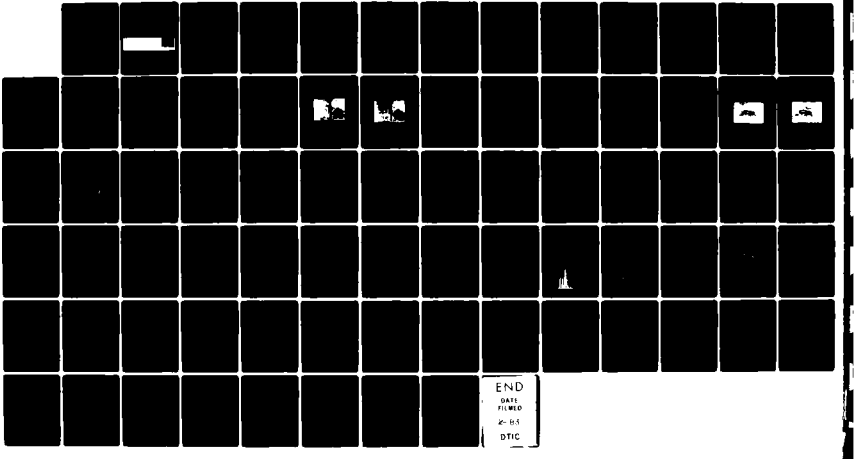
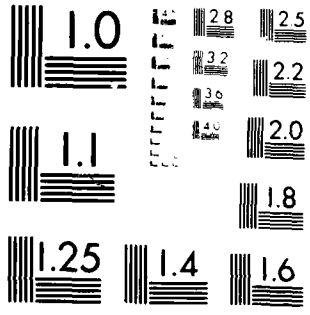
AD-A122 823 PROCESSING TRANSLATIONAL MOTION SEQUENCES(U) 1/1
MASSACHUSETTS UNIV AMHERST DEPT OF COMPUTER AND
INFORMATION SCIENCE D T LAWTON OCT 82 COINS-TR-82-22
UNCLASSIFIED N00014-75-C-0459 F/G 12/1 NL





Computer and Information Science

(12)

PROCESSING TRANSLATIONAL MOTION SEQUENCES

Daryl T. Lawton

COINS Technical Report 82-22

October 1982

ABSTRACT

This paper presents a procedure for processing real world image sequences produced by relative translational motion between a sensor and environmental objects. In this procedure, the determination of the direction of sensor translation is effectively combined with the determination of the displacements of image features and environmental depth. It requires no restrictions on the direction of motion, nor the location and shape of environmental objects. It has been applied successfully to real-world image sequences from several different task domains.

The processing consists of two basic steps: Feature Extraction and Search. The feature extraction process picks out small image areas which may correspond to distinguishing parts of environmental objects. The direction of translational motion is then found by a search which determines the image displacement paths along which a measure of feature mismatch is minimized for a set of features. The correct direction of translation will minimize this error measure and also determine the corresponding image displacement paths for which the extracted features match well.

1. INTRODUCTION

This paper develops a procedure for processing real world image sequences from a translating sensor. The procedure is not dependent on an initial matching process for finding image displacements before inferring environmental depth and camera motion. Instead, the determination of the direction of sensor translation is combined with the determination of the displacements of image features and of environmental depth. No restrictions on the shape of environmental objects are required. The procedure has been applied to real world image sequences under several different operating conditions with robust performance.

The processing consists of two basic steps: Feature Extraction and Search. The feature extraction process picks out small image areas which may correspond to distinguishing parts of environmental objects. The direction of translational motion is then found by a search which determines the image displacement paths along which a measure of feature mismatch is minimized for a set of features. The correct direction of translation will minimize this error measure and also determine the corresponding image displacement paths for which the extracted features match well.

The feature extraction process, which is presented in section

two, finds distinctive points along image contours determined by simple processes which are sensitive to image edges. It finds features which are positioned at points of high curvature along these contours. The technique utilizes contours determined by zero-crossing extraction and image thresholding.

The search process, which is presented in section three, minimizes a simple error measure. This error measure is defined with respect to 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. The error measure thus associates with a point on the unit sphere, corresponding to a particular translational axis, a number describing the total extent of 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 neighboorhood about the correct translational axis. Because of this, the search process can be quite simple.

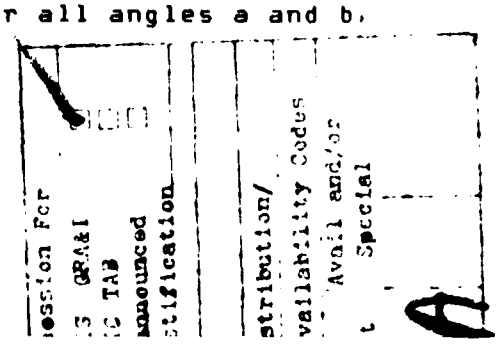
The fourth section presents several experiments showing the results of applying the procedure in several different situations. The experiments indicate that the procedure is very robust and applicable to a wide range of real world image sequences.

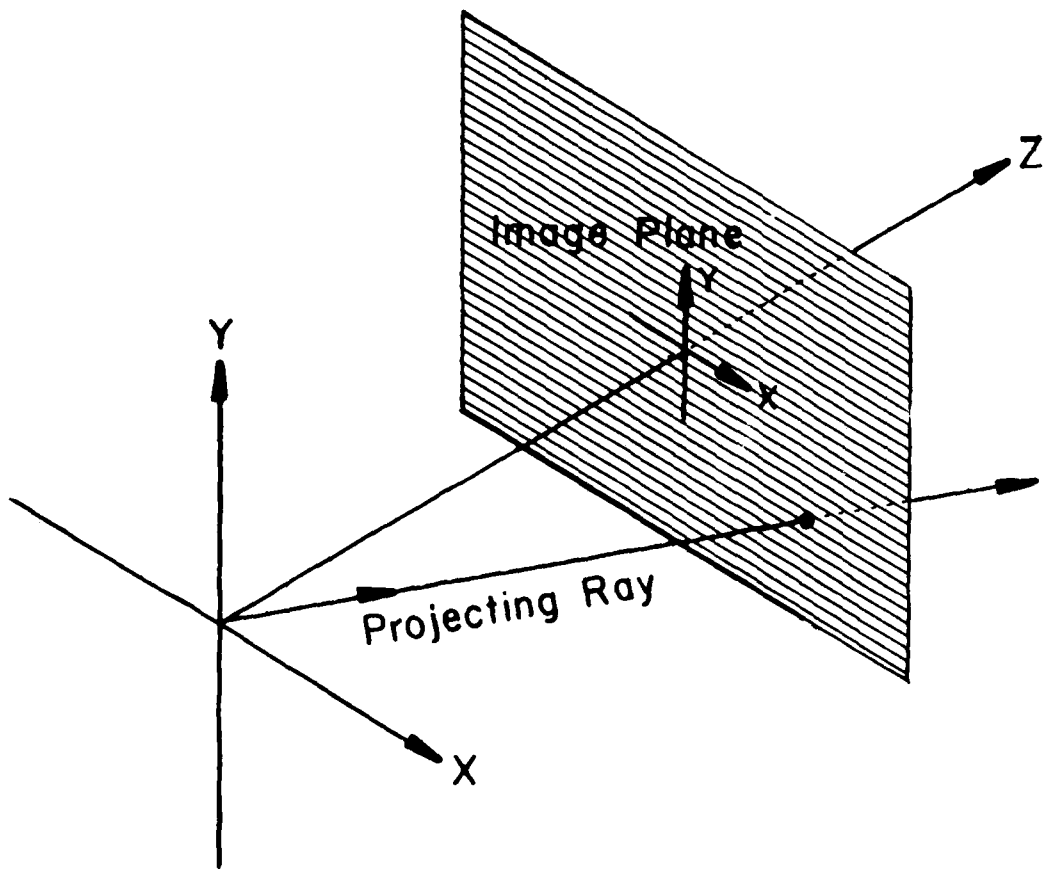
The fifth section discusses various aspects of the procedure and outlines some applications and extensions.

1.1. Coordinate System

The camera model 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 1). 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 at some distance along the Z axis. Positions in the image plane are described using a 2-d coordinate system aligned with the X and Y axes of the camera coordinate system and with the origin determined by the intersection of the image plane and the Z-axis.

The axes of translation are unit vectors based at the origin of the camera coordinate system and are described by two angles (Phi1, Phi2) (figure 2). For a unit vector, $\underline{V}$, based at the origin, Phi1 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 $\underline{V}$. Phi1 thus specifies a plane containing the X axis. Phi2 is the angle between (-1,0,0) and $\underline{V}$. Note that for all angles a and b, (a,0)=(b,0) and (a,PI)=(b,PI).





Camera Model

Figure 1.

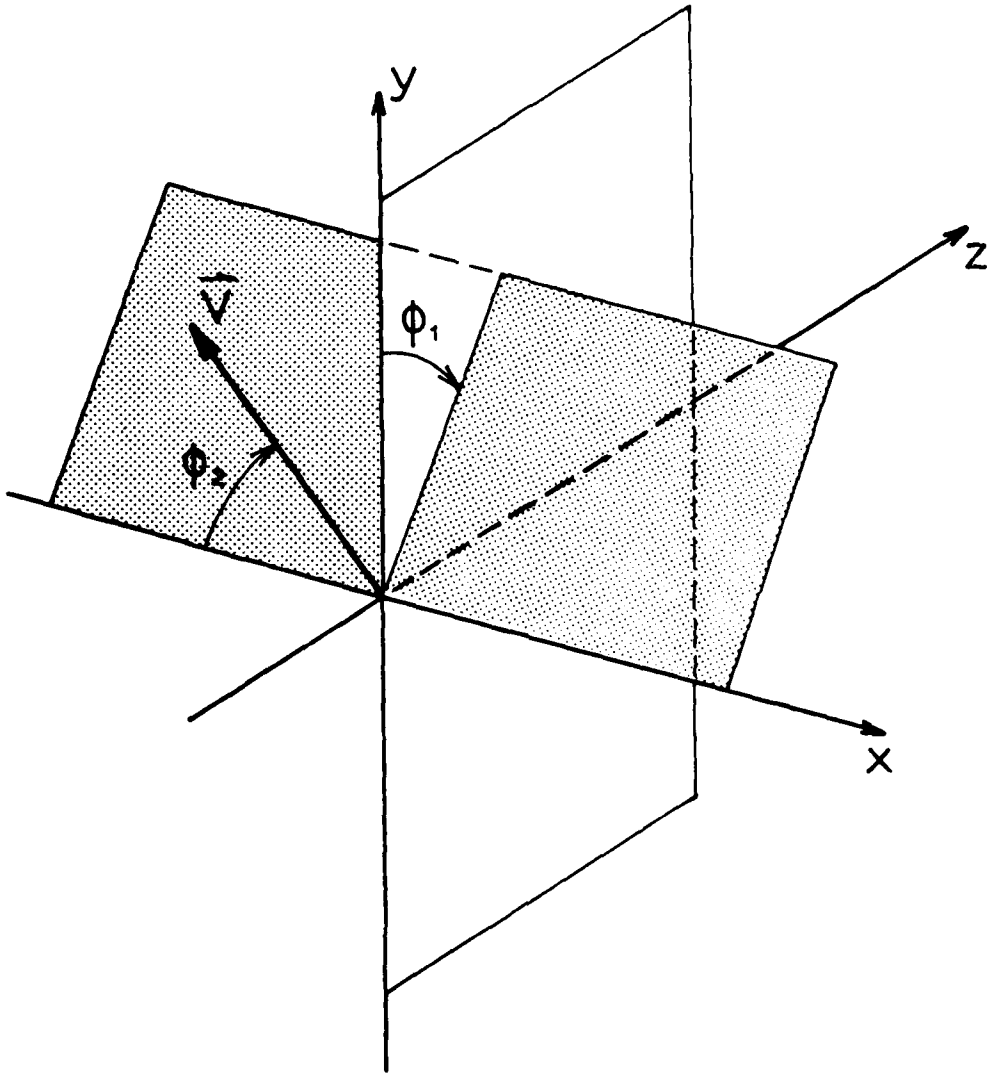


Figure 2. (Phi1, Phi2) Coordinates

1.2. Translational Motion Properties

It is necessary to have a set of 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 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 descibe 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 environmental points are moving along for particular motions.

For general camera motion, there are 5 parameters [PRA80, PRA81] that can be recovered from processing image motion without knowing absolute camera displacement or velocity

(since absolute depth is lost): two parameters for the unit vector (T1(t), T2(t)) which describes the axis of translational motion at time t; two parameters for the unit vector (R1(t), R2(t)) describing the axis of rotation at time t; and one parameter R3(t) which describes the extent of rotation about this axis at time t. Both of these axes are positioned at the origin of the camera coordinate system. The problem of processing image motion resulting from rigid body camera motion can be organized into subcases of increasing complexity, corresponding to the number of camera motion parameters that are unconstrained.

For purely translational motion, 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 camera motion towards 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 it. This corresponds to camera motion away from environmental points. The intersections of axes parallel to the image plane are points at infinity and are treated as FOEs.

Given the direction of translation and image displacements, the relative depths of points can be computed by solving the inverse perspective transform [ROG76]. Relative depth can also be 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

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

where Z is the value of the Z component of an environmental point at time t+1, del Z 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 del D is the image point's displacement from time t to time t+1. Thus, the Z value of an environmental point can be recovered from image measurements in units of del Z, or what has been termed Time-Until-Contact by Lee [LEE76].

The set of all possible translational axes describes a unit sphere called the translational direction sphere. The procedures below are defined with respect to this sphere, rather than the image plane itself, for reasons described in section 5.2.2.

1.3. Previous and Related Work

Gibson[GIB50, GIB66] was among the first to study the importance of the structure of optic flow fields in the

control of egomotion. He also pointed out the potential importance of the FOE in the translational case. This work was extended by Purdy and by Lee [LEE76, LEE80]. Lee analyzed the computation of important control information during translational motion, such as time-until-contact, braking information, and environmental depth in a natural coordinate system. However, recent work by Beverly and Regan [REG78, REG79] indicates that an alternative mechanism may be used in humans for determining the direction of translation than extraction of a FOE or FOC.

Work in processing dynamic images [HUA81, MART79, NAG81a, THO81, ULL81] can be roughly divided into a set of techniques for determining the changes in a sequence of images and a set for interpeting these transformations. Determining image motion has involved work in change detection, correlation based matching techniques [QUA71, HAN74, LEV73], relaxation based matching techniques [BAR79, PRA80], region matching [NAG78, RAD81]; image differencing [JAI81]; and spatial-temporal analysis of the image gradient under constraints of locally uniform motion [THO80] and smoothly varying image motion [HOR81, GLA81].

The interpetation of image motion can result in a variety of descriptions such as determination of the occurrence and location of change [MAR79]. These include image segmentations based on common motion [THO80]; the recovery of camera motion parameters and the shape of environmental objects; and more

qualitative descriptions at a level compatible with natural language descriptions [BAD76, TSO80, O'ROU80]. Particularly relevant here is work with the recovery of camera motion parameters from an image sequence. Ullman [ULL79] demonstrated the minimal number of observable points necessary to obtain a solution over time. Roach and Aggarwal [ROA80] investigated noise robust computations of the camera motion parameters. Prazdny developed techniques for decomposing optic flow fields [PRA80] and displacement fields [PRA81] into their rotational and translational components. The latter result has been given a particularly simple algebraic formulation by Nagel [NAG81b]. Tsai and Huang [TSA81] have investigated the simplifications in determining camera motion parameters by restricting the interpretation to planar surface patches.

Williams [WIL80] was the first to develop algorithms for interpreting complex natural 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 translational axis, could be combined.

The primary weakness of Williams' work was the necessary restriction to planar surfaces at one demonstrated orientation. Additionally, the processing is quite complex when neither environmental depth nor the direction of translation is known -- involving segmentation, resegmentation, and coordinating the processes for inferring depth and for inferring the direction of translation. The method presented here 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.

The determination of the vanishing point in a static image is closely related to determining the direction of translation because the FOE is the dynamic analogue of the vanishing point. In perspective projection, parallel lines in the environment map onto lines radiating from a 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 developed by Kender [KEN79] and Kitihashi [KIT80]. The use of the Hough transform in this work is similar to the global sampling of the error measure used below.

2. FEATURE EXTRACTION

The feature extraction process is used to determine small areas (sometimes called image points or features) in an image that are distinct from neighboring areas. This distinctiveness limits the potential matches of these image areas, and possibly reflects a correspondence to actual and significant points in the environment, such as points of high curvature on object boundaries, texture elements, surface markings, etc. (However some features, termed *false features*, will result from noise, occlusion, and light source effects and have behavior which is currently difficult to interpet). Features can be represented either as arrays of numbers extracted directly from an image, or as parameterized tokens describing local image properties. In this paper, we refer to features exclusively as small arrays of data values centered at some point in an image at some time t.

Following Moravec [MOR77,MORBO], the method of feature extraction used here is based upon finding image areas which are significantly different than their neighboring areas. Using a correlation measure 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 (see the correlation measures in section 3.1.). Selecting good features then requires finding the local maxima in the values

of the distinctiveness measure over an image

We have extended this approach somewhat by constraining the neighborhoods over which the features are selected to contours determined by other global processes, such as zero-crossing extraction and thresholding, which are sensitive to edges

2.1. 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. [HIL80, MAR80] In this processing an image is convolved with Gaussian-Laplacian masks ((del)**2g) 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 of the 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

For features which are local distinctivness maxima, we approximate the curvature along the contour by the inner product of the normalized vectors describing the relative positions of adjacent local maxima along the contour (figure 3). These values are then thresholded between 1.0 (corresponding to high curvature) and -1.0 (corresponding to low curvature).

Use of zero-crossing based features requires specification of the sizes of the convolution masks that are employed, and deciding 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. The processing described below can be applied independently to the pairs of successive images formed by convolving the successive images with two such masks. Alternatively, 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.

The images in figure 4a and figure 4b were taken from a gyroscopically stabilized movie camera held by a passenger in a car traveling down a country road in Massachusetts. They are 128x128 pixel images with 6 bits of resolution in intensity and will be referred to as the roadsign images.

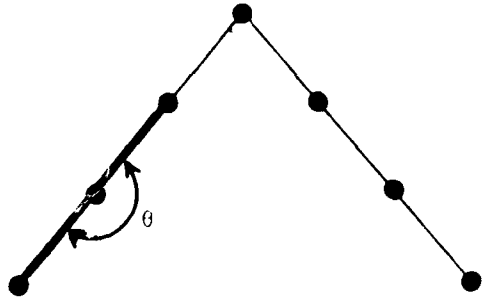


Figure 3. Curvature Approximation for Features Along an Extracted Contour

Figure 4c shows the zero-crossings extracted from the initial ROADSIGN image using a (del)**2g mask with a 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 (see section 3.1), with their 8 immediately neighboring features. The distinctiveness measure for a feature was set to 1 minus the best correlation obtained in its neighborhood, excluding itself. Figure 4d shows the local maxima in the distinctiveness measure positioned with respect to the zero-crossing contour. Note that the features are centered on pixels adjacent to the contour and not on the contour itself. Figure 4e shows the results of suppressing low-curvature points using a threshold set to -0.8 (corresponding to an angle of 143.13 degrees).

2.2. Feature Extraction Using Threshold Contours

Image contours can also be determined by thresholding. The values of the threshold can be determined in a variety of ways, such as using fixed increments, finding peaks and valleys in the image intensity histogram, or using techniques sensitive to image contrast produced using a particular threshold [KOH81, WES75].

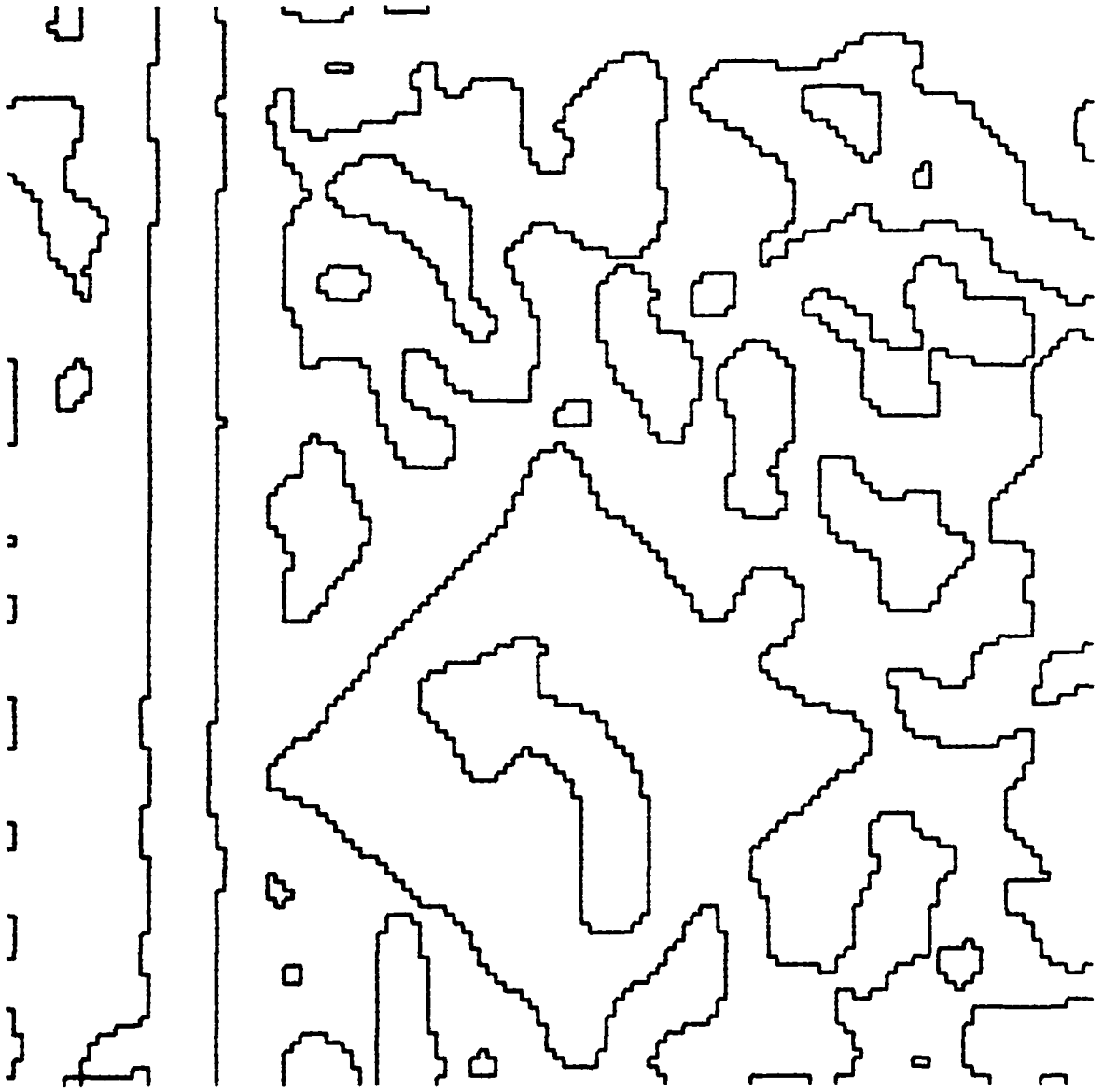


Figure 4c. Extracted Zero-Crossings of Roadsign Image 1

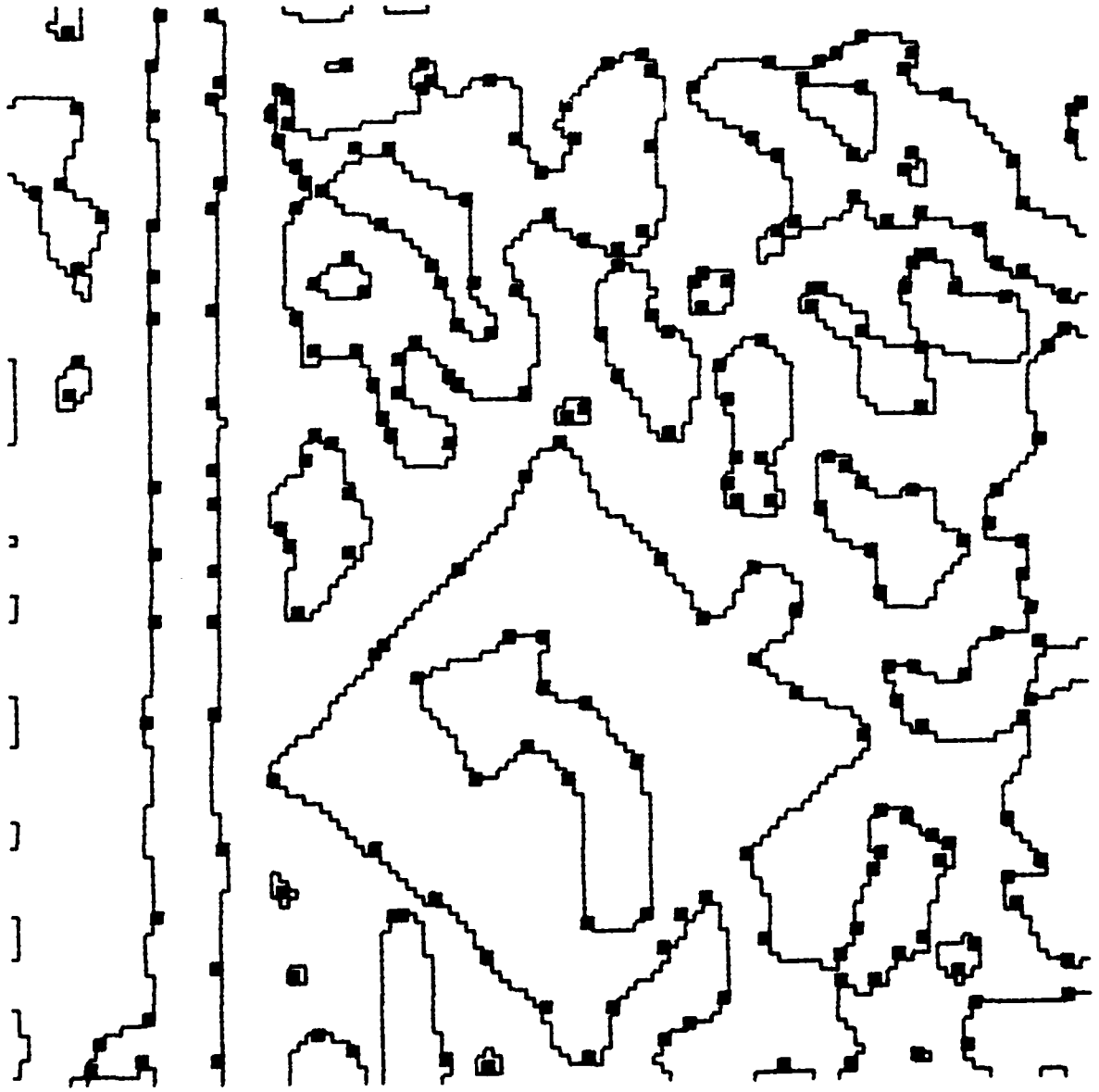


Figure 4d. Distinctive Feature Positions with Respect to Zero-Crossing Contour

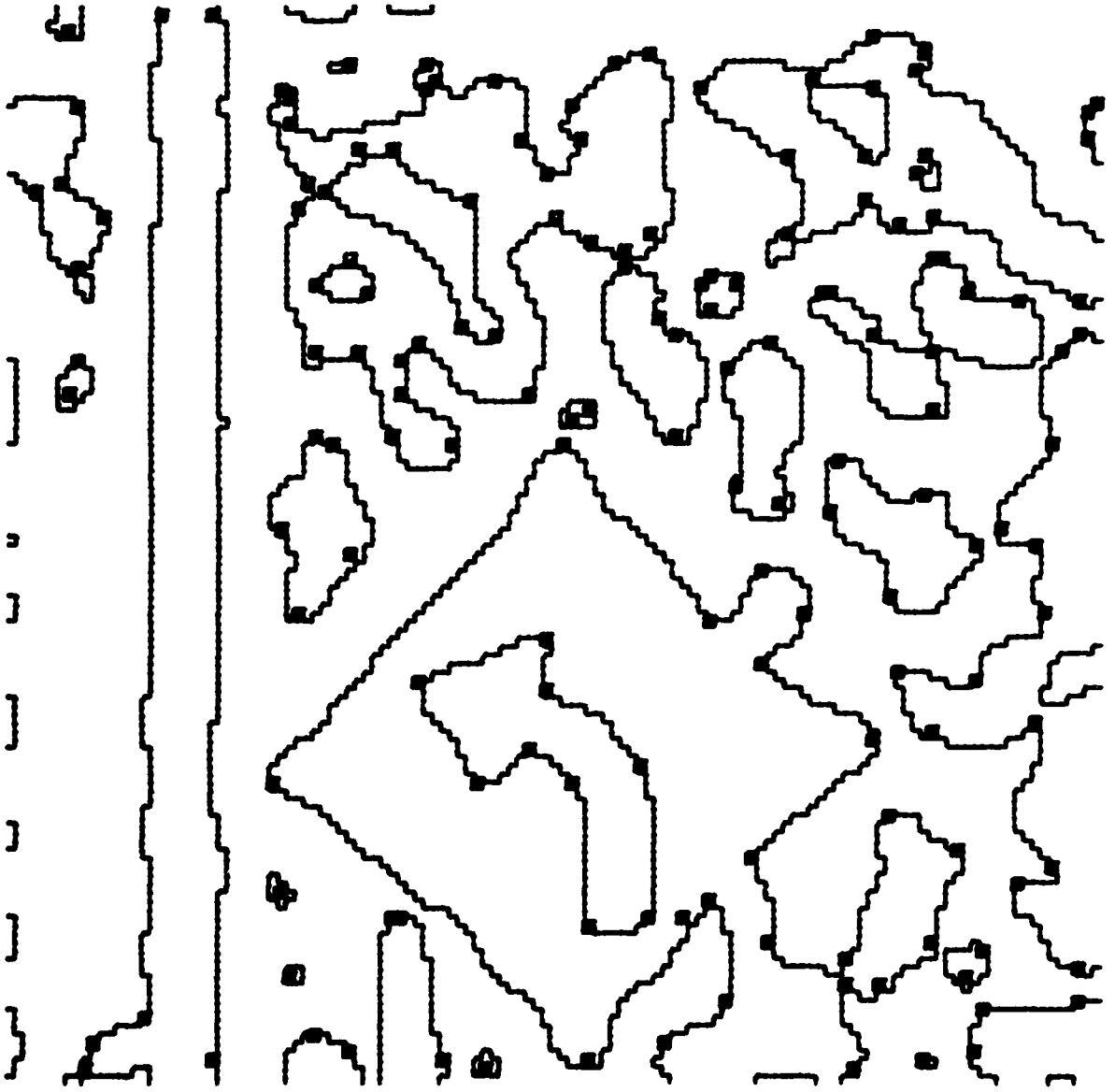
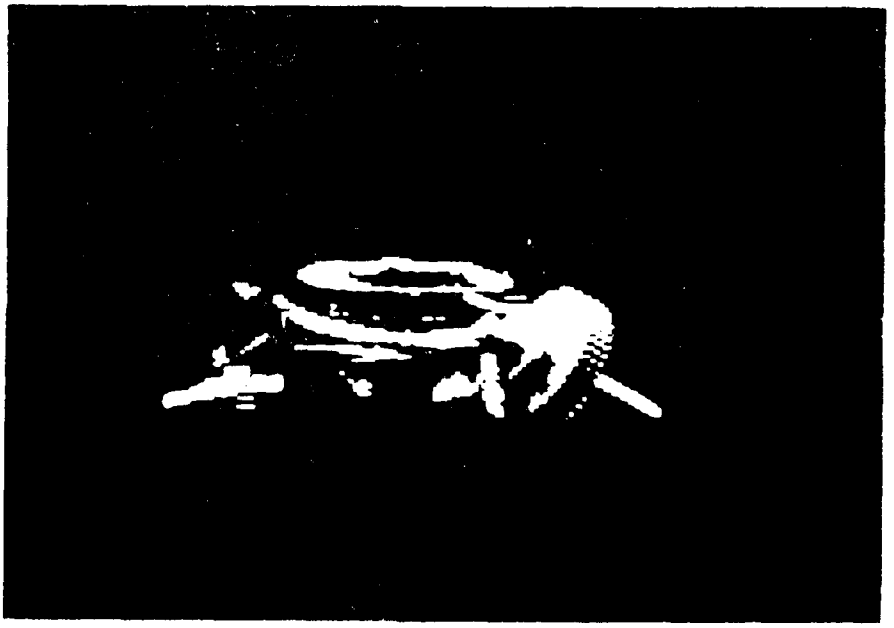
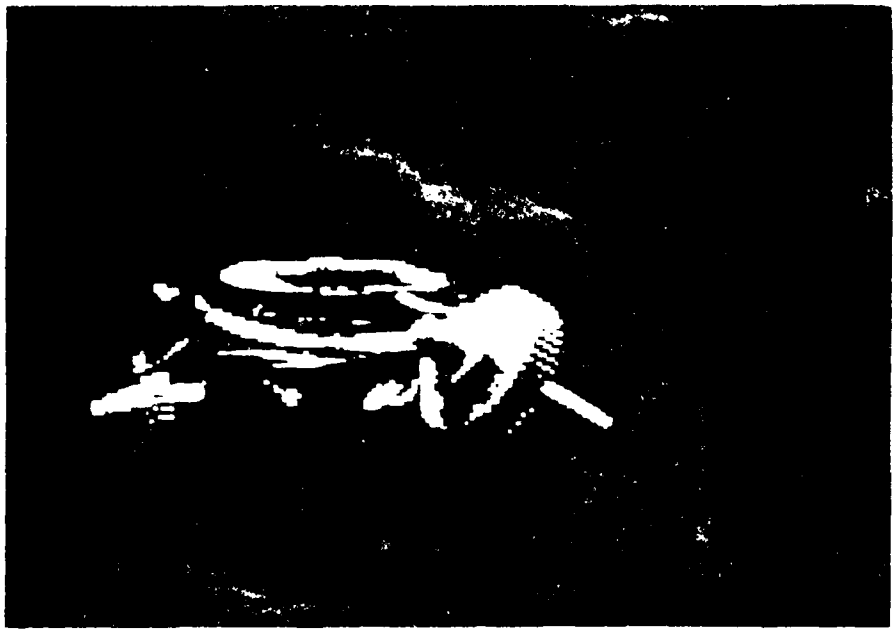


Figure 4e. Distinctive Feature Positions After Low-Curvature Suppression

The images in figure 5a and 5b were produced from a solid state camera held by a robot manipulator translating toward some industrial parts lying on a table. The images were initially 512x512 pixel images with 7 bits of intensity resolution and were averaged down to 128x128 pixel images with 6 bits of intensity resolution. These will be refered to as the industrial images. Analysis of the image intensity histogram, using the procedures described in [KOH81], determined a clear break in the histogram at an intensity level of 10 in the image. This corresponded to separating the dark background and the brighter objects in the scene. Figure 5c shows the extracted contour and figure 5d 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 5e shows the extracted feature points after low curvature suppression using a threshold set to -0.8 (corresponding to angle of 143.13 degrees).





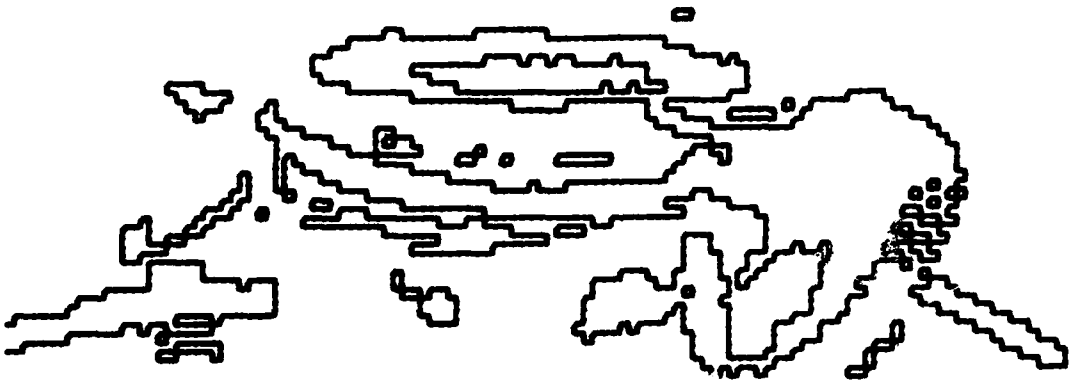


Figure 5c. Extracted Threshold Contour of Industrial Image 1

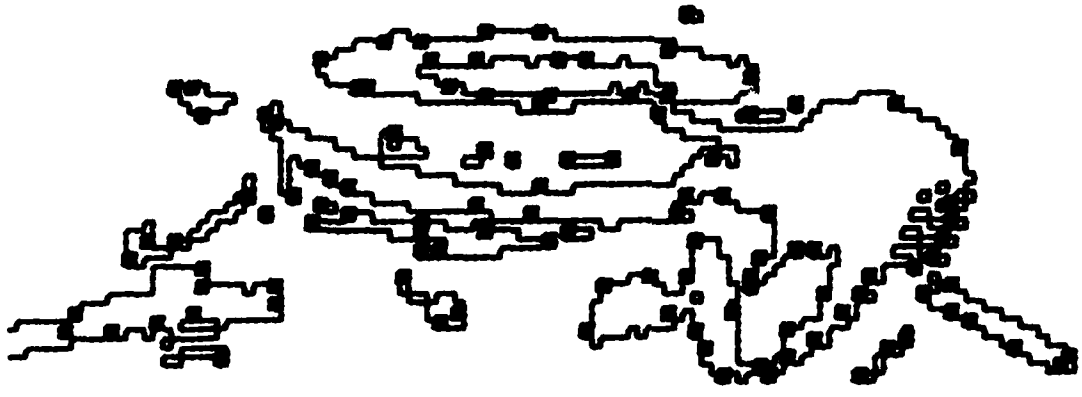


Figure 5d. Distinctive Feature Positions with Respect to Threshold Contour

Figure 5e. Distinctive Feature Positions After Low-Curvature Suppression

3. SEARCH PROCESS

The search process minimizes an error measure which describes the extent of mismatch for extracted features along the image displacement paths determined by a hypothesized translational axis. For example, figure 6 shows an FOE determined by a potential translational axis and the image displacement paths it determines for some extracted features. Also shown is the match profile for a particular feature along a segment of its displacement path with respect to features positoned in the succeeding image The adequacy of a proposed translational axis is measured by the strength of the matches that the extracted features have along these paths. This suggests finding the best match for each feature along the image displacement path determined by a translational axis and then summing the extent of error in these best matches for the error measure.

Developing this error measure requires a measure for the degree of match between features and an interpolation process for determining positions along an image displacement path. Each of these can be implemented in various ways with the choices generally involving a trade-off between the speed of evaluating the error measure and the precision with which the translational axis can be determined.

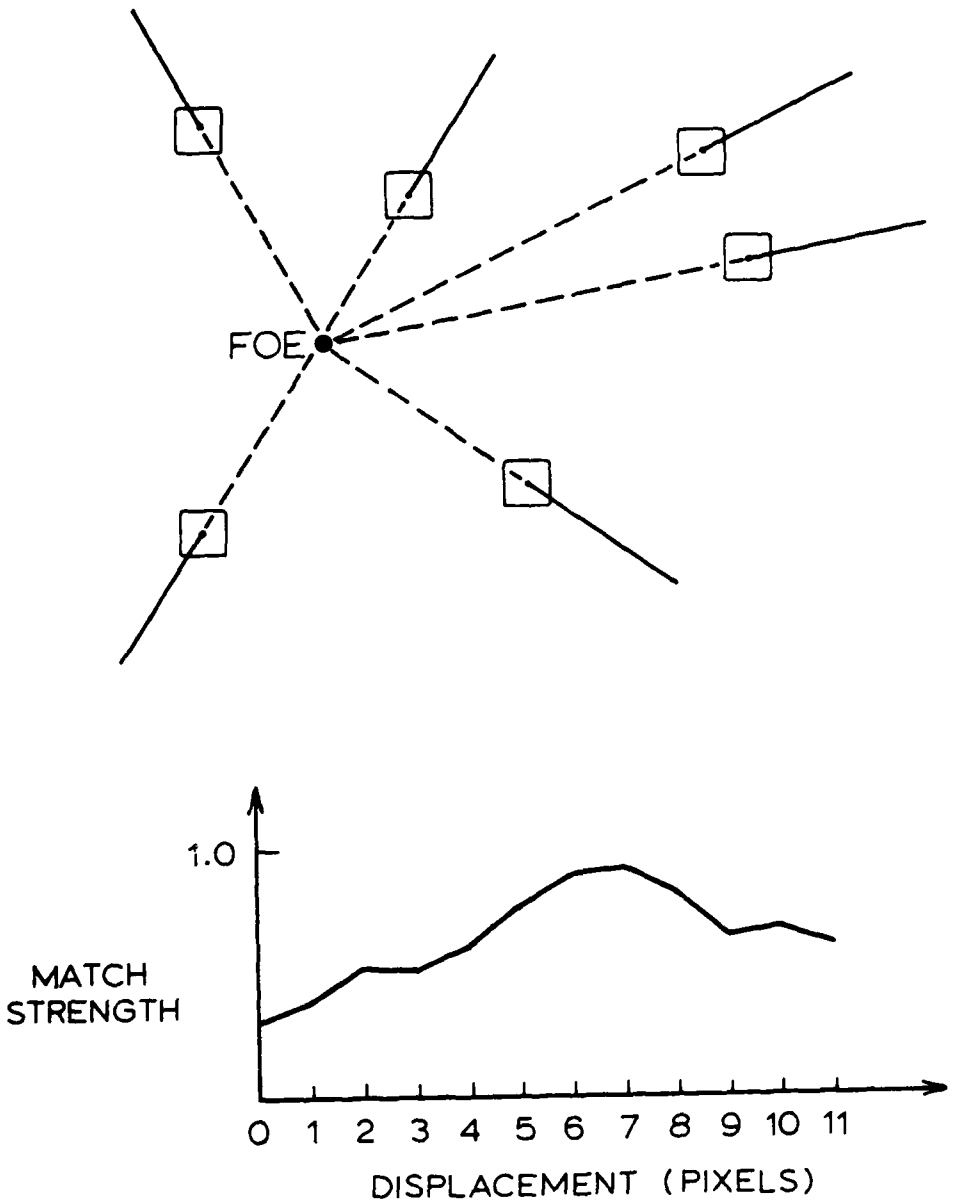


Figure 6. Feature Displacements for a Potential Translational Axis

3.1 Match Metric

There are several metrics for similarity of nxn pixel features of the form A(i,j) and B(i,j), where i ranges from 1 to n and j ranges from 1 to n. We have utilized:

Normalized Correlation

$$\frac{\sum_i \sum_j A(i,j) \times B(i,j)}{\sqrt{\sum_i \sum_j A(i,j) \times A(i,j)} \times \sqrt{\sum_i \sum_j B(i,j) \times B(i,j)}}$$

Moravec Correlation [MOR77]

$$\frac{\sum_i \sum_j A(i,j) \times B(i,j)}{(((\sum_i \sum_j A(i,j) \times A(i,j)) + (\sum_i \sum_j B(i,j) \times B(i,j)))/2)}$$

Normalized Absolute Value Difference

$$1.0 - \left(\frac{\sum_i \sum_j \text{abs}(A(i,j) - B(i,j))}{\sum_i \sum_j A(i,j) + \sum_i \sum_j B(i,j)} \right)$$

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

3.2. Interpolation Process

The interpolation process approximates the potential displacements of a feature from an initial image into a succeeding image. 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 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 a more expensive computation.

3.3. Error Measure

The error measure associates with a point on the direction of translation sphere a number describing the quality of feature matches along the image displacement paths determined by the corresponding translational axis. This error value is computed by first finding the best match for each feature along a segment of the image displacement path determined by the translational axis using one of the normalized 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 is

$$\sum_{i=1}^{n} (1.0 - \text{bestmatch}(i))$$

where bestmatch(i) is the best match value associated with feature i along the image displacement path determined for it by the hypothesized translational axis.

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 is not adequate for fine determination of the translational axis because it does not vary smoothly with respect to small changes in the position of a translational axis, due to the nearest integer approximation for feature position.

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

3.4. Search Organization

The search process used here consists of two phases. A global sampling of the error measure determines the rough shape of the error surface, then a local search determines the minimum. The local search begins at the position where the minimum value was determined by the global sampling. The procedure used for the local search is steepest descent with a diminishing step-size. That is, the steepest descent procedure begins with a initial fixed step size and determines a local minimum using it. The step-size is then reduced and the prodedure 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.

4. EXPERIMENTS

The following experiments were performed using the roadsign and industrial image sequences introduced earlier. They cover a wide range of situations. The first experiment involves determining the translational axis for the industrial image sequence using the features indicated in figure 5e. 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 small number of features that are positioned across the initial image of the sequence. The third experiment involves processing the roadsign image sequence using the features extracted at the positions indicated in figure 4e from the initial, unconvolved image. In this sequence 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 in the indicated figures.

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 PI/10 or 0.31416 radians or 18 degrees) 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. Recall that the Phi1 coordinate determines a plane containing the X-axis of the camera coordinate system and Phi2 refers to positions of unit vectors in such a plane.

The third table (table c) shows the minimal value determined by the global sampling process 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 (Phi1,Phi2) coordinates, so that translational axes computed under different situations can be compared.

4.1. Industrial Images

The procedure was applied to the industrial images using the

features extracted at the positions shown in figure 5e. Tables 1a and 1b show the global sampling of the error measure using the fast form of evaluation. Note the minima at Phi1=1.57080 and Phi2=1.25660. Table 1c shows the successive values of the local search using the precise form of evaluation. The determined translation axis is (-0.13875, -0.09887, 0.98538). The image displacements determined for these features is shown in figure 7.

4.2. Industrial Images with Selected Features

The procedure was again applied to the industrial image sequence but using features which were selected by hand from those indicated in figure 5e. The positions of these 8 features are shown in figure 8.

Tables 2a and 2b show the global sampling of the error measure using the precise form of evaluation. Note the minima at Phi1=1.57080 and Phi2=1.57080. Table 2c shows the successive position determined by the local search. The determined translational axis was (-0.15438, -0.07896, 0.98485). This corresponds to an angular difference of 0.0253 radians (1.4505 degrees) with respect to the axis determined in experiment 1.

Phi

Phi1										
		3.31416	3.62832	3.94248	4.25660	4.57080	4.88500	5.19910	5.51330	5.82740
		25.424	26.062	27.148	28.062	28.062	26.521	25.297	24.175	23.177
	3.31416	26.754	26.887	27.986	28.177	27.171	27.266	26.456	27.209	24.206
	3.62832	26.155	26.227	27.602	28.043	26.817	26.209	27.126	25.328	23.659
	3.94248	25.332	25.499	26.705	26.454	25.610	25.673	24.946	24.750	23.697
Phi1	4.25660	24.117	24.965	25.170	23.241	23.650	23.333	24.071	24.196	24.248
	4.57080	21.454	21.439	19.520	18.446	15.756	20.525	24.219	24.477	24.824
	4.88500	21.043	20.658	20.777	18.119	17.381	19.940	22.706	23.856	24.564
	5.19910	21.142	22.513	21.862	23.551	24.748	23.095	21.891	23.007	24.217
	5.51330	22.188	22.494	24.227	25.629	26.397	24.880	25.262	22.203	23.477
	5.82740	22.970	24.087	24.892	26.238	26.221	26.078	25.303	23.244	23.850

Table 1a

Phi2

Phi1		3.14159	3.45575	3.76991	4.08407	4.39823	4.71238	5.02654	5.34070	5.65486	5.96902
	3.00000	26.913	22.906	26.392	27.192	28.167	28.844	29.112	30.892	24.617	25.155
	3.31416		25.424	26.062	27.811	27.751	27.713	27.813	28.196	27.262	26.274
	3.62832		26.099	26.881	28.187	27.997	27.423	29.109	28.914	27.709	25.778
	3.94248		26.983	27.983	28.517	28.376	28.326	29.757	29.693	27.515	25.245
Phi1	4.25660		26.715	27.722	29.885	29.427	30.553	29.348	30.230	27.065	25.848
	4.57080		26.746	27.906	29.862	33.978	32.836	30.006	27.987	26.414	26.267
	4.88500		27.626	27.344	30.042	32.988	32.165	30.975	27.766	26.423	25.844
	5.19910		26.550	27.666	31.320	31.132	31.693	30.682	28.436	28.506	26.57
	5.51330		26.490	29.295	30.307	31.056	30.161	29.710	27.446	27.652	27.196
	5.82740		26.146	28.141	31.284	29.139	28.944	28.464	27.405	27.139	27.883

Table 1b

Phi1	Phi2	X	Y	Z	Error	Stepsize
1.5708	1.2566	-0.30905	0.00000	0.95105	*14.446*	
1.6708	1.4566	-0.11395	-0.09911	0.98852	3.5456	0.1
1.6708	1.4316	-0.13875	-0.09887	0.98538	3.5313	0.025
1.6708	1.4316	-0.13875	-0.09887	0.98538	3.5313	0.025

Table 1c

NOTE: The * denote that this value was determined using the fast form of evaluation. The others were determined using the precise form.

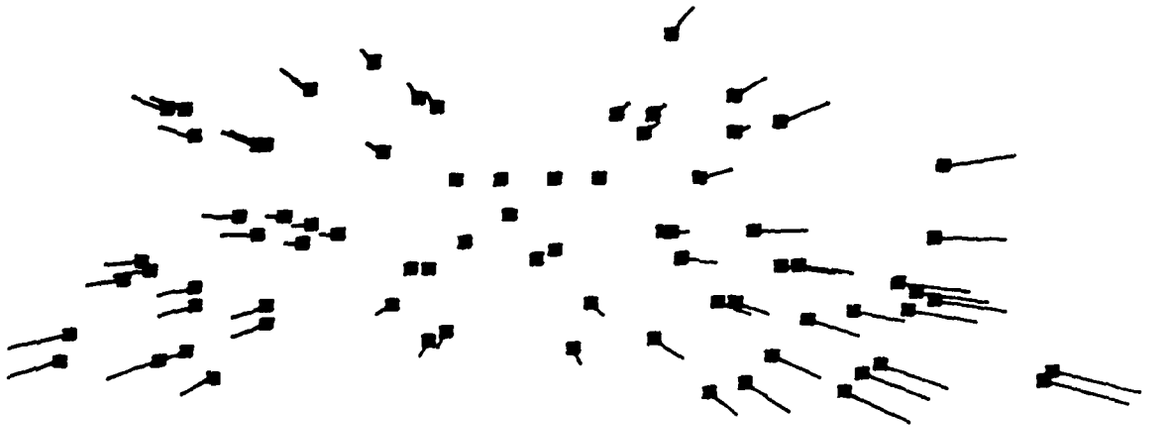


Figure 7. Image Displacements for Industrial Image Sequence

Figure 8. Selected Features for Industrial Image Sequence

Ph1

Table 1a

Ph1.2

Table 1

Phi1	Phi2	X	Y	Z	Error	Stepsize
1.5708	1.3708	1.00000	0.00000	1.00000	0.57998	
1.6708	1.3708	-0.13867	-0.09785	0.97517	0.19955	1.1
1.6458	1.4208	-0.14043	-0.07401	0.98539	0.17475	1.3
1.6508	1.4158	-0.15438	-0.07896	0.98485	0.17410	1.2

Table 2c

4.3. Roadsign Image Sequence

The procedure was applied to the roadsign image sequence using the features extracted at the positions indicated in figure 4e. Tables 3a and 3b show the global sampling of the error measure using the fast form of evaluation. Note the minima at Phi1=2.51330 and Phi2=0.62832. 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 (-0.83738, -0.42043, 0.34933) The image displacements for the feature points shown in figure 4e consistent with this translational axis are shown in figure 9.

Given the direction of translation and image displacements, the relative environmental depths of image points can be recovered by the simple relation in equation 1 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 equation 1 For this reason, it is necessary to keep track of the image displacements over several successive images with concurrent updating of the inferred depth values. This was done using a sequence of 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 4e The position of the translational axis determined from images I(t) and I(t+1) was used as the initial value in the local search for

		4.9343						6.2471	4.6419	3.2843	3.3776
								9.8010	9.8676	10.000	10.008
								9.7504	9.8316	10.007	9.9926
								9.4924	9.7982	10.020	9.9659
Phi1								8.2094	9.0514	9.8661	9.8952
								8.3138	8.8277	9.4344	9.7710
								7.7161	8.8703	9.1038	9.6692
								7.7497	8.8155	9.1473	9.6037
								8.0255	8.7509	9.0412	9.5052
								7.9031	8.8168	9.1253	9.5460

Table 3a

Phi2

		3.141.9	3.45575	3.76991	4.08407	4.39823	4.71238	5.02654	5.34070	5.65486	5.96902
	0.00000	11.201	6.4869	8.0216	9.2004	10.242	10.942	11.006	11.869	11.920	11.276
	0.31416		10.762	10.298	9.5297	8.1033	7.3353	6.5041	5.0061	3.9616	3.5378
	0.62832		10.956	10.793	9.9145	8.7336	7.2365	6.3875	5.2799	4.2700	3.8493
	0.94248		11.139	11.040	10.623	9.5921	8.2331	7.0199	5.6895	4.6911	4.2697
Phi1	1.25660		11.169	11.196	11.066	10.284	9.3429	8.1211	6.7743	5.2348	4.5115
	1.57080		11.202	11.295	11.331	10.970	10.241	9.0569	7.3829	5.6944	4.9589
	1.88500		11.200	11.335	11.639	11.157	10.921	9.4845	7.9038	6.1586	5.3944
	2.19910		11.203	11.540	11.766	11.742	10.882	9.9751	8.1577	6.8129	5.7575
	2.51330		11.231	11.687	11.919	11.713	10.939	10.347	9.0842	7.7192	6.1577
	2.82740		11.253	11.899	11.951	11.382	10.805	10.317	9.3160	8.0307	6.3139

Table 3b

Phi1	Phi2	X	Y	Z	Error	Stepsize
2.5133	0.62832	-0.40902	[illegible]	0.[illegible]4548	*0.7[illegible]*	
2.5133	0.52832	-0.36366	-0.40782	0.29628	0.21[illegible]1	[illegible]
2.4383	0.57[illegible]32	-0.[illegible]2738	-0.41[illegible]	0.[illegible]5352	0.207[illegible]	0.025
2.4483	[illegible]	-0.33738	-0.42043	0.34[illegible]	0.207[illegible]	[illegible]

Table 3c

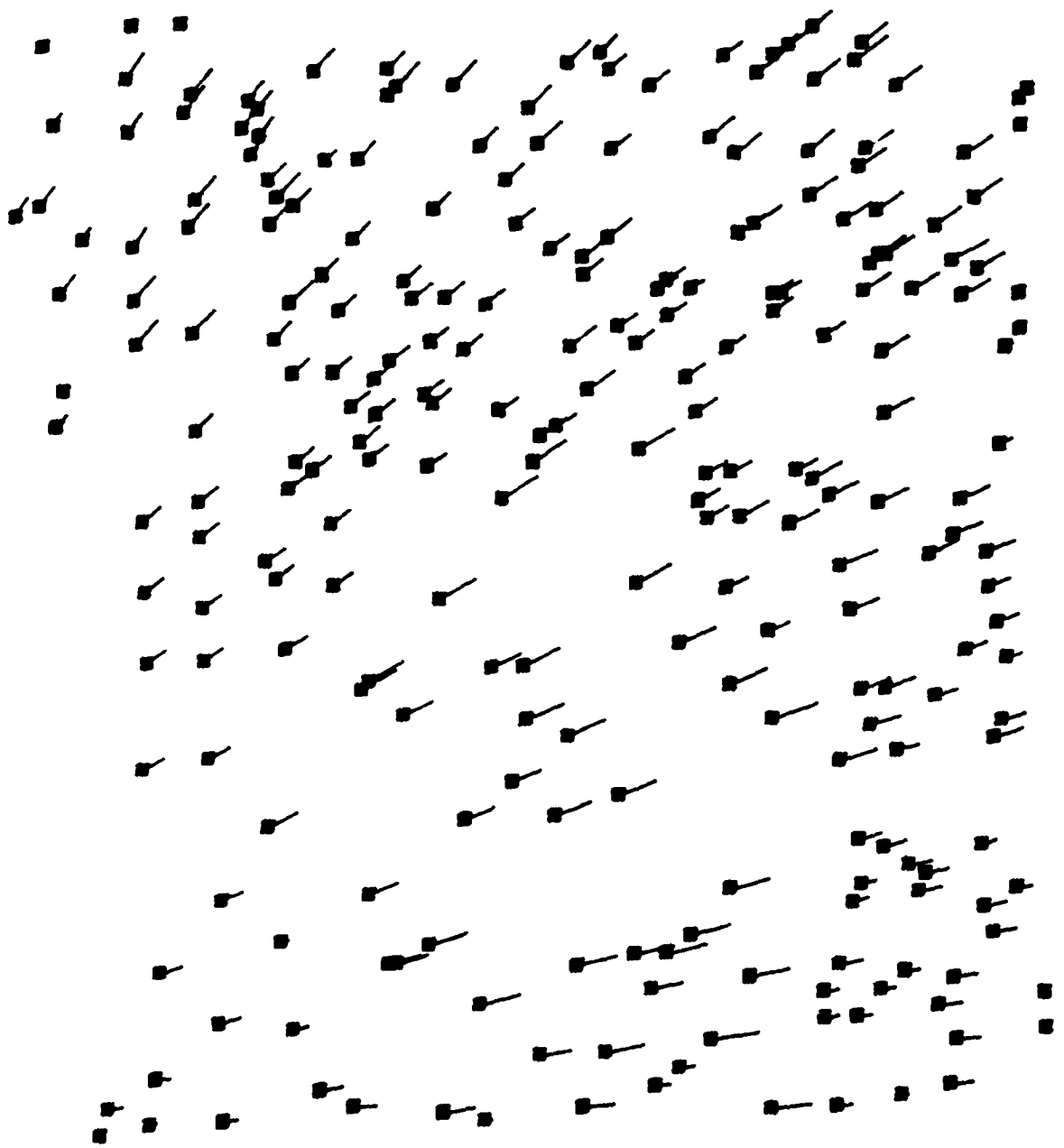


Figure 9. Image Displacements for Roadsign Image Sequence

determing the translational axis for images I(t+1) and I(t+2)

The displacements of all features along the contour in figure 4c were determined along the image displacement paths determined by the translational axis found for images I(1) and I(4). From these displacements the depth values for image points along the contour were computed using equation 1

The roadsign sequence is particularly nice for presenting depth processing results because the three environmental objects in the images are at three distinct depths This is shown in figure 10a 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 cummulative 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 4c yields: the boundary shown in figure 10b (the sign), the boundary shown in figure 10c (the pole), the boundary segment shown in figure 10d (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.

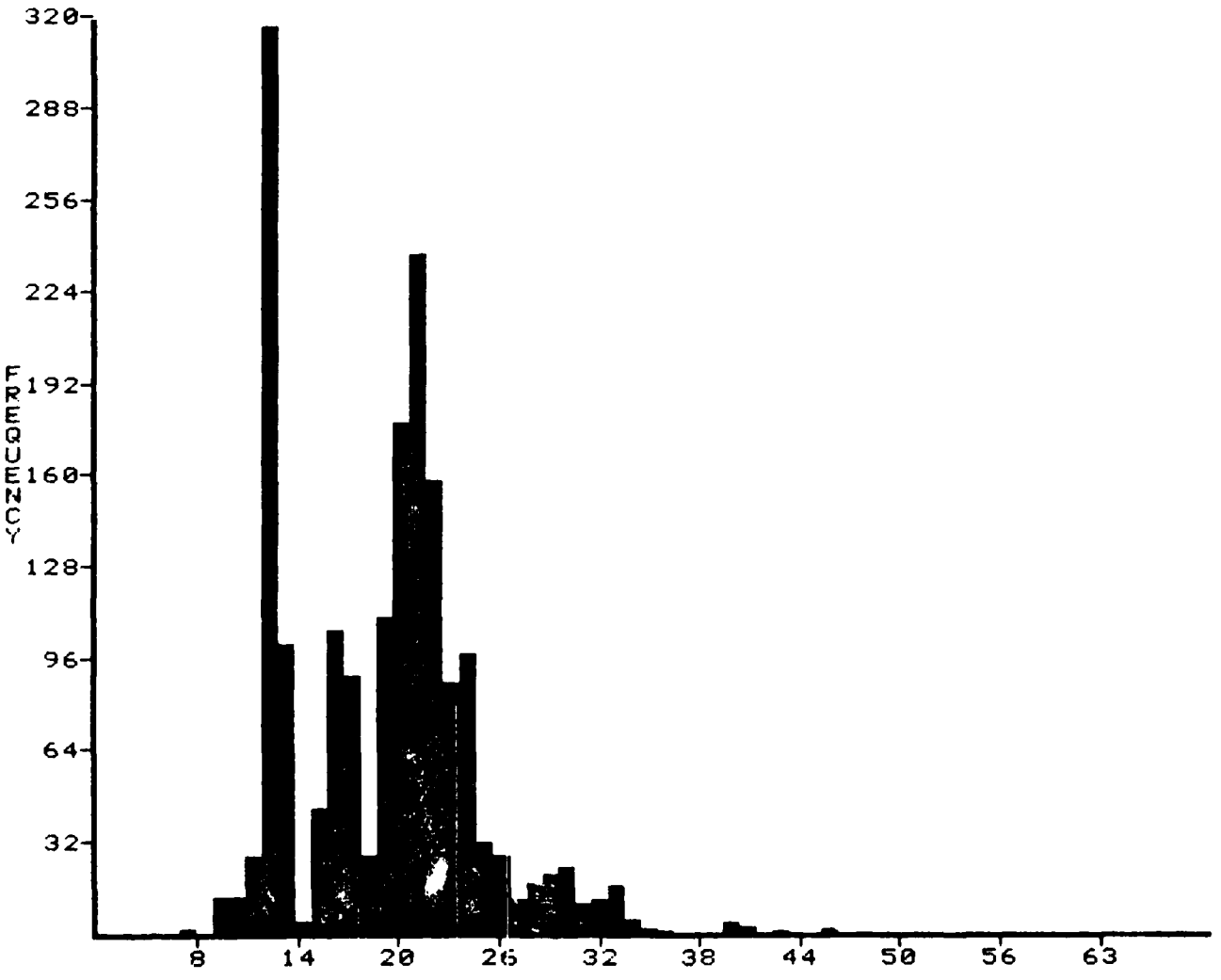


Figure 10a. Depth Histogram

Figure 10b. Image Points in Histogram Cluster Corresponding to the Sign

Figure 10c. Image Points in Histogram Cluster Corresponding to the Pole

Figure 10d. Image Points in Histogram Cluster Corresponding to the Trees

4.4 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 4d. 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 Phi1=2.51330 and Phi2=0.62832. Table 4c shows the successive values of the local search. The determined translational axis was (-0.82909, -0.42281, 0.36585). This corresponds to an angle of 0.0186 radians (1.0676 degrees) with respect to the axis determined in experiment 3.

Phi2

Phi1		[illegible]	0.31416	[illegible]	0.94248	1.25660	1.57080	1.88500	2.19910	[illegible]	[illegible]
	[illegible]	[illegible]	10.070	12.542	14.268	15.464	18.218	18.365	18.909	19.444	19.531
	0.31416		9.9415	12.448	14.147	16.669	18.318	18.343	18.943	19.546	[illegible]
	[illegible]		[illegible]	11.803	13.513	17.074	18.036	18.316	19.146	19.643	19.547
	[illegible]		9.0730	11.006	12.478	[illegible]	16.915	17.996	19.191	[illegible]	[illegible]
Phi1	1.25660		8.4125	9.4122	11.147	14.141	15.852	17.381	18.212	19.317	[illegible]
	[illegible]		7.6057	7.5620	[illegible]	12.676	15.300	16.824	17.854	[illegible]	[illegible]
	1.88500		6.6899	5.7600	8.2645	11.721	13.743	15.953	17.833	18.489	[illegible]
	[illegible]		6.0083	4.8212	6.6746	10.600	13.197	15.689	17.755	18.416	[illegible]
	[illegible]		5.5545	4.7330	6.9714	11.465	13.711	15.145	17.574	[illegible]	[illegible]
	[illegible]		5.5354	5.2061	7.5152	11.846	14.097	15.655	[illegible]	[illegible]	[illegible]

Table 4a

Phi2

Phi1		3.14159	3.45575	3.76991	4.08407	4.39823	4.71238	5.02654	[illegible]	[illegible]	[illegible]
	[illegible]	22.645	21.903	21.164	19.958	15.982	14.731	[illegible]	7.6994	[illegible]	[illegible]
	0.31416		22.057	21.512	19.503	16.250	13.012	10.821	8.1950	[illegible]	[illegible]
	0.62832		22.284	21.998	19.735	16.375	[illegible]	10.640	[illegible]	[illegible]	[illegible]
	0.94246		22.515	22.309	20.703	16.019	13.624	11.444	[illegible]	[illegible]	[illegible]
Phi1	1.25660		22.589	22.523	[illegible]	16.833	15.100	[illegible]	10.437	[illegible]	[illegible]
	[illegible]		22.647	22.736	[illegible]	17.789	15.748	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]		22.626	22.825	22.015	17.379	16.902	14.894	[illegible]	[illegible]	[illegible]
	[illegible]		22.634	23.068	22.176	18.767	16.846	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]		22.650	[illegible]	[illegible]	20.344	17.689	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]		22.681	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Table 4b

Phi1	Phi2	X	Y	Z	Error	Stepsize
0.5133	[illegible]	-0.80902	[illegible]	0.34545	*[illegible]*	
0.5133	0.52832	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
0.4133	[illegible]	[illegible]	-0.42344	0.37755	[illegible]	[illegible]
[illegible]	[illegible]	-0.82909	[illegible]	[illegible]	[illegible]	[illegible]

Table [illegible]

4.5. Roadsign Subimage

The procedure was again applied to the roadsign image sequence with features restricted to the rectangular area shown in figure 11 correponding 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 Phi1=2.19910 and Phi2=0.62832. Table 5c shows the successive values determined by the local search The translational axis is determined to be (-0.84281, -0.42928, 0.32465). This corresponds to angles of 0.0267 radians (1.5341 degrees) and 0.0439 (2.5155 degrees), with respect to the translational axes determined in experiments 3 and 4 respectively.

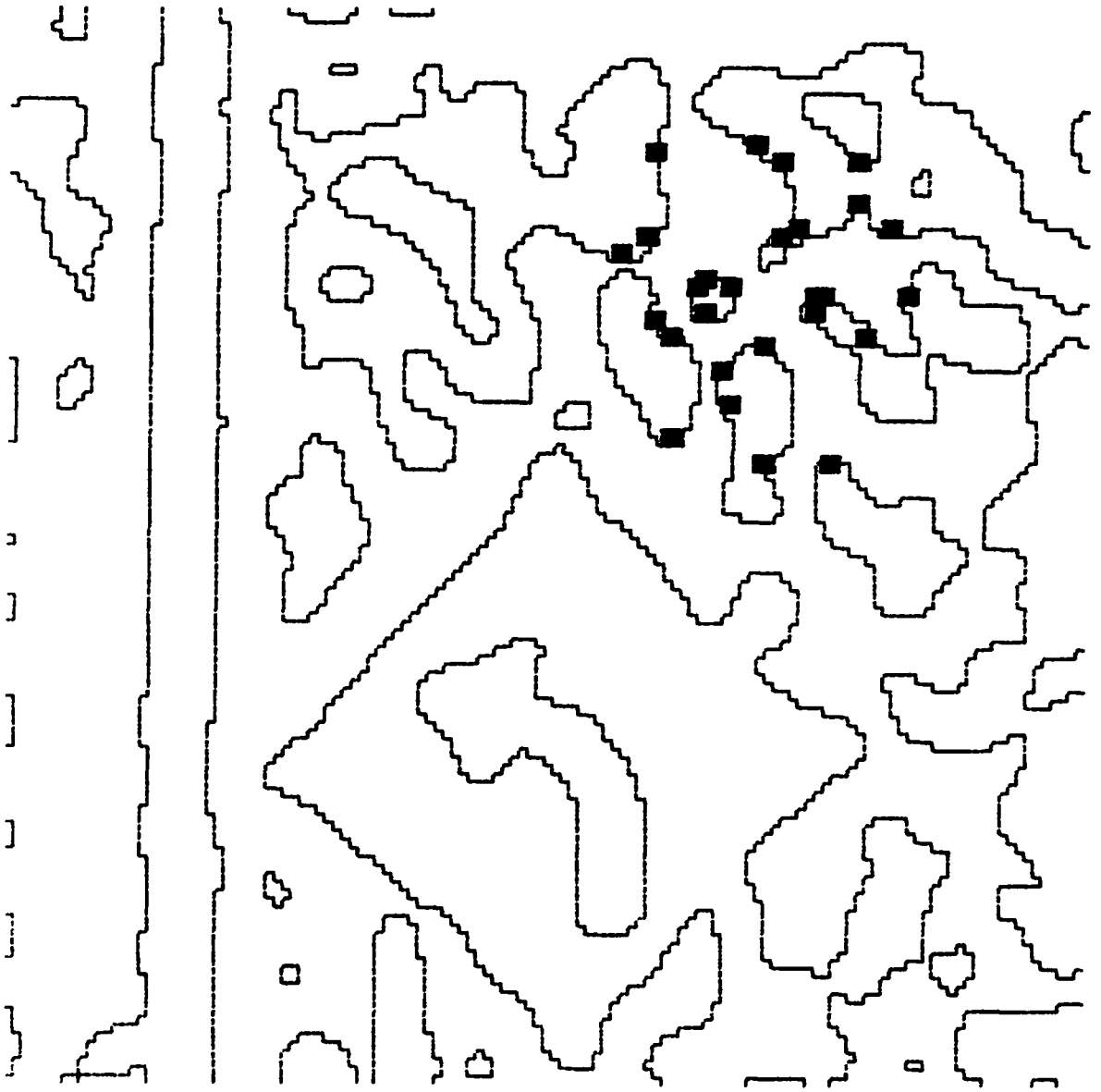


Figure 11. Features in Restricted Subarea of Roadsign Image 1

Phi2

Phi1	[illegible]	0.31416	0.62832	0.94248	1.25660	1.57080	1.88500	2.19910	2.51330	2.82740
[illegible]	[illegible]	0.24906	0.35269	[illegible]	0.48814	0.51525	0.52039	0.50438	0.50165	0.49133
[illegible]		[illegible]	[illegible]	[illegible]	0.45446	0.50514	0.52232	0.50368	0.50000	0.49018
[illegible]		[illegible]	0.25572	[illegible]	0.36971	0.46855	0.51968	0.50173	0.49779	0.48276
[illegible]		0.18373	[illegible]	0.22617	0.26425	0.31561	0.43856	0.48696	0.47778	0.47140
[illegible]		[illegible]	0.14584	0.13334	0.11565	0.12577	0.25988	0.36790	0.44363	0.45981
[illegible]		0.12784	0.09555	0.07340	0.07396	0.11472	0.24358	0.29792	0.36770	0.44106
[illegible]		0.10575	0.06857	0.06536	0.10172	0.16910	0.25533	0.29494	0.33419	0.42238
[illegible]		0.09086	0.05391	0.08389	0.13701	0.20605	0.26458	0.29662	0.32680	0.40534
[illegible]		0.08306	0.06114	0.10538	0.16614	0.23254	0.27384	0.28825	0.32911	0.40005
[illegible]		[illegible]	0.06454	0.11988	0.15651	0.25237	0.27365	0.29237	0.33585	0.40410

Table 5a

Phi2

Phi1	3.14159	3.45575	3.76991	4.08407	4.39823	4.71238	5.02654	5.34070	5.65486	5.96902
[illegible]	0.54847	0.47341	0.39276	0.33847	0.29996	0.27828	0.20602	0.12575	0.06492	0.08459
0.31416		0.49485	0.42774	0.37116	0.31139	0.27888	0.22433	0.10098	0.06289	0.09436
0.62832		0.52376	0.47798	0.44334	0.38013	0.31530	0.23509	0.10064	0.06983	0.11222
0.94248		0.54926	0.53428	0.51542	0.49406	0.45827	0.27597	0.10210	0.11508	0.14224
1.25660		0.57491	0.58046	0.58652	0.58918	0.58793	0.50301	0.32030	0.19392	0.17573
[illegible]		0.58630	0.60316	0.61143	0.61594	0.62691	0.60047	0.48204	0.30378	0.20539
[illegible]		0.59634	0.61354	0.61862	0.62469	0.62639	0.61177	[illegible]	0.38597	[illegible]
[illegible]		0.60470	0.61687	0.62440	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]		0.60822	0.61969	[illegible]	[illegible]	0.64571	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]		0.60908	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Phi1	Phi2	X	Y	Z	Error	Stepsize
2.1941	0.62832	-0.80902	-0.34549	0.47553	0.059910	
2.2991	0.62832	-0.80902	-0.39123	0.43867	0.059542	0.1
2.4741	0.57832	-0.83738	-0.42930	0.33837	0.059288	0.025
2.4941	0.56832	-0.84281	-0.42928	0.32465	0.059269	0.005

Table 5c

5. DISCUSSION

This paper presents a simple and robust procedure for determining the direction of environmental motion and image displacemensts in image sequences produced by a translating sensor. The procedure is robust in several different ways. It is resilent with respect to weak and false features. It 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.

The primary difficulty with real-time implementation is the expense of performing the interpixel interpolation for fine resolution and of performing correlations on features which are arrays of pixels. The computational limitation is the speed with which a feature can determine and evaluate its potential matches given a specified translational axis. These computations can be carried out in parallel amoung the individual features. The expense can be lessened somewhat through the use of higher resolution images (to lessen the need for interpixel interpolation), use of the absolute value norm, and by sampling images at a rate sufficient to limit the extent of feature displacements between images. Additionally, if the direction of sensor translation is changing slowly, information from processing preceeding images can be used to speed up the procedure when applied to succeeding images. It

is also possible to modify the procedure so that feautres are extracted from both images and correlations are performed only at the positions where features have been extracted in the succeeding image. This would probably affect the robustness and resolution of the procedure. We hope that real-time implementation will eventually be possible with special purpose, highly parallel architectures.

In the remainder of this section, we review some specific aspects of the procedure and outline some potential extensions and applications.

5 1. Feature Extraction

Since the procedure's performance is does not degrade severely due to the occurrance of poor features, the type of feature extraction used in probably not critical. Nonetheless, the feature extraction process used here could be extended in many ways. The low-curvature suppression could take into account boundary length along a contour between distinctiveness maxima to determine whether to suppress or generate a feature for further processing. It may also be possible to determine points of high curvature along the boundary with out having to walk along the contour by using operators which can directly measure curvature [KITC80].

Another useful extension would be to use information

determined from the extraction of the translational axis to isolate false features. This could involve removing those features which have weak matches from the error measure calculation once a translational axis has been determined and re-evaluating. Alternatively, the depth inferences could be used to isolate the positions of potential false featuresby noting discontinuities in depth along an extracted contour 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.

5.2. Search Process

5.2.1. Properties of the Error Measure

In the experiments, the error measure has a distinct global minimum at the point on the unit sphere corresponding to the correct translational axis. It is expected to have such behavior generally 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 should not be widely separated.

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 overide the lack of discrimination of weak features and the random quality of the matches of false features.

5.2.2. 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. Additionally, the resolution in the position of the translational axis varies accross the surface of the image plane. For example, the FOEs determined by translational axes seperated by very small angles will be

seperated by larger and larger distances in the plane as the intersections of the translational axes and the image plane are placed further from the visible image. The effect of this on the error measure, when it is defined over the image plane, is large flat areas for FOEs further 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 of the image plane or to an FOC off to the opposite side On the direction of translation sphere, the corresponding translational axes would be close while on the plane they are completely separated.

5.2.3. Optimization Procedure

The optimization procedure used here is very simple, and, because of the strong unimodality of the error measure and its smoothness, other techniques with more rapid convergence could be used. It is interesting to note, however, that the global component of the optimization performed here is an instance of a generalized Hough Transform [BAL81, O'ROU81] inwhich each feature scales its vote for a particular translational axis by the best match it can find consistent with the translational axis.

5.3 Extensions and Applications

5.3.1. Other Cases of Restricted Motion

The procedure developed in this paper is applicable to other cases of unknown but restricted camera motions for which it is computationally feasible to search directly through a subspace of the camera motion parameters to determine feature matches. Two particular cases are pure sensor rotation and motion constrained to a known plane.

For pure sensor rotation, there are three unknown camera parameters. Two for the axis of rotation and one for the extent of rotation about the axis. In this case, the error measure would be defined with respect to a unit sphere inwhich each point corresponds to an axis of rotation. For each rotational axis, the extent of displacement for image features is determined by different rotations about the axis. There is the *additional constraint* in the rotational case that the displacements of all features must correspond to the same extent of rotation.

During arbitrary sensor motion relative to a stationary environment, the image motion due to distant environmental points is primarily due to the rotational component of sensor motion Sensor rotation can be recovered by applying the observer rotation processing procedure to the images of such distant points. The rotation can then be subtracted out to

yield successive images related by sensor translation only. These resulting images can then be processed by the techniques here.

5.3.2. Multiple Independently Moving Objects

The processing here has been limited to a camera moving relative to a stationary environment, or a stationary camera with a stationary background and a single moving object. A useful extension would allow for several independently moving objects with different directions of translation. The technique of summation of errors in feature matching only allows a single axis of translation to be determined and will cause the analysis of the several objects in independent motion to be confounded. Due to the similarity of the global search and a generalized Hough transform noted above, the suggested techniques for decomposing generalized Hough transforms into constituient objects having different parameter values [ADI82, BAL81, O'ROU81] may be applicable.

Another approach is to segment an image into regions which potentially correspond to objects, or to arbitrarily divide the image into regular overlapping subimages and perform the translational analysis for each region or subimage independently [WIL80, NAG79]. Experiments have shown it is possible to work with small image areas, at a size comparable to extracted regions or subimage areas, and still determine

the axis of translation with a reasonable level of precision. If objects with similar translations correspond to several different regions or image subareas, then similar translational axes will be determined for these regions or subimages. If objects with different translations correspond to the same regions or subimages then there will be poor, indistinct error values for the error function. For this second case, it is necessary to resegment and redetermine a translational axis.

5.3.3. Stabilized Retina

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.

A difficulty with such a stabilized retina is that much of the environment would not be observable. 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 directon of translation which can be integrated

5.3.4. The Local Translational Decomposition

This technique can be extended to less restricted forms of sensor motion by applying the procedure for translational motion to small, overlapping areas across an image surface over a sequence of images. This approximates more general motions as consisting locally of environmental translations and interpets local image motion as resulting from environmental translations. The feasibility of this is based upon experiments showing that the direction of translation can be extracted with reasonable precision using small image areas containing a few features. The resulting description associates with a set of image points (or small image areas) the approximated direction of motion of the corresponding environmental points (or small environmental surface area). As a low level representation of environmental motion, this can considerably simplify the recovery of the sensor motion parameters [LAW82]. It can also provide qualitative information concerning the rough direction of motion of objects in a scene.

ACKNOWLEDGEMENTS

I would like to acknowledge the continuous and strong support of Ed Riseman, Al Hanson, and the UMASS VISIONS group. I thank Ed, Al, Frank Glazer, George Reynolds, and Steve Epstein for many useful discussions. Rae Ann Weymouth, Randy Ellis, Gilbert Shaw, and the reviewer made many useful suggestions for improving this paper. Dr Nelson Corby of the Machine Intelligence Laboratory of General Electric in Schenectady, New York made possible the industrial image sequences I have been working with. Tom Williams is responsible for the roadsign image sequences. This research was supported by DARPA contract N00014-82-K-0464, ONR contract N00014-75-C-0459 and NSF grant MCS-7918209.

BIBLIOGRAPHY

[ADI82] Adiv, G. "Recovering 2-D Motion Parameters in Scenes Containing Multiple Moving Objects", MS Thesis, Computer and Information Science Department, University of Massachusetts, 1982.

[BAD75] 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.

[BAL81] Ballard, D.H., "Parameter Networks: Towards a Theory of Low-Level Vision," Proc. of 7th IJCAI, Vancouver, British Columbia, August 1981, pp. 1068-1078.

[BAR80] 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.

[GIB50] Gibson, J.J., The Perception of the Visual World, Cambridge, MA: Riverside, 1950.

[GIB66] Gibson, J.J., The Senses Considered as Perceptual Systems, Boston, MA: Houghton-Mifflin, 1966.

[GLA81] Glazer, F. "Computing Optic Flow", Proc. IJCAI-7, 1981, pp. 644-647, vol II.

[HAN74] Hannah, M.J., "Computer Matching of Areas in Stereo Images," Stanford A.I. Memo 239, July 1974.

[HIL80] Hildreth, E.C., "Implementation of a Theory of Edge Detection," MIT AI Technical Report AI-TR-579, MIT, Cambridge, MA, 1980.

[HOR80] Horn, B.K.P. and Schunck, B.G., "Determining Optical Flow," Massachusetts Institute of Technology, A I Memo Number 572, April 1980.

[HUA81] Huang, T.S. (editor) Image Sequence Analysis, Berlin-Heidelberg: Springer-Verlag, 1981.

[JAI81] Jain, R. "Extraction of Motion Information from Peripheral Processes", IEEE Transactions on PAMI, PAMI-3, No. 5, September 1981, pp. 489-503.

[KEN79] Kender, J.R., "Shape from Texture: An Aggregation Transform that Maps a Class of Textures into Surface Orientation," Proc. of 6th IJCAI, Tokyo, August 1979, pp. 475-480.

[KITC80] Kitchen, L. and Rosenfeld, A. "Gray-Level Corner Detection", Tr-887, Computer Vision Laboratory, Computer Science Center, University of Maryland, College Park, MD 20742, April, 1980.

[KOH81] Kohler, R., "A Segmentation System Based on Thresholding," Computer Graphics and Image Processing, Volume 15, 1981, pp. 319-338.

[LAW82] Lawton, D. T., "Motion Analysis via Local Translational Processing", IEEE Workshop on Computer Vision: Representation and Control, Rindge, New Hampshire, August, 1982.

[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.

[LEV73] 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.

[MAR80] Marr, D. and Hildreth, E., "Theory of Edge Detection," Proc. Royal Soc. London, Volume B, 1980, pp. 187-217.

[MAR79] 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.

[MOR80] Moravec, H.P., "Rover Visual Obstacle Avoidance," Robotics Institute, Carnegie-Mellon University.

[MOR77] Moravec, H.P., "Towards Automatic Visual Obstacle Avoidance," Proceedings of the 5th IJCAI, MIT, Cambridge, MA, 1977, p. 584.

[NAG78] 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.

[NAG81a] Nagel, H., "Image Sequence Analysis: What Can We Learn from Applications," in Image Sequence Analysis (T.S. Huang, editor), Berlin Heidelberg: Springer-Verlag, 1981, pp. 19-228.

[NAG81b] Nagel, H.-H., "On the Derivation of 3-D Rigid Point Configurations from Image Sequences," IEEE PRIP-81, Dallas, Texas, August 1981

[NAK80] 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.

[O'ROB0] O'Rourke, J. and Badler, N. I. "Model-Based Image Analysis of Human Motion Using Constraint Propogation", IEEE-PAMI Volume 2, Number 6, November 1980.

[O'ROB1] O'Rourke, J. "Motion Detection Using Hough Techniques", Proceedings of PRIP 1981, pp. 82-87

[PRA79] Prager, J.M., "Segmentation of Static and Dynamic Scenes," COINS Technical Report 79-07 and Ph D Dissertation, University of Massachusetts, Amherst, 1979.

[PRA81] 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.

[PRA80] Prazdny, K., "Egomotion and Relative Depth Map from Optical Flow," Biology and Cybernetics, Volume 36, 1980, pp. 87-102.

[QUA71] Quam, L.H., "Computer Comparison of Pictures," Stanford A.I. Memo AIM-144, May 1971.

[RAD81] 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.

[REG78] Regan, D. and Beverly, K.I., "Looming Detection in the Human Visual Pathway," Vision Research, Volume 18, 1978, pp. 415-421.

[REG79] Regan, D., Beverly, K.I. and Cynader, M., "Stereoscopic Subsystem for Position in Depth and for Motion in Depth," Proc. Royal Society London, Volume B, Number 204, 1979, pp. 485-501.

[ROA80] 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.

[ROG76] Rogers, D.F. and Adams, J.A., Mathematical Elements of Computer Graphics, New York: McGraw-Hill, 1976.

[THO80] 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.

[THO81] Thompson, W.B. and Barnard, S.T., "Lower-Level Estimation and Interpretation of Visual Motion," Computer, August 1981.

[TSA81] Tsai, R.Y. and Huang, T.S., "Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch," Proc. of Pattern Recognition and Image Processing Conference, Dallas, Texas, August 1981, pp. 94-97.

[TSO80] 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.

[ULL81] Ullman, S., "Analysis of Visual Motion by Biological and Computer Systems," IEEE Transactions on Computers, August 1981, pp. 57-69.

[ULL79] Ullman, S., The Interpretation of Visual Motion, Cambridge and London: MIT Press, 1979.

[WES75] Weszka, J.S., "Threshold Selection Techniques 5," TR-349, Computer Science Center, University of Maryland, February 1975.

[WIL80] 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.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: COINS TR 82-22	2. GOVT ACCESSION NO.: AD-A122 833
	3. RECIPIENT'S CATALOG NUMBER:
4. TITLE *(and Subtitle)*: PROCESSING TRANSLATIONAL MOTION SEQUENCES	5. TYPE OF REPORT & PERIOD COVERED: INTERIM
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Daryl T. Lawton	8. CONTRACT OR GRANT NUMBER(s): ONR N00014-75-C-0459; DARPA N00014-82-K-0464
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Computer and Information Science Department, University of Massachusetts, Amherst, Massachusetts 01003	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Arlington, Virginia 22217	12. REPORT DATE: 10/82
	13. NUMBER OF PAGES: 75
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*:	15. SECURITY CLASS. *(of this report)*: UNCLASSIFIED
	15a. DECLASSIFICATION DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT *(of this Report)*

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

feature extraction
translational motion processing
environmental depth inference

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

This paper presents a procedure for processing real world image sequences produced by relative translational motion between a sensor and environmental objects. In this procedure, the determination of the direction of sensor translation is effectively combined with the determination of the displacements of image features and environmental depth.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

It requires no restrictions on the direction of motion, nor the location and shape of environmental objects. It has been applied successfully to real-world image sequences from several different task domains.

The processing consists of two basi eps: Feature Extraction and Search. The feature extraction process p s out small image areas which may correspond to distinguishing parts of environmental objects. The direction of translational motion is then found by a search which determines the image displacement paths along which a measure of feature mismatch is minimized for a set of features. The correct direction of translation will minimize this error measure and also determine the corresponding image displacement paths for which the extracted features match well.