.39	24.88	24.25	22.20	23.67
9		22.97	24.09	24.89	26.34	26.22	26.08	25.30	23.24	23.83

Table 1a. Industrial Image Error Values

	10	11	12	13	14	15	16	17	18	19
0	26.91	22.91	26.39	27.19	28.17	28.84	29.11	30.89	29.62	26.16
1		25.42	26.06	27.81	27.75	27.71	27.81	28.20	27.25	26.02
2		26.10	26.88	28.19	28.00	27.42	29.11	28.91	27.73	26.28
3		26.98	27.98	28.52	28.38	28.33	29.76	29.69	27.52	26.35
4		26.72	27.72	29.89	29.43	30.55	29.35	30.23	27.07	25.85
5		26.75	27.01	30.86	33.98	32.84	30.01	27.99	26.35	24.36
6		27.03	27.34	30.04	32.99	32.16	30.98	26.77	23.97	23.86
7		26.55	27.67	31.02	31.13	31.69	30.68	29.49	24.51	24.37
8		26.49	29.30	30.81	31.06	30.16	29.72	27.44	27.69	23.20
9		26.14	29.14	31.25	30.14	28.94	28.46	27.41	27.11	22.88

Table 1b. Industrial Image Error Values

Stepsize	ϕ_1	ϕ_2	X	Y	Z	Error
	1.5708	1.2566	-0.30905	0.00000	0.95105	*14.446*
0.1	1.6708	1.4566	-0.11395	-0.09919	0.98852	3.5456
0.025	1.6708	1.4316	-0.13875	-0.09887	0.98538	3.5313
0.005	1.6708	1.4316	-0.13875	-0.09887	0.98538	3.5313

Table 1c. Industrial Image Local Search Values

*Denotes this error value was computed using the fast evaluation form. The other values were computed using the precise evaluation form.

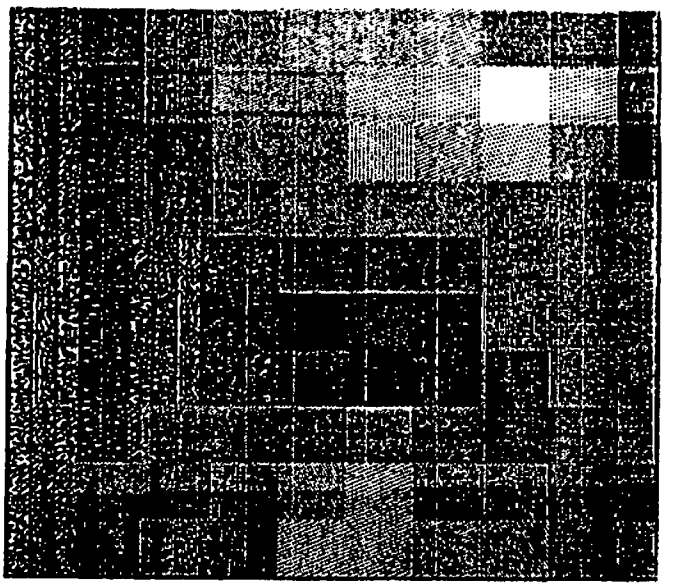


Figure 13a. Intensity plot of Table 1a.

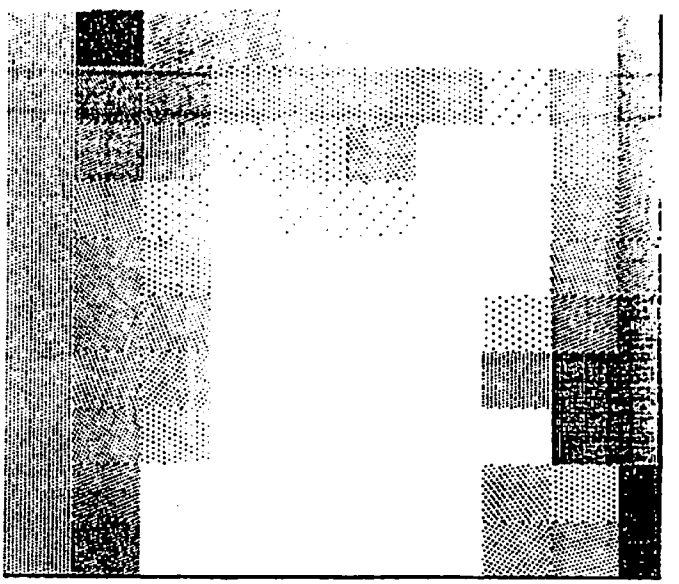


Figure 13b. Intensity plot of Table 1b.

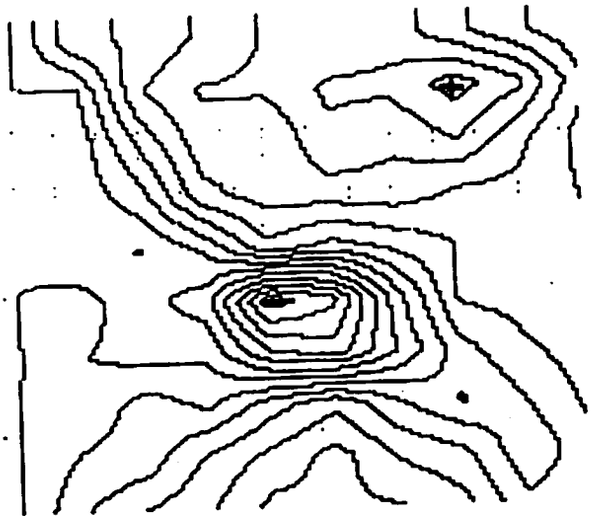


Figure 13c. Contour plot of Table 1a.

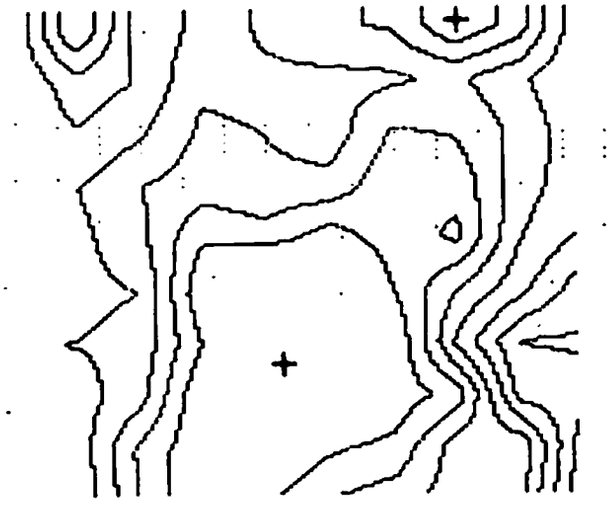


Figure 13d. Contour plot of Table 1b.

Industrial Images with Selected Features

The procedure was again applied to the industrial image sequence but using features which were selected by hand. The positions of these 8 features are shown in figure 14.

Tables 2a and 2b show the global sampling of the error measure using the precise form of evaluation. Note the minima at $(\phi_1, \phi_2) = (5\frac{\pi}{10}, 5\frac{\pi}{10})$. Table 2c shows the successive position determined by the local search. The translational axis was determined to be (-0.154, -0.079, 0.985). This corresponds to an angular difference of 0.025 radians (1.45 degrees) with respect to the axis determined in experiment 1.

Figure 14. Selected Features from Industrial Image 1.

	0	1	2	3	4	5	6	7	8	9
0	1.583	1.967	2.106	2.055	1.991	1.986	1.840	1.711	1.713	1.779
1		1.946	2.060	2.013	1.823	1.830	1.785	1.739	1.567	1.754
2		1.907	2.024	1.905	1.777	1.690	1.758	1.622	1.488	1.757
3		1.908	2.062	1.783	1.670	1.590	1.610	1.464	1.533	1.829
4		1.806	1.784	1.762	1.419	1.284	1.241	1.446	1.631	1.874
5		1.630	1.626	1.414	0.580	0.427	0.688	1.117	1.672	1.983
6		1.505	1.551	1.740	1.227	1.018	1.062	1.768	1.890	2.003
7		1.451	1.470	1.424	1.312	1.513	1.578	1.480	1.695	1.907
8		1.449	1.459	1.716	1.888	1.895	1.804	1.495	1.607	1.739
9		1.456	1.645	1.949	2.082	2.163	1.755	1.669	1.627	1.684

Table 2a. Industrial Image Selected Feature Error Values

	10	11	12	13	14	15	16	17	18	19
0	2.267	1.750	1.857	1.937	2.137	2.675	2.642	2.354	2.089	1.635
1		1.773	1.926	2.024	2.218	2.532	2.731	2.492	2.141	1.611
2		1.802	1.961	2.143	2.304	2.416	2.721	2.655	2.160	1.616
3		1.835	1.964	2.221	2.370	2.421	2.748	2.685	2.047	1.669
4		1.933	1.977	2.348	2.579	2.526	2.647	2.597	1.811	1.739
5		2.224	2.291	2.709	2.770	2.681	2.663	2.490	1.874	1.802
6		2.443	2.524	2.434	2.614	2.893	2.647	2.129	2.100	1.902
7		2.308	2.096	2.193	2.582	2.520	2.625	2.589	2.108	2.083
8		2.045	2.003	2.109	2.137	2.295	2.478	2.491	2.379	2.134
9		1.939	2.004	1.888	1.910	2.219	2.299	2.335	2.469	2.132

Table 2b. Industrial Image Selected Feature Error Values

Stepsize	ϕ_1	ϕ_2	X	Y	Z	Error
	1.5708	1.5708	0.00000	0.00000	1.00000	0.57998
0.1	1.6708	1.3708	-0.19867	-0.09785	0.97517	0.19955
0.025	1.6458	1.4208	-0.14943	-0.07401	0.98599	0.17476
0.005	1.6508	1.4158	-0.15438	-0.07896	0.98485	0.17410

Table 2c. Industrial Image Selected Feature Local Search Values

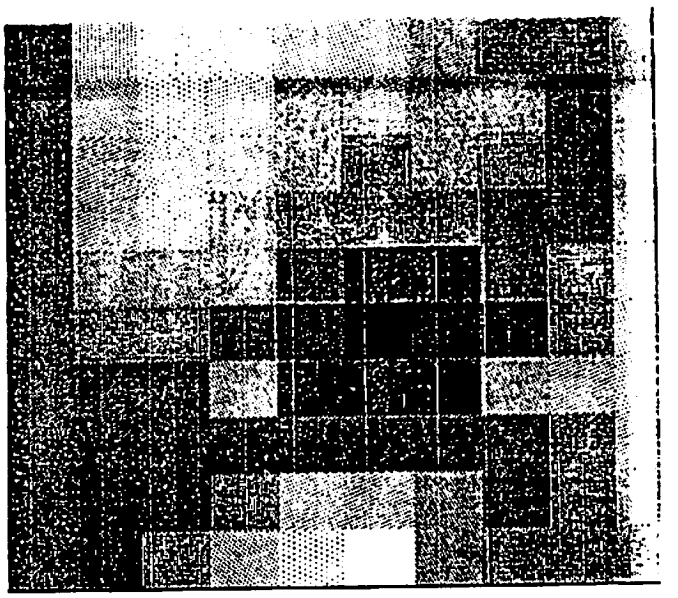


Figure 15a. Intensity plot of Table 2a.

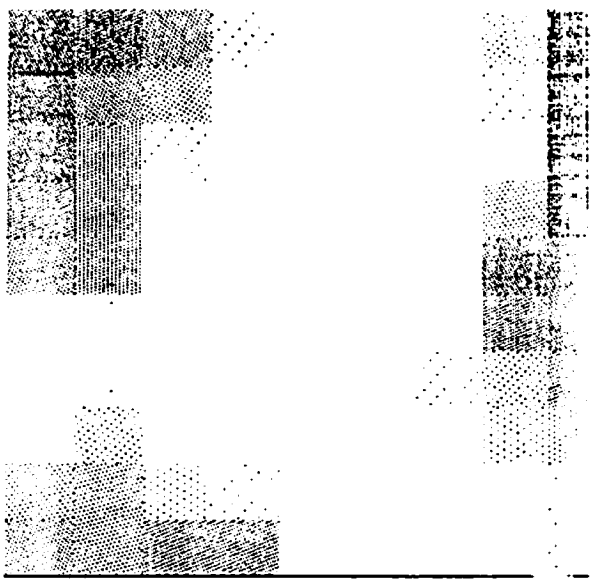


Figure 15b. Intensity plot of Table 2b.

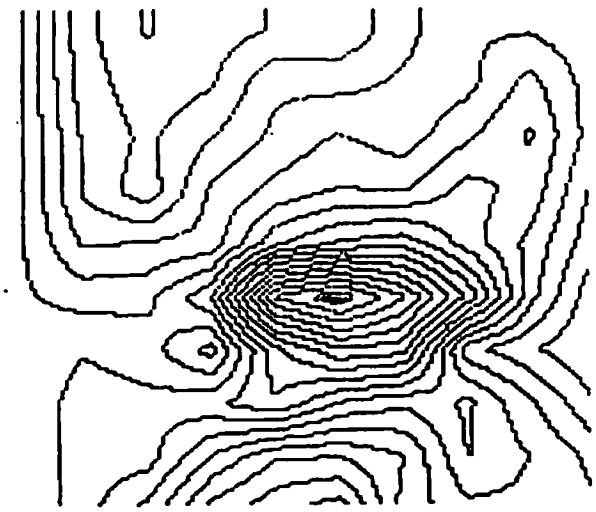


Figure 15c. Contour plot of Table 2a.

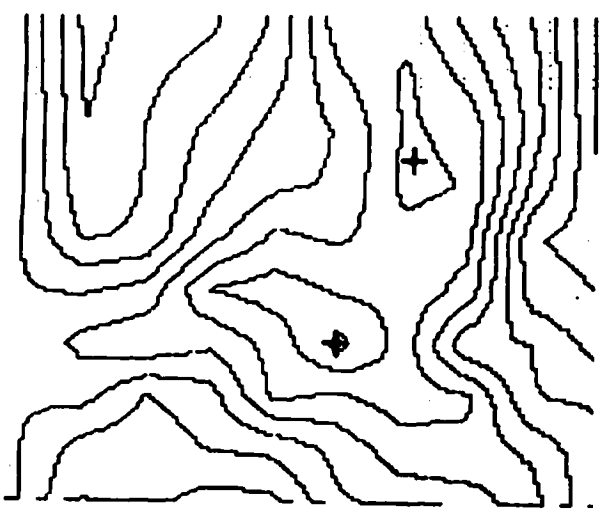


Figure 15d. Contour plot of Table 2b.

Roadsign Image Sequence

The procedure was applied to the roadsign image sequence using the features extracted at the positions indicated in figure 8e. Tables 3a and 3b show the global sampling of the error measure using the fast form of evaluation. Note the minima at $(\phi_1, \phi_2) = (8\frac{\pi}{10}, 2\frac{\pi}{10})$. Table 3c shows the successive values of the local search using the precise form of evaluation for the error measure. The translational axis determined by this process is (-0.837, -0.420, 0.349). The image displacements for the feature points shown in figure 8e that are associated with this translational axis are shown in figure 17.

Given the direction of translation and image displacements, the relative environmental depths of image points can be recovered by the simple relation in equation ten from chapter III. When image displacements are small, the inferred depth values can be quite erratic due to sensitivity to small numbers in the denominator in the left hand side of this equation. For this reason it is necessary to use image pairs for which large displacements can be determined. One way to do this for image sequences which are related by successive sensor translations is to track the FOE from a given image with respect to successive later image. This was done with four successive images from the roadsign sequence beginning with roadsign images 1 and 2 and using the features from image 1 at the positions in figure 8e. The position of the translational axis determined from images I(1) and I(t+1) was used as the initial value in the local search for determining the translational axis for images I(1) and I(t+2), where $t = 1, 2$ in this example. The displacements of all features

along the contour in figure 8c were determined along the image displacement paths determined by the FOE found for images I(1) and I(4). To compute depth along the contours, 5x5 windows, centered at each contour point, were matched along the image displacement paths and the displacement corresponding to the best match were determined. The resulting relative depth map is shown in figure 18 where depth is encoded by intensity (more distant things are brighter).

The roadsign sequence is particularly nice for presenting depth processing results because the three environmental objects in the images are at three distinct depth intervals. This is shown in figure 19 by the three distinct clusters in the histogram of the depth values calculated for the points along the contour. The units in the histogram are cumulative time-until-contact values. That is, the depth is given in units of the displacement of the camera from I(1) to I(4) along the Z-axis. From left to right, the first peak corresponds to the sign, the second to the pole, and the third to the trees. As can be seen, there is a wide range of depths associated with the trees. Mapping these clusters back onto contour points from figure 8c yields the distinct objects: the boundary shown in figure 20a (the sign), the boundary shown in figure 20b (the pole), the boundary segment shown in figure 20c (the trees). Points near the image boundary of I(1) were ignored because the processing did not take into account occlusion effects along the image boundaries.

	0	1	2	3	4	5	6	7	8	9
0	4.935	6.487	8.022	9.200	10.24	10.94	11.00	11.87	11.92	11.27
1		6.296	7.493	8.540	9.329	9.729	9.801	9.868	10.06	10.01
2		6.059	7.122	8.177	8.971	9.625	9.750	9.812	10.10	9.993
3		5.739	6.593	7.270	8.309	8.967	9.492	9.788	10.02	9.966
4		5.402	5.651	5.940	6.988	8.119	8.709	9.082	9.806	9.895
5		4.787	4.536	4.838	6.117	7.454	8.314	8.828	9.434	9.771
6		4.149	3.590	4.035	5.071	6.537	7.716	8.870	9.200	9.669
7		3.694	2.865	3.357	4.622	5.999	7.750	8.816	9.147	9.604
8		3.319	2.795	3.808	5.432	6.821	8.026	8.751	9.041	9.505
9		3.281	3.129	4.385	6.078	7.126	7.903	8.817	9.125	9.546

Table 3a. Roadsign Image Error Values

	10	11	12	13	14	15	16	17	18	19
0	11.20	9.617	9.147	8.802	7.836	7.277	6.247	4.632	3.284	3.378
1		10.76	10.40	9.530	8.103	7.335	6.504	5.006	3.962	3.538
2		10.96	10.80	9.915	8.734	7.237	6.388	5.280	4.270	3.849
3		11.14	11.04	10.62	9.592	8.233	7.020	5.690	4.691	4.270
4		11.17	11.20	11.07	10.28	9.343	8.121	6.774	5.235	4.511
5		11.20	11.29	11.33	10.97	10.24	9.057	7.383	5.694	4.959
6		11.20	11.34	11.64	11.16	10.92	9.485	7.904	6.159	5.394
7		11.20	11.54	11.77	11.74	10.88	9.975	8.158	6.813	5.758
8		11.23	11.69	11.92	11.71	10.94	10.35	9.084	7.719	6.158
9		11.25	11.90	11.95	11.38	10.81	10.32	9.316	8.031	6.314

Table 3b. Roadsign Image Error Values

Stepsize	ϕ_1	ϕ_2	X	Y	Z	Error
	2.5133	0.62832	-0.80902	-0.47554	0.34548	*2.7952*
0.1	2.5133	0.52832	-0.86366	-0.40782	0.29628	0.21031
0.025	2.4383	0.57832	-0.83738	-0.41691	0.35352	0.20767
0.005	2.4483	0.57832	-0.83738	-0.42043	0.34933	0.20760

Table 3c. Roadsign Image Local Search Values.

* Denotes this error value was computed using the fast evaluation form. The other values were computed using the precise evaluation form.

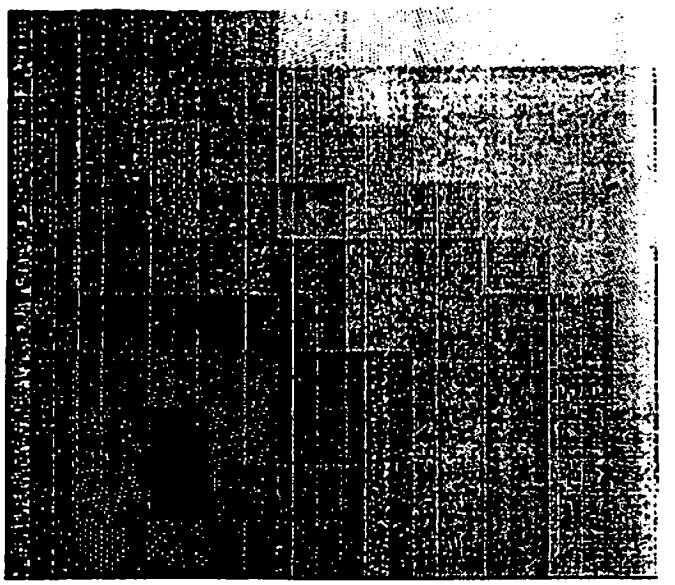


Figure 16a. Intensity plot of Table 3a.

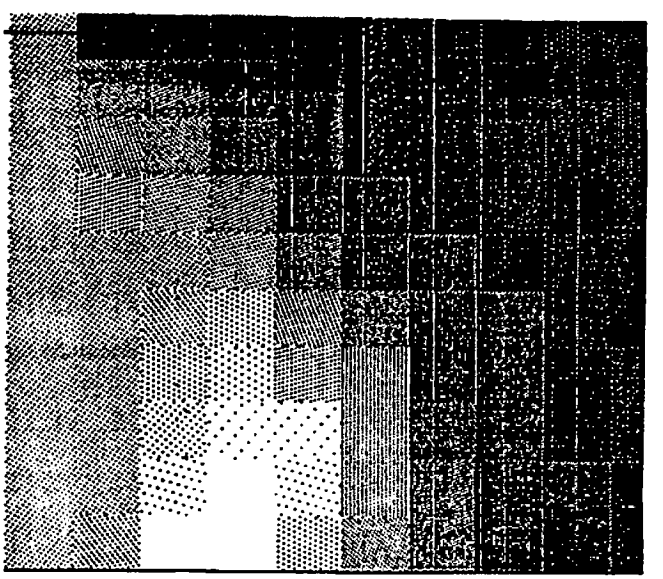


Figure 16b. Intensity plot of Table 3b.

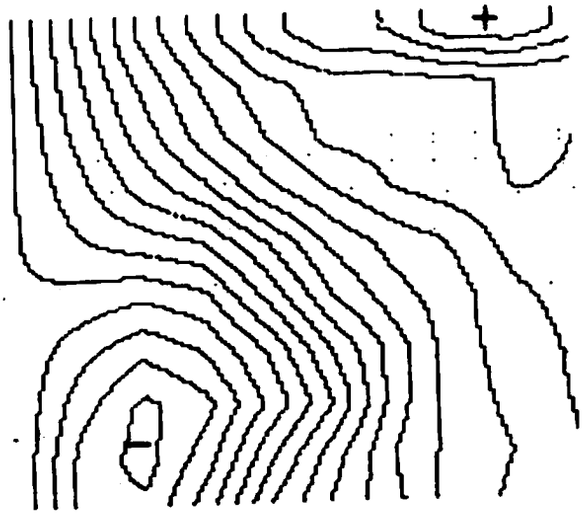


Figure 16c. Contour plot of Table 3a.

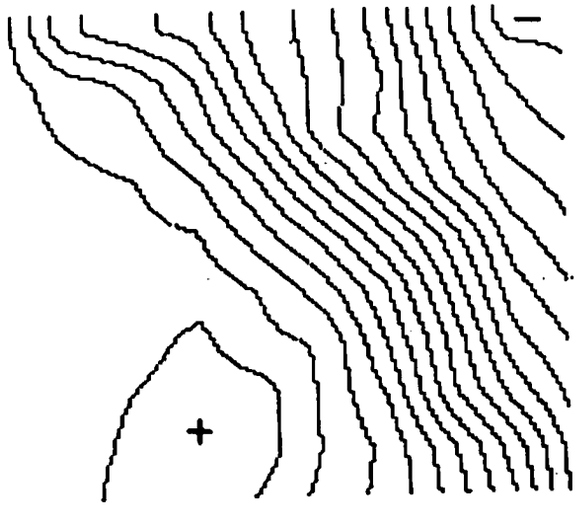


Figure 16b. Contour plot of Table 3b.

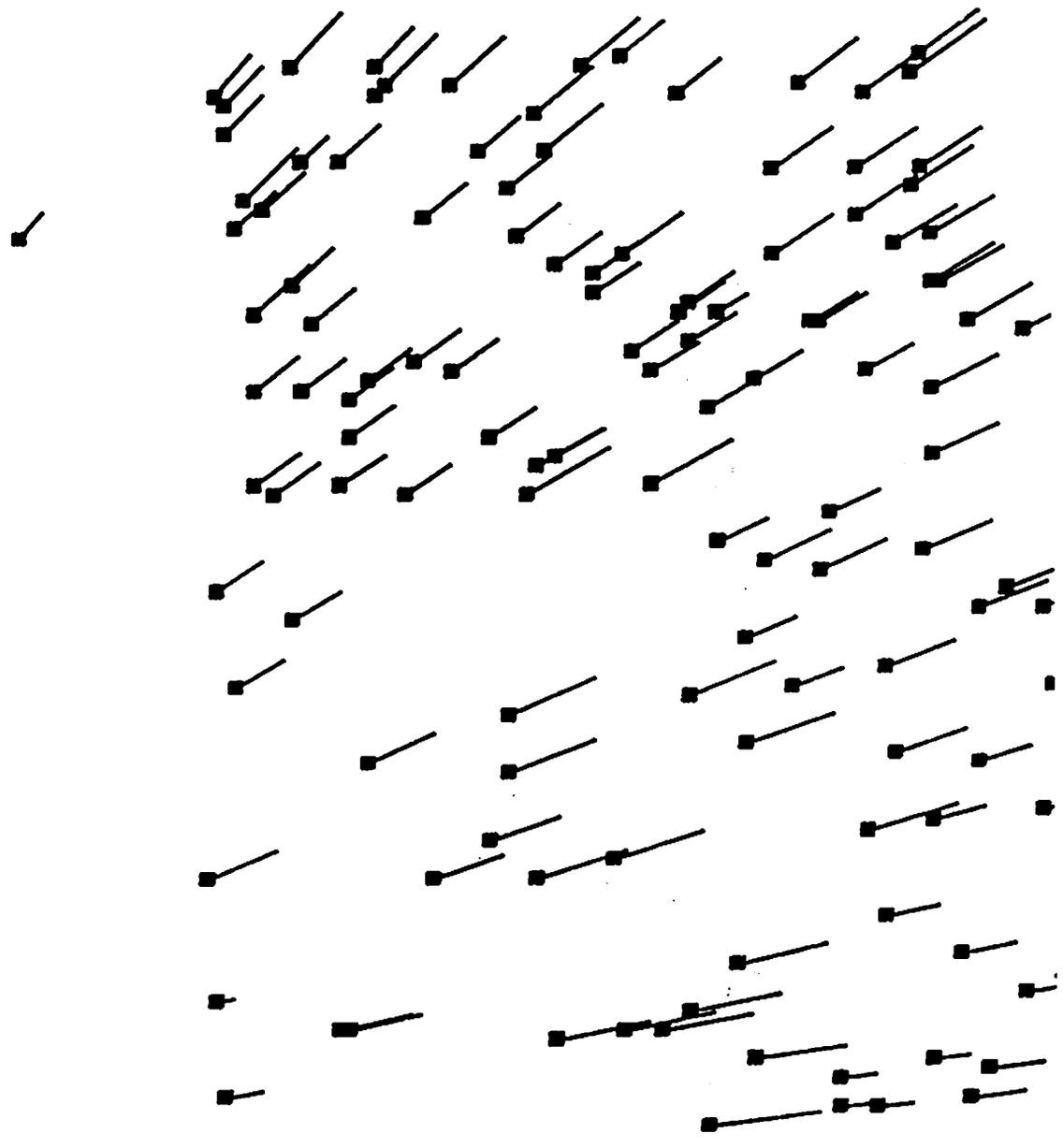


Figure 17. Displacements for Roadsign Images.

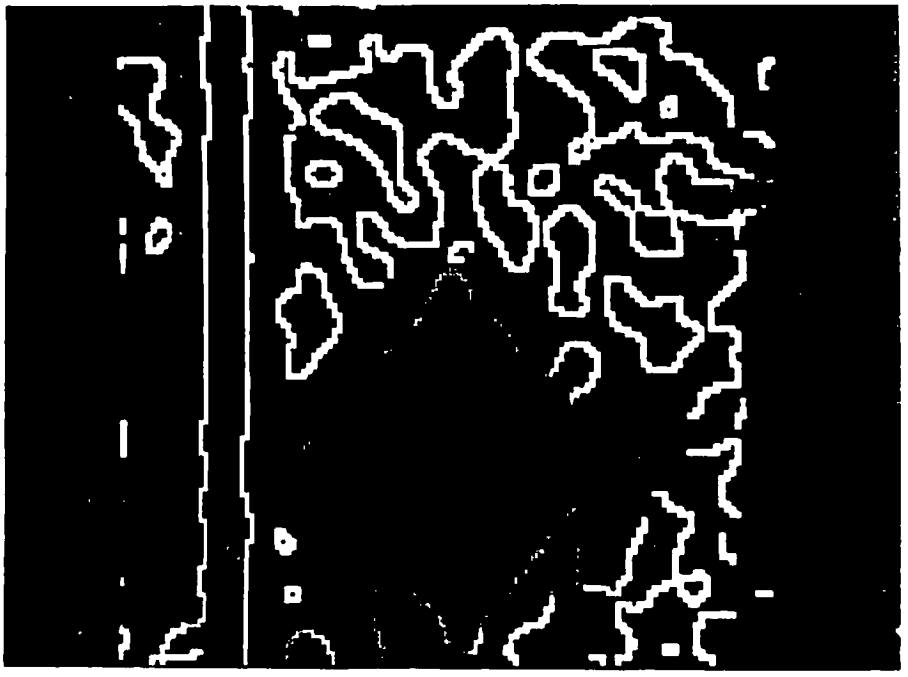


Figure 18. Depth Map. Contour depth encoded by intensity.

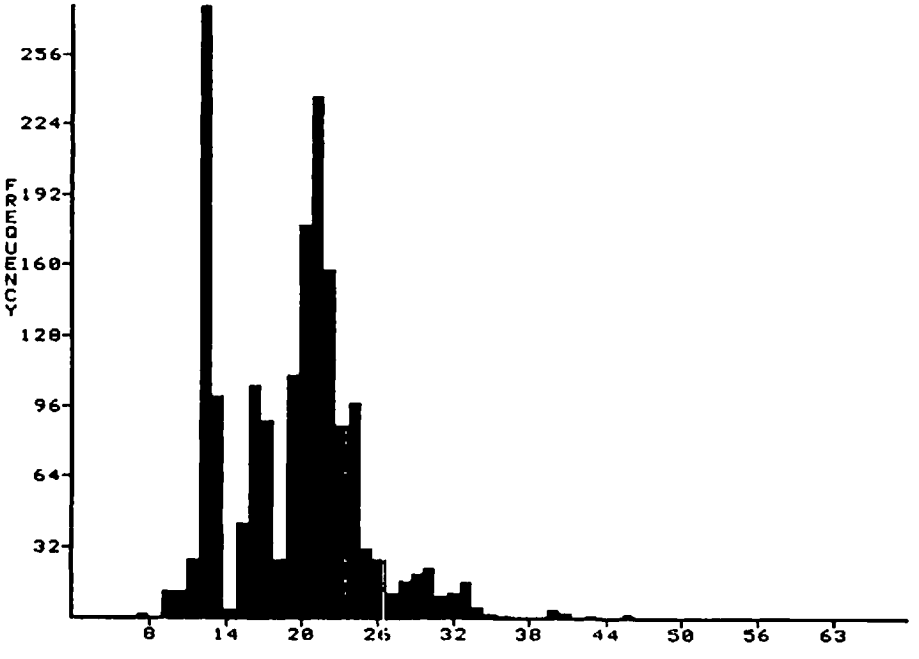


Figure 19. Depth Histogram.

Figure 20a. Sign Depth Cluster

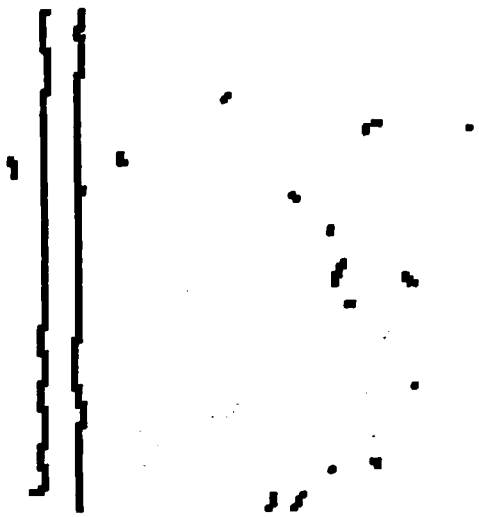


Figure 20b. Pole Depth Cluster

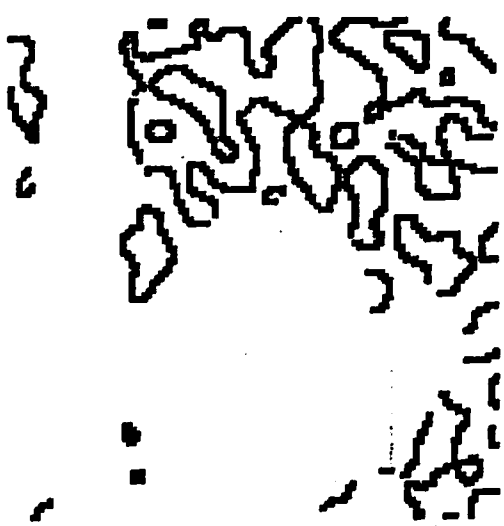


Figure 20c. Trees Depth Cluster.

Roadsign Sequence with Redundant Features

The procedure was applied to the roadsign image sequence using the features which were extracted prior to low-curvature suppression. The positions of these features is shown in figure 8d. This has the effect of including several weak and false features in the evaluation of the error measure.

Tables 4a and 4b show the values of the global sampling of the error measure using the fast form of evaluation. Note the minima at $(\phi_1, \phi_2) = (8\frac{\pi}{10}, 2\frac{\pi}{10})$. Table 4c shows the successive values of the local search. The determined translational axis was $(-0.829, -0.423, 0.366)$. This corresponds to an angle of 0.019 radians (1.068 degrees) with respect to the axis determined in experiment 3.

	0	1	2	3	4	5	6	7	8	9
0	7.777	10.07	12.54	14.27	15.46	18.22	18.37	18.91	19.44	19.53
1		9.942	12.45	14.15	16.67	18.32	18.34	18.94	19.55	19.54
2		9.593	11.80	13.52	17.07	18.04	18.32	19.15	19.64	19.55
3		9.071	11.01	12.47	16.12	16.92	17.80	19.19	19.60	19.49
4		8.412	9.412	11.15	14.14	15.85	17.38	18.21	19.32	19.40
5		7.506	7.562	9.772	12.68	15.30	16.82	17.85	18.82	19.25
6		6.690	5.760	8.265	11.72	13.74	15.95	17.83	18.49	19.15
7		6.008	4.821	6.675	10.60	13.20	15.69	17.76	18.41	19.07
8		5.555	4.733	6.971	11.47	13.71	16.15	17.57	18.30	18.97
9		5.535	5.206	7.515	11.45	14.10	15.66	17.61	18.34	19.02

Table 4a. Roadsign Redundant Feature Error Values.

	10	11	12	13	14	15	16	17	18	19
0	22.65	21.90	21.16	19.95	15.98	14.73	10.76	7.699	5.361	5.569
1		22.06	21.51	19.50	16.25	13.01	10.82	8.195	6.394	5.781
2		22.28	22.00	19.74	16.38	12.43	10.64	8.510	6.821	6.319
3		22.52	22.31	20.70	16.02	13.62	11.44	8.860	7.491	6.908
4		22.59	22.52	21.21	16.83	15.10	12.45	10.44	8.361	7.257
5		22.65	22.74	21.25	17.79	15.75	13.80	11.51	9.119	7.853
6		22.63	22.83	22.02	17.38	16.90	14.88	12.46	9.834	8.575
7		22.63	23.07	22.18	18.77	16.85	15.66	12.96	10.92	9.096
8		22.65	23.01	21.95	20.34	17.69	16.64	14.57	12.22	9.679
9		22.68	23.21	22.02	20.56	18.29	16.80	15.26	13.03	9.924

Table 4b. Roadsign Redundant Feature Error Values

Stepsize	ϕ_1	ϕ_2	X	Y	Z	Error
	2.5133	0.62832	-0.80902	-0.47554	0.34548	*4.7330*
0.1	2.5133	0.52832	-0.86366	-0.40782	0.29628	0.34143
0.025	2.4133	0.60332	-0.82346	-0.42344	0.37765	0.33771
0.005	2.4283	0.59332	-0.82909	-0.42281	0.36585	0.33693

Table 4c. Roadsign Redundant Feature Local Search Values

*Denotes this error value was computed using the fast form of evaluation. All other values were computed using the precise form.

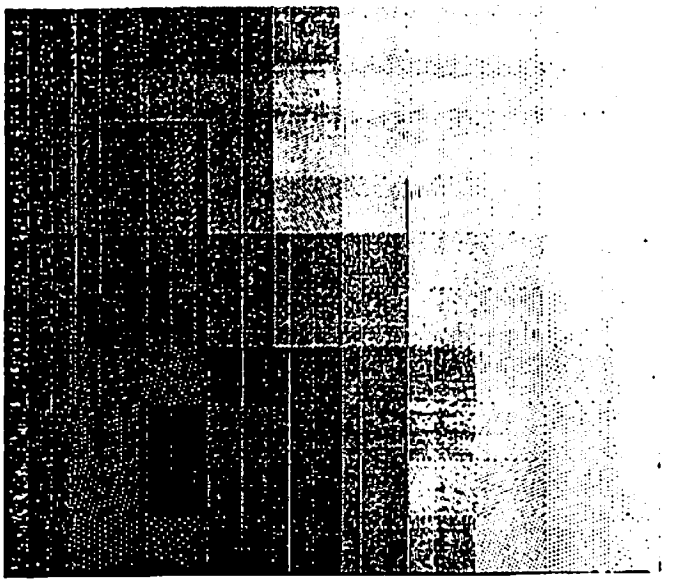


Figure 21a. Intensity plot of Table 4a.

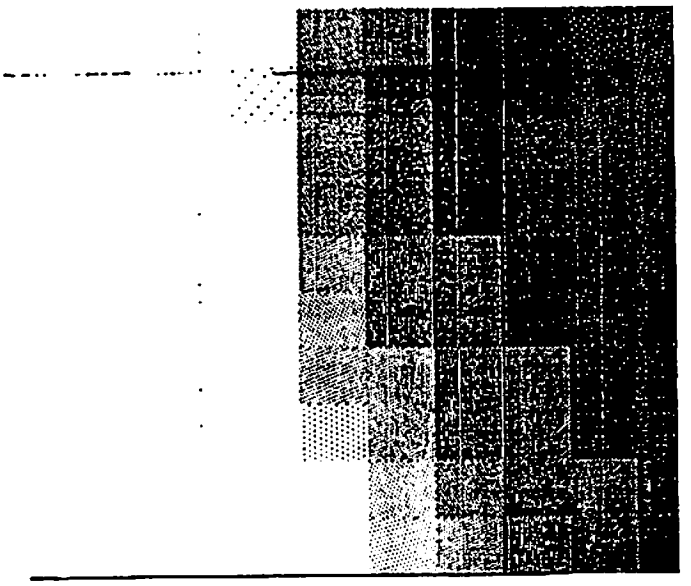


Figure 21b. Intensity plot of Table 4b.

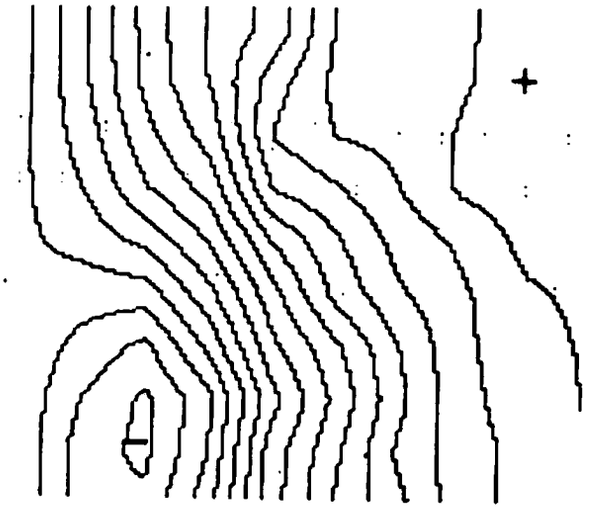


Figure 21c. Contour plot of Tables 4a.

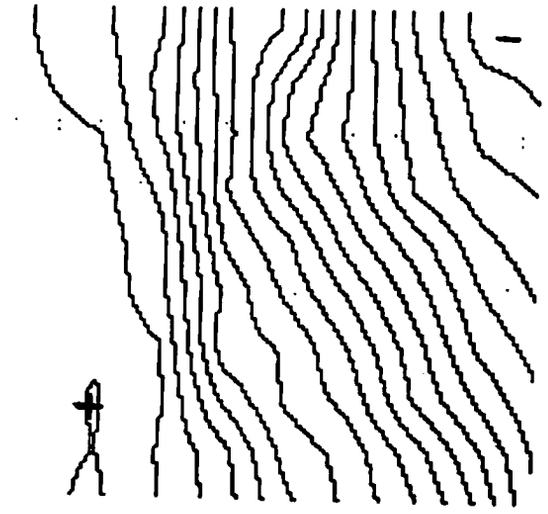


Figure 21d. Contour plot of Tables 4b.

Roadsign Subimage

This experiment was conducted to test the accuracy of the algorithm when applied to a very small area of the visual field. The procedure was applied to the roadsign image sequence with features restricted to the rectangular area shown in figure 22 corresponding to texture in the distant trees.

Tables 5a and 5b show the values of the global sampling of the error measure using the precise form of evaluation. Note the minima at $(\phi_1, \phi_2) = (7\frac{\pi}{10}, 2\frac{\pi}{10})$. Table 5c shows the successive values determined by the local search. The translational axis is determined to be $(-0.843, -0.429, 0.325)$. This corresponds to angles of 0.027 radians (1.53 degrees) and 0.044 (2.516 degrees), with respect to the translational axes determined in experiments 3 and 4 respectively.

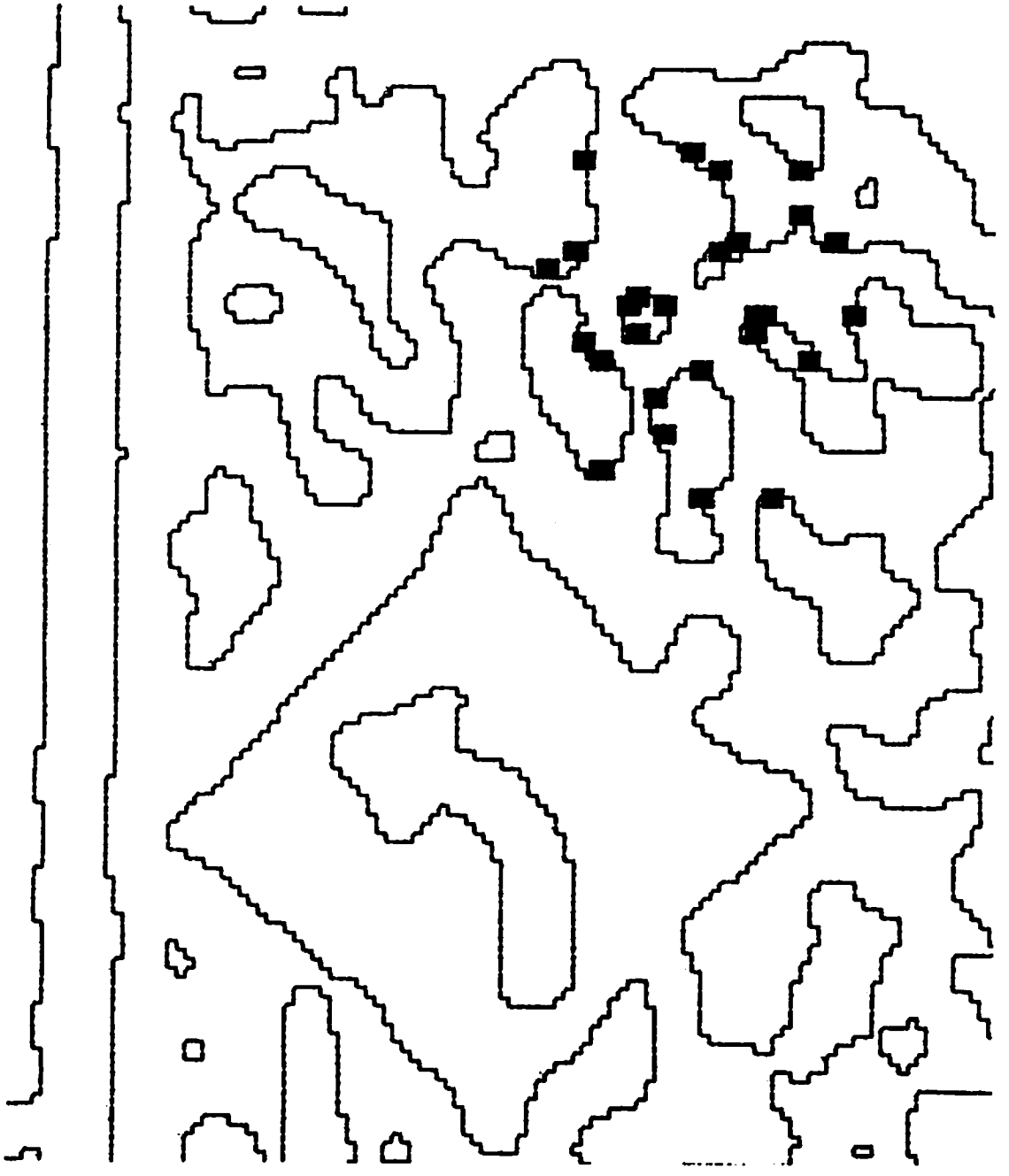


Figure 22. Roadsign Subimage Features.

	0	1	2	3	4	5	6	8	9	0
0	0.171	0.249	0.353	0.437	0.488	0.515	0.520	0.504	0.502	0.491
1		0.231	0.309	0.388	0.454	0.505	0.522	0.504	0.500	0.490
2		0.207	0.256	0.307	0.370	0.469	0.520	0.502	0.498	0.483
3		0.184	0.202	0.226	0.264	0.316	0.439	0.487	0.478	0.471
4		0.156	0.146	0.133	0.119	0.126	0.260	0.368	0.444	0.460
5		0.128	0.096	0.073	0.074	0.115	0.244	0.298	0.368	0.441
6		0.106	0.069	0.065	0.102	0.169	0.255	0.295	0.334	0.422
7		0.091	0.060	0.084	0.137	0.206	0.265	0.290	0.327	0.405
8		0.083	0.061	0.105	0.166	0.233	0.274	0.288	0.329	0.400
9		0.081	0.065	0.119	0.187	0.252	0.279	0.292	0.336	0.404

Table 5a. Roadsign Subimage Error Values

	10	11	12	13	14	15	16	17	18	19
0	0.568	0.473	0.393	0.334	0.300	0.279	0.206	0.126	0.065	0.085
1		0.495	0.428	0.371	0.319	0.279	0.224	0.121	0.063	0.094
2		0.524	0.478	0.433	0.380	0.315	0.235	0.100	0.070	0.112
3		0.549	0.534	0.515	0.494	0.458	0.276	0.102	0.115	0.142
4		0.575	0.580	0.587	0.589	0.588	0.503	0.320	0.194	0.176
5		0.586	0.603	0.611	0.616	0.627	0.600	0.482	0.303	0.205
6		0.597	0.614	0.619	0.624	0.626	0.612	0.512	0.386	0.236
7		0.605	0.617	0.624	0.623	0.636	0.610	0.520	0.417	0.256
8		0.608	0.620	0.620	0.629	0.646	0.597	0.515	0.418	0.265
9		0.609	0.621	0.620	0.639	0.639	0.573	0.498	0.397	0.262

Table 5b. Roadsign Subimage Error Values

Stepsize	ϕ_1	ϕ_2	X	Y	Z	Error
	2.1991	0.62832	-0.80902	-0.34549	0.47553	0.059910
0.1	2.2991	0.62832	-0.80902	-0.39123	0.43867	0.059542
0.025	2.4741	0.57832	-0.83738	-0.42930	0.33837	0.059288
0.005	2.4941	0.56832	-0.84281	-0.42928	0.32465	0.059269

Table 5c. Roadsign Subimage Local Search Values

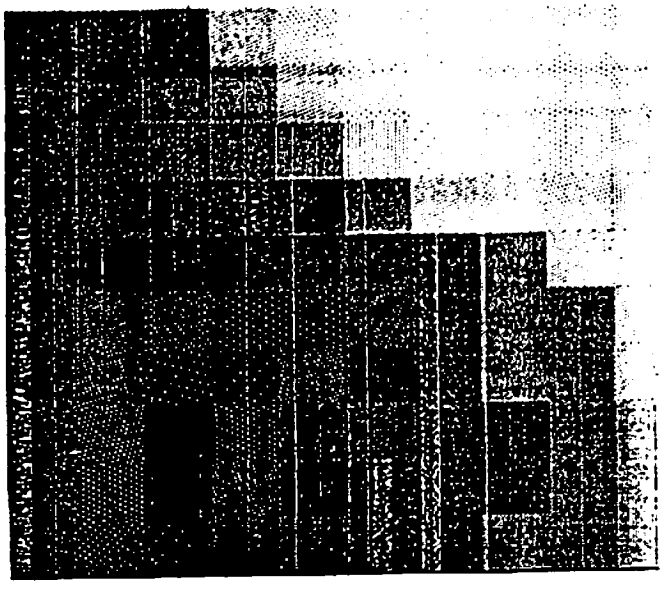


Figure 23a. Intensity plot of table 5a.

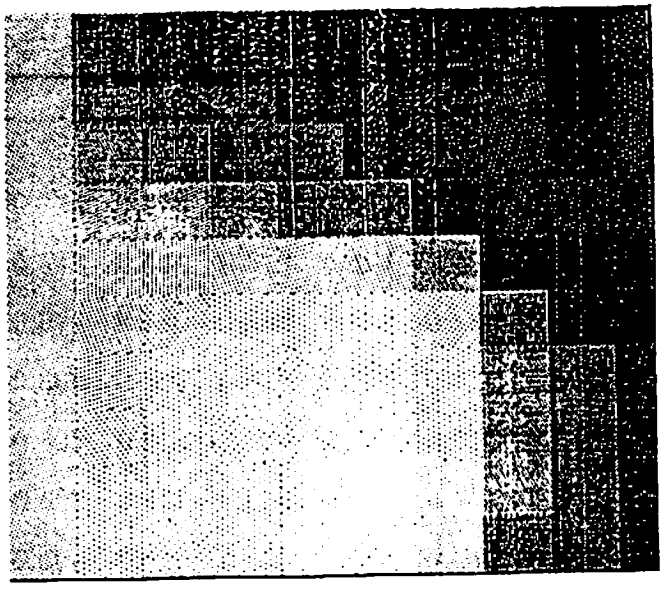


Figure 23b. Intensity plot of table 5b.

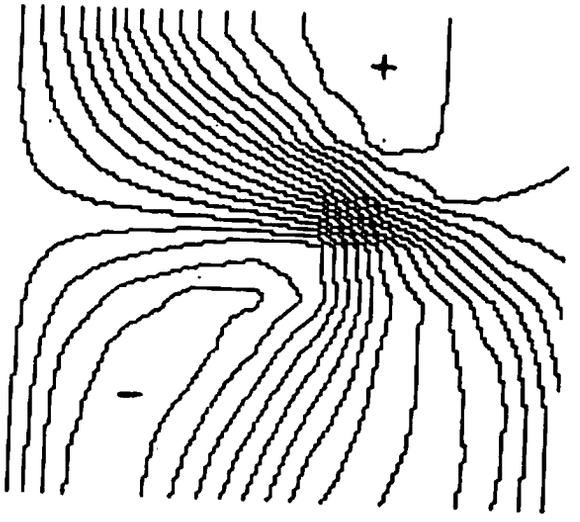


Figure 23c. Contour plot of table 5a.

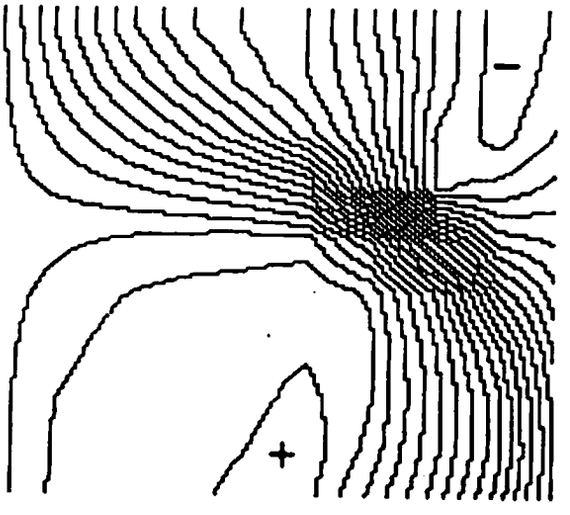


Figure 23d. Contour plot of table 5b.

Discussion

The experiments presented here, as well as others, have shown that the procedure is robust in several important ways. It is resilient with respect to weak and false features and is not dependent on identical features being extracted in successive images prior to matching. It can use a small number of features positioned across an image surface, or a small number of features from a limited area of the image.

In the remainder of this chapter, we discuss the feature extraction process and how it may be made more efficient, and the general behavior of the error measure. In the next chapter we explore several potential extensions of the translational motion procedure.

Feature Extraction

Since the procedure's performance does not degrade severely due to the occurrence of poor features, the type of feature extraction used is not critical. Nonetheless, the feature extraction process developed here could be extended in many ways. A simple one is to constrain the extraction of interesting points to positions where image contrast exceeds some minimal value. Also, other types of contour extraction can be used. For example, contours can also be determined by local application of histogram guided thresholding and segmentation. This resolves some of the problems associated with using a single threshold determined for image subparts with

significantly different brightnesses [Kohl81].

A significant question concerns the speed at which features are extracted. Locality of processing leads to the most efficient computation in array processing architectures. In the procedure here, the technique of contour walking to determine curvature is significantly non-local. Since the algorithm is robust with respect to weak features, the use of less costly methods for extraction of possibly weaker features may be acceptable. It may be possible to directly determine points of high curvature by using corner finders [Kitc80, Zuni83].

Another alternative to the contour walking is to simply use a threshold on the distinctiveness measures, with or without the determination of local maxima in distinctiveness. Examination of the local maxima along the telephone pole in figure 2c, reveals that these are local maxima with very small distinctiveness measures. This has been observed in general.

An additional speed-up can be obtained when features are selected from contours determined by segmentation procedures (such as thresholding or zero-crossing extraction) which produce binary images where pixel values may be represented by 1 or -1. In this case there is no need to normalize the correlation measure used to determine distinctiveness because each image subarea of equal size has identical constant image energy [Duda73]. Thus, the normalizing terms in the correlation measures become constants and the arithmetic operations are restricted to products or additions over the set $(1, -1)$. When the distinctiveness measures are determined along the contours of binary images followed by a threshold on distinctiveness and local maximal extraction, very rapid rates of feature extraction can be achieved

in the particular architectures we have explored, on the order of a fraction of a millisecond [Lawt84].

The binary image in Figure 24 was determined by thresholding at zero the initial roadsign image with the $\nabla^2 G$ mask used above. Figure 25 shows the interesting points extracted from the binary image in figure 24 using a threshold on distinctiveness set to 0.1 followed by local maxima extraction. The results are reasonable, although mistakes can occur if the neighborhoods over which local maxima are computed contain points of high curvature from distinct regions. This could be remedied by restricting the calculation of distinctiveness for points only along contours of the same region (which would then require the determination of region labels via a connected components algorithm).

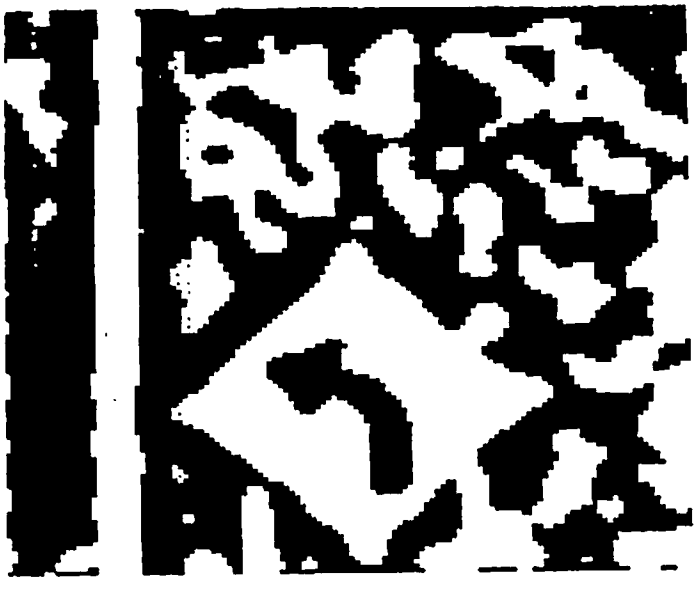


Figure 24. Binary Roadsign Image.

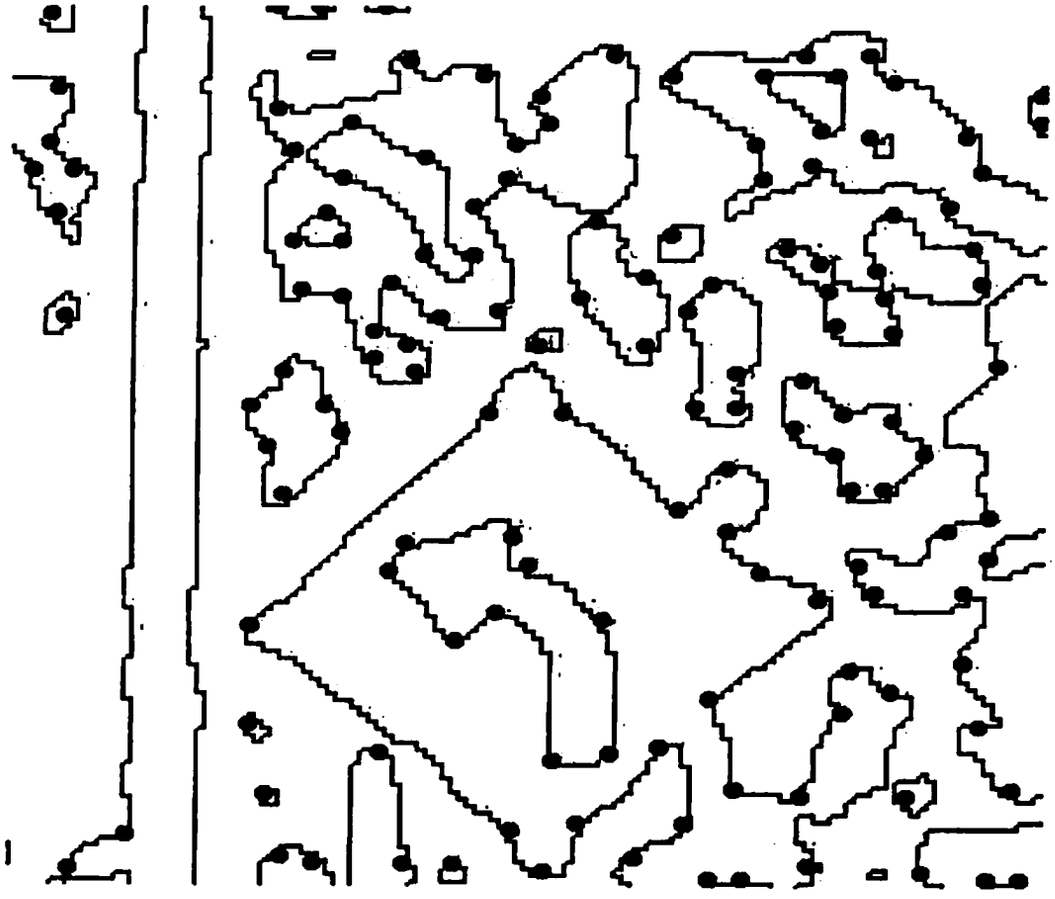


Figure 25. Interesting Points along Contours.

It would also be useful to incorporate information determined from the extraction of the translational axis to isolate false features. This could involve removing from the error measure those features which have weak matches once a translational axis has been determined, and re-evaluating to refine the FOE. Such a filtering pro-

cess would be particularly helpful when the total minimum error was not sufficiently low thereby casting doubt on the correctness or accuracy of the solution. Alternatively, the depth inferences could be used to isolate the positions of potential false features by noting discontinuities in depth along an extracted contour. Such features tend to be associated with vertices generated by surface occlusion. Such extracted features could be removed from the re-evaluation of the error measure if they are at or near such positions.

Another type of feature which can affect the evaluation of the error measure are those near an FOE or FOC which is contained in a visible portion of the image. Such features tend to move very small amounts along their image displacement paths and hence require fine interpolation to determine their best matches. The depth inference associated with such points tend to be highly erratic since their use in the inference relation from chapter III involves dividing a small number by another small number.

Properties of the Error Measure

In the experiments presented, the error measure has a distinct global minimum at the point on the unit sphere corresponding to the correct translational axis. It is generally expected to have such behavior because it is very unlikely that translational axes that are far from the correct position will define image displacement paths that simultaneously allow good matches for many features. Thus, competing candidates for the global minimum are not expected to be widely separated. This

reasoning implies strong unimodality and smoothness of the error measure over a large neighborhood and this has been confirmed empirically. Therefore, the optimization procedure used here could be replaced by other techniques which generally have faster convergence.

The error measure is affected by both non-distinctive and false features. Non-distinctive features will match well for many different translational axes. Large numbers of these weak features will flatten the response of the error measure. False features will also distort the error measure since they will often have their best matches with incorrect translational axes.

The effects of these poor features should be compensated by the agreement of good features. Every one of the good features will tend to have a bad match for the incorrect translational axis and their unanimity is expected to override the lack of discrimination of weak features and the random quality of the matches of false features. However, there is a limit in the percentage of weak and false features before the algorithm will degrade. This limit has not been explored, but our experience suggests that it may be quite high, with perhaps as many as 50 percent of the features being ineffective.

CHAPTER V

EXTENSIONS TO TRANSLATIONAL MOTION PROCESSING

Introduction

In this chapter we discuss several extensions to the translational motion procedure. We begin by formulating the computation hierarchically. This significantly increases the computational speed of the procedure and the extent of image displacements that can be processed. We then show how to process the blur paths of nearby textured surfaces when prolonged exposures are used during translational motion. We note the implications of this case, both for processing computed translational displacement fields, and for using blur to determine image displacements in general. The third extension to our algorithm considers different approaches for processing image sequences containing multiple, independently translating objects. One of these is based upon generalized Hough techniques to decompose the error measure response into the effects of the different objects. The others are based upon local application of the procedure to image subareas determined by segmentation or image subdivision. Finally, we consider the use of translational motion processing for autonomous vehicle navigation by using devices to stabilize the sensor or to obtain the rotational parameters directly.

Hierarchical Computation

A basic paradigm in computer vision is the use of hierarchical representations and processes [Burt82, Glaz83a, Glaz83b, Hans80, Tani80, Uhr78]. This allows different magnitudes and scales of image events to be handled uniformly. Additionally, the consistent agreement among hierarchically organized processes is a basic control strategy for a wide range of high and low level interpretation tasks. Hierarchical processing can produce significant computational reductions, wherein results from processing performed rapidly at lower resolutions of image information are used to direct and constrain more detailed and extensive processing of higher resolution image information.

The processing of translational motion can be developed in a hierarchical fashion with the primary benefits being increased speed and the ability to deal with larger image displacements. This requires specifying the hierarchical representations of the successive images and the extracted features, and specifying how processing at different levels of image resolution are related.

Hierarchical Representation of Images and Features

In the initial work described here, images have been represented in the VISIONS image operating cone structure [Hans80]. This consists of a sequence of images $I_0, I_1, I_2, ...I_n$, where the successive sizes of the images are $1 \times 1, 2 \times 2, 4 \times 4, ..., 2^n \times 2^n$. The value n is the level of the image in the cone. Each pixel in the i-th image, except for the first and last images, has a connected neighborhood of immediate descendants in the $i+1$ image and a parent in the $i-1$ image. The

size and shape of the immediate descendant neighborhood can be arbitrary and the immediate descendent neighborhoods of adjacent pixels may or may not overlap.

There are several ways to reduce the resolution of an image in the VISIONS cone [Hans80] and other pyramid architectures [Burt82, Tani80, Uhr78]. These techniques involve smoothing the image with some operator and then sampling at a reduced interval, or by using a reduction operator which is some function of the pixels in the immediate descendent neighborhoods. The results of reducing image resolution by averaging using Gaussian masks over 5x5 pixel immediate descendent neighborhoods at successive levels of the roadsign image 1 is shown in figures 26a-d.

The positions of extracted features can also be represented in the cone structure at different levels of resolution. There are several alternatives for doing this. First, it may not be necessary to extract features at all and simply apply the procedure uniformly to features at each position, relying on the increased speed of hierarchical computation or potential architectures to make this possible. One approach is to apply the feature extraction process for each image at each level of image resolution. Another technique is to extract features in the highest resolution image and then treat the ancestors of these in the lower resolution images to be features. In this case, the immediate descendent neighborhoods should not overlap (so each feature has unique ancestors). A feature is then positioned at a parent pixel if any of its descendants are at positions where a feature has been extracted. These approaches may interact in interesting ways if the strength of a feature is expressed as a function of its own distinctiveness and that of its descendants. We have thus far utilized the approach based upon extracting features at the highest image resolutions, though general problems with this should be noted. Features that are separated at higher resolutions become adjacent at lower resolutions. Thus, the inferred features at the lower resolutions may not be meaningful, especially since the information is not

uniform across the range of spatial frequencies represented in the different image resolutions. The benefit of this technique is that there are explicit and unique links between features at different image resolutions so that displacements determined at coarse levels can be used to initialize the estimates of displacements at finer levels.

Figures 27a-d show the features resulting for roadsign image 1 at different levels of resolution by using the feature positions determined from the highest level of image resolution (figure 8e in chapter IV) at the corresponding positions in the lower resolution images.

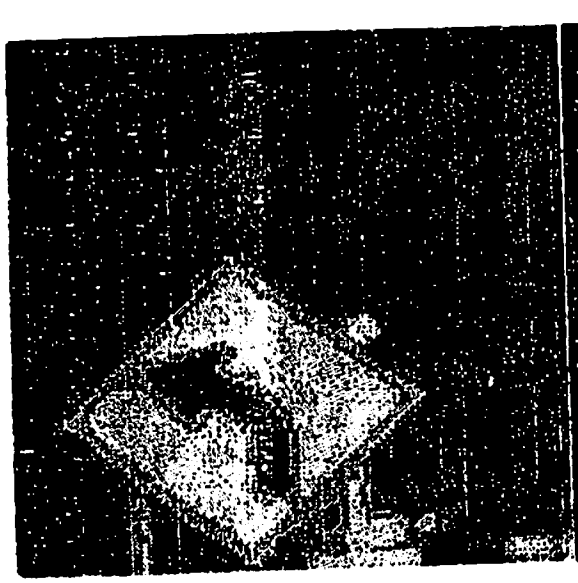


Figure 26a. 128 × 128 Resolution.

Figure 26b. 64 × 64 Resolution

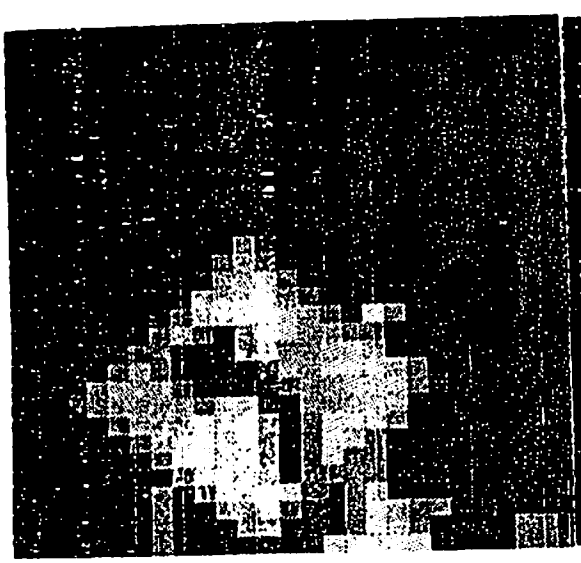


Figure 26c. 32 × 32 Resolution.

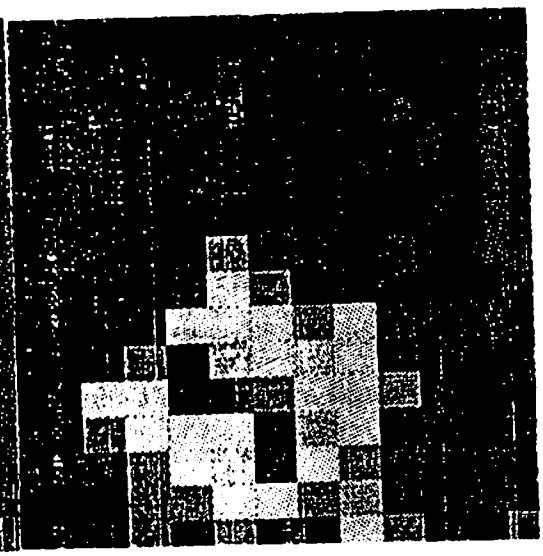


Figure 26d. 16 × 16 Resolution

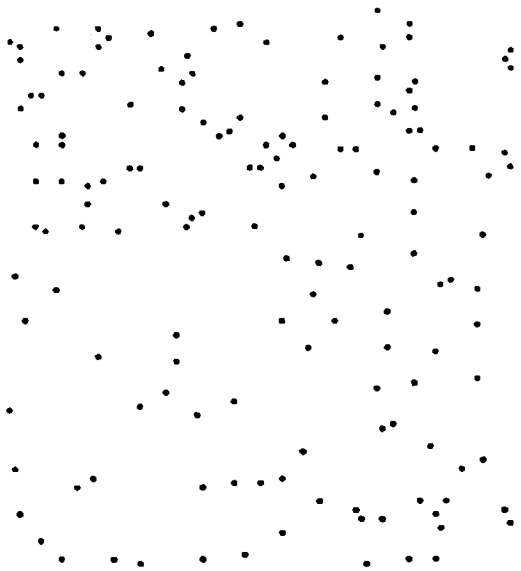


Figure 27a. 128 x 128 Resolution.

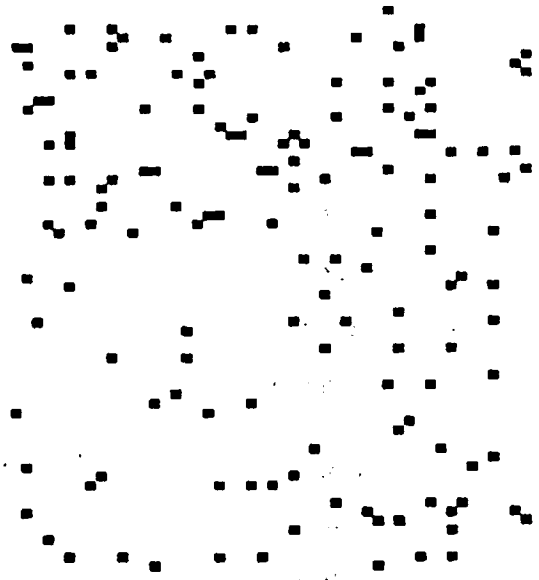


Figure 27b. 64 x 64 Resolution

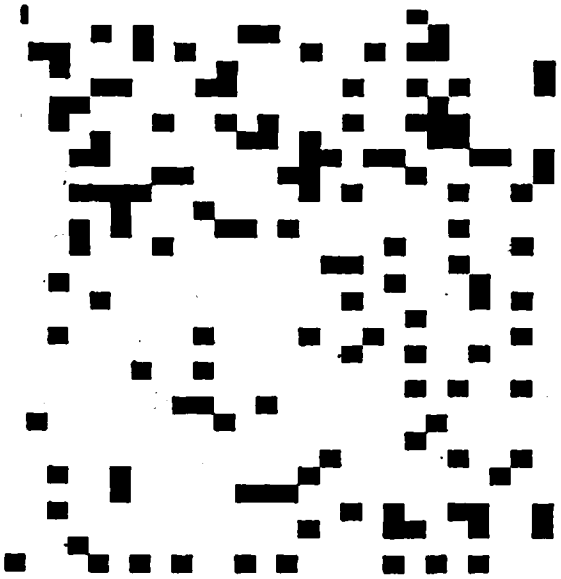


Figure 27c. 32 x 32 Resolution.

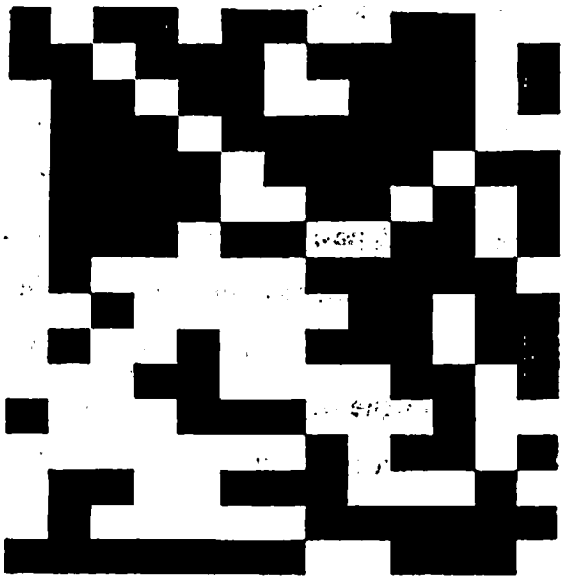


Figure 27d. 16 x 16 Resolution

Translational Processing at Different Resolutions

The translational processing can be applied to successive images at any level of resolution for which features have been extracted from the initial image. The basic questions concern how processing at one level affects processing at another level. In particular, how do processing results at a coarser level of resolution constrain the processing at finer levels of resolution? At what level in the cone can processing be meaningfully initialized? How do the various parameters involving feature window size, displacement resolution along a flow path, and resolution of the optimization procedure change at different levels of the cone?

Let us present our first effort to deal with these issues. For a given pair of images at level i in the cones formed from successive images, the translational error measure will be minimized for the set of features determined at level i (using the ancestors of features determined from the highest resolution version of the initial image). The position of the minimum error in the translational axis at level i is then used to constrain the optimization of the error function for the images and feature positions at the $i+1$ level in the cone. In addition to constraints on the position of the error function minimum, processing higher in the cone constrains the evaluation of the potential displacements of extracted features along their displacement paths. Figure 28 shows flow paths at different levels of resolution. For each displacement determined at level i only three positions have to be evaluated at level $i+1$. Thus, not only is the minimum of the error function passed on, but also the displacements of parent features which are then used to constrain the evaluation of the displacements of descendent features [Glaz83b].

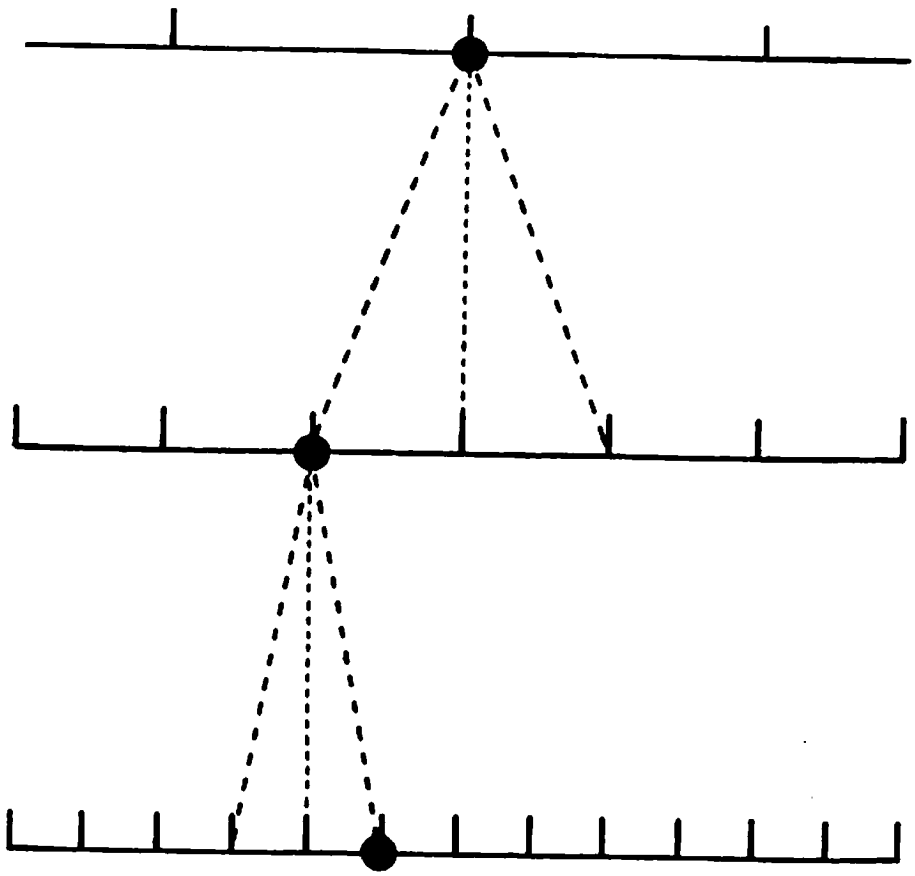


Figure 28. Relations between displacements at different resolutions.

There are a wide range of possibilities for relating the error function minimization across the different image resolutions. One strategy that has been employed involves the use of different step sizes in the error function evaluation correlated with particular image levels. That is, as processing moves to higher image resolutions, the stepsize of the error function evaluation decreases. Alternatively, a complete search could be done at a given level before proceeding to the higher resolutions. Feature size can also change as processing goes down the cone since at higher levels

a given window size corresponds to an increased area with respect to the image. At a high level of resolution, features described by small image areas may not be distinctive enough to match well.

In the experiments in figures 29a-d processing was initialized at level 4 by performing the global sampling of the error measure at the same density as the experiments in chapter IV (a separation of $\frac{\pi}{10}$ radians in the coordinate system for the direction of translation sphere). The resulting flow field is shown in figure 29a. The first step of the local processing was initialized at the minimum determined in the global sampling and used a stepsize equal to 0.1 radians for the images and features at level 5. The resulting flow field is shown in figure 29b. At level 6, the stepsize was reduced to 0.025 and the local search initialized at the minimum determined by the processing done at level 5. At level 7, the stepsize was reduced to 0.005 and the search was initialized at the minimum determined at level 6. 5x5 windows were used at each level. The procedure converged to the same results as in experiment three in chapter IV.

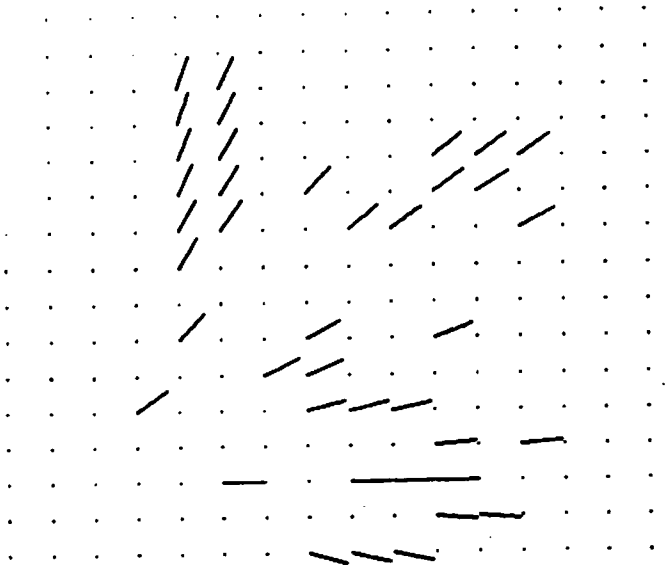


Figure 29a. Image displacements at 16 × 16 resolution

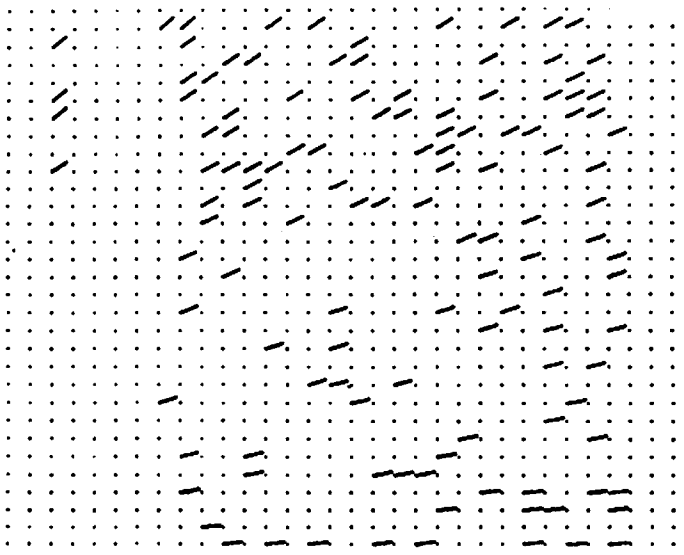


Figure 29b. Image displacements at 32 × 32 resolution

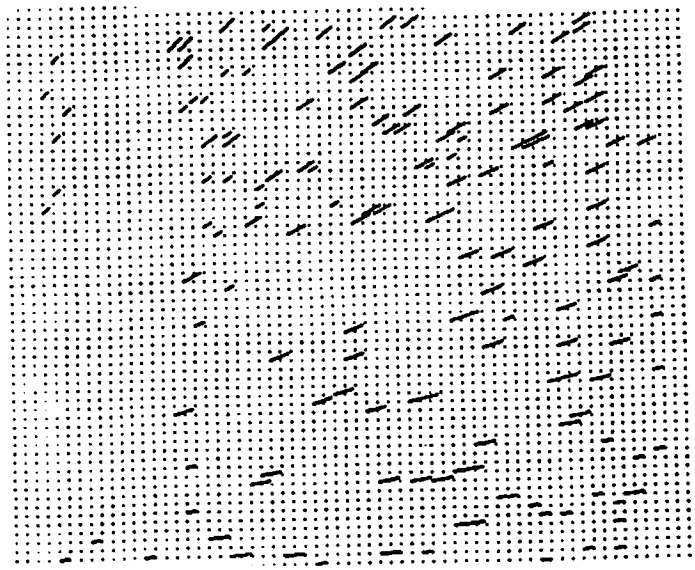


Figure 29c. Image displacements at 64×64 resolution

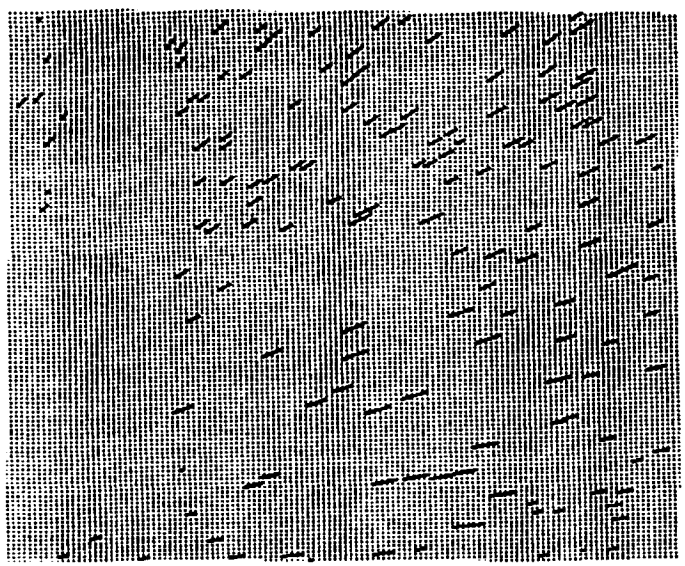


Figure 29d. Image displacements at 128×128 resolution

Some Problems

A reasonable change to the procedure described here would be the use of band-passed filtered images instead of the smoothed ones used here. Work by Burt [Burt82] and Glazer et. al. [Glaz83b] indicates that the matches of features from successive bandpassed images are much more distinctive than using features from low-pass images. Another important question which has not been addressed in any detail concerns the image level at which to begin processing. One criteria could be the level at which significant changes in image values occur as determined by an average difference value. Another could be the response of the error function. This would involve determining the level at which the error function has a distinct minimum.

A particular problem in hierarchical matching schemes occurs at occlusion boundaries. Here, features on different sides of an occlusion boundary can have a common ancestor, but will themselves have different displacements. Therefore, the displacement value inherited from the parent may be incorrect for one of the features and that feature should have its potential displacements re-evaluated along it's displacement path. A possible criterion to determine the need for re-evaluation of the displacements of a feature is if its match value is ever less than some threshold or is less than the match strength of its parent. It may be sufficient simply to not evaluate such features if they are found, and to then determine their displacements or occlusion after the more certain image displacements have been found for other image points.

Translational Blur Path Extraction

Blur streaks are commonly produced when the shutter mechanism of a camera remains open while the camera is moving relative to a textured surface. The streaks are produced by the successive positions of the image projections of the texture elements. Recent work [Harr80, Shep83] indicates that blur streaks may be a very common motion effect in the human visual system.

For translational camera motion, the blur streaks will correspond to the image displacement paths: straight line segments radiating from a common intersection point. In the analysis of translational blur paths, some information is lost concerning the direction (from an FOE or towards an FOC) and magnitude of the displacements of image points over time. Nonetheless, the techniques developed in chapter IV can be easily modified for the extraction of translational blur paths. First, it is necessary to compute the gradient of the blurred image. The image gradient will be perpendicular to the translational blur paths at positions where image blur occurs. Thus, the error measure can be expressed as

$$\sum_{i=1}^{N} \| \cos \theta_i \| \tag{15}$$

where i is an index over image positions, and θ_i is the angle between the image gradient at point i and the translational displacement path corresponding to a particular translational axis. The same evaluation techniques can be used for this error function as above, except that there is no need to distinguish between FOEs and FOCs. Thus, the evaluation of the error measure need only occur on a hemisphere. It should be noted that a variant of this error measure can be used for processing translational motion sequences for which image displacements have been determined. In this case, the image displacement vectors will lie along (not

perpendicular to) the correct translational displacement paths. The corresponding error measure becomes $\sum_{i=1}^{N} 1.0 - |\cos\theta_i|$.

The results of a preliminary experiment are shown in Figures 30-33. Figure 30 shows an image taken from a car traveling down a straight road. The shutter was kept open for a prolonged exposure and blur streaks resulted from the texture elements in the nearby tree. Figure 31a-c shows the gradient magnitude of the image and its normalized row and column components. Figure 32a-b show intensity and contour plots of the error function at points on the direction of translation sphere roughly corresponding to the potential positions of FOEs. Darker corresponds to less error in the intensity plot. In the contour plot, a " - " is used to indicate the position of a local minima of the error function and a " + " is used to indicate the position of a local maxima. The error function is unimodal due to wrap around on the direction of translation sphere because the FOEs and FOCs along a particular line of translation are not distinguished. Figure 33 shows the set of translational blur paths that were determined.

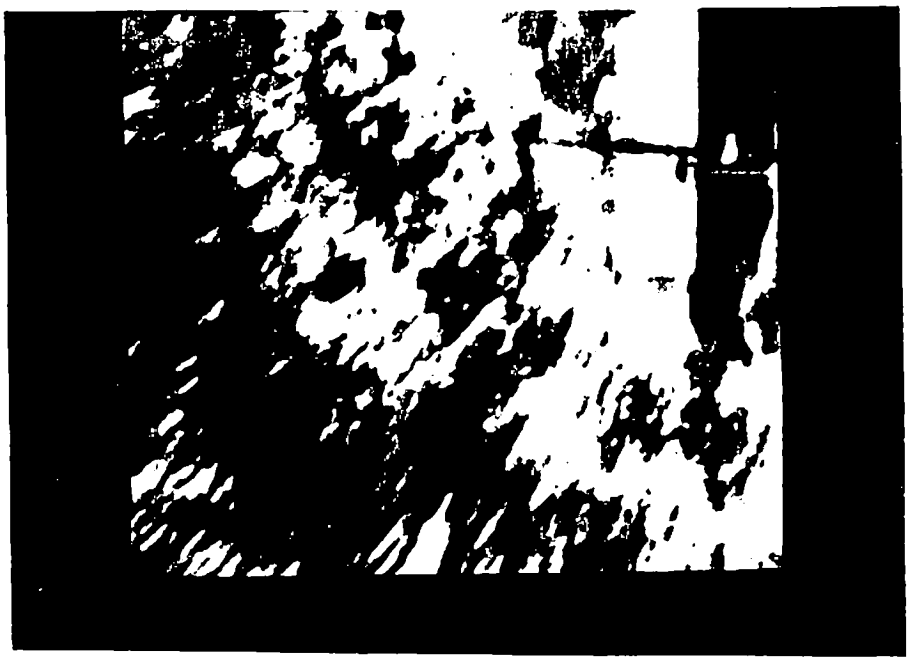


Figure 30. Blur image.

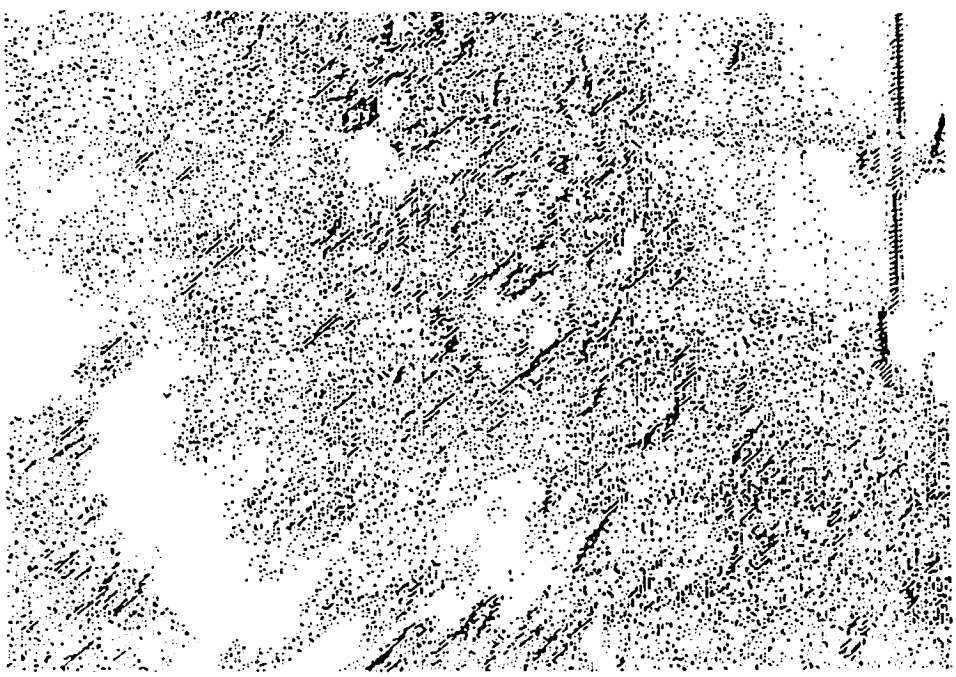


Figure 31a. Magnitude of gradient of Blur image.

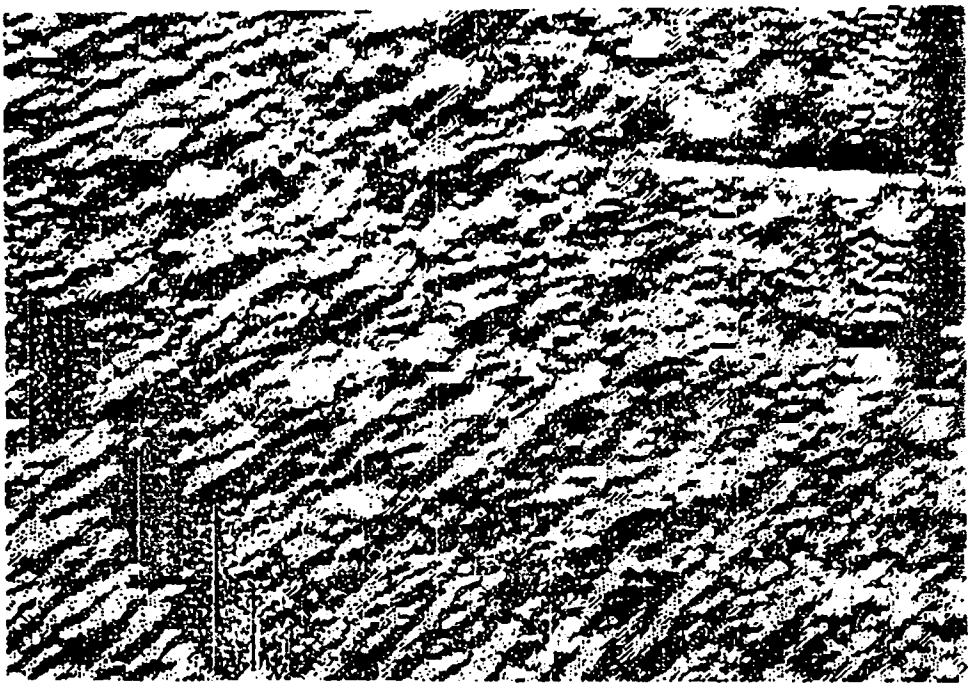


Figure 31b. Row component of normalized gradient.

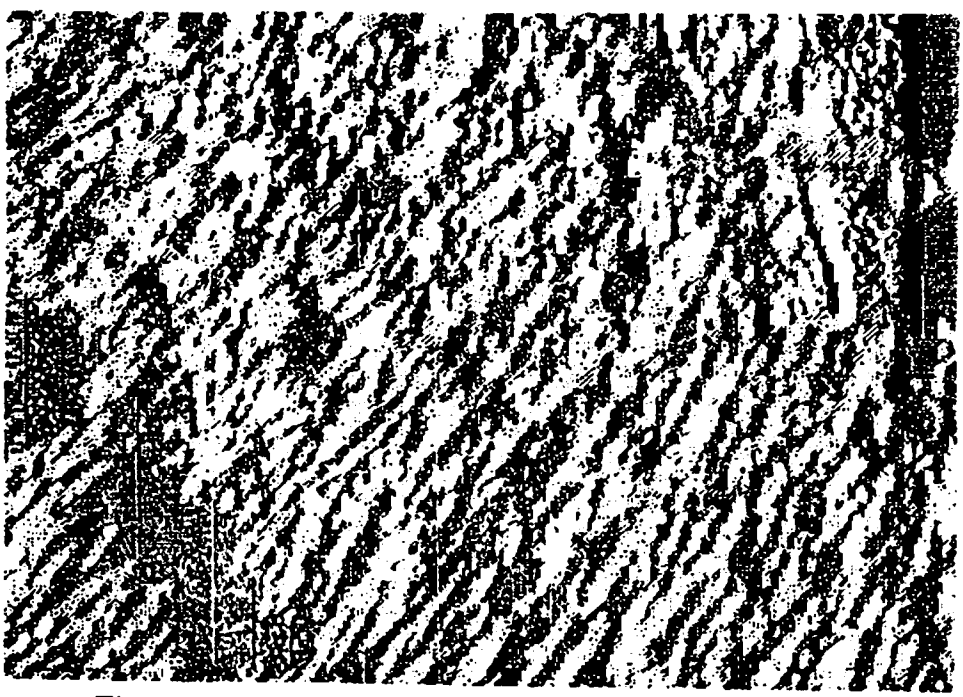


Figure 31c. Col component of normalized gradient.

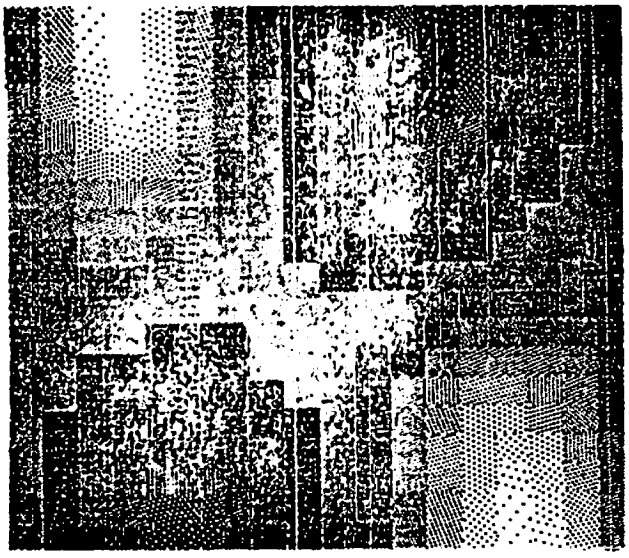


Figure 32a. Intensity plot of error function.

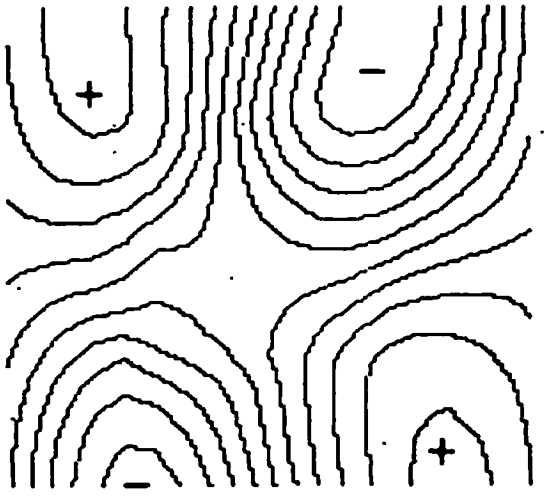


Figure 32b. Contour plot of error function.

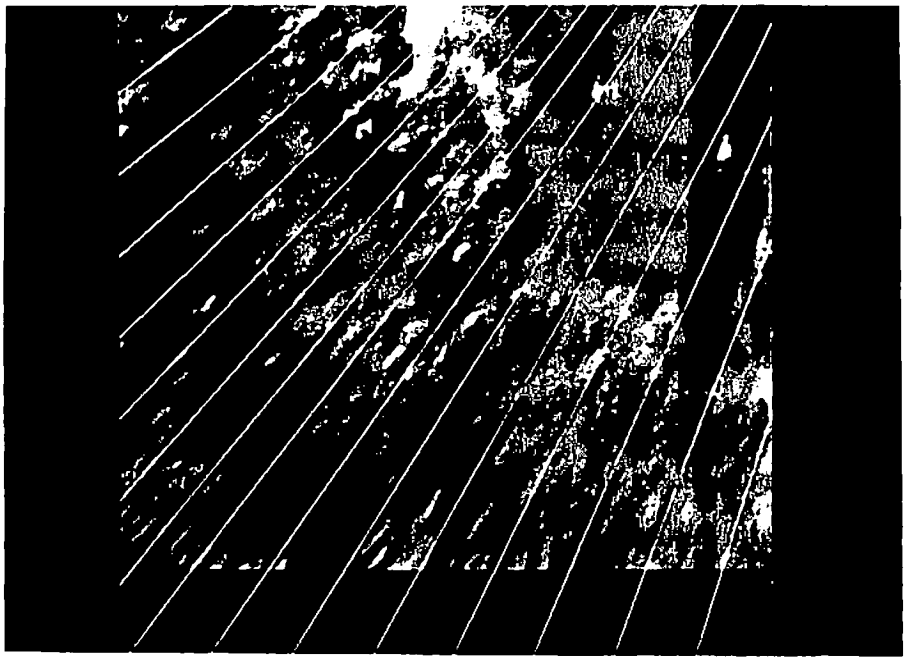


Figure 33. Determined translational blur paths.

It may be useful to use multiple versions of the same image sequence, each formed using a different exposure rate. Those formed with short exposure times would have very little blurring and their gradients would correspond to static edges. By subtracting the images formed with very short exposure rates from those formed during the same interval but with longer exposure rates, it may be possible to suppress edges in the blurred images which are non-blur related. Of more general importance in such a representation is the potential ability to relate blur streaks to the displacements of features extracted from the static images.

The extraction of translational blur paths is also similar to the extraction of vanishing points and lines from static images. The same procedure can be applied,

without the initial extraction of edges: the determination of edges can occur concurrently with the extraction of the vanishing point. However, vanishing point analysis is typically more difficult because only small portions of the image are rudely organized with respect to the potential vanishing points. Determination of these areas, or finding a way not to have the 'noise' from the rest of the image dominate the analysis, are the key difficulties. In this case, the error measure may need to be extended to incorporate information concerning edge length or connectedness along the radial paths determined by a particular vanishing point.

Approaches for Multiple Independently Moving Objects

The procedure developed here assumes a sensor moving relative to a stationary environment, or a single object moving relative to a stationary sensor. A useful extension would allow the presence of multiple, independently moving objects, while maintaining the ability to determine image displacements concurrently with the direction of translation. There are at least three techniques which could make this possible. One is to utilize generalized Hough transform techniques [Ball81, O'Rou81] for decomposing the responses in a error measure into the corresponding image structures or segments. The other two constrain the analysis to independent limited image areas over which the procedure can successfully function.

We begin by noting that the global component of the optimization process used in chapter IV is an instance of a generalized Hough transform in which each feature scales its vote against a particular translational axis as a function of the best match it can find that is consistent with the translational axis. With only a minor change, instead of using an error measure, we could use an optimization measure by which each feature scales its vote for a particular translational axis by the extent of the

best match it can find that is consistent with the axis. The problem then becomes a typical one for Hough transforms: how to associate labels corresponding to the resulting peaks in the histogram with image points or features. The general form of this processing is to find the translational axis with the greatest response in the histogram, associate a label with it, and then associate this label with image features which match above some threshold along the image displacement paths determined by the corresponding translational axis. The resulting set of features are then removed and a new histogram is produced. The peak in this new histogram and the process is repeated until there are no more distinct peaks in the resulting histograms, or all image features are labeled [Adiv83].

This procedure will have difficulties with weak or homogeneous feature points which have strong matches consistent with several distinct translational axes. Thus, when rehistogramming occurs it is necessary to establish which image features already labeled are consistent with the newly extracted peak. An alternative, is to proceed in the conventional manner and determine a set of labels corresponding to translational axes for which there is evidence. Each feature is then labeled with each translational axis from this set with which it is consistent. Note that a given feature could have several labels. A unique consistent labeling is then obtained by using other information: segmentation-grouping using other image attributes, depth consistency with neighbors, and common magnitude of image displacements. Additionally, this disambiguation can occur over several successive images. In fact, a potentially significant aspect of generalized Hough techniques may be the correlation of histograms from successive instants to bring out structures that are moving consistently.

Two basic questions have to be addressed in this use of Hough techniques: what is the required density of translational axes in the transform and what is the minimal

match threshold. In general, the higher the density, the better.

An alternative approach is to break the image into subparts and then locally apply the procedure to associate a translational axis with each subpart. In one scheme, this would be done using regular image areas (as in a grid) at multiple levels of resolution. Techniques similar to this are used in chapter seven to determine the local directions of environmental motion. In another scheme, the subparts are determined by some segmentation procedure, and the translational axis is determined from image features within or lying along the boundary of the extracted segments. Segments for which the error function response is indistinct are resegmented or their features are associated with the translational axes determined for adjacent image subparts.

Hybrid Sensor Systems

Translational processing is sufficient for vision-based navigation in a stationary environment if the orientation of the optic sensor can be fixed relative to the environment over time. In this case, sensor motion amounts to a sequence of translations in possibly different directions over time. There has been much recent work on sensor stabilization, notably by researchers at McDonnell Douglass Aerospace Corporation in suspending electro-optical systems in a magnetic field, and elsewhere using more conventional gimbel-based stabilization.

A difficulty with such a stabilized retina is that it is not able to rotate to focus on particular parts of the environment. This can be corrected by using a set of such stabilized retinas arranged to yield a complete view of space. There would then be no need to rotate the sensor to view a particular environmental point. A possible arrangement of retinal surfaces is a cubical one. One of the retinal planes

will always contain an FOE and another will always contain an FOC (unless the direction of motion is right on an edge of the cube and the focal length has not been properly adjusted). There will also be several independent estimates of the direction of translation which can be integrated. Figure 34 shows such a proposed arrangement of optic sensors attached to a Cartesian robot manipulator so such a complete, stabilized view of a workspace is produced at all times.

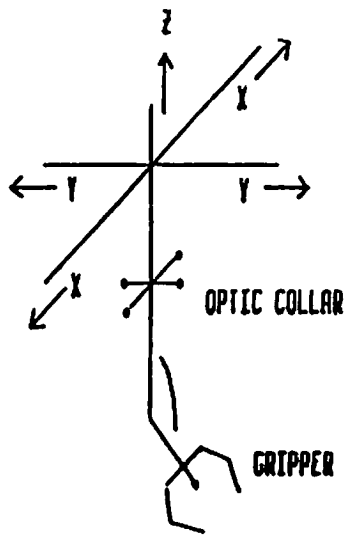


Figure 34. Cartesian Manipulator with attached optic devices.

Alternatively, if the sensor cannot be stabilized, there are other devices which can at least determine the rotational parameters of sensor motion. The rotational effects can then be removed from successive images, reducing them to translational sequences which can be processed by the techniques here. A particular technology which is very attractive for this use is that of fiber optic rotation sensors [Esek82] (figure 35). These sensors are expected to be the low-cost gyroscope of the near

future since they are small, cheap, and precise. Because they have no moving elements, they are not as affected by rapid accelerations as conventional gyroscopes. There are currently slow drift problems when sensor orientation is considered over long periods of time. In our processing though, we would be concerned with measurements of rotation over much shorter periods. Additionally, when such sensors are coupled with an image processing system for guidance and navigation, the effects of such long term drifts could be recognized and accounted for by noting the position of specified landmarks.

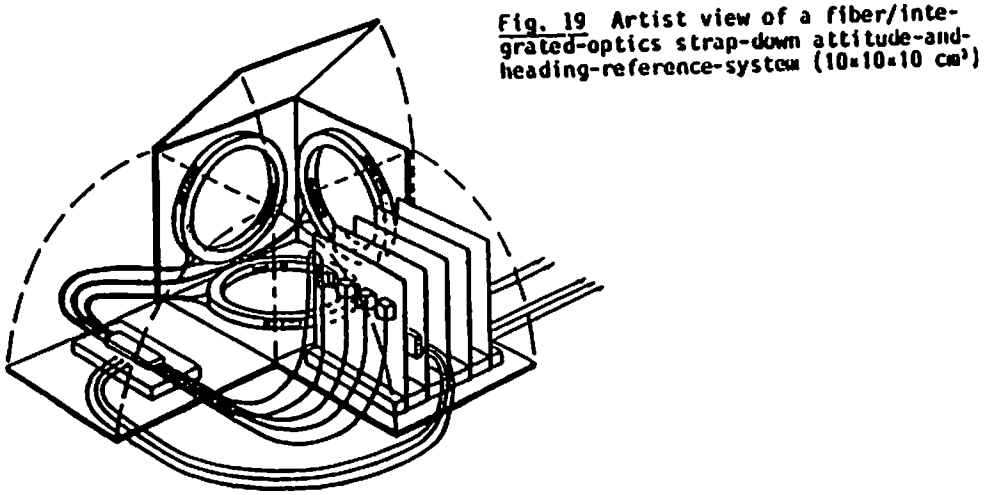


Figure 35. Layout of Fiber Optic Rotation Sensor (from [Ezek82]).

CHAPTER VI

PROCESSING RESTRICTED SENSOR MOTION

Introduction

The techniques used for translational motion can also be applied to other cases of restricted motion. The issue is the computational feasibility of a search through a subspace of sensor motion parameters for values that are consistent with image feature displacements. In this chapter we briefly consider two such cases, pure sensor rotation and motion constrained to a known plane.

Processing Pure Sensor Rotation

For processing pure sensor rotation, the error measure can again be defined with respect to a unit sphere with each point corresponding to an axis and a direction of rotation. We use the (ϕ_1, ϕ_2) coordinate system from chapter IV for referring to these positions. In addition to these two parameters for specifying an axis of rotation, there is a third corresponding to the extent of rotation. The extent of rotation is defined relative to the orientation of a given axis and encoded with positive values denoting rotation in a clockwise direction. Thus, on the unit sphere the points (x, y, z) and $(-x, -y, -z)$ will lie along the same axis of rotation but correspond to different directions of rotation.

As in the case of translation, we utilize the error of matches of selected features along their respective image displacement paths. However, there are a few basic differences with the translational procedure. First, feature displacements are not

measured in image units, but in the extent of angular displacement about the axis of rotation. Second, the displacement along the image displacement path can cause significant reorientation and expansion in a feature, especially for large rotations. For this reason, each pixel of the feature array has its position interpolated independently (figure 36). If motion is restricted to small rotations only, this may not be necessary.

For a rotational field, the extents of angular displacements for all the features must be identical. This yields a constraint which can be incorporated into the evaluation of a particular set of rotational parameters in different ways. The evaluation can be done as in the translational case where the best match of each feature along its displacement path is determined independently of the other features. This results in two different error measures: one based on the summed error values of the best matches and the other based on the variance of the extent of displacements corresponding to these matches. Alternatively, the feature displacement determination can be restricted such that they all evaluate the same extent of displacements simultaneously.

We have tried these three error measures on a simple image pair and found that they all give roughly the same result. The variance of the extent of displacements was minimized at the correct value, but was very jagged and rough. The summed error values for the best matches and the direct 3-D search were very smooth and had a distinct global minimum in a very large neighborhood.

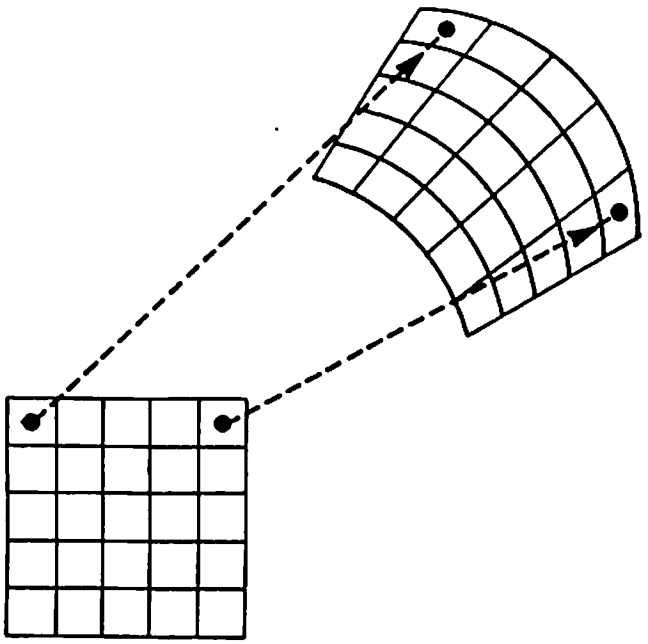


Figure 36. Determining Individual Pixel Displacements of a Feature.

Figure 37a and 37b show successive images formed with the image generation system MOVIE BYU and are referred to as the House Sequence 1. The motion was a rotation of 2 degrees (0.035 radians) about the $(0,-1,0)$ axis. The field of view was 45 degrees. Image contours for application of the interest operator were determined by a threshold selection algorithm which produces boundaries with maximum average contrast [Kohl81]. The resulting contour and the extracted features are shown in figure 37c. The interesting points were extracted by finding the local maxima in the distinctiveness measure values which were also greater than a minimal threshold. Both the features and the neighborhoods over which local maxima were determined were 3x3 pixel areas. This small neighborhood size caused the feature extraction process to be sensitive to the notches along the contours as can be seen by the number of extracted features along the bush boundary. Figure 37d shows the displacements determined for these features.

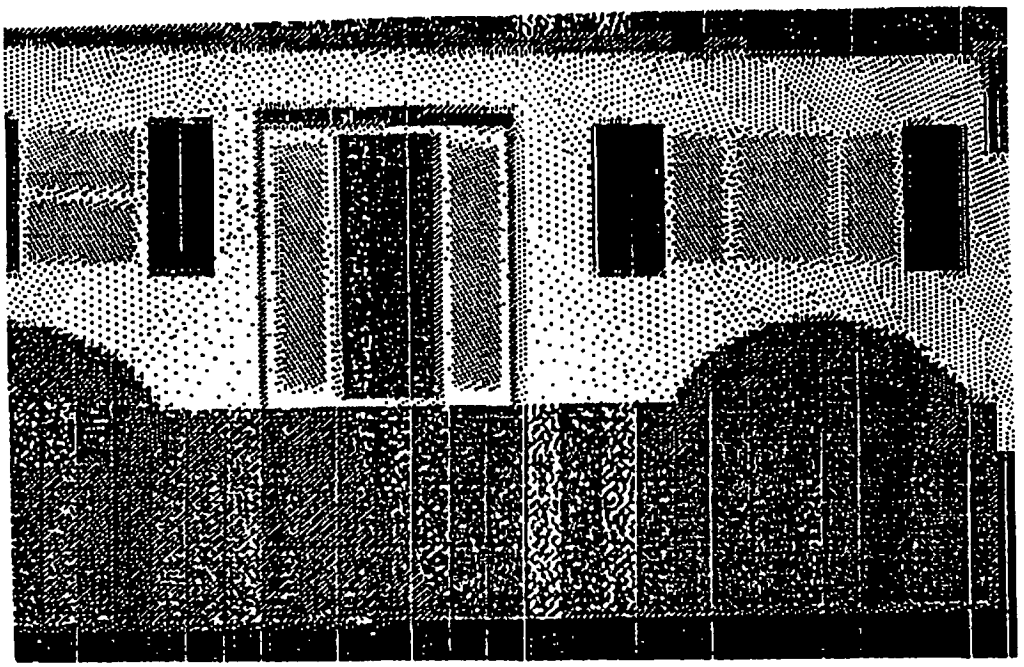


Figure 37a. House Sequence 1 Image 1

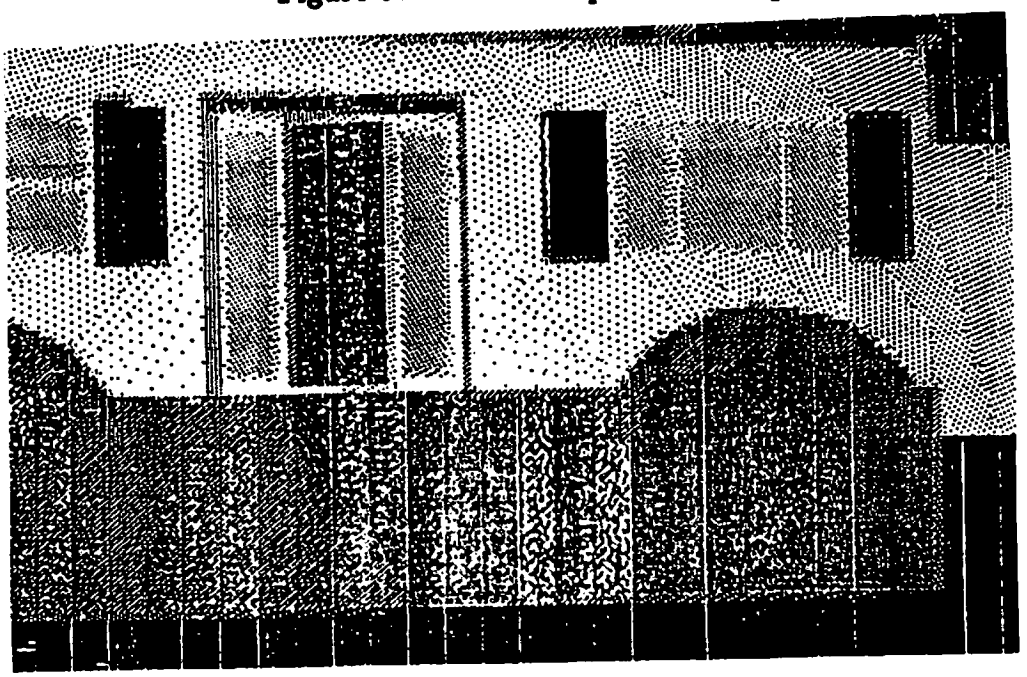


Figure 37b. House Sequence 1 Image 2

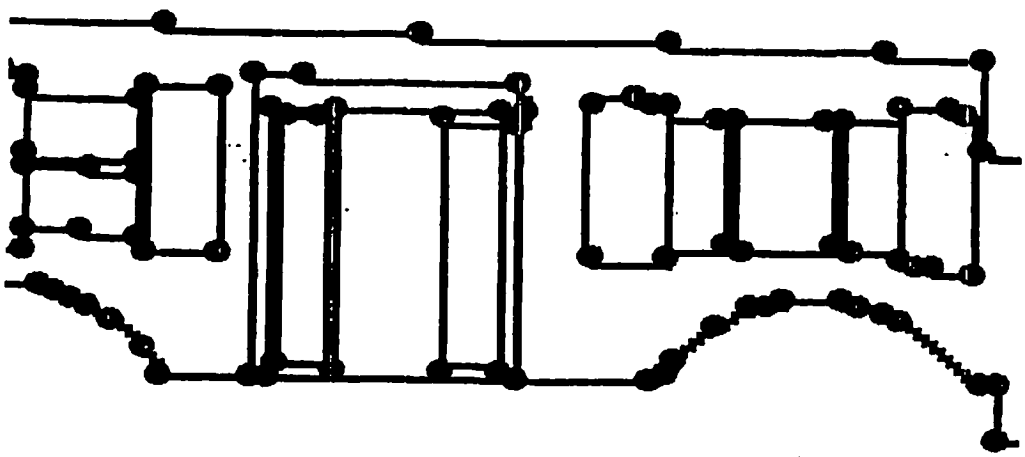


Figure 37c. Extracted Contour and Features.

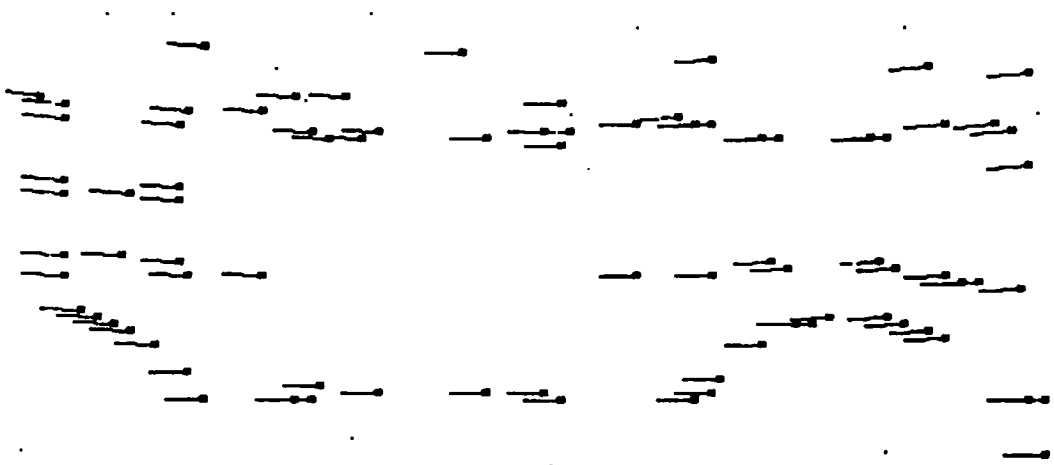


Figure 37d. Determined Displacements

The evaluation of the error measure based on the extent of feature mismatch is presented as in chapter IV using the (ϕ_1, ϕ_2) values in two tables. The first table (table 6a) basically corresponds to those axes of rotation on the positive Z portion of the unit sphere. The second (table 6b) basically corresponds to axes on the negative Z portion. Axes for which Z is equal to zero and Y is positive are represented in the first row of the first table while axes for which Z is equal to zero and Y is negative are represented in the first row of the second table. The tables are shown as intensity plots in which darker corresponds to less error and also as contour plots in figures 38a and 38b. There is a distinct global minimum at the position corresponding to the $(0, -1, 0)$ axis. Nearly all the features had displacements corresponding to a rotation of 0.035 radians for this axis. This was also the best axis and extent of rotation determined by the local search using the extent of feature mismatch for features restricted to evaluating the same displacements simultaneously.

	0	1	2	3	4	5	6	7	8	9
0	6.407	6.966	6.252	5.557	5.241	5.384	5.208	5.252	5.810	6.377
1		7.004	6.507	5.961	5.472	5.479	5.430	5.577	6.096	6.391
2		7.064	6.985	6.606	6.387	6.258	6.285	6.283	6.364	6.441
3		7.103	7.229	7.418	7.725	7.684	7.359	6.981	6.593	6.404
4		6.658	7.081	7.457	7.940	8.125	8.006	7.128	6.539	6.298
5		6.294	6.083	6.008	5.879	6.009	5.870	5.862	5.823	6.057
6		5.985	5.496	5.109	4.159	3.465	3.775	4.809	5.401	5.676
7		5.666	5.094	4.189	2.767	1.586	2.219	3.913	4.935	5.492
8		5.566	4.753	3.762	2.347	0.998	1.966	3.327	4.632	5.444
9		5.473	4.485	3.647	2.170	0.611	1.944	3.279	4.399	5.386

Table 6a. House Sequence 1 Error Values

	10	11	12	13	14	15	16	17	18	19
0	6.311	5.415	4.457	3.352	2.057	0.469	2.140	3.651	4.541	5.492
1		5.475	4.791	3.664	2.306	0.634	2.272	3.775	4.861	5.576
2		5.641	5.175	4.525	3.109	1.318	2.660	4.507	5.286	5.744
3		5.907	5.572	5.277	4.613	3.949	4.283	5.392	5.710	5.978
4		6.540	6.677	6.410	6.549	6.431	6.506	6.427	6.400	6.334
5		6.594	6.860	7.159	7.558	7.741	7.859	7.367	7.023	6.816
6		6.598	6.723	6.662	6.702	6.575	6.873	7.040	6.999	6.862
7		6.530	6.325	5.925	5.646	5.568	5.940	6.309	6.698	6.878
8		6.444	5.943	5.399	5.184	5.285	5.351	5.821	6.371	6.908
9		6.390	5.736	5.187	5.149	5.366	5.242	5.515	6.222	6.926

Table 6b. House Sequence 1 Error Values

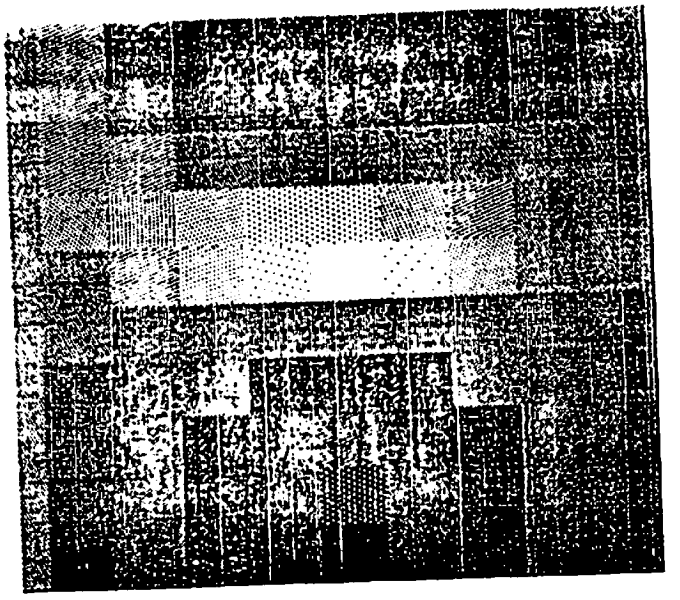


Figure 38a. Intensity plot of Table 6a

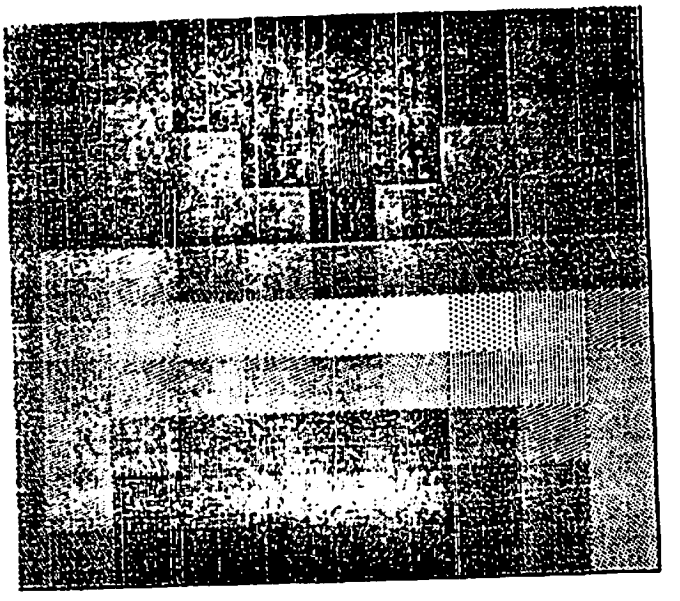


Figure 38b. Intensity plot of Table 6b

Motion Constrained to a Known Plane

If motion is constrained to a known plane, the translational axis must lie on a plane perpendicular to the rotational axis which contains the focal point. Therefore, the FOE/C in the images are restricted to lie along the line determined by the intersection of this plane and the image plane. There are two parameters to recover: the extent of rotation about the axis that is perpendicular to the plane at the focal point, and the position of the translational axis in this plane. Both of these are expressed as angles: θ_1 for the extent of rotation and θ_2 for the orientation of the translational axis (figure 39a).

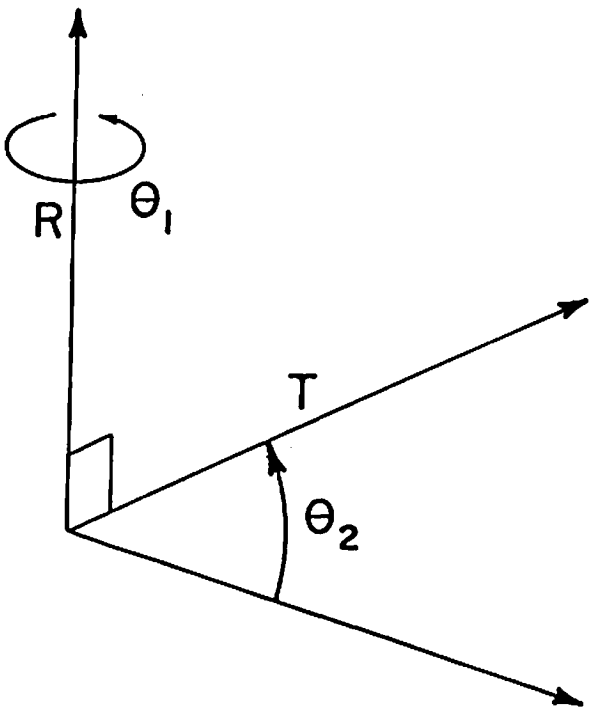


Figure 39a. θ_1 , θ_2 parameters for describing planar motion.

The error measure for this case combines the computation for rotation and translation. For the rotation and translation corresponding to particular (θ_1, θ_2)

values, a feature is first positioned along its rotational displacement path using bilinear interpolation for each pixel and then displaced along the translational displacement path at equal increments to determine its best match. As in translational processing, the interpolation for individual pixels is not performed for the translational displacement (figure 39b). The minimal match errors for each feature are then summed. The error function in this case can be thought of as being mapped on a cylinder with the θ_2 parameter, corresponding to the direction of translation, wrapping around.

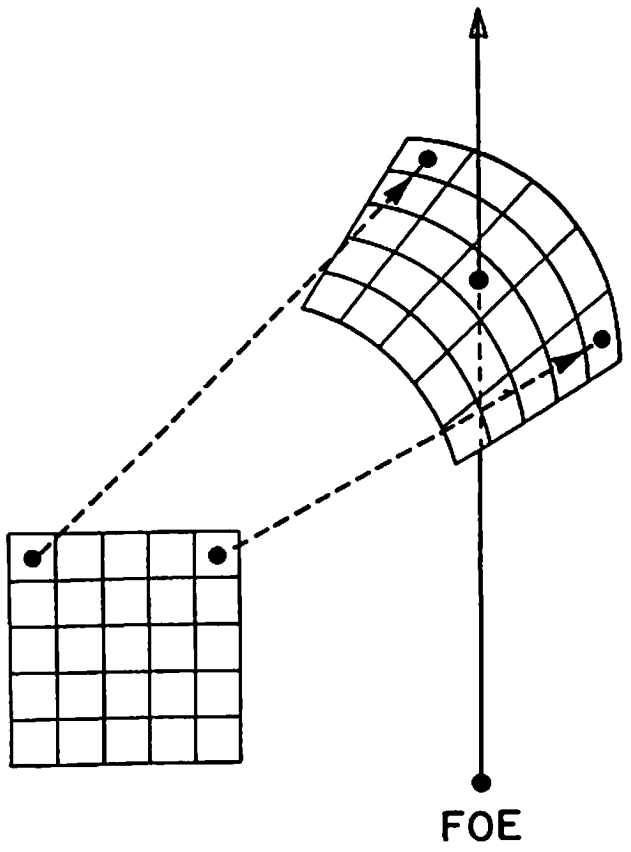


Figure 39b. Evaluation of image displacements corresponding to θ_1, θ_2 values.

Figures 40a and 40b show the grass sequence 1. The image in figure 40b of sample grass texture was produced from figure 40a by rotating 0.1 radians about the $(0,0,1)$ axis and then translating along the $(0,1,0)$ axis. Figure 41a shows 50 points which were selected at random from image positions where contrast exceeded a minimal value. Figure 41b shows the displacements determined for these points. Figure 42a shows the resulting error function in terms of θ_1 and θ_2 coordinates as an intensity plot. Figure 42b shows the error function as a contour plot with " - " indicating the local minima and " + " indicating the local maxima. θ_1 ranges from -0.15 to 0.15 radians in 0.01 radian increments. θ_2 ranges from 0.0 to $2 \times \pi$ radians with 0.0 corresponding to the position of the translational axis at (-1,0,0). The minimum in the error function corresponded to the correct values of the rotation and translation.

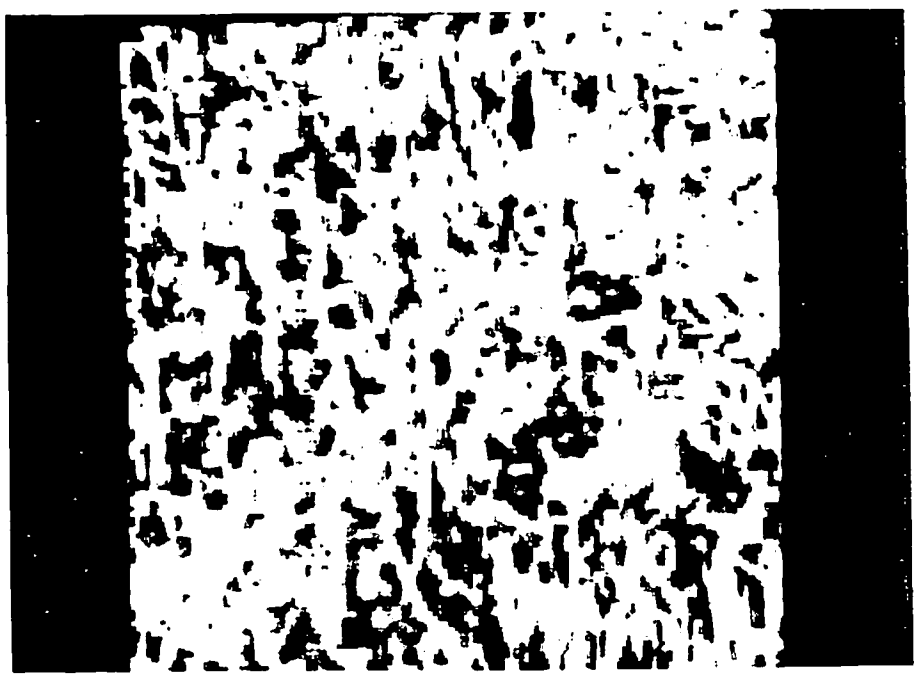


Figure 40a Grass Sequence 1 Image 1

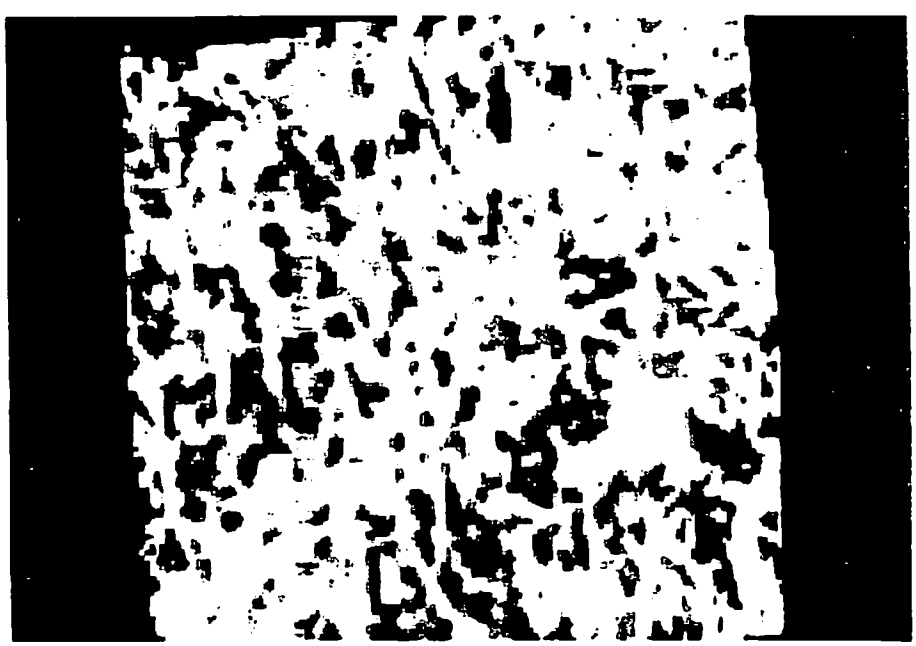


Figure 40b Grass Sequence 1 Image 2

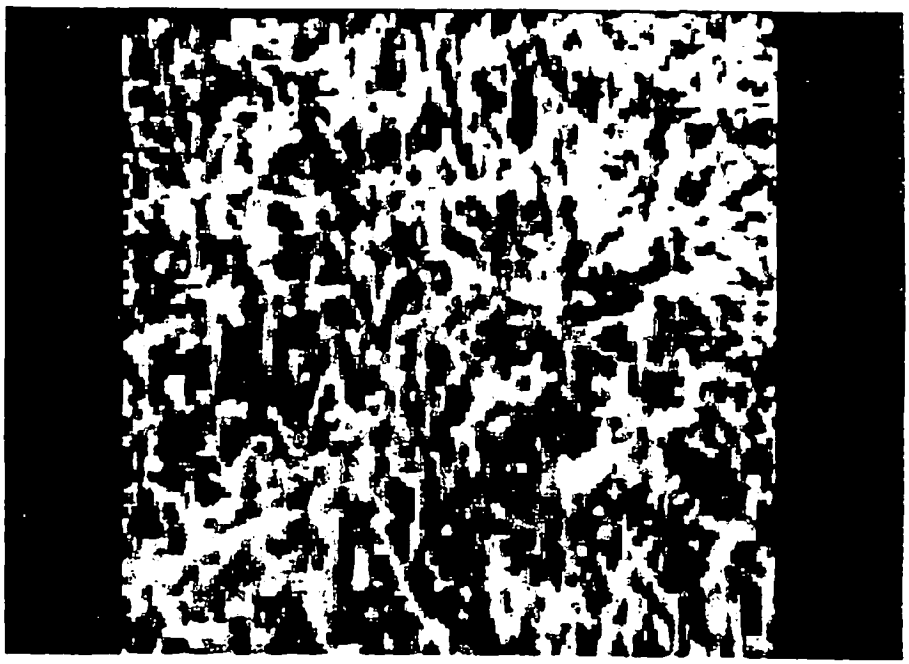


Figure 41a. Selected Features.

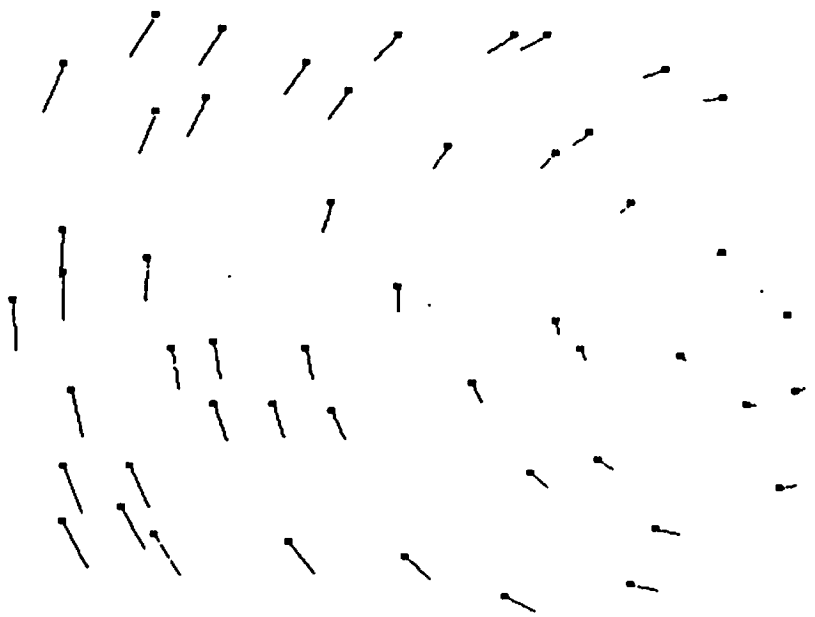


Figure 41b. Determined Displacements.

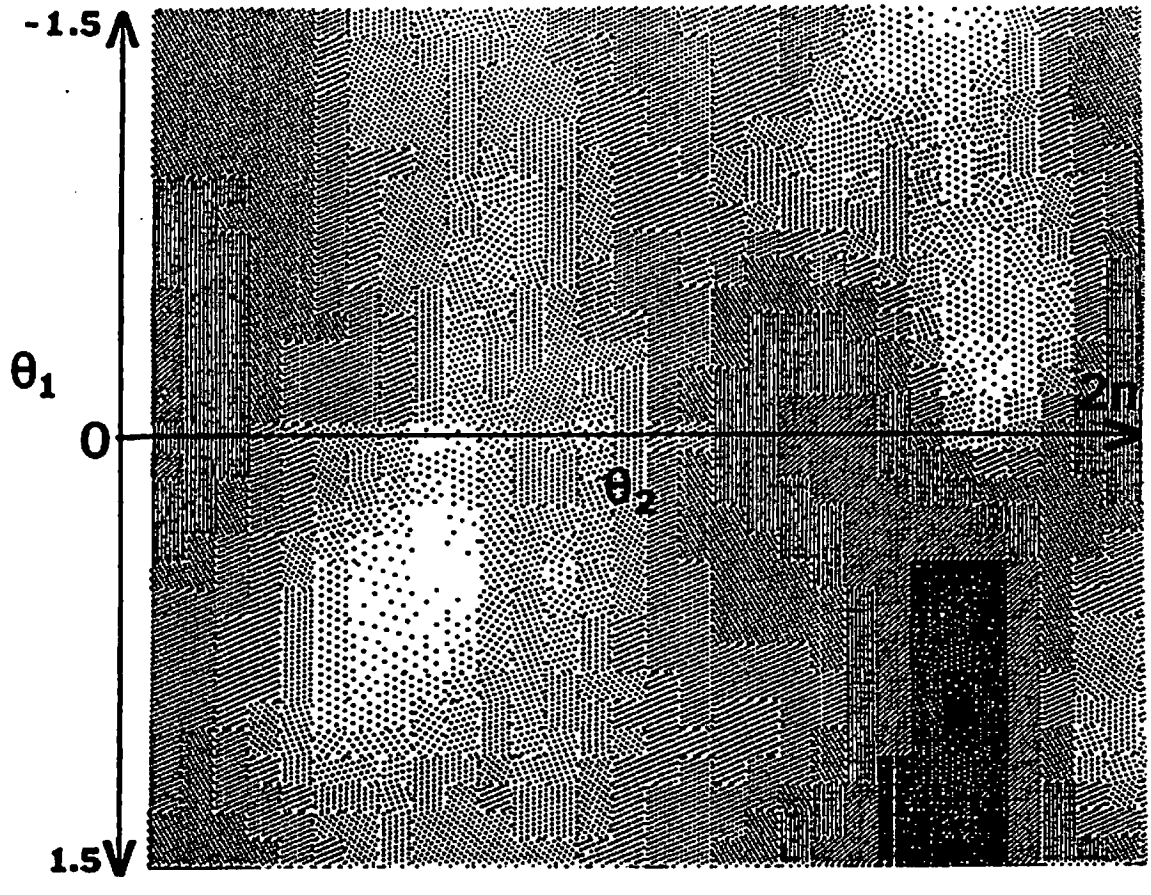


Figure 42a. Intensity plot of Error Measure.

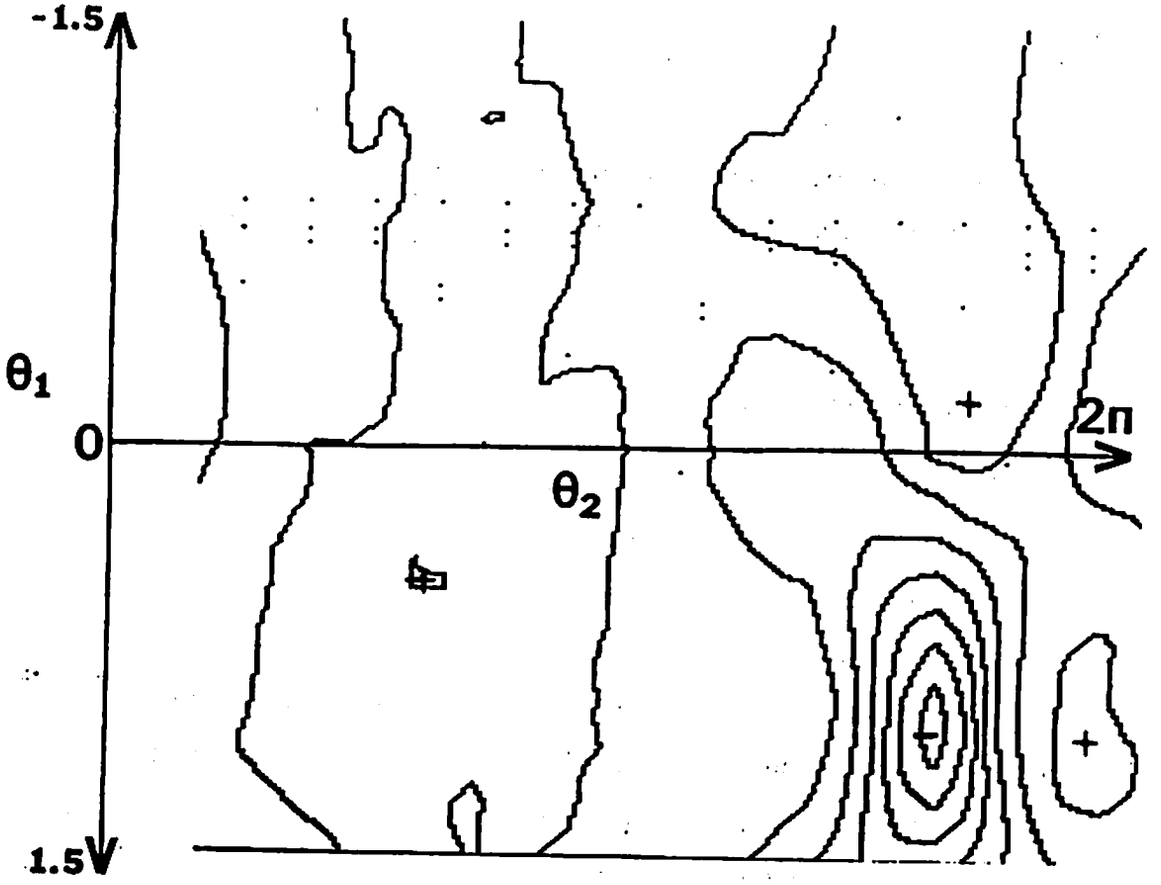


Figure 42b. Contour plot of Error Measure.

Known Planar Motion with Determined Image Displacements

To process known planar motion for image sequences for which image displacements have been computed, we use the error measure based on the properties of composite image motions discussed in chapter III to describe the consistency of a given set of image displacements with particular values of θ_1 and θ_2.

Referring to Figure 6b in chapter III, for a given image displacement from image point I_{mi} to I_{ni}, its consistency with particular values of θ_1 and θ_2 is determined by first applying the rotation specified by θ_1 to obtain a displacement from I_{mi} to J_{mi} (figure III.6b). The angle between the vector $J_{mi} - I_{ni}$ and the translational displacement path line determined by the FOE/C corresponding to θ_2 and J_{mi} reflects the degree of consistency. We actually use one minus the cosine of this angle. By summing these values for a set of image displacements, the consistency of the entire field is determined.

This procedure has to be extended slightly to deal with pure rotations. In this case, the difference vector between the image displacement vector and the correct rotational displacement vector will be quite short and behave erratically with respect to the determination of the angle with the corresponding translational field line. Pure rotational fields have two properties which we utilize to detect their occurrence. First, when rotational fields having the same axis but different extents of rotation are subtracted from each other, the variance of the length of the difference vectors tends to be small. Secondly, the correct rotational field will minimize average length of these difference vectors. Thus, a purely rotational field is indicated when the variance of the length of the difference vectors is small with respect to one of the rotational fields generated by the axis of rotation corresponding to the known plane of motion, or the average length of the difference field is small. The

correct extent of rotation is that which minimizes the total length of the difference vectors.

Ambiguities in Planar Motion

We have noted an ambiguity that occurs in the case of motion constrained to a known plane when the focal length is relatively long and the axis of rotation is roughly parallel to the image plane. In this case, the rotational component field is very similar to a translational field with the FOE/C at infinity in the image plane. The extent of displacements are also nearly identical. The effect of this is to displace the translational component by some amount proportional to the direction and extent of rotation. As a result, the composite field looks like a translational field which could result from a wide range of translations and compensating rotations (figure 43). The effect of this on the error measure is a trough of low error values.

Figures 44a-b are successive images formed using MOVIE BYU and are referred to as House Sequence 2. 44a is identical to 37a while 44b was generated by translating along the (0,0,1) axis after the rotation shown in images 37a and 37b. Figures 45a and 45b show the error measure with θ_1 ranging from -0.05 to 0.05 radians and θ_2 ranging from 0.0 to $2 \times \pi$. The trough of low error values is apparent.

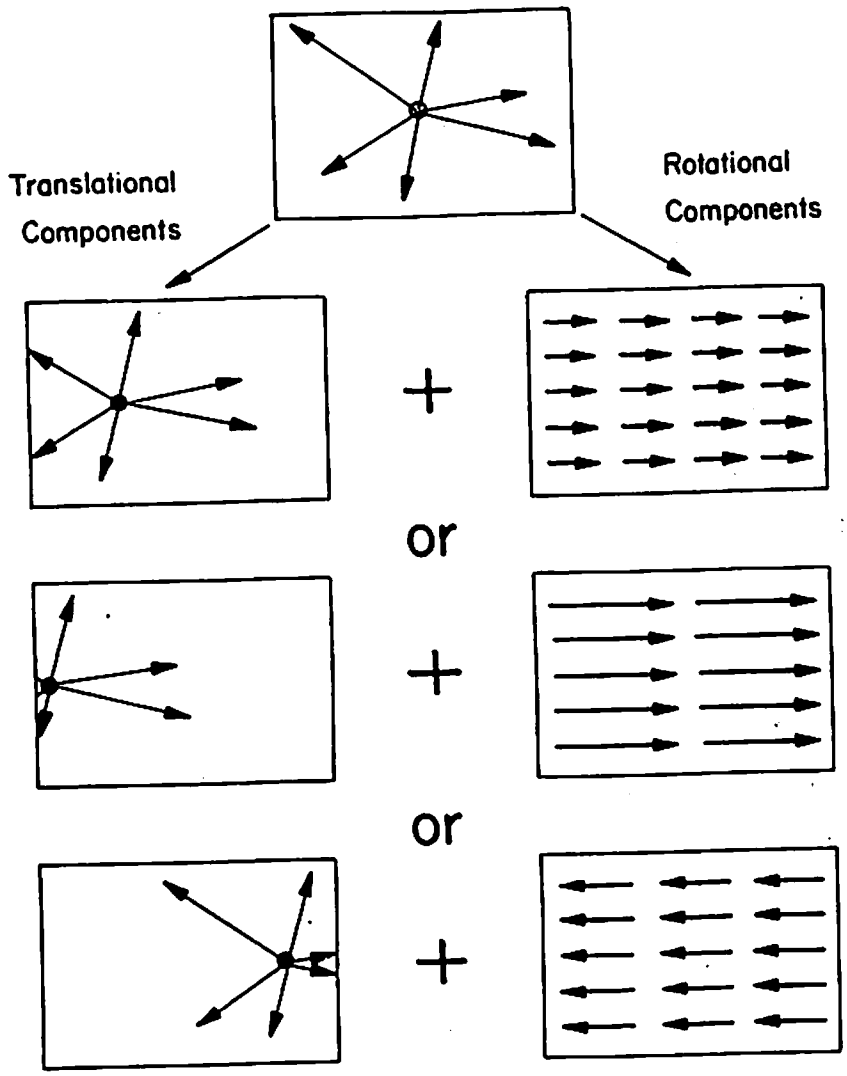


Figure 43. Ambiguity in a case of Planar Motion.

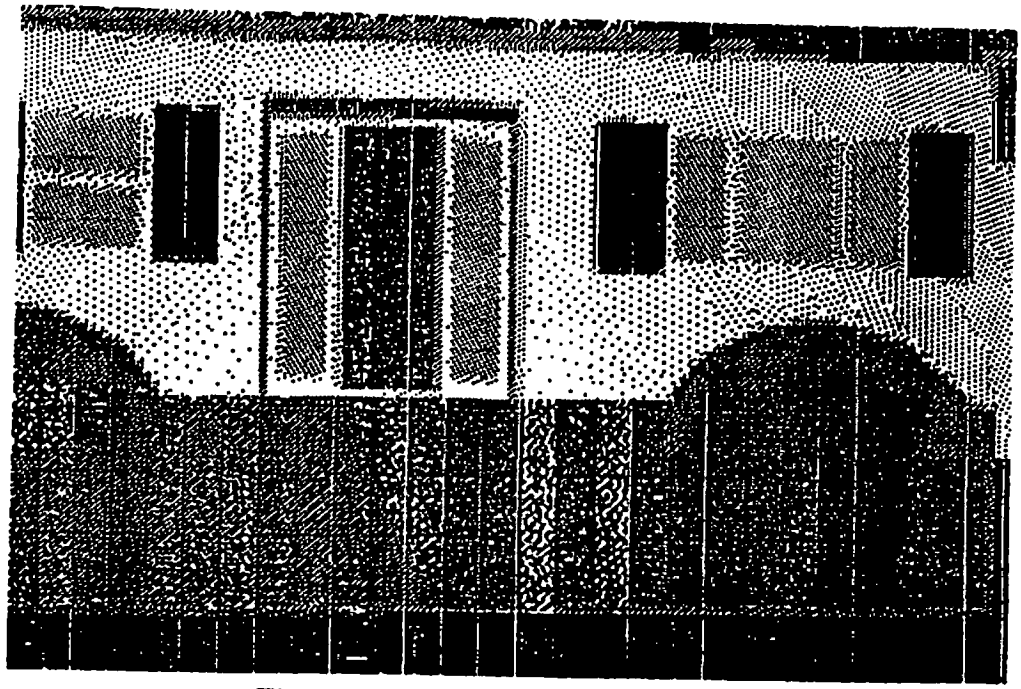


Figure 44a. House Sequence 2 Image 1

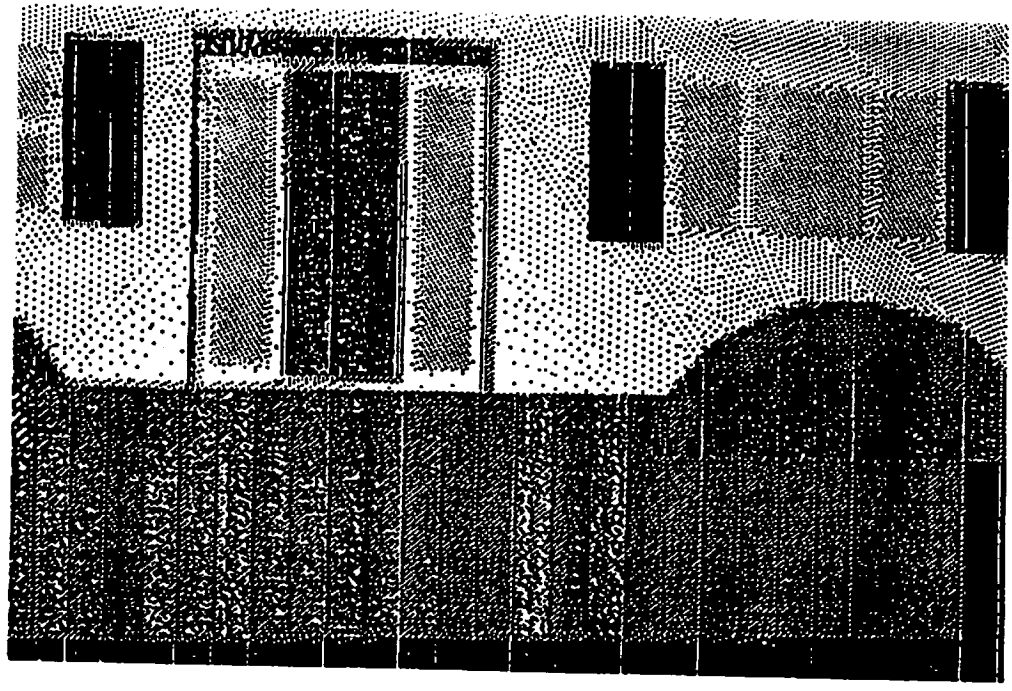


Figure 44b. House Sequence 2 Image 1

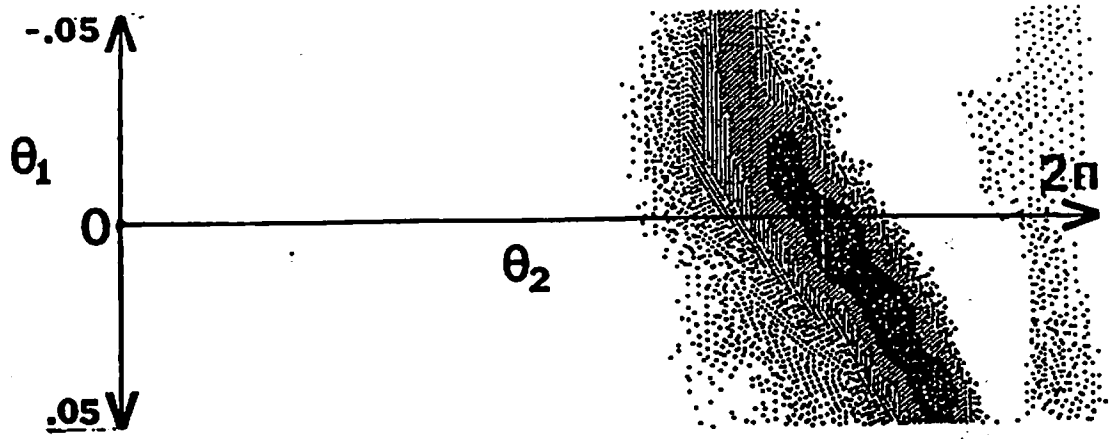


Figure 45a. Intensity plot of Error Measure.

Figure 45b. Contour plot of Error Measure.

Discussion

All of the extensions discussed for translational processing – hierarchical processing, blur path extraction, independently moving objects — should be directly applicable to the pure rotational case. There are some specific differences however. The blur path extraction is more complex in the rotational case because the structure of the image displacement paths are conics instead of straight lines; the necessary expression for the tangents to the image displacement paths in the rotational case were derived in chapter III. While independently moving objects may not frequently move in trajectories corresponding to rotation about an axis positioned at the focal point, there is a related phenomena which may be of some use in decomposing arbitrary motion. The image displacements of very distant, stationary objects or environmental features (like the horizon, the moon, the stars) will primarily be a reflection of the effects of the rotational sensor motion. Thus, if image features whose displacements are dominated by rotational motion could be detected, the rotational parameters could be extracted, the image corrected, and the translational parameters inferred by the procedures in chapter four.

These extensions should also be applicable to the case of pure planar motion though with some complications. The blur paths are more difficult to characterize in the planar case. The error function response also seems to have large flat areas which would especially affect the processing of planar motion in restricted portions of an image. Finally, the cases for which planar motion is ambiguous would be serious for any of the discussed extensions and may require processing over several frames.

CHAPTER VII

THE LOCAL TRANSLATIONAL DECOMPOSITION

Introduction

In this chapter we utilize the procedure for translational motion to process image sequences produced by other classes of restricted and arbitrary sensor motion. This is accomplished via application of the translational procedure to small image areas. This approximates more general motion as an array of local environmental translations, and interprets local image motions as if they resulted from translational motion of the corresponding portions of the environment. The feasibility of this approach was demonstrated in chapter IV where the direction of translation was extracted with reasonable precision from small image areas containing a few features. The resulting description of motion is an approximation to what we term the Environmental Direction of Motion Field (EDMF) which associates with a set of image points (or small image areas) the relative direction of motion of the corresponding environmental points (or small environmental surface areas). This is a low level representation of environmental motion which considerably simplifies the recovery of the sensor motion parameters.

This chapter consists of four parts. The first considers computing the Environmental Direction of Motion Field when image displacement vectors have or have not been initially computed. The second section describes EDMF properties for different cases of sensor motion. In the third section, these properties of the local translational decomposition are used to process image sequences produced by sensor motion constrained to an unknown plane in textured environments. In the

fourth section, we develop a set of equations for environmental depth inferences from image displacements based upon an assumption of environmental rigidity. We then show how these equations may be solved using the EDMF.

Computing the Environmental Direction of Motion Field

The Environmental Direction of Motion Field (EDMF) is a low level description of environmental motion which associates with each feature, or small image area, a three dimensional unit vector describing the direction of motion of the corresponding feature (or small surface area) in the environment relative to the observer. In the continuous case, the EDMF can be thought of as a description of environmental motion where only the orientations of tangents along the environmental displacement paths are known. We consider first how to compute the EDMF and then how it can be used to recover sensor motion parameters and environmental depth.

Analysis of Raw Image Sequences

The procedure for translational motion described in chapter IV yields a set of image displacements consistent with a determined translational axis. Application of this procedure to a small area of an image containing extracted features will yield a set of image displacements consistent with an interpretation of the local image motion as a relative translation of that corresponding part of the environment. Note that where the translational approximation is poor there will be a large value of the error measure reflecting the weaker confidence in the validity of the approximation. It is also necessary to incorporate information concerning the number and distribution of the feature points in the local image areas for this evaluation. For example,

if there is only one feature in a small area or the features are bunched together, then the translational approximation would be suspect. The further processing of the EDMF should not utilize local areas which do not have satisfactory characteristics.

This use of the translational procedure can be seen as a local constraint on the determination of image displacements. Typically, most such constraints are based upon smoothness of the resulting displacement field [Barn80, Glaz81, Horn80], where image displacements are computed under the constraint of being a local average of the displacements in their surrounding neighborhood. In our case, image displacements are determined such that the corresponding environmental motion can be interpreted locally as being translational. Note that this constraint does not necessarily imply local smoothness in the displacement field.

Computing the EDMF from raw image sequences depends upon how the images are divided into subareas. The image could be divided into small, regular, square subareas across the image and the procedure for determining the axis of translation is applied to each subarea independently. Alternatively, the procedure could be applied to individual regions determined by some segmentation procedure. In our work to date, we have used another approach in which the image subareas are neighborhoods centered on single features and the computation is applied independently over the neighborhood of each feature.

Computing the EDMF can be expensive for such feature-based neighborhoods since the feature displacements of many points are being determined simultaneously for different, overlapping, image subareas. An approximation is used to simplify this computation. For each feature, its best match and corresponding displacement along each of a set of radial directions are determined from one image into the next. These values are then stored in a 1-D array where each index corresponds to a particular radial direction centered at the feature and the associated best match

value for the corresponding direction (figure 46).

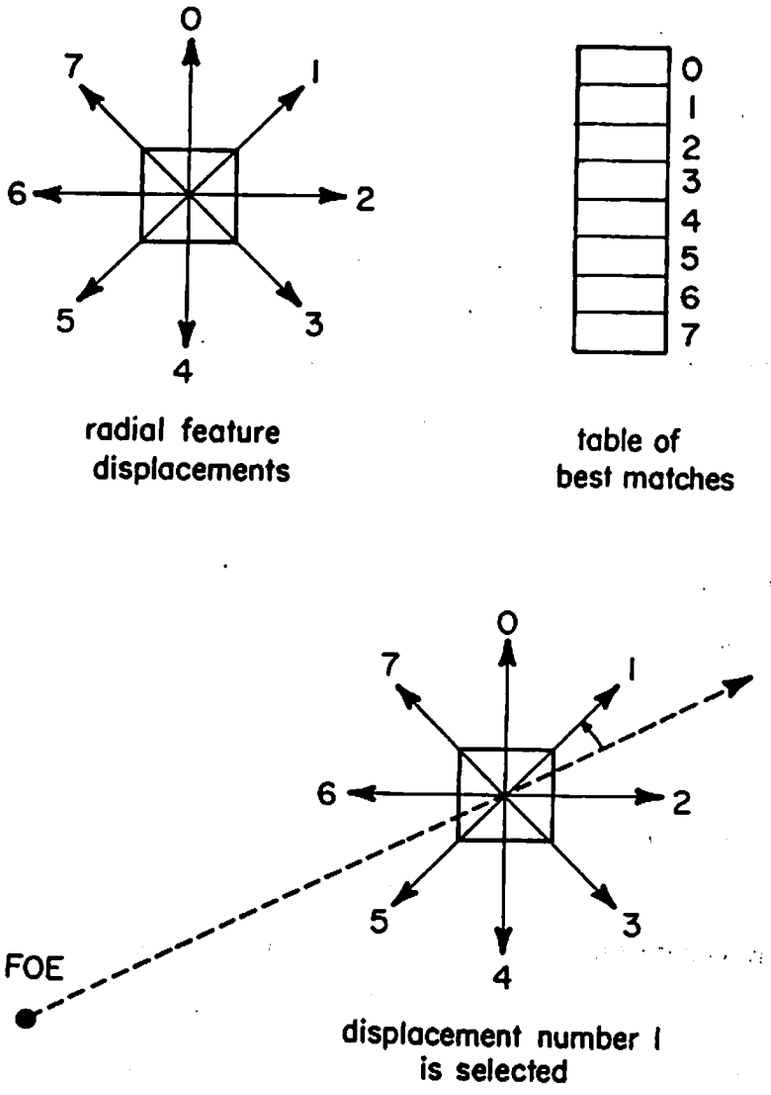


Figure 46. Approximating Match Values Along Translational Flow Paths.

This set of values is then used for all the translational computations employing this feature in its various neighborhoods. To determine the value of a particular translational axis with respect to a the neighborhood of a feature, each feature in the neighborhood finds its best match along the direction closest to that determined by the translational axis and the resulting values are then summed up. In this way, redundant evaluations of feature matches are avoided.

Figures 47a-b are referred to as the Grass Sequence 2. Figure 47a is a 128x128 pixel image of some grass texture with seven bits of intensity. Figure 47b was derived from figure 47a by applying a rotation of 0.1 radians about the Y axis of the camera coordinate system described in chapter III. The focal length was set to one and bilinear interpolation was used. Features were selected from the image in figure 47a by determining image points where the contrast was greater than 20 intensity levels and which were also local maxima in the distinctiveness values associated with 5x5 pixel square features centered at those points. The resulting feature positions are shown in figure 48.

The direction of translation was determined for 11x11 pixel neighborhoods centered at each feature in figure 48. Each feature determined its best displacements in 256 evenly spaced directions for distances of up to 10 pixels. The image displacement associated with a feature was the displacement that was consistent with the FOE/C determined by the translational approximation for the feature's neighborhood. The resulting image displacement field is shown in figure 49. As can be seen from the discussion in chapter III, it has the correct form for rotational motion about the Y-axis.

Figure 50a-c show the (X, Y, Z) components of the EDMF for the corresponding image points. The values in the EDMF are between 1.0 and -1.0 since it consists of unit vectors. Note that all the features have displacements in the same X direction

(Figure 50a) because the camera rotation about Y induces all points to move left or right. The Y displacements were all very close to zero (consistent with motion constrained to planes parallel to the Y-axis). The mean Y displacement was -0.003 (figure 50b). The Z components are positive for the right half and negative for the left half of the image (figure 50c. The scale of the display has also been increased). This motion occurs in pure rotation about Y because the environmental motions lie on circular paths with one side going away from the observer and the other side going towards the observer.

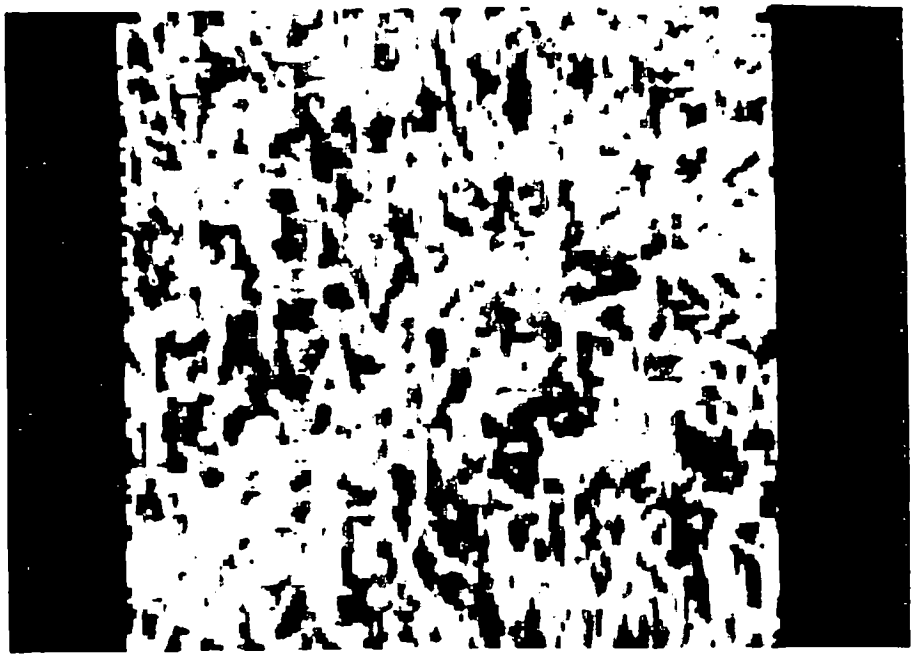


Figure 47a. Grass Sequence 2 Image 1.

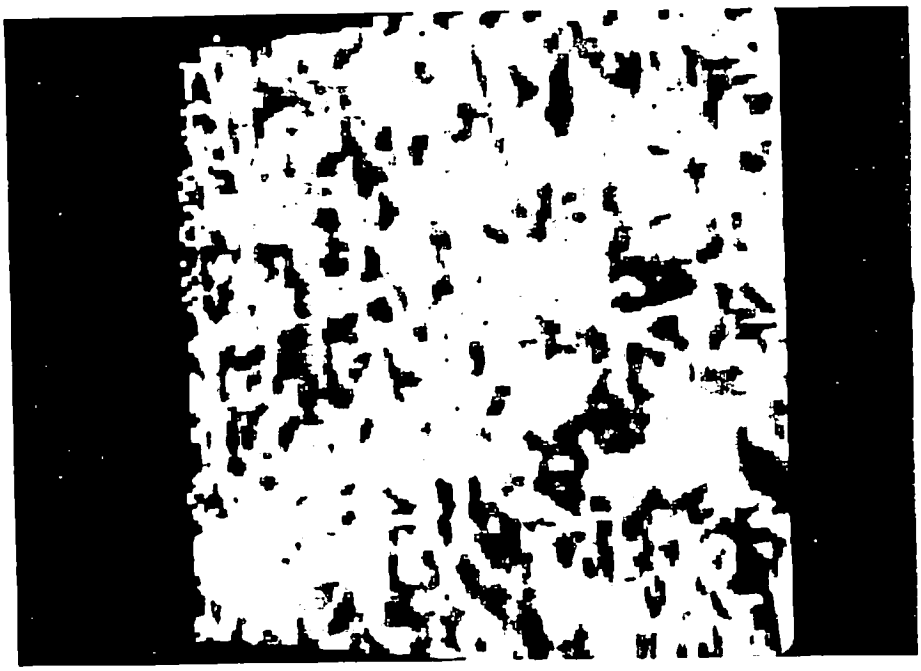


Figure 47b. Grass Sequence 2 Image 2.

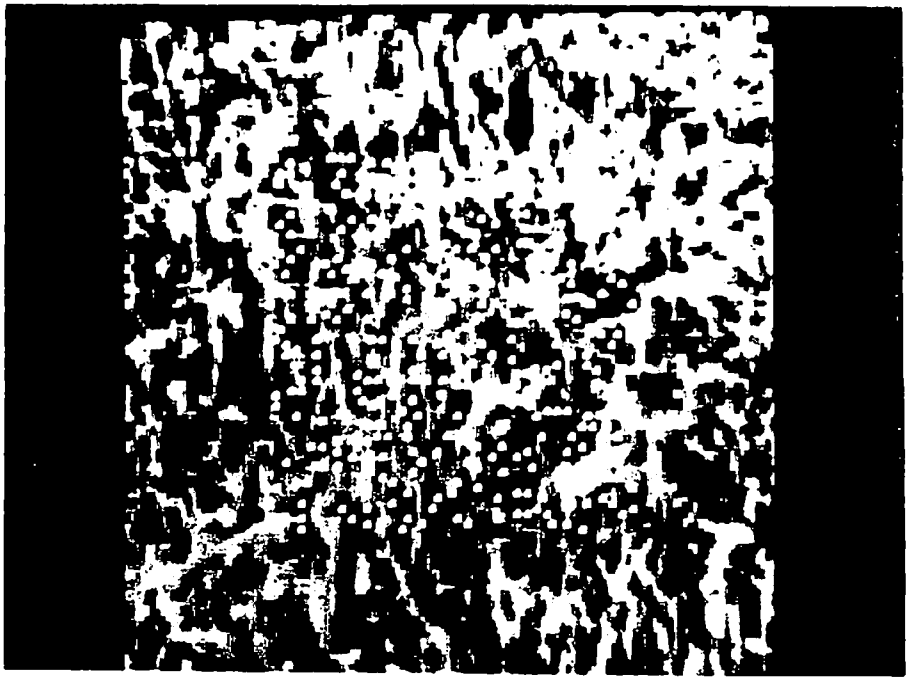


Figure 48. Selected Features.

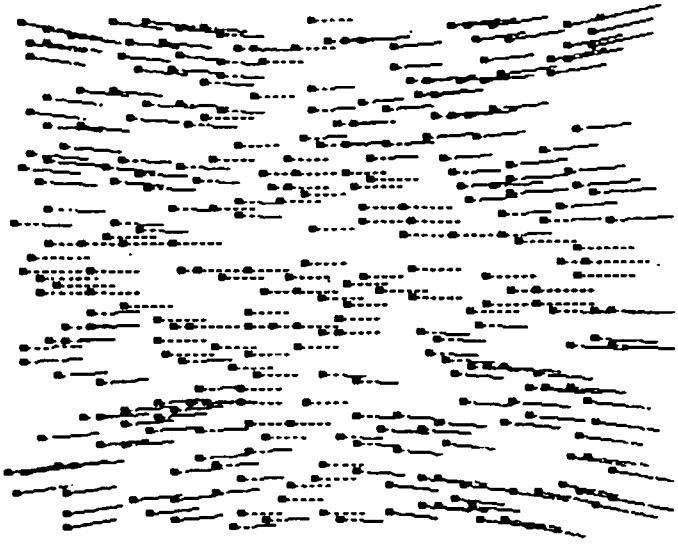


Figure 49. Determined Image Displacements.

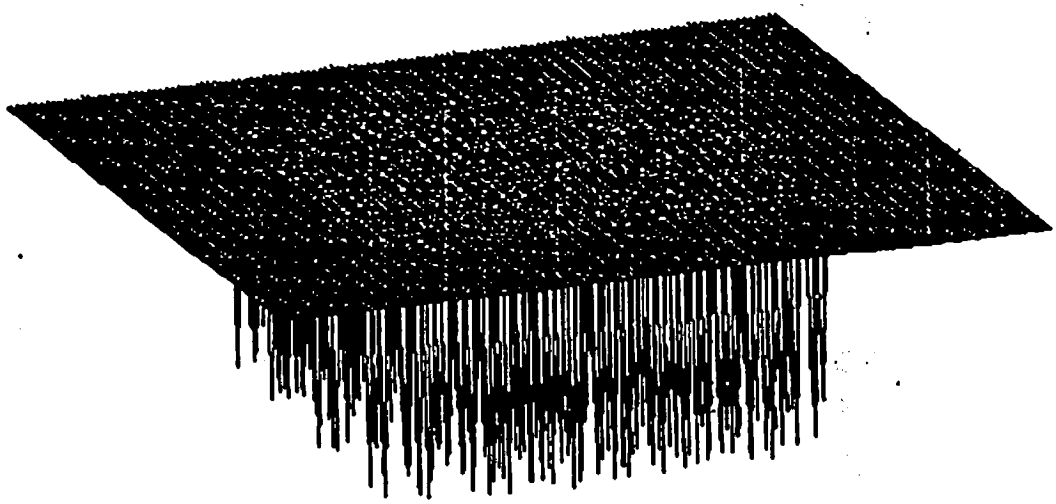


Figure 50a. Computed X Component of EDMF.

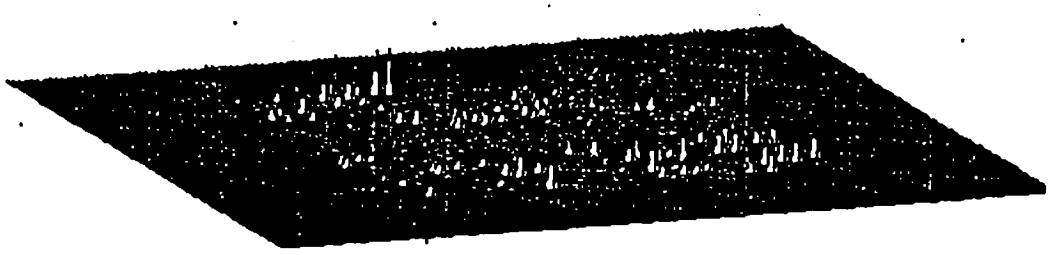


Figure 50b. Computed Y Component of EDMF.

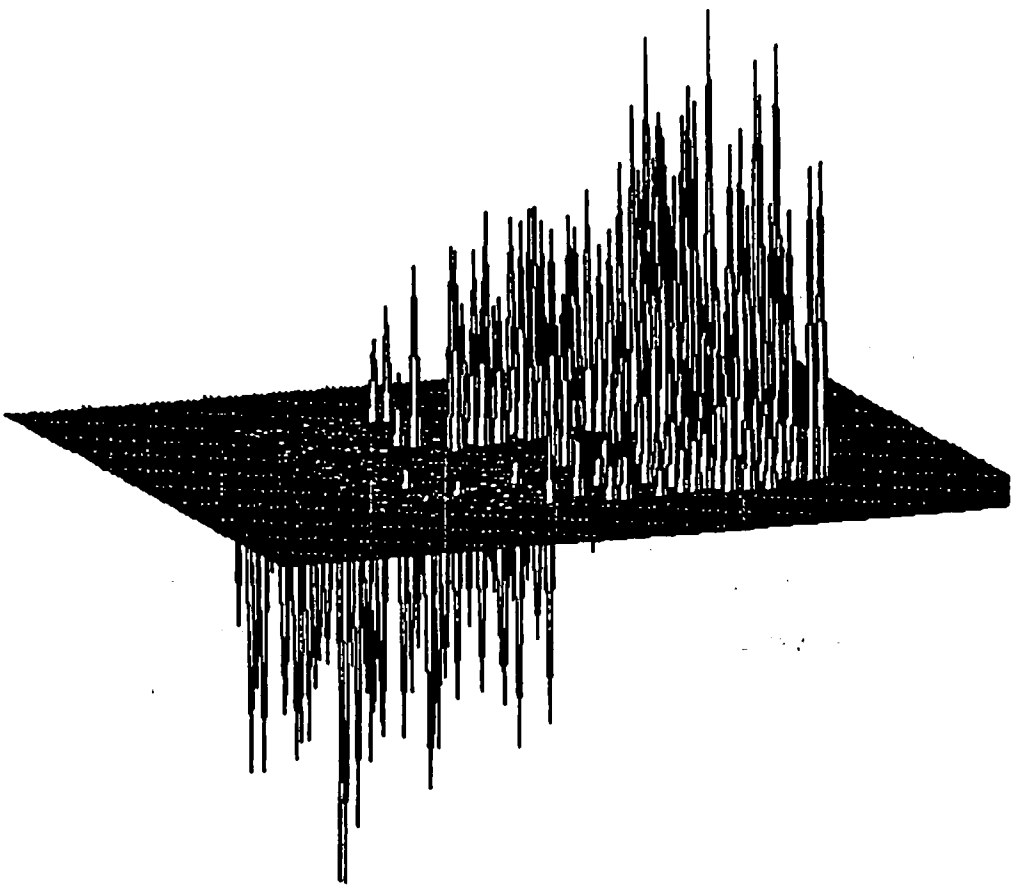


Figure 50c. Computed Z Component of EDMF.

Analysis of an Existing Displacement Field

To compute the EDMF from image sequences for which image displacements have already been determined, it is necessary to use the modification of the error

measure from chapter IV discussed in the section on the processing of translational blur paths. The error associated with a particular translational axis is a function of the angles between the image displacement paths determined by the FOE and the image displacement vectors. The function employed is the sum of one minus the cosine of each such angle, $\sum_i^N (1.0 - \cos\theta_i)$. To compute the EDMF, the translational axis is determined by applying this error measure, minimized as in chapter IV, to local areas of a computed displacement field.

Figure 51 shows a 32x32 image displacement field produced using a spherical distribution of environmental points about the Z-axis. The observer is looking into the interior of a sphere with noise modulation added to the depth values of the points in this figure. This noisy sphere was rotated 0.1 radians about an axis tangent to a point on the back of it along the $(1, 1, 1)$ axis. Note that this field was generated by an axis of rotation that was not positioned at the origin of the camera coordinate system. Each image point was the center of a 5x5 neighborhood over which the translational procedure, using the adapted error measure, was applied. Figure 52a-c show the X, Y, Z components of the computed EDMF and the correct EDMF, encoded as intensity with -1 being darkest, 1 the brightest and the neutral gray intensity along the border is 0. Figure 53 shows the values of the error of the translational approximation. Note how the approximation is poor where the field has a rotational character with vectors at very different orientations in a small area.

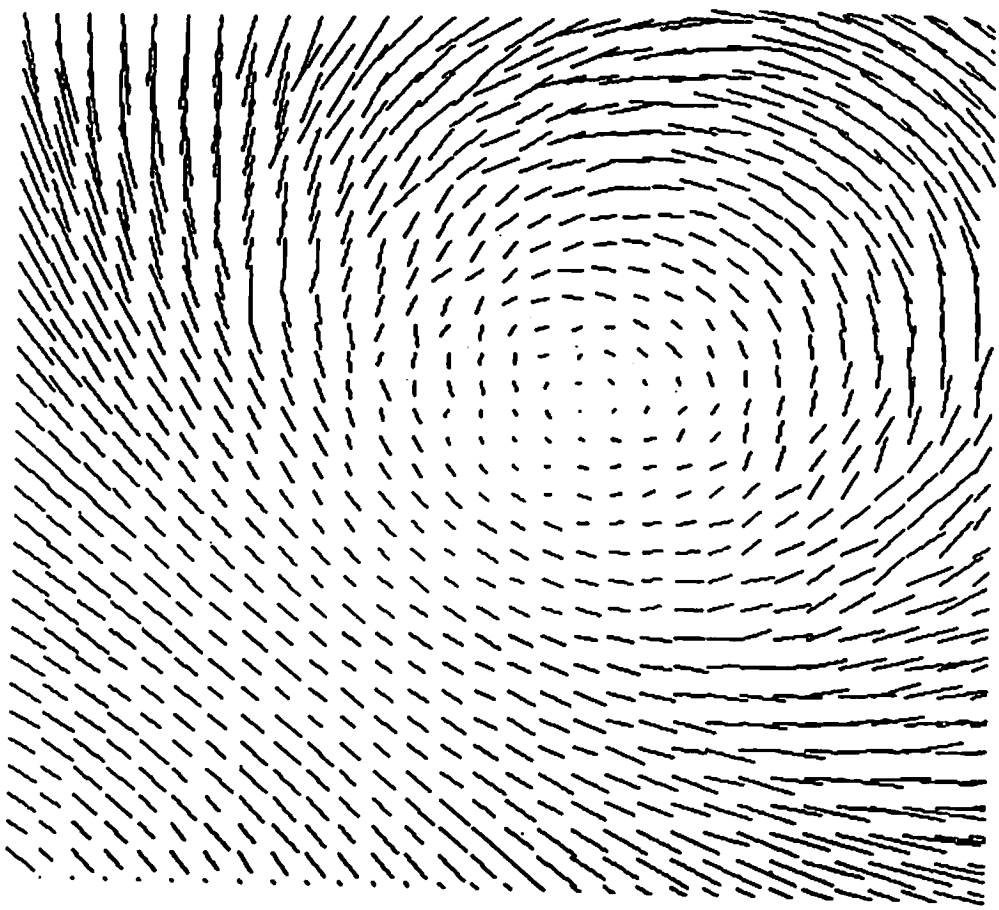


Figure 51. Simulated Flow Field.

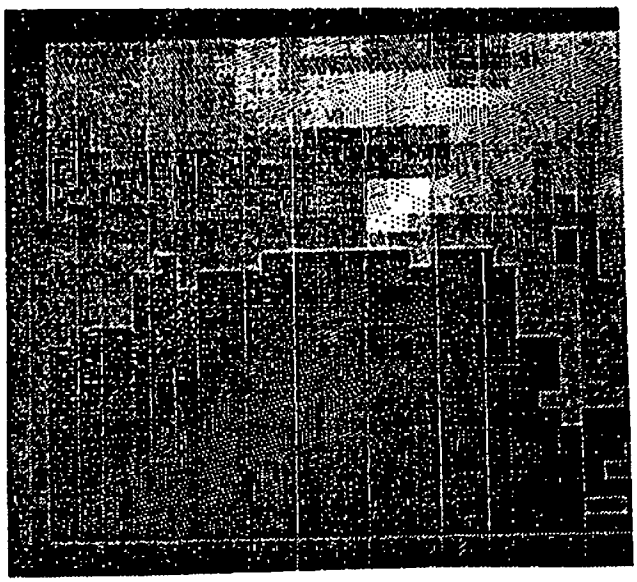


Figure 52a. Computed X Component of the EDMF.

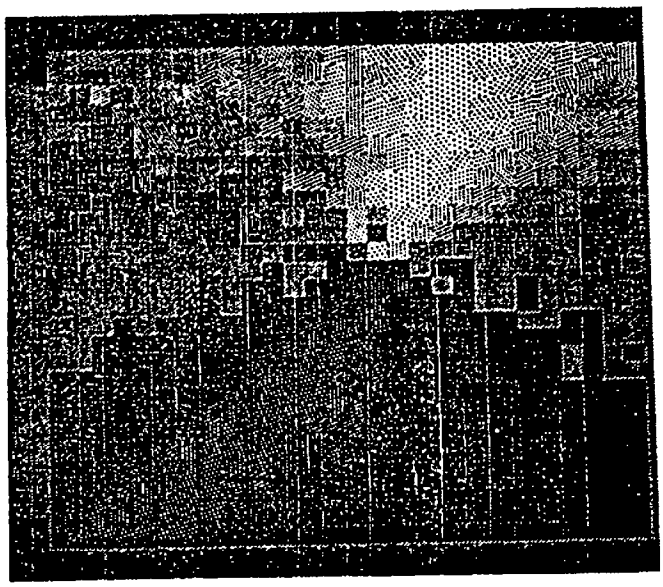


Figure 52b. Correct X Component of the EDMF.

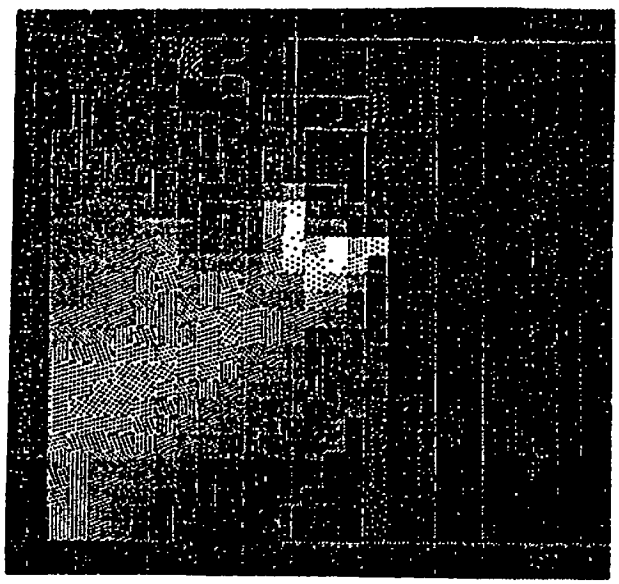


Figure 52c. Computed Y Component of the EDMF.

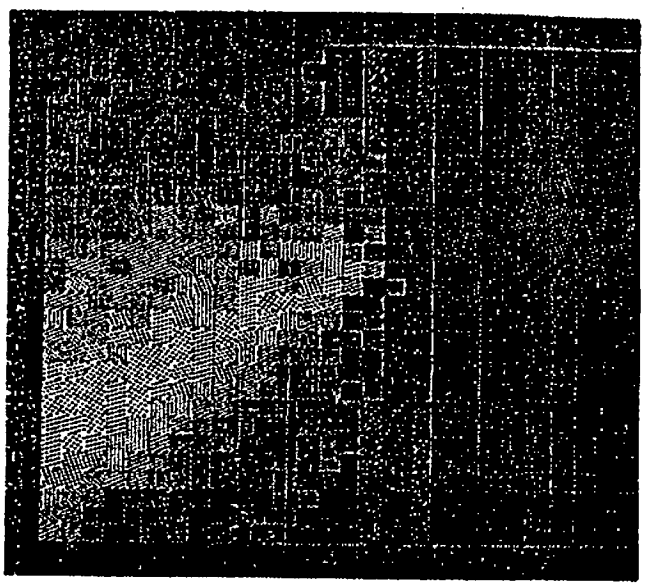


Figure 52d. Correct Y Component of the EDMF.

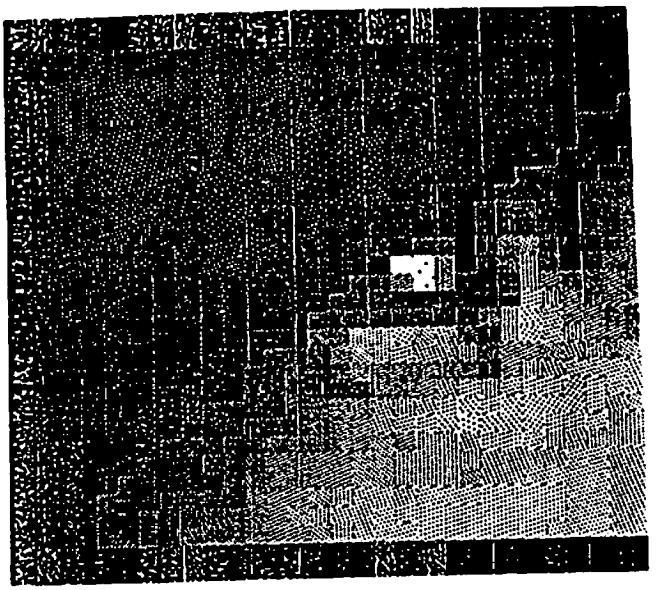


Figure 52e. Computed Z Component of the EDMF.

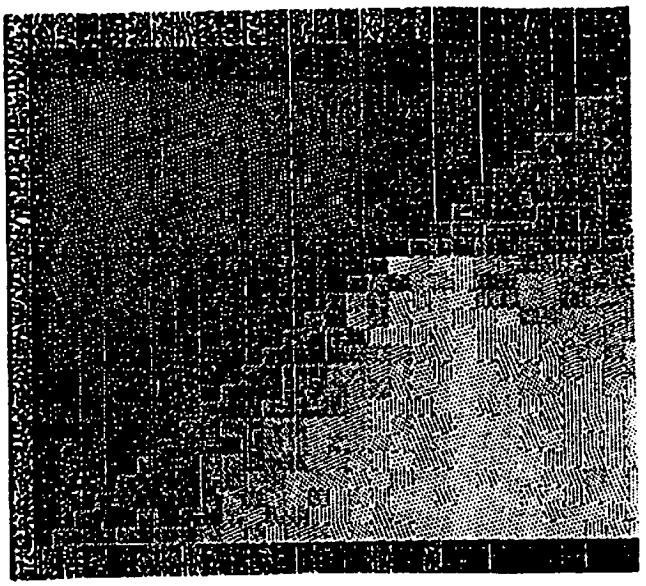


Figure 52f. Correct Z Component of the EDMF.

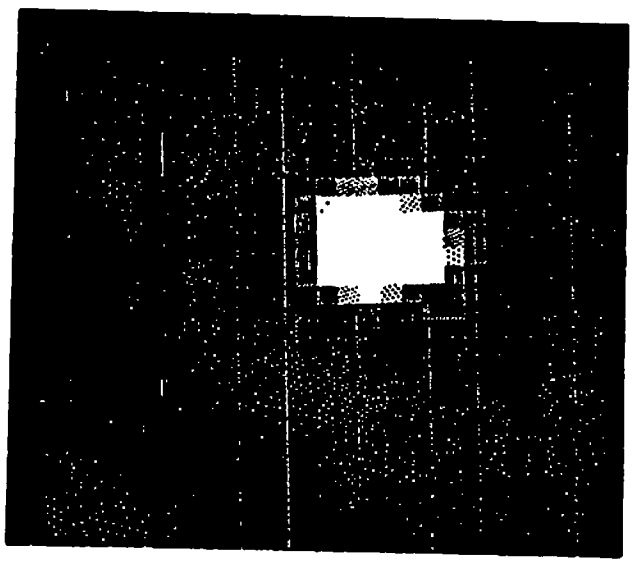


Figure 53a. Intensity plot of Error of Approximation.

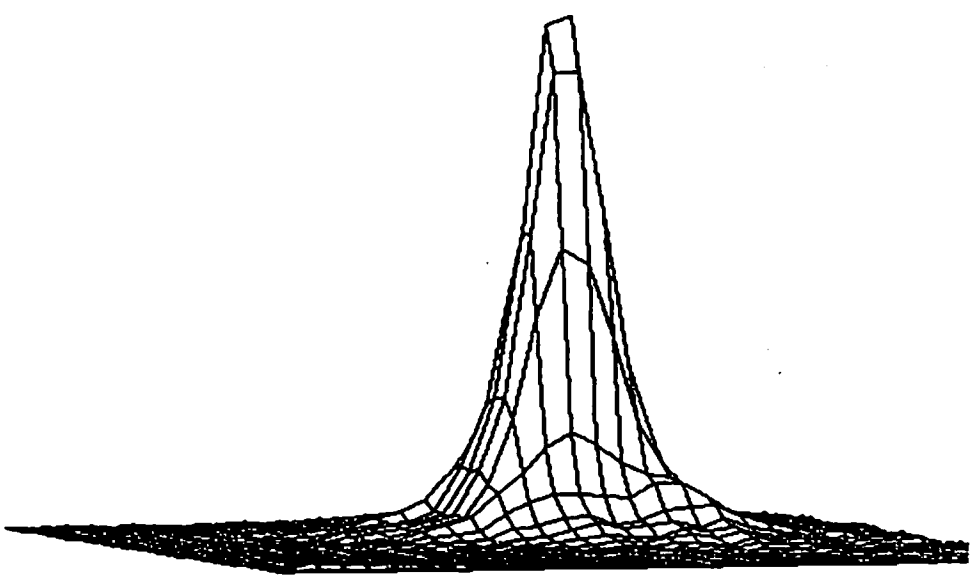


Figure 53b. Surface plot of Error of Approximation.

Computing the EDMF From Sparse Flow Fields

It may be possible to compute the EDMF from sparse displacement fields by applying an interpolation process [Glaz83b, Grim81, Terz82, Terz83] to produce a field of adequate density and then applying the techniques above. Some initial experiments have been performed to test this possibility, and they have shown a correlation between field density and the reliability of the approximation. The primary difficulty with very sparse fields is that the interpolation processes produce large areas of parallel displacements about the given image displacement vectors upon which the interpolation is based. This resulting flow field can be very different than the actual flow field from which the points were sampled, and therefore result in a poor approximation to the actual EDMF.

EDMF Properties for Different Cases of Motion

To describe EDMF properties for different cases of motion, it is useful to map all the EDMF vectors onto the direction of translation sphere. In Chapter IV, the direction of translation sphere was used as the domain of the error measure. Here it is used in a manner similar to a histogram. Each EDMF vector votes for a particular point on the direction of translation sphere. Processing then involves finding certain patterns in the distribution of the EDMF vectors.

EDMF Properties of Pure Translational Motion

As discussed previously the image displacement paths for translational motion

are straight lines intersecting at a point. The environmental displacement paths are straight, parallel lines. All the vectors in the EDMF are identical and map onto a single point on the direction of translation sphere corresponding to the translational axis.

EDMF Properties of Pure Rotational Motion

For pure rotational motion of the camera, the image displacement paths are conic sections determined by the intersection of the image plane with the nested family of cones aligned with the axis of rotation based at the origin of the camera coordinate system. The environmental displacement paths are circles about the axis of rotation and are contained in planes perpendicular to it. When mapped onto the direction of translation sphere, the EDMF vectors will lie upon a great circle contained in a plane perpendicular to the axis of rotation.

EDMF Properties of Motion Constrained to an Unknown Plane

For this case, the environmental displacement paths are circles in planes perpendicular to the axis of rotation, but the axis does not necessarily contain the origin of the coordinate system (see the discussion of kinematics in chapter 1 of [Whit44]). As for the rotational case, the EDMF vectors will lie on a great circle in a plane perpendicular to the axis of rotation when mapped onto the direction of translation sphere.

EDMF Properties of Arbitrary Motion

For arbitrary motion, the image displacement paths cannot be easily described. However, the environmental displacement paths are helices about an axis which does not necessarily contain the origin (since a screw displacement is the most general form of a rigid body motion [Coxe61, Whit44]).

The set of normalized tangent vectors to a helix, when based at a common origin, will generate a cone which we term the tangent cone. The orientation of this cone specifies the axis of rotation. The set of tangent cones determined by a rigid body motion for all points in space will all have the same orientation. Note that the difference vectors between any vectors of a tangent cone will lie in a plane perpendicular to the axis of rotation. Thus, the EDMF produced during arbitrary motion has a particularly nice property if the rigid body motion is constant over two or more intervals. For such motion there will be successive environmental direction of motion vectors associated with each image point, and the difference vectors between these successive EDMF vectors will lie in the same plane, perpendicular to the axis of rotation, for all image points.

In general, by mapping the EDMF onto the direction of translation sphere, the local differential properties of the EDMF are not being utilized. Such things as the extent of rotation can be recovered, or at least strongly constrained, by analyzing the local changes in the orientation of the EDMF vectors either spatially (over a small area of an image) or temporally (over successive inter-image intervals). Consider the case where the parameters of motion remain constant over successive intervals. Here, the angle between the successive EDMF vectors associated with an image point will be equal to the angle of rotation. This angle will be the same for all points in the image sequence and suggests a potentially robust technique for

determining the extent of rotation by finding the mean angle between successive EDMF vectors. For a single EDMF and image displacement field, this technique could be extended by predicting the EDMF vector for a point in the next interval by interpolating the value in the EDMF at the position determined by the head of the image displacement vectors.

Processing of Motion Constrained to an Unknown Plane

The EDMF produced by motion constrained to an unknown plane leads to a particularly simple algorithm. For this case there is one constraint on the inference of sensor motion parameters: the axis of rotation is perpendicular to the axis of translation. This corresponds to inferring four independent parameters: the rotational axis, the extent of rotation and the position of the translational axis in the plane perpendicular to the axis of rotation. All of the EDMF vectors are constrained to lie in a plane which is parallel to the plane of environmental motion. By calculating the EDMF vectors and fitting a plane to them, the plane of motion and thus the axis of rotation can be recovered. If the motion occurs over several successive instants and remains constrained to the same plane, then the vectors in the successive EDMFs are also constrained to lie in a plane parallel to it and containing the origin on the direction of translation sphere. Thus, more and more values for the fit can be collected over time, thereby increasing the accuracy of the processing. The extent of rotation can then be recovered by techniques for processing motion restricted to a known plane described in chapter VI. The processing is further simplified since the image displacements have already been computed or were determined from computing the EDMF.

The best planar fit to the EDMF vectors can be found using any of a number

of plane fitting routines. In the experiments here, an eigenvector fit procedure (described in [Duda73] pp. 332-335) is used, having been adapted for planes containing the origin. Once the plane of motion is determined, the algorithm for processing known planar motion from a computed displacement field is used. We now consider some examples.

The grass sequence 2 from this chapter involving pure rotation is a case of motion constrained to a plane since the environmental displacement paths all lie in planes perpendicular to the axis of rotation. Using the EDMF determined for the grass texture sequence described above, the normal to the best plane fit was (.003,.999,-.014). This is in error by .015 radians, or .836 degrees, from the correct rotational axis.

Using all the EDMF vectors determined for the flow field in figure 51 in the plane fitting procedure, the normal to the plane of motion is determined to be (.647, .544, .534). This deviates from the correct axis by .089 radians or 5.078 degrees. This fit can be improved by removing vectors from the EDMF for which the corresponding local FOE/C yields a large error, and therefore a poor translational approximation. For the EDMF vectors computed from the flow field in figure 51, the error value is equal to the sum of the angles between the flow vectors in each 5x5 neighborhood over which the EDMF vector was determined and the displacement paths corresponding to the translational axis which minimized the error measure. We can thus express the validity of a computed EDMF vector by the sum of these deviation angles. Figure 53 shows the error values in the translational fit proportional to image darkness. Note that the greatest errors occur where the image displacement vectors have a rotational character. By restricting the planar fit to EDMF vectors for which the sum of the deviation angles corresponds to less than some threshold (90 degrees in this example) of error relative to the determined

translational field lines over the 5x5 pixel neighborhoods, the normal is determined to be (.579462, .583347, .569148). This deviates by .010380 radians or .594798 degrees from the correct rotational axis. Thus, the high error measure values have been used to remove the rotational-like displacements in the center of the image. The error histogram derived from the flow field in figure 51, assuming motion to be constrained to this plane, is shown in figure 54a and 54b. In the contour plot (figure 54b) a " - " indicates a local minimum and a " + " indicates a local maximum. The correct rotation was selected from the histogram; (the rotational parameter was varied from -0.15 to 0.15 radians in 0.1 radian increments). The determined rotational field is shown in figure 55a and the translation field which results from subtracting the determined rotational field from the original displacement field is shown in figure 55b.

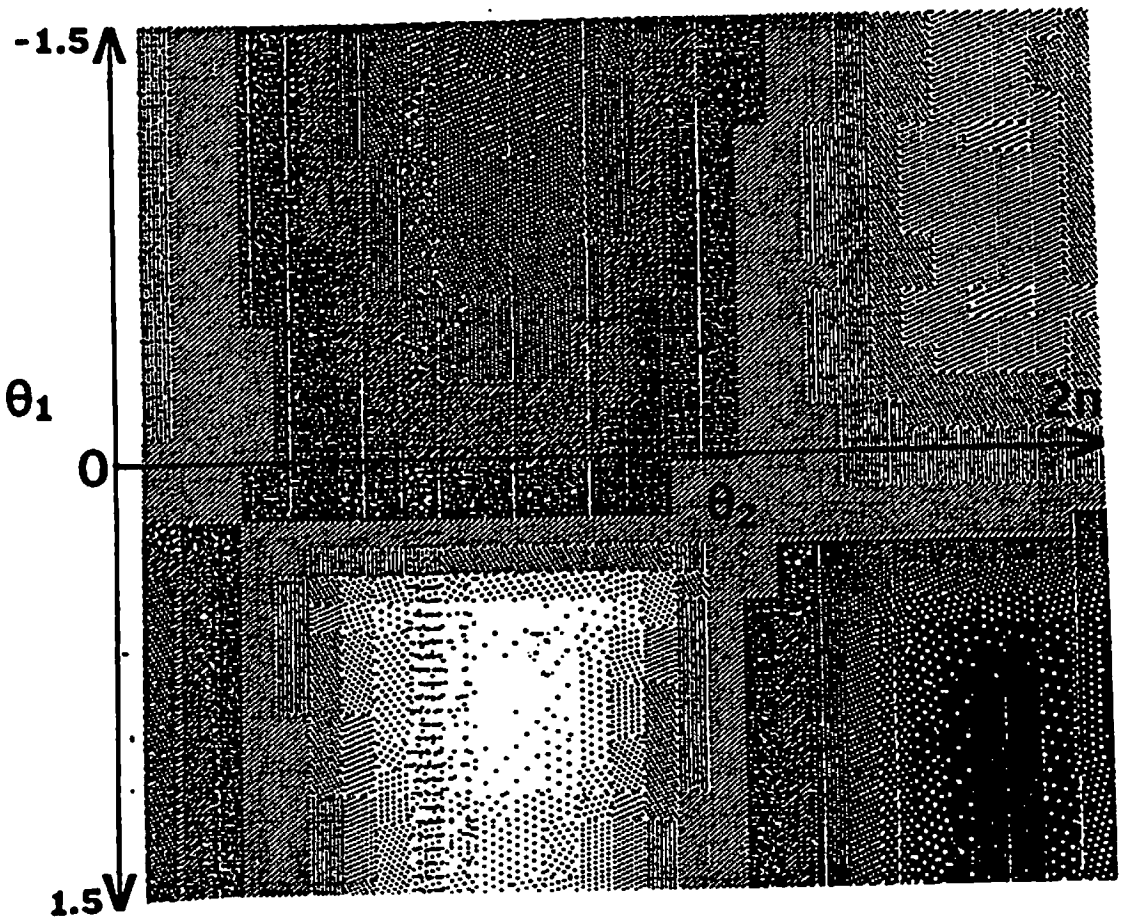


Figure 54a. Error Histogram for Simulated Flow Field

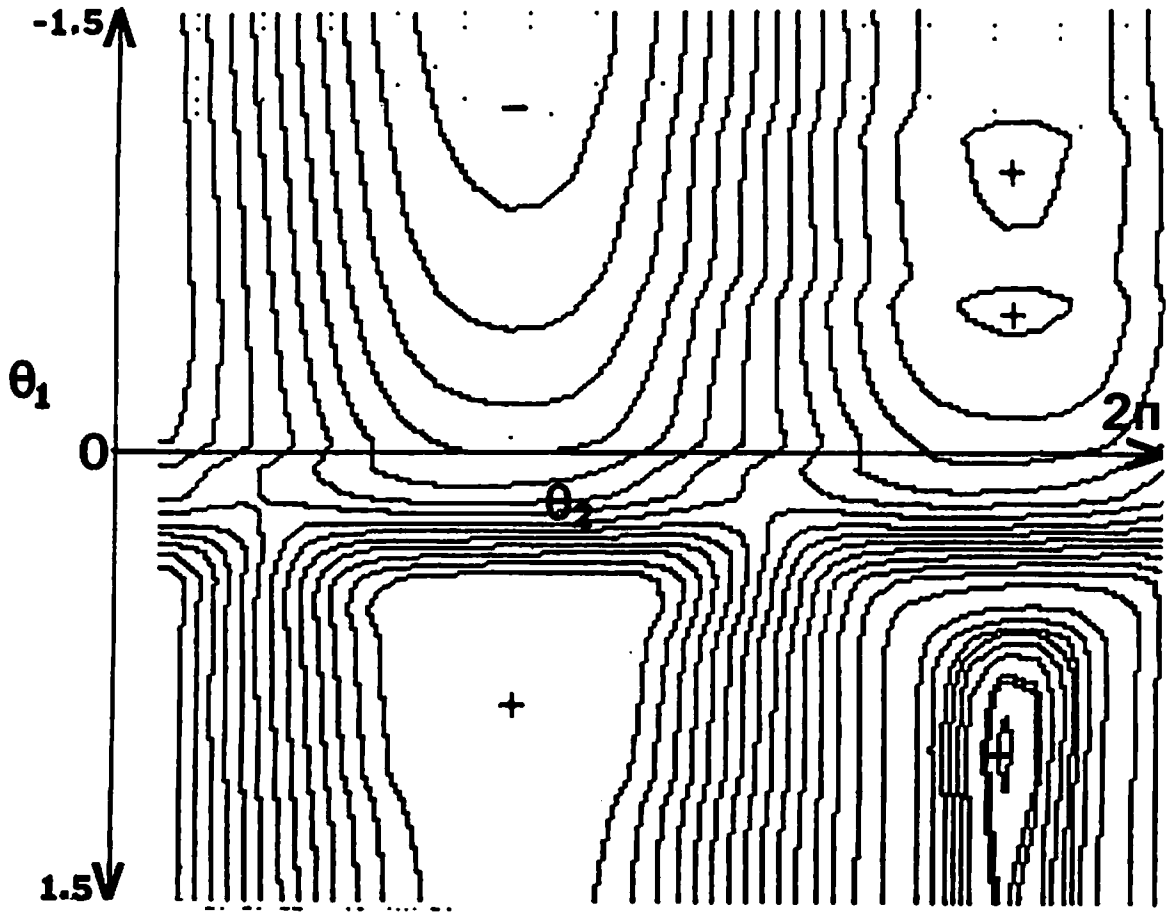


Figure 54b. Contour Plot of Error Histogram

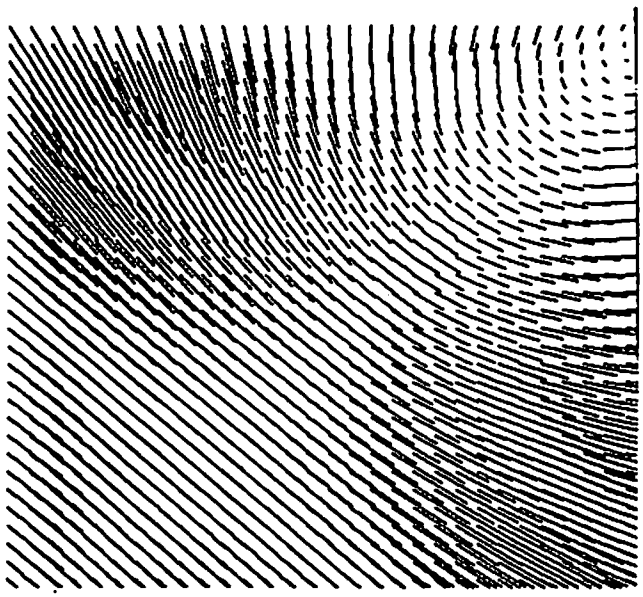


Figure 55a. Determined Rotational Field

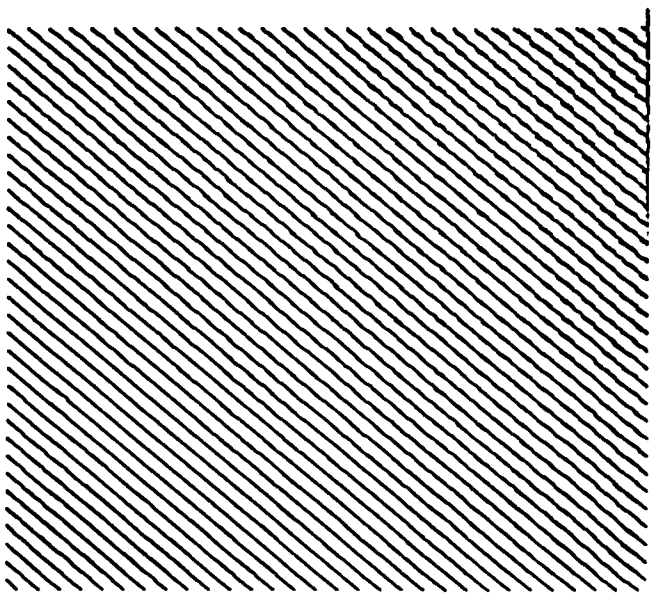


Figure 55b. Determined Translational Field

Environmental Inference via EDMF and Rigidity Constraints

A basic paradigm in computer vision is to take an environmental property and express it in terms of the constraints it imposes on resulting image structures [Barr81]. These constraints are then expressed as equations whose solution determines an interpretation of image events consistent with the assumed environmental properties. In this section, we utilize the constraint of environmental rigidity to derive a set of equations whose solution determines a set of environmental depths that are consistent with given image displacements. We show the conditions under which solutions to these equations are possible [Lawt80, Meir80, Ullm79, Webb81] for general motion and how these conditions are affected for restricted cases of motion. We then show how the equations for unrestricted motion are significantly simplified when information concerning the direction of environmental motion is also utilized.

Development of Rigidity Constraints

For this development, we refer to the camera model described in chapter III. Equation 1 from chapter III can be used transform expressed relations between environmental points into a set of equations in terms of image position vectors and unknown Z values which correspond to the environmental depth values. Solutions to the resulting equations yield a set of Z values which provide a consistent interpretation over time for the positions of the corresponding set of environmental points.

The basic relation for interpreting environmental motion is the assumption of

rigidity which reflects the invariance of distance between environmental points during motion. For two points i and j on a rigid body at times m and n, this preservation of distance is expressed as

$$\|P_{mi} - P_{mj}\| = \|P_{ni} - P_{nj}\| \tag{16}$$

which can be expanded, by using the substitution specified by equation 1 from chapter III and squaring both sides, into the image-based equation

$$\begin{aligned} & z_{mi}^2(I_{mi} \cdot I_{mi}) + z_{mj}^2(I_{mj} \cdot I_{mj}) \\ & -2z_{mi}z_{mj}(I_{mi} \cdot I_{mj}) - z_{ni}^2(I_{ni} \cdot I_{ni}) \\ & -z_{nj}^2(I_{nj} \cdot I_{nj}) + 2z_{ni}z_{nj}(I_{ni} \cdot I_{nj}) = 0 \end{aligned} \tag{17}$$

where the inner-product terms in parentheses are constants determined from the positions of image points. To determine a solution, we will find the minimum number of points and frames for which the number of independent constraints (in the form of equation 17) equals or exceeds the number of unknown Z values. It is then necessary to solve the resulting set of simultaneous equations. Note that each such constraint is a second degree polynomial in 4 unknowns.

We begin with the number of unknown Z values. For N points in K frames (where $N > 2$ and $K > 1$), there are $(NK - 1)$ unknown Z values. The decrease by one in the number of unknowns reflects the loss of absolute scale information.

Thus, one of the Z-values can be set to an arbitrary value which can be recovered from the actual sensor displacement if such absolute measurements are available.

The number of rigidity constraints generated by a set of N points in K frames is the product of $3 \times (N-2)$ and $(K-1)$. The first term is the minimum number of unique distances which must be specified between pairs of points, in a body of N points with no three points being collinear, to assure its rigidity. Thus, 4 points require 6 pairwise distances (all that are possible). For configurations of more than 4 points, it is necessary to specify the distance of each additional point to only 3 other points to assure rigidity. The second term is the number of interframe intervals, with each interval providing a set of additional constraining points. Each distance specified must be maintained over each interframe interval.

A solution is possible when the number of constraints is greater or equal to the number of unknowns. This occurs when:

$$2NK - 6K - 3N + 7 \geq 0 \tag{18}$$

Thus, minimal solutions can be found when $N = 5$ and $K = 2$, producing nine constraint equations or when $N = 4$ and $K = 3$ producing 12 constraint equations.

Rigidity Constraints Applied to Known Planar Motion. As one would expect, the rigidity constraints are simplified by adding restrictions on allowable motions of environmental points. For example, consider motion constrained to a plane. For simplicity, we will assume that it is parallel to the XZ plane of the camera coordinate system, but an appropriate transformation can be applied so that the results are valid for motion constrained to an arbitrarily oriented, but known, plane.

Here, the Y component of an environmental point is assumed to remain constant over time. For a point i at times m and n, this is expressed as

$$y_{mi} = z_{mi} b_{mi} = z_{ni} b_{ni} = y_{ni} \tag{19}$$

and solving for z_{ni} yields

$$z_{ni} = z_{mi}(b_{mi}/b_{ni}) \tag{20}$$

This allows a substitution for points i and j in equation 17 which simplifies (at least in terms of the number of unknowns) the rigidity constraint to

$$\begin{aligned} & z_{mi}^2((I_{mi} \cdot I_{mi}) - (\frac{b_{mi}}{b_{ni}})^2(I_{ni} \cdot I_{ni})) \\ & + z_{mj}^2((I_{mj} \cdot I_{mj}) - (\frac{b_{mj}}{b_{nj}})^2(I_{nj} \cdot I_{nj})) \\ & + z_{mi} z_{mj}(2((\frac{b_{mi}}{b_{ni}})(\frac{b_{mj}}{b_{nj}})(I_{ni} \cdot I_{nj}) - (I_{mi} \cdot I_{nj}))) \\ & = 0 \end{aligned} \tag{21}$$

The planarity constraint has removed two unknowns. Note that the bracketed expressions are again constants that can be determined from the locations of the image points. This equation can be solved given two points in two frames. Thus, for

points i and j at times m and n with the corresponding unknown depth values z_{mi}, z_{mj}, z_{ni}, z_{nj}, equation 21 reduces these to a system of 2 unknowns, z_{mi} and z_{mj}. One of these variables, say z_{mi}, can be set to an arbitrary value, reflecting scale independence, allowing z_{mj} to then be determined by solving the quadratic in terms of z_{mi}.

Rigidity Constraints Applied to Translational Motion The constraint imposed by translational motion of points i and j on a rigid body at times m and n is expressed by

$$P_{mi} - P_{mj} = P_{ni} - P_{nj} \tag{22}$$

which is similar to equation 16 except the operation is vector subtraction reflecting the preservation of length and orientation under translation. Setting z_{mi} to a constant value 1, to reflect scale independence in equation 22, yields 3 simultaneous linear equations in 3 unknowns

$$(a_{mi}, b_{mi}, 1) = z_{mj}(a_{mj}, b_{mj}, 1) + z_{ni}(a_{ni}, b_{ni}, 1) - z_{nj}(a_{nj}, b_{nj}, 1) \tag{23}$$

Thus, not surprisingly, environmental inference from translation requires 2 points in 2 frames.

Solving the Rigidity Constraints using the EDMF

The rigidity constraints can be significantly simplified when they are integrated with information concerning the environmental direction of motion from the local translational decomposition. To do this the EDMF is used first to find consistent relative depths for single points over successive images. Consistent relative depths for several points are then determined by scaling the particular depth values for the individual points using the rigidity constraint.

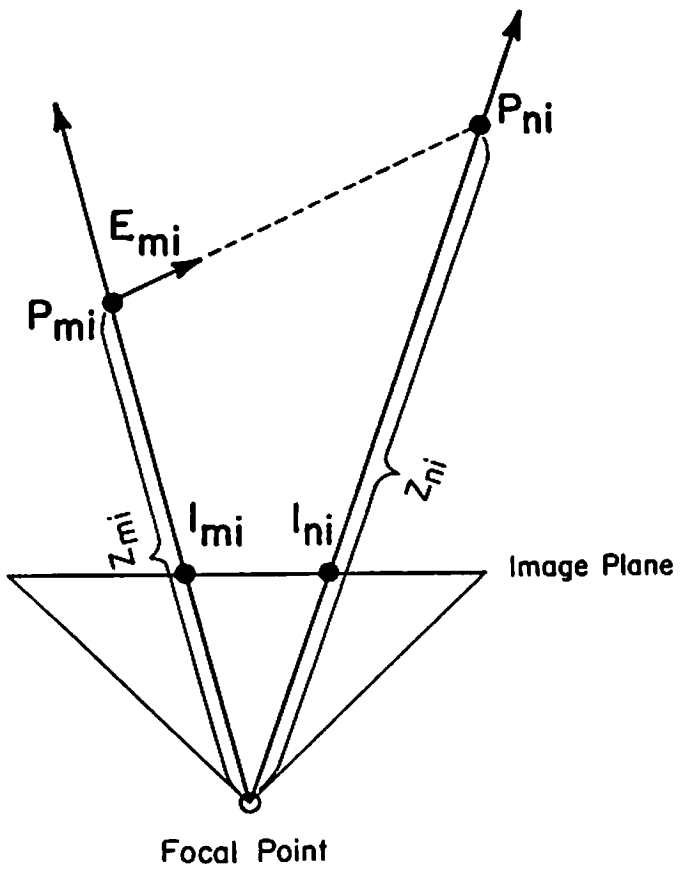


Figure 56. Relative Depths for a point over time from the EDMF.

We first examine the use of the EDMF in the determination of consistent relative depths for a single point over time. Consider the image position vectors I_{mi} and I_{ni} (for the successive image positions of point i at times m and n) and the environmental direction of motion associated with point i at time m, E_{mi}. (Figure 56). Assuming the ideal case, in which there is no error in any of these quantities, the EDMF vector E_{mi} will lie in the plane determined by I_{mi} and I_{ni}. Thus, given a depth z_{mi} along the ray of projection corresponding to I_{mi}, one can find a depth value z_{ni} along the ray of projection associated with I_{ni} from the intersection of the lines $P_{mi}+tE_{mi}$ and $z_{ni}I_{ni}$. In the usual case of error in these measurements, these lines will not intersect because they are skewed in three dimensions. In these instances we can solve for the line segment which is perpendicular to both of these lines. Let us express the point along the ray of projection determined by I_{ni} which is closest to the line determined by the point $P_{mi} = z_{mi}I_{mi}$ and the direction of motion E_{mi} from the EDMF:

$$\begin{aligned}((z_{mi}I_{mi} + tE_{mi}) - (z_{ni}I_{ni})) \cdot E_{mi} = 0 \\ ((z_{mi}I_{mi} + tE_{mi}) - (z_{ni}I_{ni})) \cdot I_{ni} = 0\end{aligned} \tag{24}$$

which simplifies to

$$\begin{aligned}t(E_{mi} \cdot E_{mi}) - z_{ni}(E_{mi} \cdot I_{ni}) = -z_{mi}(I_{mi} \cdot E_{mi}) \\ t(E_{mi} \cdot I_{ni}) - z_{ni}(I_{ni} \cdot I_{ni}) = -z_{mi}(I_{mi} \cdot I_{ni})\end{aligned} \tag{25}$$

These equations can be expressed in terms of the ratio of the relative distances

along the successive rays of projection consistent with the environmental direction of motion E_{mi} (and treating t as a dummy variable)

$$t(E_{mi} \cdot E_{mi}) - r_{mni}(E_{mi} \cdot I_{ni}) = (I_{mi} \cdot E_{mi})$$
$$t(E_{mi} \cdot I_{ni}) - r_{mni}(I_{ni} \cdot I_{ni}) = (I_{mi} \cdot I_{ni}) \quad (26)$$

where

$$r_{mni} = \frac{z_{mi}}{z_{ni}}$$

This yields the relative depths of a single point over time. We now use the rigidity constraint to detetrmine the appropriate scaling of each of these ratios for all of the points.

Assume we have two points i and j at times m and n. Let z_{mj} be set to an arbitrary value. Then, z_{nj} may be obtained by the product $z_{mj} \times r_{mnj} = z_{mj} \times \frac{z_{nj}}{z_{mj}}$ where the ratio r_{mnj} is obtained through the relation expressed in equation 26. This yields the environmental points P_{mj} and P_{nj}. We can now use the rigidity constraint to determine a scale factor expressing $P_{mi} = z_{mi}I_{mi}$ and $P_{ni} = z_{ni}I_{ni} = z_{mi}r_{mni}I_{ni}$ in terms of P_{mj} and P_{nj} Substitution into the rigidity constraint yields

$$\|z_{mi}I_{mi} - P_{mj}\| = \|z_{mi}r_{mni}I_{ni} - P_{nj}\| \quad (27)$$

where z_{mi} is the scale factor. Equation 27 can be expanded as

$$z_{mi}^2((I_{mi} \cdot I_{mi}) - (r_{mni}I_{ni} \cdot r_{mni}I_{ni}))$$

$$+((P_{mj} \cdot P_{mj}) - (P_{nj} \cdot P_{nj}))$$

$$-2z_{mi}((I_{mi} \cdot P_{mj}) - (r_{mni}I_{ni} \cdot P_{nj})) = 0 \qquad (28)$$

The resulting equation is quadratic in one unknown. Thus, given successive depth values determined for a particular point from its EDMF vector, consistent depths can be determined for every other paur of successive depth values by solving this equation for each resulting pair of points.

In summary, given a flow field and an EDMF, a pair of depth values for each image point at successive instants m and n can be found which are consistent with the determined EDMF vectors describing motion from time m to n. These are relative depth values, and hence may be scaled arbitrarily and inferred from equation 26. Once these relative, successive depth values are determined for each point, they may then be scaled relative to a selected point whose depth is arbitrarily set by solving equation 28 for each point paired with this selected point. There is a great deal of redundancy for optimization procedures to exploit. Several depth maps can be computed (one for each selected image point) and the certainty of a particular depth inference would be based upon agreement in the relative depth values in all the resulting depth maps. If there are further spatial constraints, such as motion relative to a planar surface, all the determined depth maps would have to be in agreement with respect to the shape. For example, all the determined depth maps for a plane would have to correspond to a single plane at the same orientation.

This work shows that if the EDMF can be reliably computed, it is a very useful low level representation for rigid body motion analysis. This is possible for densely

textured image sequences for which the camera motion parameters to be recovered correspond to motion constrained to an unknown plane. The local translational decomposition may also be applicable to inferring qualitative descriptions of non-rigid motions by noting certain patterns in the relative directions of motion as would typify such motions as expanding or twisting.

CHAPTER VIII

SUMMARY AND FUTURE WORK

We summarize the major contributions of this thesis and many of the questions it raises for further study. We shall conclude with a consideration of two major areas for future research that are intimately related to motion processing: architectures for real-time processing and image interpretation in the domain of dynamic road scenes.

Summary

The review of work in dynamic image processing in chapter II stressed a basic problem in motion research. There has been a discrepancy between the precision and reliability with which image displacements can be determined and the sensitivity of the environmental and sensor motion inference procedures to such noise and resolution errors. In addition, there are open questions about the stability of the inference procedures themselves. We noted that this has limited the practical applications of dynamic image processing in domains where its use is fundamental.

In chapter IV we developed a procedure for processing translational motion. The most important feature of this procedure is that the determination of the image displacements, the direction of sensor motion, and environmental depth are combined into a single, mutually constraining computation. The procedure consists

of two basic steps: Feature Extraction and Search. The feature extraction process finds small image areas which may correspond to distinguishing, and therefore trackable, parts of environmental objects. The direction of translational motion is then found by a search which minimizes an error measure defined over a unit sphere, with each point on the sphere corresponding to a different direction of sensor translation. A given direction of translation constrains the motion of extracted image features to straight lines which radiate from or converge onto a single point in the image plane. Thus, the error measure associates a point on the unit sphere, corresponding to a particular translational axis, with a number describing the degree of total feature mismatch along the set of displacement paths determined by the translational axis. Experience has shown this error measure to be smooth and with a distinct minimum in a large neighborhood about the correct translational axis. This allows simple search procedures to be effective. Experiments were presented which indicated that the algorithm was robust in a variety of ways. It could function effectively with weak or false features, with a small numbers of features, and even with a small number of features in limited portions of an image.

Many extensions and possible areas of further work were also discussed, and we mention two, here, that are of particular interest. First, the procedure should be developed to extend over multiple frames. The determined translational axis, image displacements, and environmental depth values should be used to constrain further processing and feature extraction in a manner that will allow refinement in the accuracy of sensor motion parameters and the environmental depth map. Second, a theoretical formulation is necessary to develop a more complete, analytical understanding of the robustness of the procedure.

In chapter V we considered other extensions to the translational procedure including its embodiment as a hierarchical computation; processing translational blur

paths; dealing with multiple independently translating objects; and using the translational procedure for autonomous vehicle control by having a stabilized sensor or associated devices to determine the rotational parameters. The hierarchical extension was found to significantly increase the speed of the procedure, since it reduces the number of feature correlations that are necessary along potential translational displacement paths. There are still a variety of alternatives to be investigated before the most effective implementation of the hierarchical computation will be thoroughly understood. We showed that the processing of translational blur paths could be performed by a simple extension of the error measure used in chapter IV. The extensions discussed for multiple, independently moving objects were based upon the similarity of the translational procedure to generalized Hough transforms and the limited image areas necessary for the procedure to function. Finally, the incorporation of the procedure with sensor stabilization and rotational displacement sensing devices has exciting implications for passive-sensing based autonomous vehicles.

In chapter VI we successfully processed other simple cases of restricted motion, pure sensor rotation and motion constrained to a known plane, for which it was computationally feasible to search through the subspace of the sensor motion parameters for values that are consistent with image feature displacements. For pure sensor rotation the dimensionality of the search increased over the translational case, but was compensated for by the additional constraint that the extents of all feature displacements were identical. We noted a typical case of planar motion, quite common to terrestrial motion, which is inherently ambiguous.

In chapter VII we showed how to process sensor motion by applying the procedure for translational motion to local areas of images. This yields a low level description of motion that we termed the Environmental Direction of Motion Field

(EDMF) which associated a relative direction of environmental motion between features from restricted image subareas and the sensor. We showed how to process the case of motion constrained to an unknown plane using the constraint that all the EDMF vectors are constrained to lie in this plane. This constraint forms the basis of a robust computation to recover the parameters of sensor motion in this case. We discussed the recovery of the parameters of sensor motion from the EDMF for general sensor motion. We developed the rigidity constraints which express the inference of environmental depth from displacement fields by exploiting the preservation of object rigidity during motion. We showed that these constraints were directly solvable for restricted cases of motion and that this was also possible for arbitrary motion when information from the EDMF was incorporated with the rigidity constraints.

There are several aspects of the work in chapter VII that require further exploration. The processing of unrestricted motion should be evaluated with respect to the required accuracy of the set of direction vectors in the EDMF. It may be possible to derive qualitative inferences more robustly. This is also related to the way in which the EDMF is computed. We investigated only one of the techniques that were discussed, the case where image subareas are centered on individual features. In another of the suggested techniques, the subareas are formed by dividing the image into regular, nonoverlapping subareas and applying the translational procedure over each of these. In this case, the EDMF would not be associated with a particular environmental point, but with a larger environmental area, thereby reducing the resolution in the EDMF.

Direct solutions to the rigidity constraints should also be studied further, since our formulation of the rigidity constraints was developed some years ago [Lawt80] and was not explored beyond noting that the equations were tractable using simple

iterative optimization techniques and that the solutions were multimodal in the cases of minimal numbers of points and image frames. What, for example, are the effects of using multiple images and a greater number of points? Additionally, there has been interest in using optimization procedures based on simulated annealing [Kirk83] to solve these equations. These techniques have shown an ability to deal with multimodal error surfaces in very high dimensional spaces.

Future Work

Architectures for Translational Motion Processing

The translational procedure that we have developed offers an attractive possibility for real-time implementation of a motion processing system. The architecture is a straightforward design consisting of multiple independent processors, each associated with a unique, disjoint set of features. Each processor determines the displacement and extent of error for its features along the translational displacement paths specified by a given FOE/C. The processors are then coordinated by a global search executive which specifies a particular FOE/C, sums up the error responses of the multiple processors, and determines which translational axis to be evaluated next. The critical parameters for effective implementation are the speed with which a feature's displacement can be determined along its displacement path by its associated processor and the number of times the error function must be evaluated to determine the translational axis to sufficient accuracy. Experiments with the translational procedure indicate that, outside of pathological cases, fewer than 50 evaluations of the error function will be sufficient and even fewer when the translational axis has been initialized by previous processing. Preliminary timing

studies using Motorola 68000 processors (10 megahertz minor cycle time) to determine feature displacements indicate that the necessary processing rates are feasible [Levi83].

Image Interpretation of Dynamic Road Scenes

Research often advances by the stimulating problems that are found in a wisely chosen task domain. The VISIONS system [Hans78] used outdoor house scenes as a guiding incentive to develop procedures and representations necessary for complex imagery. A domain that we feel would be challenging, yet one in which achievable results would be possible, is the interpretation of outdoor road scenes along highways and country roads as seen from a moving vehicle. This domain is quite tractable under assumptions consistent with a variety of the algorithms presented in this thesis. The assumed constraints might include the vehicle constrained to translational motion or constrained to a plane; a stabilized sensor or knowledge of the rotational parameters; sensor and object motions constrained to slowly changing translations; or motion of independently moving objects constrained to a roughly determined plane. This domain forces us to address interesting questions such as how to achieve dynamic segmentations using the temporal behavior of complex image structures over time, the incorporation of object-specific semantics into recognition using environmental depth and image motion information, and predicative processing from a model which is established by temporally extended inferences. Thus, a whole new set of issues arise as a full road scene interpretation system is developed which integrates motion and static interpretation into a goal oriented perceptual system in a dynamic environment.

BIBLIOGRAPHY

[Adiv83] Adiv, G., Recovering Motion Parameters in Scenes Containing Multiple Moving Objects, IEEE Computer Vision and Pattern Recognition, June 19-23, 1983, Washington D.C., pp. 399-400.

[Adiv84] Adiv, G. The Interpretation of Optical Flow Fields, Ph. D. Thesis, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts, in preparation.

[Agga81a] Aggarwal, J.K. and Martin, W.N., Analyzing Dynamic Scenes Containing Multiple Moving Objects, in Image Sequence Analysis (T.S. Huang, editor), Berlin Heidelberg: Springer Verlag, 1981, pp. 355-380.

[Agga81b] Aggarwal, J.K., Davis, L.S. and Martin, W.N., Correspondence Processes in Dynamic Scene Analysis, Proceedings of the IEEE, Volume 69, Number 5, May 1981, pp. 562-572.

[Badl75] Badler, N.I., Temporal Scene Analysis: Conceptual Descriptions of Object Movements, Ph.D. Dissertation, University of Toronto, TR 80, February 1975; also University of Pennsylvania Technical Report 80.

[Ball81] Ballard, D.H., Parameter Networks: Towards a Theory of Low-Level Vision, IJCAI-7, Vancouver, British Columbia, August 1981, pp. 1068-1078.

[Barn80] Barnard, S.T. and Thompson, W.B., Disparity Analysis of Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume PAMI-2, Number 4, July 1980, pp. 333-340.

[Barr72] Barrow, H.G., Ambler, A.P., and Burstall, R.M., Some Techniques for Recognizing Structures in Pictures, Frontiers of Pattern Recognition, S.

Watanabe, (ed.), Academic Press, 1972, pp. 1-29.

[Barr81] Barrow, H.G. and Tenenbaum J.M., Interpreting Line Drawings as Three Dimensional Surfaces, Artificial Intelligence, 17, 75-117.

[Brad83] Brady, M., Criteria for Shape Representations, in Human and Machine Vision, Beck J. and Rosenfeld A. (eds.), Academic Press, 1983.

[Burr77] Burr, D., A System for Stereo Computer Vision with Geometric Models, Ph.D. Dissertation, Coordinated Science Laboratory, CSL, University of Illinois, 1977.

[Burt76] Burt, P.J., Stimulus Organizing Processes in Stereopsis and Motion Perception, Ph.D. thesis and COINS Technical REport 76-15, University of Massachusetts, Amherst, 1976.

[Burt82] Burt, P.J., Pyramid-Based Extraction of Local Image Features with Application to Motion and Texture Analysis, SPIE Conf. on Robotics and Industrial Inspection, San Diego, 1982.

[Burt83] Burt, P.J., Yen, C. and Xu, X., Multi-Resolution Flow-Through Motion Analysis, IEEE Computer Vision and Pattern Recognition, June 19-23, 1983, Washington D.C., pp. 246-252.

[Chen82] Cheng, J.K. and Huang, T.S., Recognition of Curvilinear Objects by Matching Relational Structures, Proc. of the Pattern Recognition and Image Processing Conference, Las Vegas, Nevada, 1982, pp. 343-348.

[Coxe61] Coxeter, H., Introduction to Geometry. New York, Wiley, 1961.

[Davi82] Davis, L.S., Wu, Z. and Sun, H., Contour-based Motion Estimation, TR-1179, Center for Automation Research, University of Maryland, College

Park, MD, 20742, 1982.

[Davi83] Davis, L. S., Kitchen, K., Hu, F.P., and Hwang, V., Image Matching using Generalized Hough Transforms, CAR-TR-27 or CS-TR-1335, Center for Advanced Automation Research, University of Maryland, College Park, MD 20742, October, 1983.

[Dres81] Dreschler, L. and Nagel, H.-H., Volumetric Model and 3-D Trajectory of a Moving Car Derived from Monocular TV-Frame Sequences of a Street Scene, IJCAI-7, Vancouver, Canada, August 1981.

[Duda73] Duda, R.O., and Hart, P.E., Pattern Classification and Scene Analysis, John Wiley and Sons, New York, 1973.

[Ezek82] Ezekiel, S. and Arditty, H.J., Fiber Optic Rotation Sensors and Related Technologies, New York, Springer-Verlag, 1982.

[Fang83a] Fang, J.Q. and Huang, T.S., Estimating 3-D Movement of a Rigid Object: Experimental Results, IJCAI-83, Karlsruhe, West Germany, pp. 1035-1037.

[Fang83b] Fang, J.Q. and Huang, T.S., Solving Three Dimensional Small-Rotation Motion Equations, IEEE Computer Vision and Pattern Recognition, June 19-23, 1983, Washington D.C., pp. 253-258

[Fenn79] Fennema, C.L. and Thompson, W.B., Velocity Determination in Scenes Containing Several Moving Objects, Computer Graphics and Image Processing, Volume 9, Number 4, April 1979, pp. 301-315.

[Gibs50] Gibson, J.J., The Perception of the Visual World, Cambridge, Mass: Riverside, 1950.

[Gibs66] Gibson, J.J., The Senses Considered as Perceptual Systems, Boston, MA: Houghton-Mifflin, 1966.

[Gibs79] Gibson, J.J., The Ecological Approach to Visual Perception, Houghton Mifflin, Boston, 1979.

[Glaz81] Glazer, F., Computing Optic Flow, IJCAI-7, Vancouver, B.C. Canada, August, 1981, pp. 644-647.

[Glaz83a] Glazer, F., Multilevel Relaxation in Low Level Computer Vision, in Multiresolution Image Processing and Analysis, A. Rosenfeld (Ed.), Springer-Verlag, 1983.

[Glaz83b] Glazer, F., Reynolds, G. and Anandan, P., Scene Matching by Hierarchical Correlation, IEEE Computer Vision and Pattern Recognition, June 19-23, 1983, Washington D.C., pp. 432 - 441.

[Grim79] Grimson, W.E.L., Differential Geometry, Surface Patches and Convergence Methods, MIT AI Memo 510, February 1979.

[Grim81] Grimson, W.E.L., From Images to Surfaces: A Computational Study of the Human Early Visual System, MIT Press, Cambridge, 1981.

[Hann74] Hannah, M.J., Computer Matching of Areas in Stereo Images, Stanford A.I. Memo 239, July 1974.

[Hans78] Hanson, A.R. and Riseman, E.M., Visions: A Computer System for Interpretating Scenes, In: Computer Vision Systems, A.R. Hanson and E.M. Riseman (Eds.), Academic Press, New York, 1978, pp. 303-334.

[Hans80] Hanson, A.R. and Riseman, E.M., Processing Cones: A Computational Structure for Image Analysis, in Structured Computer Vision, Tanimoto,

S. and Klinger, A. (eds.), It's Academic Press, New York, 1980.

[Hara78] Haralick, R.M., Scene Analysis, Arrangement and Homomorphisms, in Computer Vision Systems, Hanson, A.H. and Riseman, E.M. (eds.), Academic Press, 1978, pp 199-212.

[Hara82] Haralick, R.M., Zero-Crossings of Second Directional Derivative Edge Operator, SPIE Symposium in Robot Vision, Washington, D.C., May 1982

[Hara83] Haralick, R.M., Ridges and Valleys on Digital Images, Comp. Vision Graphics and Image Processing, 22 (1), 28-39.

[Harr80] Harrington, T. L., Harrington, M. K., Wilkins, C. A. and Koh, Y. O. Visual Orientation by Motion-Produced Blur Patterns: Detection of Divergence, Perception and Psychophysics, vol. 28, pp. 293-305, 1980.

[Hild80] Hildreth, E.C., Implementation of a Theory of Edge Detection, MIT AI Technical Report AI-TR-579, MIT, Cambridge, MA, 1980.

[Hild82] Hildreth, E.C., The Integration of Motion Information Along Contours, Proceedings of the IEEE Workshop on Computer Vision: Representation and Control, August, 1982, pp. 83-91.

[Horn80] Horn, B.K.P. and Schunck, B.G., Determining Optical Flow, Massachusetts Institute of Technology, A.I. Memo Number 572, April 1980.

[Huan81] Huang, T.S. (editor) Image Sequence Analysis, Berlin-Heidelberg: Springer Verlag, 1981.

[Jaco80] Jacobus, C.J., Chien, R.T. and Selander, J.M., Motion Detection and Analysis by Matching Graphs of Intermediate-Level Primitives, IEEE

Transactions on Pattern Analysis and Machine Intelligence, Volume PAMI-2, Number 6, November 1980, pp. 495-510.

[Jeri83] Jerian, C. and Jain, R., Determining Motion Parameters for Scenes with Translation and Rotation, SIGGRAPH/SIGART Interdisciplinary Workshop MOTION: Representation and Perception, Toronto, Canada, April 4-6, 1983, pp. 71-77.

[Kear82] Kearney, J.K., Thompson, W.B., and Boley, D.L., Gradient Based Estimation of Disparity, Proc. of the Pattern Recognition and Image Processing Conference, Las Vegas, Nevada, 1982, pp. 246-251.

[Kend79] Kender, J.R., Shape from Texture: An Aggregation Transform that Maps a Class of Textures into Surface Orientation, IJCAI-6, Tokyo, August 1979, pp. 475-480.

[Kirk83] Kirkpatrick, S., Gelatt, C.D., and Vecci, M.P., Optimization by Simulated Annealing, Science, May, 1983.

[Kitc80] Kitchen L. and Rosenfeld A., Gray-Level Corner Detection, TR-887, Computer Science Center, University of Maryland, College Park, Md. April, 1980.

[Kohl81] Kohler, R., A Segmentation System Based on Thresholding, Computer Graphics and Image Processing, Volume 15, 1981, pp. 319-338.

[Lawt80] Lawton, D.T., Constraint-Based Inference from Image Motion, Proc. AAAI-80, Stanford, California, August, 1980, pp 31-34.

Lawt83] Lawton, D.T. and Rieger, J.H., The Use of Difference Fields in Processing Sensor Motion, DARPA Workshop on Image Processing, June 1983,

Washington D.C., pp 77-83.

[Lawt84] Lawton, D.T., Levitan, S., Weems, C., Iconic to Symbolic Mapping using a Content Addressable Array Processor, Department of Computer and Information Science, University of Massachusetts, Amherst, Massachusetts, in preparation.

[Lee76] Lee, D.N., A Theory of Visual Control of Braking Based on Information About Time to Collision, Perception, Volume 5, 1976, pp. 437-459.

[Lee80] Lee, D.N., The Optic Flow Field: The Foundation of Vision, Philosophical Trans. Royal Soc. London, Volume B, Number 290, 1980, pp. 169-179.

[Lee82] Lee, H. C., and Fu, K. S., The GLGS Image Representation and its Application to Preliminary Segmentation and Preattentive Visual Search, Proc. of the Pattern Recognition and Image Processing Conference, Las Vegas, Nevada, 1982, pp. 256-261.

[Levi73] Levine, M.D., O'Handley, D.A. and Yagi, G.M., Computer Determination of Depth Maps, Comput. Graphics Image Processing, Volume 2, 1973, pp. 131-150.

[Levi83] Levitan, S., Personal communication, Department of Computer and Information Science Department, University of Massachusetts Amherst, Mass., 1983.

[Long80] Longuet-Higgins, H.C. and Prazdny, K., The Interpretation of Moving Retinal Image, Proceedings of the Royal Society B, 1980.

[Long81] Longuet-Higgins, H. C., A Computer Algorithm for Reconstructing a Scene from two Projections, Nature, vol. 293, 1981, pp. 133-135.

[Luca81] Lucas, B.D., and Kanade, An Iterative Image Registration Technique with Application to Stereo Vision, DARPA Image Understanding Workshop, pp. 121-130, 1981.

[Marr80] Marr, D. and Hildreth, E., Theory of Edge Detection, Proc. Royal Soc. London, Volume B, 1980, pp. 187-217.

[Marr79] Marr, D. and Ullman, S., Directional Selectivity and Its Use in Early Visual Processing, Artificial Intelligence Laboratory Memo No. 524, MIT, June 1979.

[Mart79] Martin, W.N. and Aggarwal, J.K., Computer Analysis of Dynamic Scenes Containing Curvilinear Figures, Pattern Recognition, Volume 11, 1979, pp. 169-178.

[Medi83] Medioni, G.G., Matching Regions in Aerial Images, IEEE Computer Vision and Pattern Recognition, June 19-23, 1983, Washington D.C., pp. 364-365.

[Meir80] Meiri, A.Z., On Monocular Perception of 3-D Moving Objects, IEEE Trans. PAMI-2, 1980, pp. 582-583.

[Mora81] Moravec, H.P., Robot Rover Visual Navigation, UMI Research Press, Ann Arbor, Michigan, 1981.

[Mora77] Moravec, H.P., Towards Automatic Visual Obstacle Avoidance, Proceedings of the 5th IJCAI, MIT, Cambridge, MA, 1977, p. 584.

[Nage77] Nagel, H.-H., Analyzing Sequences of TV-Frames, Proceedings of the 5th IJCAI, MIT, Cambridge, MA, 1977, August 1977, p. 626.

[Nage78] Nagel, H., Formation of an Object Concept by Analysis of Systematic

Time Variations in the Optically Perceptible Environment, Comput. Graphics Image Processing, Volume 7, 1978, pp. 149-194.

[Nage81] Nagel, H.-H., On the Derivation of 3-D Rigid Point Configurations from Image Sequences, IEEE PRIP-81, Dallas, Texas, August 1981.

[Nage83] Nagel, H.H., Constraints for the Estimation of Displacement Vector Fields from Image Sequences, IJCAI-83, Karlsruhe, Wesr Germany, pp. 945-951.

[Naka80] Nakatani, H., Kimura, S., Saito, O. and Kitahashi, T., Extraction of Vanishing Point and its Application to Scene Analysis Based on Image Sequence, Proc. of 5th International Conference on Pattern Recognition, Miami Beach, Florida, December 1980, pp. 370-372.

[Naka74] Nakayama, K. and Loomis, J.M., Optical Velocity Patterns, Velocity-Sensitive Neurons and Space Perception, Perception, Volume 3, 1974, pp. 63-80.

[O'rou81] O'Rourke, J. Motion Detection Using Hough Techniques, Proceedings of PRIP 1981, pp. 82-87.

[Prag79] Prager, J.M., Segmentation of Static and Dynamic Scenes, COINS Technical Report 79-07 and Ph.D. Dissertation, University of Massachusetts, Amherst, 1979.

[Praz81] Prazdny, K., Determining the Instantaneous Direction of Motion from Optical Flow Generated by a Curvilinearly Moving Observer, Proc. of the Pattern Recognition and Image Processing Conference, Dallas, Texas, August 1981, pp. 109-114.

[Praz80] Prazdny, K., Egomotion and Relative Depth Map from Optical Flow,

Biology and Cybernetics, Volume 36, 1980, pp. 87-102.

[Quam71] Quam, L.H., Computer Comparison of Pictures, Stanford A.I. Memo AIM-144, May 1971.

[Radi81] Radig, B.M., Image Region Extraction of Moving Objects, in Image Sequence Analysis (T.S. Huang, editor), Berlin Heidelberg: Springer-Verlag, 1981, pp. 311-354.

[Rieg83] Rieger, J.H. and Lawton, D.T., Sensor Motion and Relative Depth from Difference Fields of Optic Flows, IJCAI-83, Karlsruhe, West Germany, pp. 1027-1031.

[Roac79] Roach, J.W. and Aggarwal, J.K., Computer Tracking of Objects Moving in Space, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume PAMI-1, 1979, pp. 127-134.

[Roac80] Roach, J.W. and Aggarwal, J.K., Determining the Movement of Objects from a Sequence of Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume PAMI-2, Number 6, November 1980, pp. 554-562.

[Roge76] Rogers, D.F. and Adams, J.A., Mathematical Elements of Computer Graphics, New York: McGraw-Hill, 1976.

[Rose83] Rosenfeld, A., Multiresolution Image Processing and Analysis, Springer-Verlag, 1983.

[Schu83] Schunck, B.G., Motion Segmentation and Estimation, Ph.D. Thesis, AI Lab, MIT, 1983.

[Shap82] Shapiro, L. G. and Haralick, R.M., Organisation of Relational Models for

Scene Analysis, IEEE-PAMI, Vol. PAMI-4, No. 6, November 1982, pp. 595-602.

[Shaw83] Shaw, G.B., Determining Motion Parameters using a Perturbation Approach, Technical Report 83-30, Department of Computer and Information Science, University of Massachusetts, Amherst, Mass, 01003, September, 1983.

[Shep83] Shepard,R. N. and Zare,S. L., Path-Guided Apparent Motion, Science, vol. 220, pp. 632-634, 1983.

[Stee83] Steenstrup, M.E., Lawton, D.T., and Weems, C., Determination of the Rotational and Translational Components of a Flow Field using a Content Addressable Parallel Processor, IEEE Computer Vision and Pattern Recognition, June 19-23, 1983, Washington D.C., pp. 401-404.

[Tani80] Tanimoto, S. and Klinger, A. (Editors), Structured Computer Vision: Machine Perception through Hierarchical Computation Structures, Academic Press, New York, 1980.

[Terz82] Terzopolous,D., Multi-Level Reconstruction of Visual Surfaces: Variational Principles and Finite Element Representation, AI Memo-671, MIT AI Lab, Cambridge, MA, 1982.

[Terz83] Terzopolous,D., The Role of Constraints and Discontinuities in Visible-Surface Reconstruction, IJCAI-83, Karlsruhe, West Germany, pp.1073-1077.

[Thom80] Thompson, W.B., Combining Motion and Contrast for Segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 2, Number 26, 1980, pp. 543-549.

[Thom81] Thompson, W.B. and Barnard, S.T., Lower-Level Estimation and Interpretation of Visual Motion, Computer, August 1981.

[Thom83] Thompson, W. B., Mutch, K. M., and Berzins, V. A., Determining Qualitative Spatial Properties from Optical Flow Fields, Proceedings from the Workshop on Vision, Brain, and Cooperative Computation, held at University of Massachusetts, Amherst, MA, 01003, May 9-11, 1983. Also Availiable as a technical report from the Department of Computer Science at the University of Minnesota.

[Tsai82] Tsai, R.Y. and Huang, T.S., Uniqueness and Estimation of Three Dimensional Motion Parameters of Rigid Objects with Curved Surfaces, Proc. of the Pattern Recognition and Image Processing Conference, Las Vegas, Nevada, 1982, pp. 112-118.

[Tsot80] Tsotsos, J.K., Mylopoulos, J., Convey, H.D. and Zucker, S.W., A Framework for Visual Motion Understanding, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume PAMI-2, Number 6, November 1980, pp. 563-573.

[Uhr78] Uhr, L., Recognition Cones and some Test Results, In: Computer Vision Systems, A.R. Hanson and E.M. Riseman, (Eds.), Academic Press, New York, 1978, pp 363-372.

[Ullm81] Ullman, S., Analysis of Visual Motion by Biological and Computer Systems, IEEE Transactions on Computers, August 1981, pp. 57-69.

[Ullm79] Ullman, S., The Interpretation of Visual Motion, Cambridge and London: MIT Press, 1979.

[Webb81] Webb, J.A. and Aggarwal, J.K., Visual Interpretation of the Motion of

Objects in Space, Proc. of Pattern Recognition and Image Processing Conference, Dallas, Texas, August 1981, pp. 516-521.

[Wesz75] Weszka, J.S., Threshold Selection Techniques 5, TR-349, Computer Science Center, University of Maryland, February 1975.

[Whit44] Whittaker, E.T., A Treatise on the Analytical Dynamics of Particles and Rigid Bodies, New York: Dover Publications, 1944.

[Will80] Williams, T.D., Depth from Camera Motion in a Real World Scene, IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume PAMI-2, Number 6, November 1980, pp. 511-516.

[Wong78] Wong, R.Y. and Hall, E.L., Sequential Hierarchical Scene Matching, IEEE Tr. Comp, 27(4), pp. 359-366, 1978.

[Yen83] Yen, B.L. and Huang, T.S., Determining 3-D Motion Parameters of a Rigid Body: A Vector-Geometrical Approach, Proc. ACM Workshop, Motion: Representation and Perception, Toronto 1983, pp. 78-90.

[Zuni83] Zuniga, O. A. and Haralick, R.M., Corner Detection using the Facet Model, IEEE Computer Vision and Image Processing, June, 1983, Washington, D.C., pp 30-37.