MOTION ANALYSIS VIA LOCAL TRANSLATIONAL PROCESSING

Daryl T. Lawton

COINS Technical Report 82-27

October 1982

This paper has appeared in the Proceedings of the Workshop on Computer Vision: Representation and Control, IEEE Computer Society Press, pp. 59-72, 1982.

MOTION ANALYSIS VIA LOCAL TRANSLATIONAL PROCESSING

Daryl T. Lawton
Computer and Information Science Department
University of Massachusetts
Amherst, Massachusetts 01003

ABSTRACT

The first part of this report 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.

In the second part we extend this procedure to less restricted cases of rigid body motion. Part of the robustness of the technique is that it can work with reasonable precision even when applied to small image areas containing a few features. This allows more general image motion to be locally approximated as translations of small areas in the environment. Given such an approximation, we then show how to recover the parameters of camera motion.

I. INTRODUCTION

I.A. Definitions

Our analysis is restricted to image sequences formed by a sensor moving relative to a stationary environment. The t-th image of an image sequence is referred to as I(t). Motion of the sensor from one image to the next is characterized by a camera motion parameter vector M(t), whose six dimensions describe the displacement and reorientation of the sensor from time t to t+1.

An Image Displacement Vector is 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 have 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.

The Environmental Direction of Motion Field (EDMF) associates with each image point a unit vector describing the three dimensional direction of motion of its corresponding environmental point. Note that for a particular motion, the vectors of the EDMF approximate the tangents of the corresponding environmental points along their Environmental Displacement Paths

I.B. 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 z-axis (figure 1). The x and y axes correspond to the gravitationally intuitive horizontal and vertical directions. 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.

This research was supported by grants DARPA/ONR N00014-75-C-0459 and NSF MCS-7918209

CH1793-9/82/0000/0059$00.75©1982 IEEE

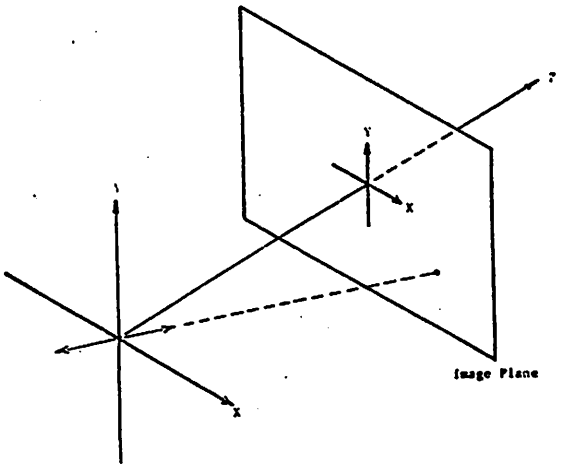

Fig 1. Camera Coordinate System

I.C. Recovery of Camera Motion Parameters

There are 5 parameters [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 the axis of rotation 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.

II. PROCESSING TRANSLATIONAL MOTION

In this section, we begin with a review of the properties of translational displacement fields and an overview of the procedure for processing them. This is followed by a more detailed description of the components of the procedure: feature extraction, error measure computation, and optimization. We then present some experimental results showing the effectiveness of the method and discuss some extensions.

II.A. Translational Motion Properties

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.

The translational axis alone does not completely determine an image displacement field. It constrains the direction of motion of image features, but not the magnitude of their displacements, which are a simple function of both feature position in the image and the depth of the corresponding environmental points.

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 II.D.5.

II.B. Overview

Processing translational motion consists of determining the axis of translation and finding the extent of image feature displacements along the paths determined by the corresponding FOE or FOC. The direction of camera translation from an image sequence is computed in two basic steps: Feature Extraction and Search. The feature extraction process picks out small image areas which potentially correspond to distinguishing parts of environmental objects. The search process optimizes an error measure which reflects the validity of a hypothesized translational axis by evaluating the matches of extracted features along the image displacement paths determined by the hypothesized translational axis. The search process consists of two basic steps: a global sampling of the error measure to determine the rough position of the minimum followed by a search based on local evaluation of the error measure gradient.

The procedure requires specification of 1) the feature extraction process; 2) the form and computation of the error measure; and 3) the organization of the search process.

II.C. Feature Extraction

The feature extraction process is used to determine small areas (sometimes called image points) in an image that are distinct from neighboring areas. This distinctiveness limits the likelihood of 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 difficult to analyze). Features can be represented 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,MOR80], 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 normalized 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 respect to its immediately neighboring areas. 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 which are sensitive to image edges. For the results in section II.F., these contours were determined using zero-crossings.

II.C.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 corresponding to local occlusion vertices and weak maxima will also be extracted.

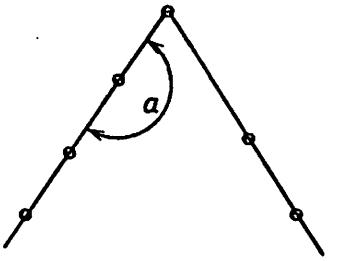

Fig 2. Curvature Approximation

Many of the weak features can be removed by suppressing those which are at points of low curvature along the zero-crossing contours. The curvature of a feature on a contour is approximated by the inner product of the normalized vectors describing the relative positions of the features adjacent to it along the contour. These values are then thresholded between 1 (corresponding to high curvature) and -1 (corresponding to low curvature). (the cosine of angle alpha in figure 2)

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. In general, it is beneficial to use masks of various positive 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 del**2g masks of different positive widths. 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 features can move significant distances from where a person would generally place them with respect to the original image.

II.D. Error Measure

The error measure is used to evaluate the validity of a translational axis with respect to successive images. It reflects the quality of the matches of extracted features along the image displacement paths determined by a potential translational axis. It is expected that most features will have their best matches along the image displacement paths determined by the correct translational axis. This will tend to be violated by false features and those features affected by occlusion.

For example, a sketch of several of the image displacement paths determined by the intersection of a potential translational axis and the image plane is shown for a set of extracted features in figure 3a. If the hypothesized translational axis is correct, the majority of features will tend to have good matches along these paths. Figure 3b shows the match profile for a particular feature along its displacement path with respect to the succeeding image. The units of displacement are pixels.

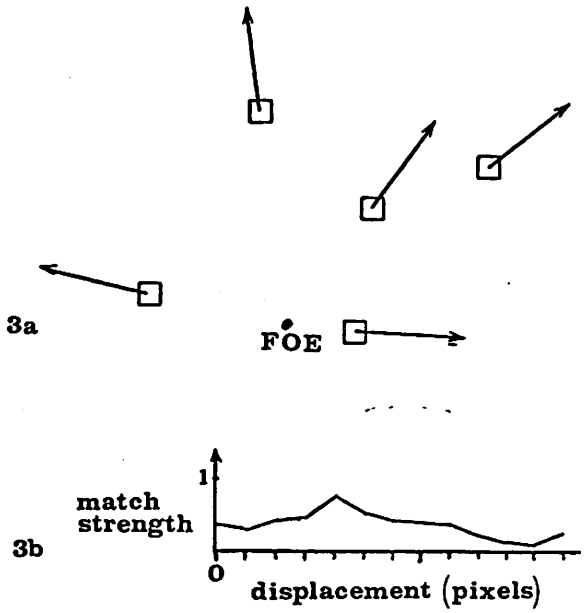

Fig 3. Constrained Feature Displacements

The development of an 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.

II.D.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 (1)

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

Moravec Correlation [MOR77] (2)

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

Normalized Absolute Value Difference (3)

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

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.

II.D.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 a) roughly by setting the coordinates of the feature's position to the nearest integer value; or b) more accurately by performing a subpixel interpolation of the feature at each of a set of selected positions along the image displacement path with respect to the succeeding image. The basic trade-off is between speed and accuracy, with subpixel interpolation being a more expensive computation.

II.D.3. Error Measure

The error measure associates with a point on the direction of translation sphere a value describing the quality of image feature matches along the image displacement paths determined by the corresponding translational axis. This value is computed by determining the best match for each feature along the image displacement path determined by the hypothesized translational axis and then summing the normalized error values (using one of the metrics above) for all of the image feature points. 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

$$\sum_{i=1}^{n} (1.0 - \text{bestmatch}(i)) \qquad (4)$$

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

II.D.4. Properties of the Error Measure

The error measure should have a distinct global minimum at the point on the unit sphere corresponding to the correct translational axis. It is expected to be well behaved globally 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, we do not expect competing candidates for the global minimum to be widely separated, and the experiments we have performed confirm this expectation.

The error measure will be 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 FOE and their unanimity is expected to overide the lack of discrimination of weak features and the random quality of the matches of false features.

II.D.5. 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 displacement vectors 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.

II.E. Search Organization

The search process used here consists of two phases: A global sampling of the error measure to determine its rough shape followed by a local search to determine the minimum. The local search is initialized 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 to 0.025 and 0.005 radians.

The form of the error function for several different translational sequences is smooth, with a single minimum in a large neighborhood around the correct translational axis. Thus, the global sampling could be quite sparse or the initial step size of the local search quite large.

II.F. Experiments

Figures 4a and 4b (128x128 pixels, 64 intensiy levels, black and white) show successive images taken from a car driving down a country road in Massachusetts. Figure 5a shows the extracted zero-crossings using a mask of positive width equal to five pixels. Figure 5b shows the interesting points extracted along these contours and figure 5c shows the set of interesting points after low-curvature suppression (see section II.2.C.) was applied using an inner product threshold set to -0.75. Features were 5x5 pixel arrays. For this experiment, the extracted feature positions were applied relative to the raw image.

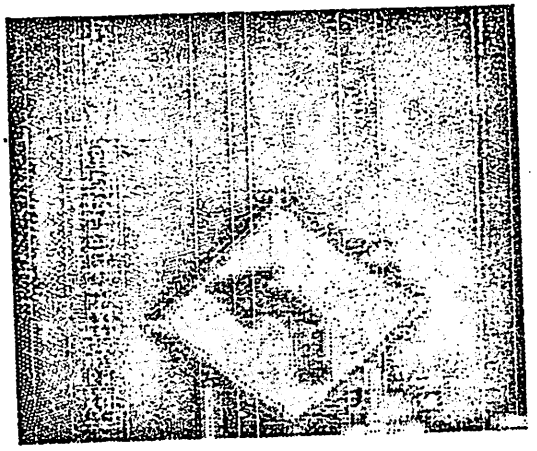

Fig 4a. Road Image 1

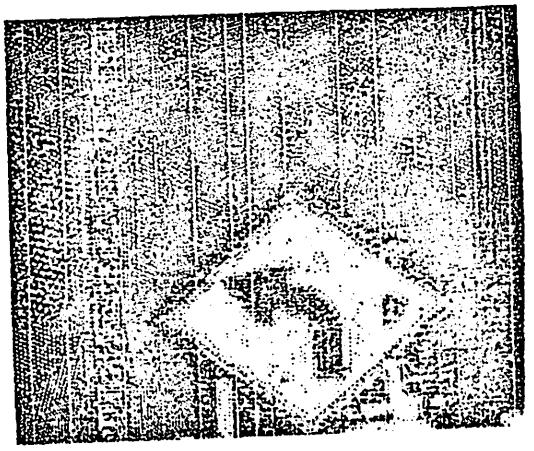

Fig 4b. Road Image 2

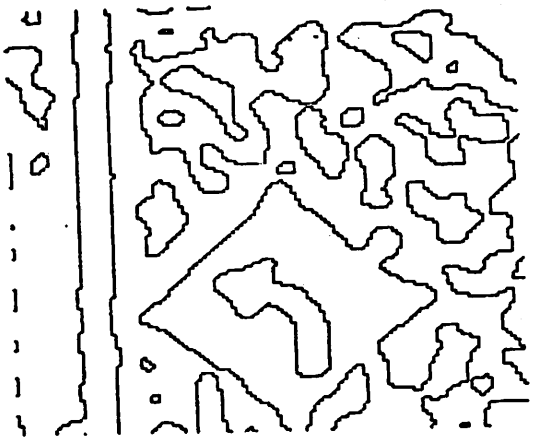

Fig 5a. Extracted Zero Crossings

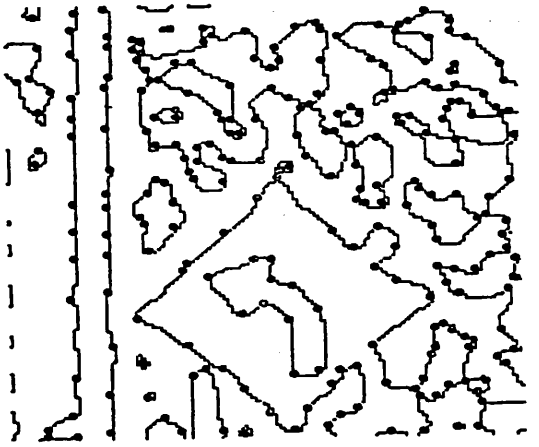

Fig 5b. Distinctive Image Points

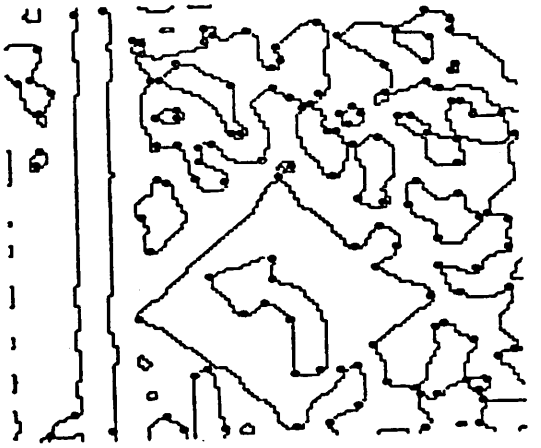

Fig 5c. After Low Curvature Suppression

The global search used the absolute value norm and nearest integer interpolation. The sampling increment corresponded to the vectors on the direction of motion sphere being separated by .314157 radians from each other. Maximal Image displacements along the hypothesized image displacement paths was set to 10 pixels. Features were centered at the positions shown in figure 5c. The global sampling determined a minimum in the error function at the unit vector (-.80902, -.47554, .34548) on the direction of translation sphere.

The local search used the Moravec norm and bi-linear interpolation. The determined translational axis was (-.83738, -.42043, .34933). The displacements of the feature points from figure 5b for this translational axis are shown in figure 6.

Fig 6. Image Displacements

The procedure was repeated, but using features at the positions from figure 5b (those prior to low curvature suppression). This has the effect of introducing weak and false features into the computation. The translational axis extracted was (-.82909, -.42281, .36585) This is a difference of 0.01863 radians or 1.06765 degrees from that determined using the features indicated in figure 5c.

The procedure was also applied using the features from the restricted subarea shown in figure 7, corresponding to some faint tree texture. Using these features, the translational axis extracted was (-.84281, -.42928, .32465). This is a difference of 0.02677 radians or 1.53418 degrees with the translational axis determined using the feature centered at the positions indicated in figure 5c.

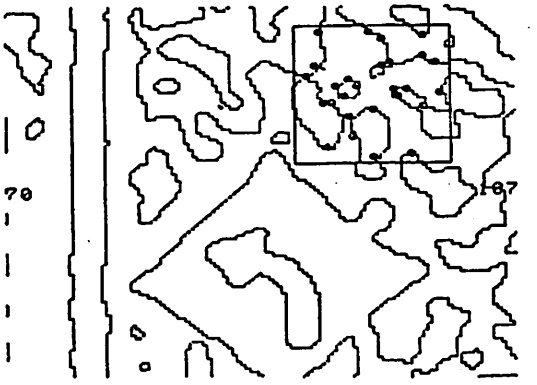

Fig 7. Image Subarea

Given the direction of translation and image displacements, relative environmental depth can be recovered by the simple relation [LEE80]

$$\frac{D}{\Delta D} = \frac{Z}{\Delta Z} \qquad (5)$$

where Z is the value of the Z component of an environmental point at time t+1, delta 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 delta D is the image point's displacement from time t to time t+1. Z can be recovered in units of Delta Z without knowledge of the actual extent of camera displacement. When Delta D is 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 5. For this reason, it is useful 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 of the roadsign. In this processing, 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 determing the translational axis for images I(t+1) and I(t+2).

Given the image displacements determined from I(1) to I(4) of the sequence, the depth values for image points along the contour in figure 5a were computed using equation 5. This sequence is especially nice for presenting depth processing results since the three environmental objects in the images are at three distinct depths. This is shown in figure 8a by the three distinct clusters in the histogram of the depth values calculated for the points along the contour in figure 5a. Mapping feature labels from these clusters back onto contour points from figure 5a yields: the boundary shown in figure 8b (the sign), the boundary shown in figure 8c (the pole), the boundary segment shown in figure 8d (the trees). Points in a 10 pixel wide margin along the boundary of I(1) were ignored since the processing did not take into account occlusion/disocclusion effects along the image boundaries.

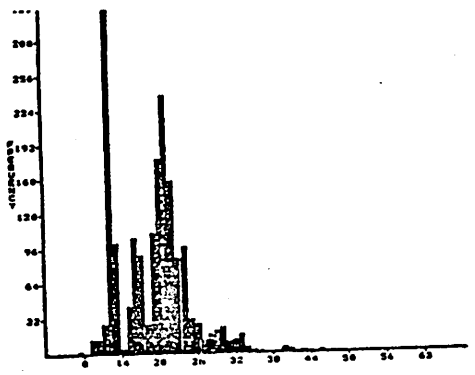

Fig 8a. Depth Histogram (Z component)

Fig 8b. Sign Segments

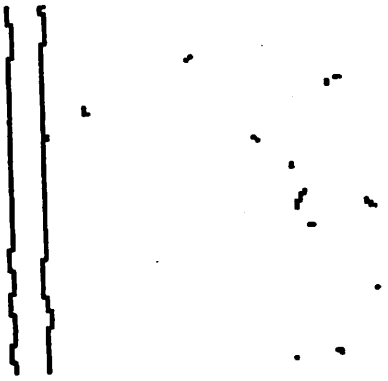

Fig 8c. Pole Segments

Fig 8d. Tree Segments

II.G. Summary and Extensions

This work demonstrates a simple and robust procedure for determining the direction of environmental motion and image displacements in real-world image sequences produced by observer translation. It is not dependent on an initial matching process prior to the inference of camera motion. Instead, features are extracted from an initial image and their displacements are determined concurrently with the inference of direction of sensor motion. Thus complications in matching that arise from an individual feature being extracted in one image and not in the next are reduced. The process is also relatively insensitive to weak and false features. It has been successfully applied to image sequences produced by a car translating down a road, by a camera attached to a robot manipulator in an industrial environment, and to artificially generated sequences. We now consider some extensions.

II.G.1. Other Cases of Restricted Motion

The procedure developed in this paper should be 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. Two particular cases are pure sensor rotation and motion constrained to a known plane.

With pure sensor rotation, the unknown camera parameters are constrained to R1(t), R2(t), and R3(t). In this case, the error measure from section II.D.3. would be defined with respect to the direction of rotation sphere where each point corresponds to an axis of rotation. For each rotational axis, the extent of displacement for image features is determined by different values of R3(t). There is the additional constraint in the rotational case that the displacements of all features must correspond to the same value of R3(t).

For motion constrained to a known plane, the rotational axis is known to be perpendicular to that plane and the translational axis is constrained to lie in that plane. Thus, only R3(t) and one translational parameter can vary and the error measure can be computed with repect to these two parameters. The global sampling in this case amounts to evaluating a set of translational axes for each of a set of potential rotations.

II.G.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.

One 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,NAGI79]. 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.

II.G.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.

III. THE LOCAL TRANSLATIONAL DECOMPOSITION

We now extend the translational case to less restricted forms of sensor motion by applying the procedure for determining the direction of translational motion to small, overlapping areas across an image surface over a sequence of images. The motivation is to approximate general motions as consisting locally of environmental translations and to interpet 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 of motion is an approximation to the Environmental Direction of Motion Field (EDMF) (section I.A.) which associates with a set of image points (or small image areas) the direction of motion of the corresponding environmental point (or small environmental surface area). As a low level representation of environmental motion, this considerably simplifies the recovery of the sensor motion parameters.

This section is divided into three parts. In III.A., the properties of the EDMF for different sensor motions are summarized. The cases considered are pure rotational motion; motion constrained to an unknown plane; and arbitrary motion. This analysis shows how to recover the axis of rotation from the EDMF for these cases.

Techniques for computing the EDMF from image sequences are presented in section III.B. There are two cases considered: 1) sequences for which image displacement vectors have not been determined; and 2) sequences for which image displacement vectors have been determined. In the first case, computing the EDMF also determines image displacements.

Section III.C. demonstrates the use of the local translational decomposition for processing image sequences that are produced by sensor motion constrained to an unknown plane in highly textured environments. There are indications that this processing is quite robust. We also note the effect of coupling the EDMF and environmental rigidity constraints for the recovery of relative depth.

III.A. EDMF Properties for Different Types of Camera Motion

Before discussing the computation and use of the EDMF, it is necessary to describe some of its basic properties for different classes of motion. To describe these properties, it is useful to map the EDMF vectors onto the direction of translation sphere. In section II, the direction of translation sphere was used as the domain for 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.

III.A.1. Pure Translational Motion of the Camera

As discussed above, for translational motion the image displacement paths 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.

III.A.2. Pure Rotational Motion of the Camera

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. The EDMF vectors will lie upon a great circle contained in a plane perpendicular to the axis of rotation when mapped onto the direction of translation sphere.

III.A.3. 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 [WHI44]). 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.

III.A.4. Arbitrary Motion

For arbitrary motion, the image displacement paths cannot be easily described. But 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 [COX61,WHIT44]).

The set of normalized tangent vectors to a helix, when based at a common origin, will generate a cone, called 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. Because of this, 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.

III.B. Computing the EDMF

III.B.1. From an Unprocessed Image Sequence

The translational processing procedure described in section II yields a set of image displacements consistent with a determined translational axis. Applying this procedure to a small area of an image containing extracted features finds a set of image displacements consistent with interpreting the local image motion as if it were produced by a translation of the corresponding part of the environment. Note that where the translational approximation is poor (for example, image areas near the intersection of the axis of rotation and the image plane) there will be a large value of the error measure describing the validity of the translational axis. Thus, the error measure can serve to validate 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 the area or the features are bunched together, the translational approximation will be poor. Processing is not applied to local areas which do not satisfy these requirements.

Figure 9a is a 128x128 pixel image of some grass texture with seven bits of intensity. Figure 9b was derived from figure 9a by applying a simulated rotation of 0.1 radians about the Y axis of the camera coordinate system (the focal length was set to one). Features were selected from the image in figure 9a by first determining image points where the contrast was greater than 20 intensity levels, and then finding local maxima in the distinctiveness values (section II.C.) associated with the 5x5 pixel square features centered at those points. The resulting feature positions are shown in figure 10.

Using the translational processing procedure, the direction of translation was determined for 11x11 pixel neighborhoods centered at each feature in figure 10. The image displacement associated with a feature was that determined by the best translational approximation for the feature's neighborhood. The resulting image displacement field is shown in figure 11.

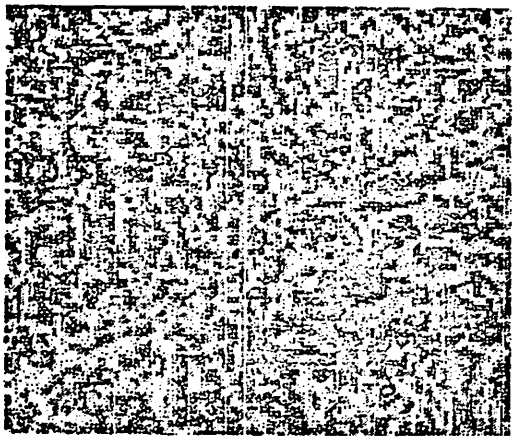

Fig 9a. Grass Texture Image 1

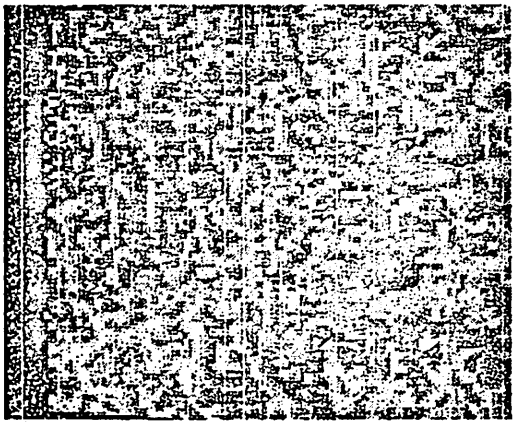

Fig 9b. Grass Texture Image 2

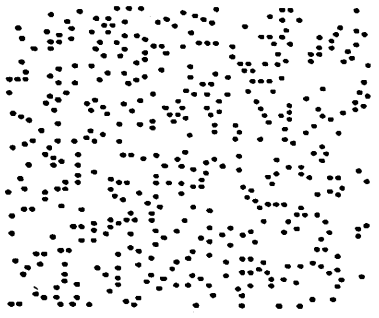

Fig 10. Extracted Feature Positions

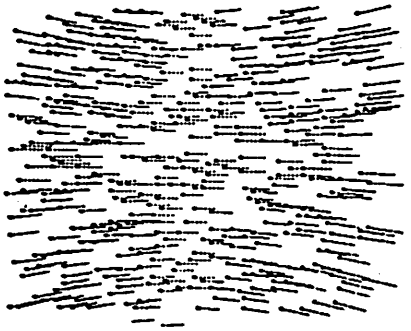

Fig 11. Image Displacements

III.B.2. From a Computed Displacement Field

An error measure can be developed to evaluate translational axes for image sequences for which image displacements have been determined. The error, with respect to an image displacement field, can be calculated for a translational axis by summing the angles between the image displacement vectors and the image displacement paths from the FOE or FOC determined by the translational axis (figure 12). Similarly, the sum of one minus the cosine of each angle could be used. To compute the EDMF, the translational axis is determined applying this error measure to local areas of the computed displacement field.

It may be possible to determine the EDMF from sparse image displacement fields by filling the image displacement field to an adequate density by a smoothing or averaging procedure which treats the sparse determined image displacements as boundary conditions and then locally applying the translational processing procedure using the adapted error measure.

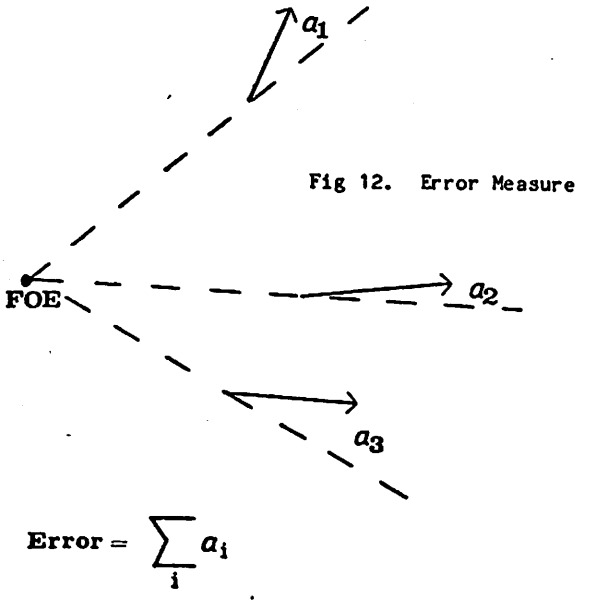

Fig 12. Error Measure

III.C. Processing the Computed EDMFs

III.C.1. Processing Arbitrary Planar Motion

For arbitrary planar motion, all the environmental displacements are constrained to lie in planes perpendicular to the axis of rotation. In this case four of the five recoverable camera parameters are unconstrained: the axis and extent of rotation are arbitrary, and the translational axis is constrained to be perpendicular to the rotational axis. When the ideal EDMF vectors are mapped onto the Direction of Translation Sphere, they will lie on a great circle in a plane perpendicular to the axis of rotation. Thus, processing consists of determining the EDMF and finding the best planar fit of the EDMF vectors which also contains the origin. This may be done using any of a number of plane fitting routines. In the experiments here, the eigenvector fit procedure described in [DUD73] pp. 332-335 is used, having been adapted for planes containing the origin.

Note that 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 the plane parallel to it and containing the origin. Thus more and more values for the fit can be collected over time, thereby increasing the accuracy of the processing.

For example, using the EDMF determined for the grass texture sequence described in section III.B.1., the normal to the best plane fit was determined to be (.002518,.999893,-.0143709). This is off by .014592 radians or .836053 degrees from the correct rotational axis. Figure 13 shows a histogram of the computed EDMF vectors in a polar coordinate system (for the unit vectors (X,Y,Z) on the direction of translation sphere, Phi1 = arctan(Y/X), Phi2 = arccos(Z)). The number of vectors at a particular location is encoded by darkness. Note the orientation in a plane perpendicular to the Y axis.

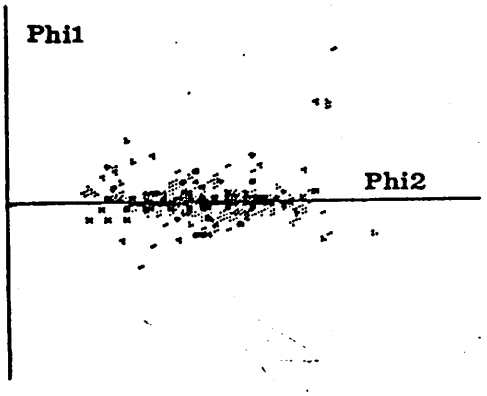

Fig 13. EDMF Histogram

Figure 14 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. A rotation of 0.1 radians occurs about an axis whose orientation is parallel to (.577350, .577350, .577350) and positioned at the back of the sphere. The local direction of translation was determined at all positions across the displacement field for 5x5 pixel windows, using the measure described in section III.C.2. Using all the determined EDMF vectors in the plane fitting procedure, the normal is determined to be (.647159, .543663, .534429). This deviates from the correct axis by .088631 radians or 5.078184 degrees. Figure 15 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 plane fit to EDMF vectors which have low associated error values for the translational approximation, the determination of the axis of rotation is improved. By using EDMF vectors for which the determined error measure is less than 90 degrees over the 5x5 pixel areas, 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.

Once the axis of rotation has been determined, processing has been reduced to the case of known planar motion. This could be solved directly via the suggested adaptation of the translational technique to known planar motion (section II.G.1.). Alternatively, the inference techniques of Prazdny [PRA81] and Nagel [NAG81a,NAG81b] could be applied to the image displacement field determined concurrently with the EDMF. In these techniques a composite image displacement field (one produced by combined camera rotation and translation) is decomposed into its translational and rotational components by searching through the three-dimensional space of rotational parameters to find a rotational displacement field which, when subtracted from the composite field, yields a translational displacement field. By having determined the axis of rotation via analysis of the EDMF, this search has been reduced to a single bounded dimension corresponding to the extent of rotation.

For the case of planar motion, the FOE or FOC is further constrained to lie along a line in the image plane determined by the intersection of the image plane and the plane perpendicular to the axis of rotation and containing the focal point. Because of this, the decomposition procedure is simplified. When the correct rotational field is subtracted from the composite field, the resulting field should have an FOE or FOC along the line. Thus, it is only necessary to evaluate the distribution of the intersections of the image displacement vectors resulting from subtraction of a hypothesized rotational displacement field with this line.

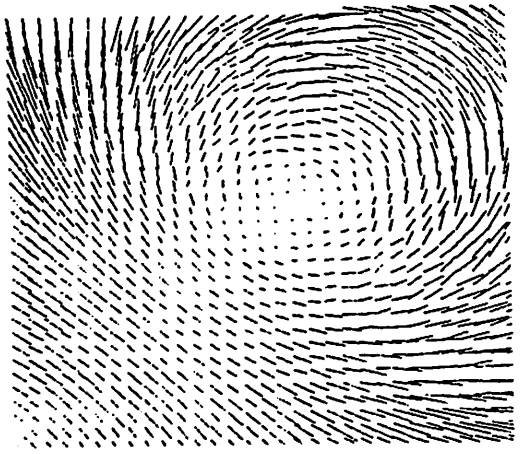

Fig 14. Image Displacement Field

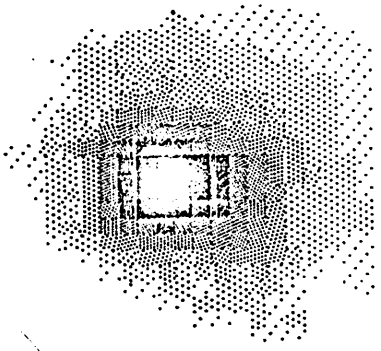

Fig 15. Translational Approximation Error

Note that by mapping the EDMF onto the direction of translation sphere, the local differential properties of the EDMF are not being utilized. We suspect that 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). If this is so, processing could be directly reduced to the purely translational case by removal of the determined rotational component.

Let us 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.

III.C.2. Coupling the approximated EDMF and Rigidity Constraints

There are several formulations for the recovery of environmental depth and camera motion parameters based upon environmental rigidity [LAW80, NAG81, MER80, ROA80, ULL79, WEB81]. Solving these constraints is simplified when information about the direction of environmental motion is incorporated into them. In particular, the number of points in successive frames that is necessary to infer their relative depth is reduced from five to two.

For two points in successive images there are four unknowns to be recovered corresponding to the depths of the two points at instants t and t+1. One of the depths at time t can be set arbitrarily since only relative depth can be recovered. Both depths of the points at time t+1 can be determined from their image displacement vectors, their depths at time t, and their corresponding EDMF vectors. (To see this (figure 16) , note that given 1) succesive rays of projection P1 and P2; 2) a depth D for the corresponding environmental point along P1 and 3) the direction of environmental motion for the point along P1, the depth of the environmental point along P2 can be determined by selecting the point on P2 that is closest to the (dotted) line determined by the environmental point along P1 and its direction of motion). Thus, the depth of one of the points can be set arbitrarily and the other depth determined based on satisfaction of the rigidity constraint over successive instants.

Each point can thus assign relative depths to all other image points. This suggests a consistency computation wherein agreement between the relative depth maps determined by each point are used to find a globally consistent depth map.

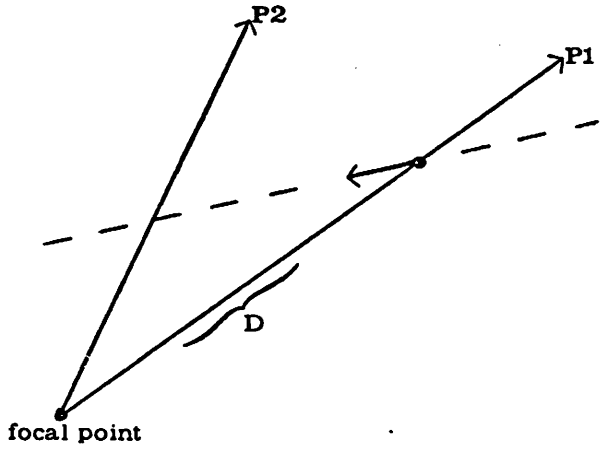

Fig 16. Depth Inference

III.D. Discussion

This work shows that if the EDMF can be reliably computed, it is a very useful low level representation for rigid body motion analysis. There are strong indications that this is possible for densely textured image sequences and that camera motion parameters can be recovered for cases of motion of complexity corresponding to motion constrained to an unknown plane.

Techniques for processing arbitrary motion have been suggested in section III.A.4. (finding the best planar fit to the difference vectors of successive EDMF vectors) and in section III.C.2 (solving rigidity constraints coupled with information in the EDMF). The primary question concerns the robustness of processing when using a noisy, approximated EDMF in the arbitrary case.

ACKNOWLEDGEMENTS

I thank Ed Riseman for his strong and continuous support. Ed, Randy Ellis, Steve Epstein, Frank Glazer, and George Reynolds have been exceptionally fine people to talk with.

BIBLIOGRAPHY

[COX61] Coxeter, H., Introduction to Geometry. New York. Wiley 1961.

[DUD73] Duda, R.O. and Hart, P.E., Pattern Classification and Scene Analysis, New York. Wiley 1973.

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

[LAW80] Lawton, D.T., "Constraint-Based Inference from Image Motion", Proceedings of the First Annual National Conference on Artificial Intelligence, Stanford University, August 18-19, 1980.

[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.

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

[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.

[MEI80] Meiri, A.Z., "On Monocular Perception of 3-D Moving Objects", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume PAMI-2 Number 6, November, 1980.

[NAG81a] Nagel, H.-H., "On the Derivation of 3-D Rigid Point Configurations from Image Sequences," IEEE PRIP-81, Dallas, Texas, August 1981.

[NAG81b] Nagel, H.-H. and Neumann, B., "On 3-D Reconstruction from Two Perspective Views," Int'l Joint Conference on Artificial Intelligence, Vancouver, Canada, August 1981.

[NAGI79] Nagin, P. A., "Studies in Image Segmentation Algorithms based on Histogram Clustering and Relaxation", Ph.D. Dissertation, COINS Technical Report, Computer and Information Science Dept., September, 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.

[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.

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

[WEB81] 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.

[WHI44] Whittaker, E.T., A Treatise on the Analytical Dynamics of Particles and Rigid Bodies, New York: Dover Publications, 1944.

[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 82-27	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) MOTION ANALYSIS VIA LOCAL TRANSLATIONAL PROCESSING		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 14
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)

Translational Motion Processing
Arbitrary Motion Processing
Local Translational Decomposition

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

The first part of this report 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,

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)

nor the location and shape of environmental objects. It has been applied successfully to real-world image sequences from several different task domains.

In the second part we extend this procedure to less restricted cases of rigid body motion. Part of the robustness of the technique is that it can work with reasonable precision even when applied to small image areas containing a few features. This allows more general image motion to be locally approximated as translations of small areas in the environment. Given such an approximation, we then show how to recover the parameters of camera motion.