THE UNIVERSITY OF MICHIGAN

COMPUTING RESEARCH LABORATORY[1]

DYNAMIC SCENE ANALYSIS

Ramesh Jain

CRL-TR-6-84

JANUARY 1984

Room 1079, East Engineering Building
Ann Arbor, Michigan 48109
USA
Tel: (313) 763-8000

[1] This work was supported by the National Science Foundation, under the Grant No. NSF-G-MCS-8219739. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the funding agency.

To appear in Progress in Pattern Recognition, Vol.2, Ed. L. Kanal and A. Rosenfeld

DYNAMIC SCENE ANALYSIS

Ramesh Jain
Department of Electrical and Computer Engineering
The University of Michigan
Ann Arbor MI 48109

Abstract

This paper presents an overview of dynamic scene analysis. It discusses techniques for change detection, segmentation, computing optical flow and extracting information therefrom, recovering 3-Dimensional information about the objects and the motion, representation of motion in terms of low level concepts and the verbalization of motion. The emphasis is on an overview of the state of the art, rather than an exhaustive survey.

1. Introduction

The World is dynamic. Most biological vision systems have evolved to cope with the changing world. The past decade has seen the emergence of computer vision systems [BaB82]. For a computer vision system engaged in the performance of non-trivial real world operations and tasks, the ability to cope with moving and changing objects and viewpoints is vital. Though early computer vision systems were concerned with static scenes, the last few years have seen an ever-increasing interest in computer vision systems capable of analyzing dynamic scenes.

The input to a dynamic scene analysis system is a sequence of frames of a changing world. The camera may also be moving. Each frame represents image of the scene at a particular time instant. The changes in a scene may be due to the motion of the camera or the motion of the objects, illumination changes, or the changes in the structure, size, or shape of an object. Usually it is assumed that the changes in a scene are due to motion of camera and/or objects and that the objects are either rigid or quasi-rigid; other changes are not allowed. The task of the system is to detect changes, to find motion characteristics of the observer and the objects, to recover the structure of the objects, to characterize the motion using high level abstraction, and to recognize moving objects. It is also possible that future systems will be required to observe a scene, then describe the events taking place in a language understandable by a possibly naive user of the system, who may not know anything about computers.

A scene usually contains several objects. An image of the scene at a given time represents a projection of a part of the scene, which depends on the position of the camera. It is necessary to consider separately the four possible relations between camera and scene:

1. Stationary Camera, Stationary Objects (SCSO).

2. Stationary Camera, Moving Objects (SCMO),

3. Moving Camera, Stationary Objects (MCSO), and

4. Moving Camera, Moving Objects (MCMO).

This work was supported by the National Science Foundation, under the Grant No. NSF-G-MCS-8219739.

The SCSO, static scene analysis, is of only peripheral interest in this paper, though it is clear that many techniques developed for static scene analysis will play an important role in the dynamic scene analysis. In many applications it may be necessary to process a single image to obtain required information and it may be possible to extract such information from the image. It appears, however, that there are many applications which require the information extracted from a dynamic environment. Some applications require understanding of a dynamic process, such as cell motion, using vision techniques. Clearly, availability of a frame sequence offers more information for the understanding of a scene, but at the cost of a significant increase in the amount of the data to be processed by the system. Applying static scene analysis techniques to each frame of the sequence for the analysis of dynamic events requires an enormous amount of computation and suffers from all the problems of the static scene analysis. Fortunately, research in the field of dynamic scene analysis has shown that the extraction of information in dynamic scenes may be easier than in static scenes. In most cases the increase in the amount of the data is not a problem. On the contrary, the total computational effort may be significantly less and the performance better for some tasks (eg., the segmentation of a scene).

In dynamic scene analysis research [MaA78, Nag78b, Nag82c], SCMO scenes have received maximum attention. In such scenes, usually it is desired to detect the motion, extract masks of the moving objects with the aim to recognize them, and compute their motion characteristics. The MCSO scenes received attention recently, particularly because of optical flow research, but the case of moving objects and moving camera, the MCMO, has received very little attention [MaA78]. Clearly, the MCMO is the most general, and possibly, the most difficult situation in dynamic scenes. SCMO and MCSO find application in many situations and have been studied by researchers in different contexts under different assumptions. Many techniques developed for stationary camera are not applicable if the camera is allowed to move. Similarly techniques developed for moving camera have assumed stationary scenes and are not applicable if the objects were allowed to move. Only recently [Jai83b, Jai83c, Jai83d] are techniques being developed that will be applicable to all dynamic scenes.

Some researchers [MaA78,Jai81a] view the process of analyzing a dynamic scene in three phases: *peripheral*, *attentive*, *and cognitive*. The peripheral phase is concerned with extraction of information that is imprecise but is very helpful for later phases of analysis. The information extracted in the peripheral phase gives an indication of the activities in the scene and helps in deciding which parts of the scene need careful analysis. The attentive phase concentrates its analysis in the active part of the scene and extracts information that may be used for the recognition of objects, analysis of motion of objects, preparing a history of events taking place in the scene and other related phenomena. The cognitive phase applies the knowledge about the objects, motion verbs, and other application dependent concepts to analyze the scene in terms of objects and events taking place in a scene. A conceptual system based on these three phases is described by Jain and Haynes in [JaH82]; Jerian and Jain discuss some cognitive level issues related to the recovery of 3-Dimensional information in the design of such a system in [JeJ83b].

In this paper we present an overview of the research in dynamic scene analysis. Our aim is to discuss various aspects of the analysis of dynamic scenes, with an emphasis on the analysis of motion, and to outline briefly some approaches proposed by researchers. We will not try to give an exhaustive survey of the field. The field of dynamic scene analysis is very active; many different approaches for various aspects of the problem are being presented. For a good idea of the changes in the field see [MaA78, Nag78b, Nag82c]. The organization of this paper is influenced by the currently active areas of research, which are by no means disjoint. Many related areas, such as tracking [Sch79, Sch82, ScM81, LeY82], motion compensated coding [NeR79,NeR80, SNR80, NeS79], architecture for the dynamic scene analysis [AgJ82, GGF80], and applications are not discussed in this paper.

The section 2 discusses methods for change detection. Methods for the segmentation of dynamic scenes to extract images of moving objects are discussed in section 3. The segmentation of a dynamic scene may be only on the basis of motion to extract images of moving objects or may involve segmentation of the static components of the scene also. Optical flow has attracted researchers in psychology since Gibson [Gib79, Lee80] suggested that pilots and birds use it. The presence of information in the optical flow, methods for determining optical flow, and techniques to extract information from optical flow are the subject of section 4. Motion has been considered a potent clue for determining depth and structure of objects. Various methods for the extraction of structure of objects and for the determination of motion parameters are discussed in section 5. If the aim is to describe the events in a scene then the representation of complex motions in simpler forms and the abstraction of the temporal motion characteristics using motion verbs will be required. The high level processes used for such a description are the subject of section 6. Finally, we review the relationship between various aspects and methods inorder to integrate the outputs of different techniques and to suggest future directions for research in section 7.

1.1. Terminology

The input frame sequence is represented as $F(x,y,t)$ where x and y are the spatial specification of an element in a frame representing the scene at time t. The value of the function represents the intensity of the pixel and is generally quantized into 256 levels. It is assumed that the image is obtained using a camera located at the origin of the 3-dimensional coordinate system. This means that the coordinate system is observer centered. The projection used may be perspective or orthogonal. Orthogonal projection approximates the camera when the objects are located far away from the camera as compared to the focal length of the camera. The 3-D coordinate of a point will be denoted by (X,Y,Z) and the projection of the point will be denoted by (x_p,y_p). The line of sight, or the optical axis of the camera, is assumed to be along the Z axis. An example of an imaging geometry is shown in Figure 1.

Figure 1 The imaging geometry employs the left hand coordinate system and planar perspective projections.

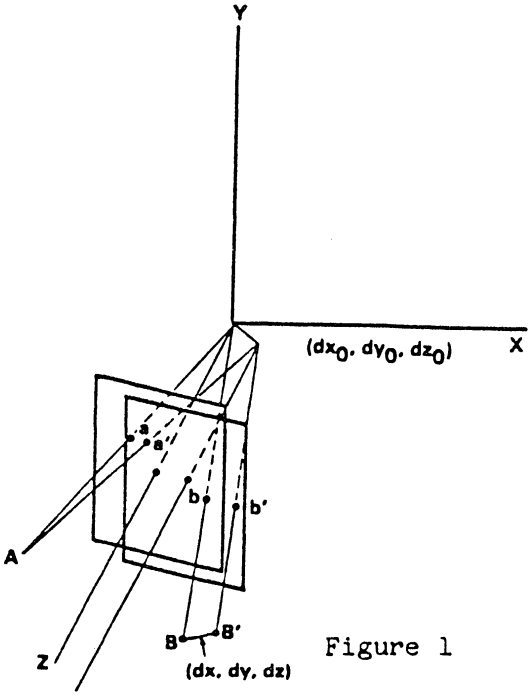


Figure 1

Since the frames are usually taken at regular interval, we will assume that t represents the t^{th} frame of the sequence, rather than the frame taken at absolute time t.

2. Change Detection

The result of a perceptible motion is some change in the frames of the sequence. By detecting such changes one may analyze motion characteristics. If the motion is restricted to a plane that is parallel to the image plane, then a good quantitative estimates about the motion components may be obtained; in case of 3-D motion only qualitative estimates are possible. Historically, early methods for dynamic scene analysis were based on change detection in a frame sequence. By analyzing frame-to-frame changes the global analysis of the sequence was performed. The changes were detected at different levels: pixel, edge, and region. The changes detected at pixel level can be aggregated to obtain useful information for constraining the computational requirements of the later phases [Jai81a]. In this section we discuss various methods developed for change detection and their application in dynamic scene analysis.

2.1. Pixel Level

The most obvious method of change detection in two frames is to compare the corresponding pixels of the frames to determine whether they are same or different. The subtractive TV used in [OHO73] is one of the application using this method. In the simplest form a binary difference picture $DP_{jk}(x,y)$ for two frames $F(x,y,j)$ and $F(x,y,k)$ is defined as:

$$DP_{jk}(x,y)=1 \quad \text{if} \quad \left|F(x,y,j)-F(x,y,k)\right| > \tau$$
$$=0 \quad \textit{otherwise}$$

In the difference picture all pixels with value 1 are considered the result of the motion of objects. Clearly, this method assumes that the frames are properly registered and the illumination in the image remains constant. In real scenes such a simple test for change detection, usually, results in unsatisfactory results due to noise. A simple size filter may be used [JMN77] for ignoring noisy pixels which do not form a connected cluster by using a simple criterion that only those difference picture that belong to a 4-connected (or 8-connected) component of size above a threshold size will be attributed to motion. The motivation behind this filter is the fact that usually noise entries are isolated and the changes due to motion of surfaces form connected clusters in a difference picture. The filter is very effective in reducing noise but, unfortunately, it also filters some desirable signals, eg. those from slow or small moving objects.

Nagel [Nag78a] modified Yakimovsky's method of region growing to compare corresponding areas of two frames. By considering corresponding areas of two frames using the likelihood ratio

$$\lambda=\frac{\left[\frac{\sigma_1+\sigma_2}{2}+(\frac{\mu_1-\mu_2}{2})^2\right]^2}{\sigma_1 * \sigma_2}$$

(where μ and σ denote the mean grayvalue and the variance for the sample areas from the frames.)

one may obtain the areas where changes are taking place, again by using a threshold. A minor problem in the application of the likelihood ratio is that it can be applied to the areas, not to single pixels. This problem may be solved by considering the corresponding areas of frames. It was suggested that the corresponding areas of the frames may be the super-pixels formed by combining nonoverlapping rectangular areas of the frames comprising m rows and n columns. The values of m and n are selected to compensate for the aspect ratio of the camera. The likelihood ratio test combined with the size filter discussed above works quite well [JMN77, JaN79] for removal of noise. The problem of small and slow moving objects is exacerbated, however, since super-pixels effectively raise the threshold for the detection of the motion of such objects. Jain, Militzer, and Nagel [JMN77] introduced the concept of accumulative difference picture for removing this limitation. In place of comparing two frames and forming a difference picture, they compared every frame of the sequence to a reference frame and increased the entry in the accumulative difference picture by 1 whenever the likelihood ratio for the area exceeded the threshold. Thus

$$\begin{aligned} ADP_n(x,y) &= ADP_{n-1}(x,y)+1 \quad \text{if } \lambda > \tau \\ &= ADP_{n-1}(x,y) \quad \textit{otherwise}. \end{aligned}$$

where τ is a preset threshold.

The first frame of the sequence was usually considered the reference frame and the accumulative difference picture ADP_0 was initialized to 0. An ADP allows for the detection of small and slow moving objects.

The likelihood test discussed above was based on the assumption of the uniform 2^{nd} order statistics of a region. Nagel has recently suggested the use of likelihood tests based on the approximation of intensity values of pixels belonging to the super-pixels using surfaces such as facets and quadratic surfaces [HNR82, Nag82a, NaR82]. These higher order approximation allow for the better characterization of intensity values and result in more robust change detection. He showed that performance of these tests is far superior to the test based on the linear intensity approximation of intensities.

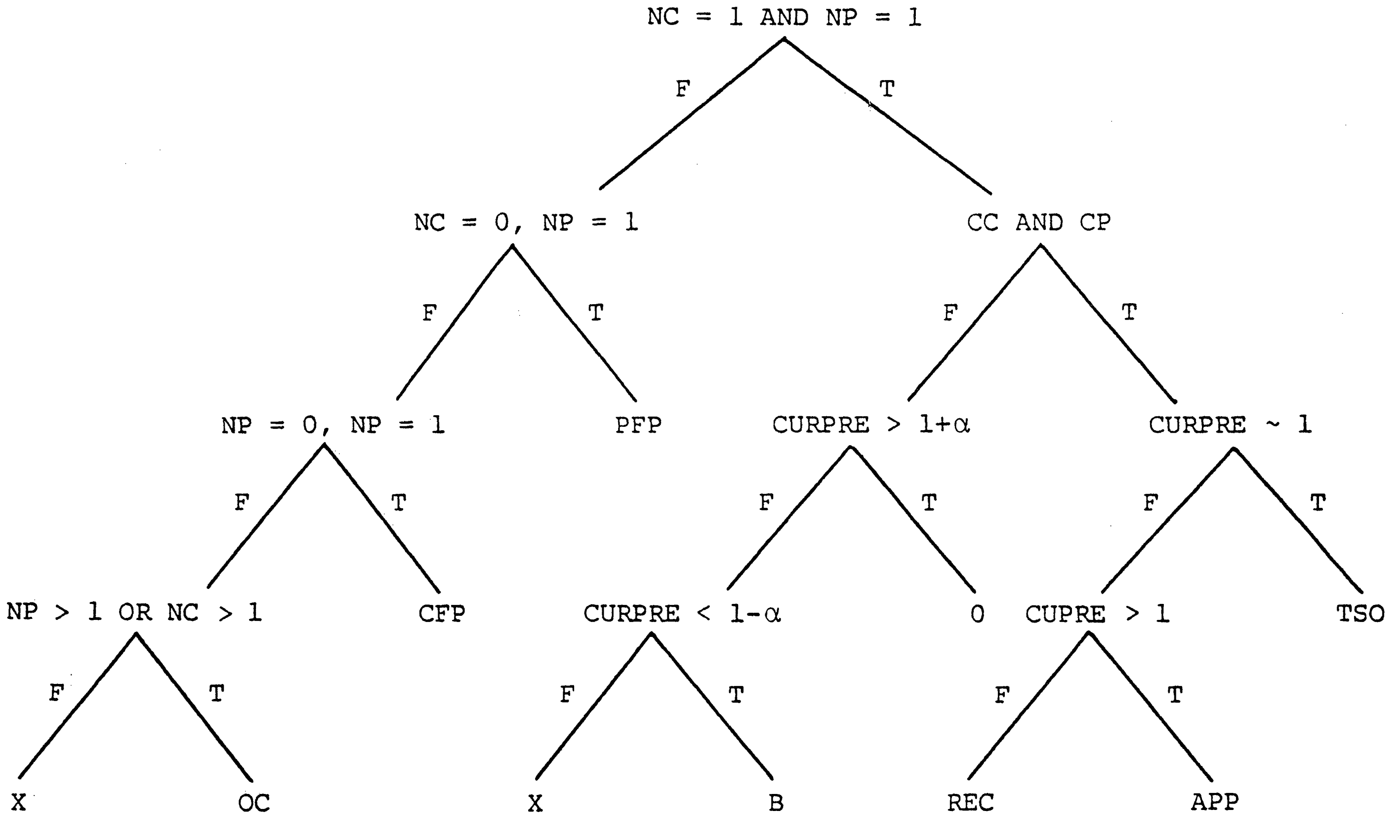
Note that the tests based on likelihood ratio result in the dissimilarity detection at super-pixel level. Moreover, since the tests are based on likelihood ratio, we can only decide whether the areas under consideration have similar greylevel characteristics or not; information about the relative intensities is not retained. Yachida et al [YAT78] propose the use of positive and negative difference pictures by using the sign of the difference. Jain [Jai83b, Jai83d] has shown that the use of the positive, negative, and absolute difference and accumulative difference pictures allows for a simple segmentation and motion characteristics extraction. These will be discussed in the section on the segmentation. It has also been proposed that more information about the objects and their motion can be obtained by combining consecutive difference pictures [Jai83d, LeG82].

Many systems have used the concept of difference picture in practical applications [LeG82, ReJ83]. Clearly, the most attractive aspect of difference picture is their simplicity. In the simplest form the use of the difference picture appears to be noise-prone in the analysis. The changes in the illumination and registration, in addition the electronic noise of the camera, may result in many false alarms in non-trivial real world scenes. By using a likelihood ratio and the size-filter most of the camera noise may be eliminated. The changes in illumination will create problems for any intensity based approaches and can only be handled at a symbolic level. The misregistration of the frames results in assigning false motion components. If misregistration is not severe, accumulative difference pictures can eliminate it.

It should be emphasized here that by using any dissimilarity measure at pixel level, we are detecting only intensity changes. In a dynamic scene analysis this is the lowest level of analysis. After such changes have been detected, some other processes are required to interpret these changes. Experience has shown that the most efficient use of difference picture is in peripheral processes for directing the attention of interpretation processes to the areas of the scene where some *activity* is taking place. Jain [Jai81a] developed a decision tree, shown in Figure 2, using simple features of difference pictures that enables extraction of approximate information about activities taking place in a dynamic scene. Several systems have been developed that start with the difference pictures in motion understanding[JaN79, JaJ83, TSR82].

Figure 2

A decision tree to extract motion information in peripheral phase from difference pictures. Using some features of the difference pictures, approximate information about events in the scene may be extracted.



O	→	Covering of background by a moving object
B	→	Uncovering of background by a moving object
OC	→	Occlusion
TSO	→	Translation of one moving object
APP	→	Approaching object
REC	→	Receding object
PFP	→	Previous frame position of a totally displaced object
CFP	→	Current frame position of a totally displaced object
X	→	Uncertain

Recently, it has been demonstrated that difference and accumulative difference pictures may be used for moving observer also [Jai83b]. Changes detected in frames at the pixel level may be used to extract useful information about the motion of objects and for the segmentation, as discussed in the following section, of a dynamic scene where camera and object motions are not restricted.

2.2. Static Segmentation and then Matching

Segmentation is the task of identifying semantically meaningful components of an image and grouping the pixels belonging to such a component. Segmentation of an image has attracted many researchers [BaB81]. It is not necessary that segmentation be performed in terms of objects; some predicate based on intensity characteristics may also be used. Usually the predicates based on intensity characteristics are called features. If an object or a feature appears in two or more images, then it may be required to segment the images to identify it in images. The process of identifying an object, or a feature, in two or more frames is called the correspondence process. In the following discussion, segmentation refers to the identification of objects and partial segmentation to feature extraction.

It is possible to segment, or at least partially segment, each frame of a sequence using static scene analysis techniques and then use matching to solve correspondence and detect changes in the location of the corresponding segments for the detection of motion. Crosscorrelation and features in the Fourier domain are used for the detection of cloud motion in [ALR75]. Aggarwal and Duda [AgD75] detect edges in images and then used the notion of false and real vertices for matching in simulated cloud images. Several systems have been developed that consider each frame of the sequence and segment, at least partially, each frame to find regions, corners, edges or some other features in each frame [BaT79, ChJ75, ChW77, WaC80, Law81, Law82, MaA79, Pot74, Pra79, RoA79, Wil80]. The next step is to try to match these features in consecutive frames to detect whether these features have been displaced. Some economy in matching can be achieved by using a prediction based on the displacements in the previous frames.

Price and Reddy [PrR77] show that using region features, such as the size, location, elongatedness, color, area of the bounding rectangle, orientation, images of a city scene taken from two different viewpoints can be matched succesfully. Roach and Aggarwal [RoA79] use multilevel matching for tracking objects in a sequence. Their objects are polyhedral and can move in 3-D space resulting in occlusion. They allow multiple interpretations and carry uncertainty until it can be resolved. The features used in matching vary from velocity values to local pictorial features. Initially, two frames of the sequence are completely segmented and only pictorial features are used for matching. In subsequent frames the results of matching from earlier frames can be utilized. Knowledge about the block-worlds is extensively used by the system. In [AYT81] the matching of vertices for the analysis of the motion is performed using the junction properties of the block-world. The extension of such a matching approach to real world scenes is a non-trivial task.

Barnard and Thompson [BaT79] propose a method for computation of disparities in images. Their approach is based on the theory that the discretness, similarity, and consistency properties can interact to enhance the performance of a matching process. They first detect discrete points in images by applying an interest operator to each image. An interest operator locates points in an image that are significantly different from the points in its neighborhood. Most interest operators exploit the variance in the intensity at the point. After finding interesting points in both images a set of possible matches for each point of image 1 is created by considering that a point may match any point from the other image within a window of limited size. A relaxation approach is used to refine the set of matches for each point. It is shown that using this approach good results can be obtained for stereo images or for two images of a sequence.

Grosky and Jain [GrJ83] approximate a region corresponding to an object by an elliptical paraboloid using least squares fit to the intensity values. The parameters of the parabolaoid are used first to establish correspondence and then to obtain the translation and the rotational component of the surface. It is shown that the parameters of the paraboloid may be computed recursively in a pyramidal data structure while finding connected components using linking in pyramids. This fact may allow implementation of this matching on special architecture.

The major problem in these approaches is the segmentation step. If one statically segments each frame completely, the computational cost becomes prohibitive. Moreover, segmentation is a complex operation. If features are used then the problem is what features should be used. Simple methods, such as Moravec's interest operator [Mor81], usually give too many features, making matching very expensive and unreliable; complex features, such as corners are computationally expensive and their precise location is difficult to obtain.

Ullman [Ull79] argues that human visual system performs a matching at intermediate level, such as features. He posits that intensity based methods may be used only by the peripheral processes.

The segmentation of a static scene from its image has been a difficult problem. Most efforts to segment a real world complex scenes indicate that even after Herculean efforts such system can give *good* segmentation only in Utopian world. It is widely believed that motion can be used for segmentation, in contrast to the previously described methods which segment each frame and then use matching for motion detection. It is interesting to note that most approaches that suggest that each frame be segmented individually and then correspondence be established, usually demonstrate their approach considering synthetic scenes or a very few frames of a sequence.

2.3. Time Varying Edge Detection

Considering the importance of edge detection in static scenes, it is reasonable to expect that techniques for time-varying edge detection may play a very important role in dynamic scenes. In segment-and-match approaches efforts are wasted by applying detection and matching operations to features many of which may be static and hence obstacles for the extraction of motion information. By detecting only moving features one may enhance matching substantially. In fact Dreschler and Nagel[DrN81] first extract moving objects and then detect feature points for matching in their approach for the recovery of the 3-D structure. If in each frame of a sequence only those edges that are moving could be detected, the subsequent processes may use only these edges.

It appears obvious to apply 3-D edge detectors [ZuH81] to dynamic scenes by considering that the frames are stacked. Haynes and Jain[Hay82,HaJ83] showed that 3-D edge detectors do not give good results when applied to dynamic scenes. The major problem with the application of the 3-D edge detectors in a frame sequence is that the motion of the objects may result in frame-to-frame displacement of the edges that is inappropriate for the resolution of the 3-D detector. When this happens, the 3-D edge detectors can not be applied because the surface coherence implicitly required by these detectors is not satisfied.

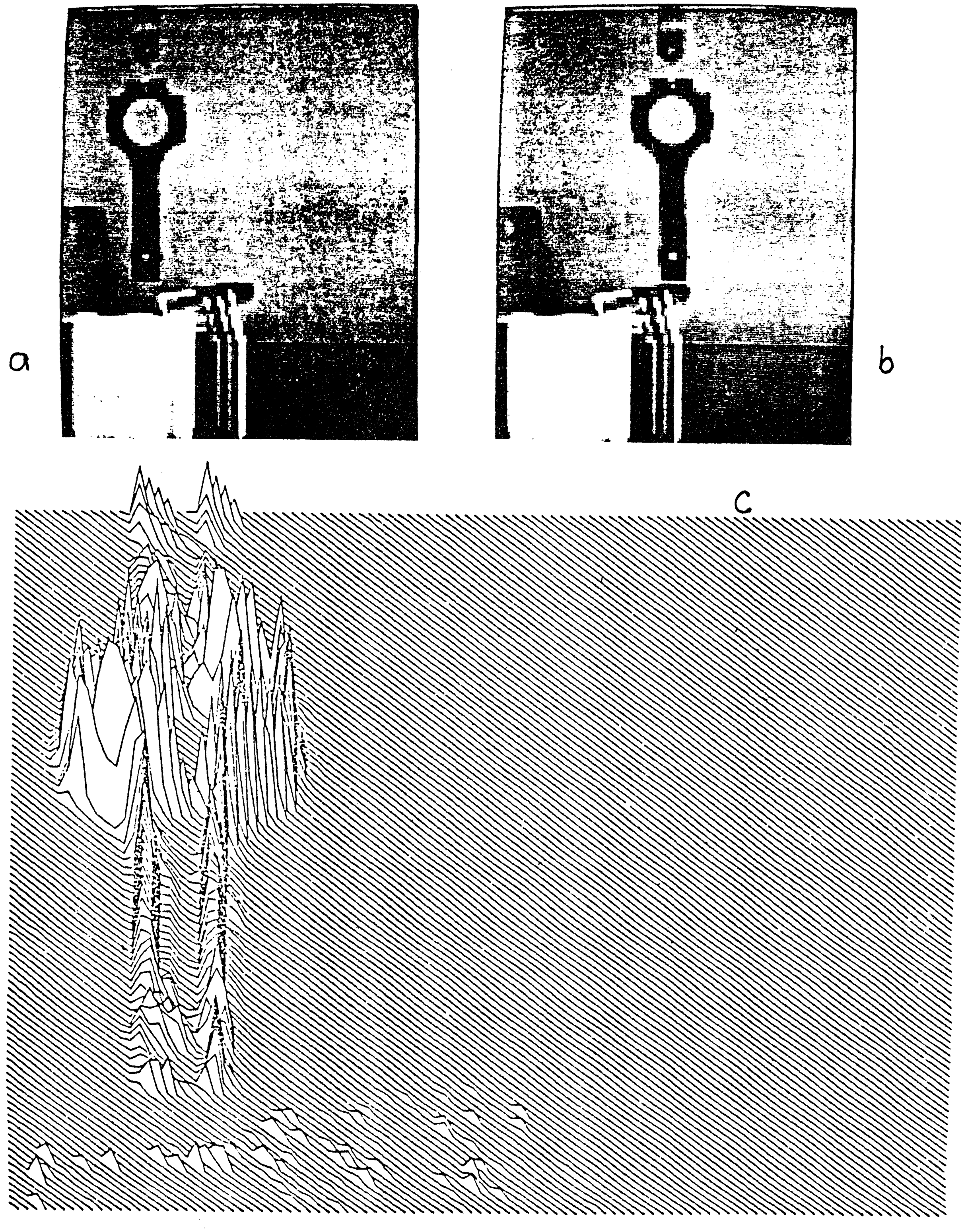
A moving edge is an edge in a frame *and* it moves. Haynes and Jain[HaJ83] argue that moving edges can be detected by combining temporal and spatial gradient using an *and* operator. For *and* they suggest the use of the product. Thus the time varying edginess of a point in a frame, $E(x,y,t)$ is given by

$$E(x,y,t) = \frac{dF(x,y,t)}{dS} * \frac{dF(x,y,t)}{dt}$$

where $\frac{dF(x,y,t)}{dS}$ and $\frac{dF(x,y,t)}{dt}$ are the magnitudes of the spatial and temporal gradients of the intensity of the point (x,y,t). In their experiments they use different conventional edge detectors to compute the spatial gradient and simple difference is used as temporal gradient. These simple operators were applied to many complex scenes, both the SCMO and the MCSO type. It was observed that this edge detector works effectively in most cases. By applying the threshold to the product rather than first differencing and then applying edge detector, as suggested by Jain [Jai81a], or by detecting edges and then computing their temporal gradient, as in [MaU79], this method overcomes the problem of slow moving and weak edges, see Figure 3 and 4. As shown in

Figure 4, this edge detector will respond to slow moving edges that have good edginess and to poor edges that are moving with appreciable speed. Another important fact about this detector is that there is no assumption of small displacement. The performance of the detector is satisfactory even when the motion is very large.

Figure 3 In 3a and 3b we show 2 frames from a sequence and in 3c the edges detected using the edge detector as in [HaJ83]. Compare the edges shown in 3c with those obtained using a 3-D detector, shown in 3d, and using a Sobel edge detector in the frame of 3a, shown in 3e.



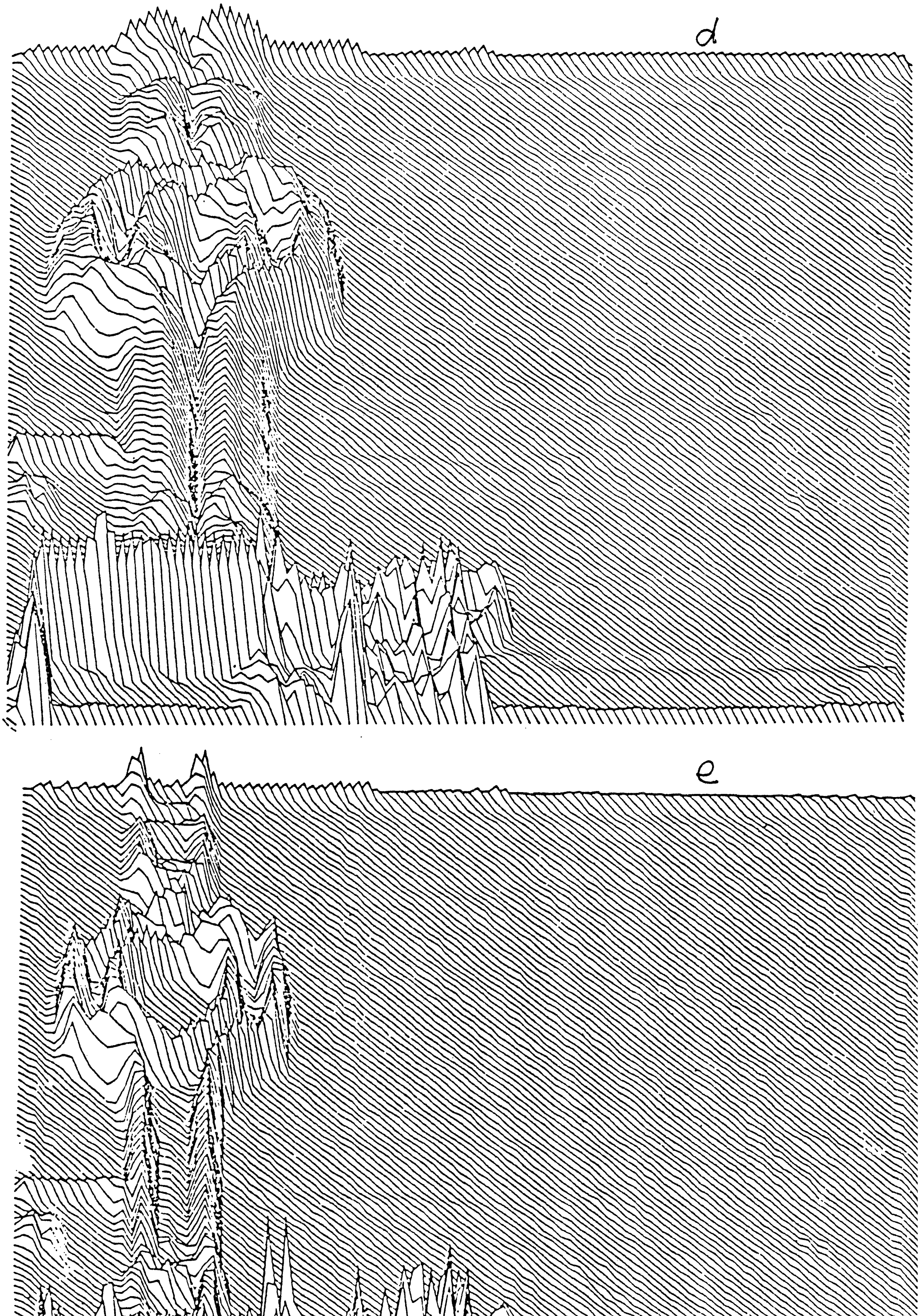
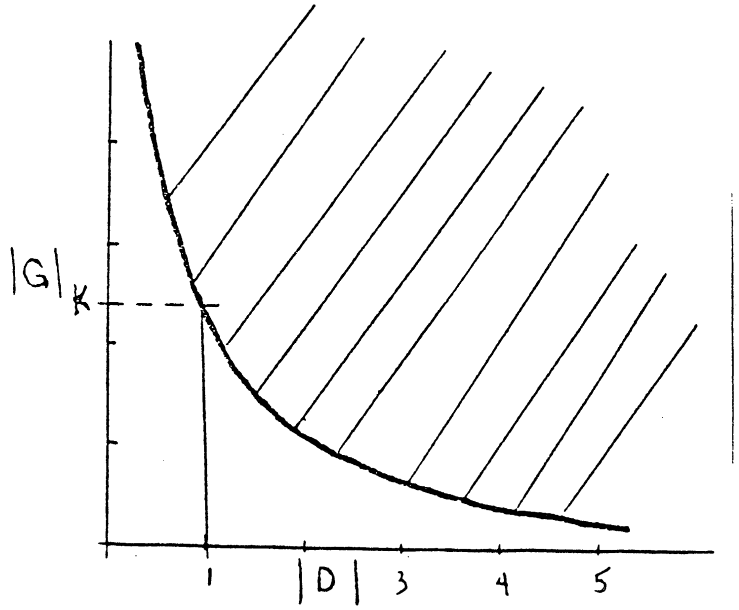

d
e

Figure 4

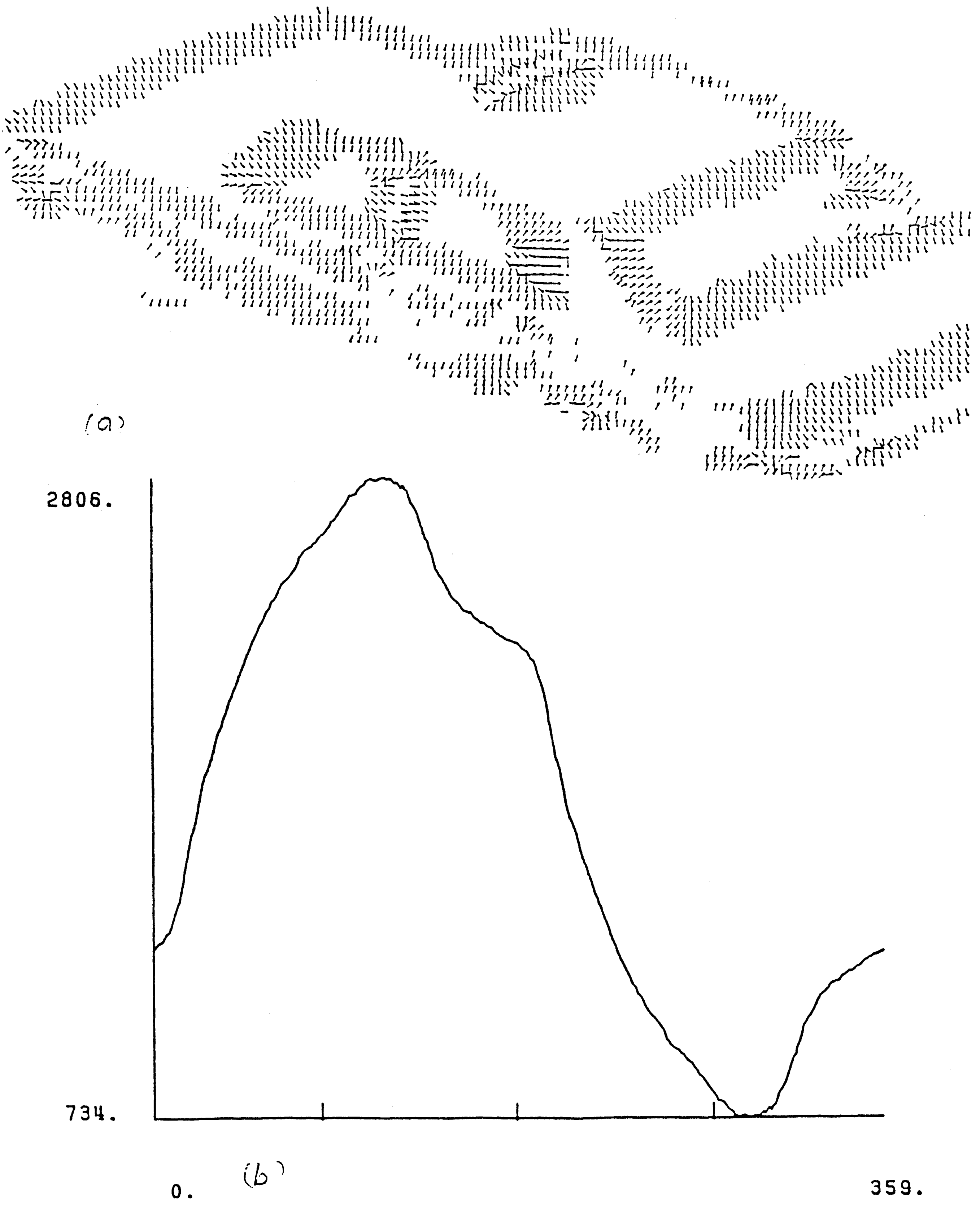
The curve showing the performance of the edge detector. Note that the slow moving edges will be detected if they have good contrast and poor contrast edges will be detected if they move well.



Though the direction of motion of an object can not be determined from the computations at a point, the direction can be determined by integrating the direction information along the edges for an object. The direction of motion computed at a point is within 180° of the direction of the spatial gradient at the point. If we use an accumulator array for an object and for each moving edge point of the object increase the count for all possible directions, then the peaks in the array will give the direction of motion of the object. In Figure 5 we show the direction of motion of an object as determined by this approach.

Figure 5

The direction of motion of every point on the image of a moving object are shown in the Figure 5a. To obtain the direction of motion of the object we used an accumulator array. The peak in the accumulator array indicates the direction of the motion of the object, as shown in the Figure 5b.



Marr and Ullman [MaU79] first find the zero crossings of $\nabla^2 G(x,y) * F(x,y,t)$ to detect edges in each frame of the sequence and then measure the derivative $\partial \frac{[\nabla^2 G(x,y) * F(x,y,t)]}{\partial t}$ at the zero crossings. They show that these measurements constrain the the local direction of motion to within 180°. They argue, as is done by Haynes and Jain [Hay82, HaJ83], that the direction of motion of an object can be determined in the second stage by combining the local evidence at every point of the object. Hildreth [Hil82] has proposed an optimization approach for computing the direction of motion of an object by using components at the boundaries.

3. Using Motion for the Segmentation

In many dynamic scene analysis systems the goal is to recognize moving objects and to find their motion characteristics. If the scene is acquired using a stationary camera, then segmentation generally refers to the separation of moving from stationary components of the scene and identification of individual moving objects based on velocity or some other characteristics. For MCSO and MCMO segmentation task may be same as above or may also include further segmentation of the stationary components of the scene by exploiting the motion of the camera. Most research efforts for the segmentation of dynamic scenes have been concerned with the extraction of the images of the moving objects observed by a stationary camera. It has long been argued by researchers in perception [Gib79, Ull79, Lee80] that motion cues help segmentation. The computer vision techniques for segmenting the SCMO dynamic scenes perform well in comparison to those for the segmentation of stationary scenes. The segmentation of moving camera scenes, into their stationary and nonstationary components, has received attention only recently [Jai82, Jai83b, KoK80, Law81, Law82, Wil80]. The major problem in the segmentation of moving observer scenes is that every surface in the scene shows motion. For separation of moving object images, the motion component assigned to various stationary surfaces in the images due to the motion of the camera, should be removed. The fact that the image motion of a surface depends on its distance from the camera and the surface structure, complicates the situation.

The segmentation may be performed either using region based approaches or edge based approaches. In this section we discuss some approaches for segmentation of dynamic scenes.

3.1. Stationary Camera

3.1.1. Using Difference Pictures

Jain, Militzer, and Nagel [JMN77] proposed an approach based on accumulative difference pictures for the separation of nonstationary component of a frame sequence. Jain and Nagel [JaN79] combine the properties of the ADPs and the DPs (they call them the first order difference picture and the second order difference picture, respectively) for the extraction of images of moving objects. In their approach several properties of a region of an ADP are computed to estimate the direction of motion, and the time for the object image to be displaced completely from its projection in the reference frame. They compute some properties of difference pictures obtained by comparing the reference frame with the current frame for the classification of a region into one of the following four classes:

1. static - where the object is in the reference frame,

2. mobile - where it is in the current frame,

3. o_grow - the part of the projection in the current frame, and

4. b_grow - the part of the projection in the reference frame.

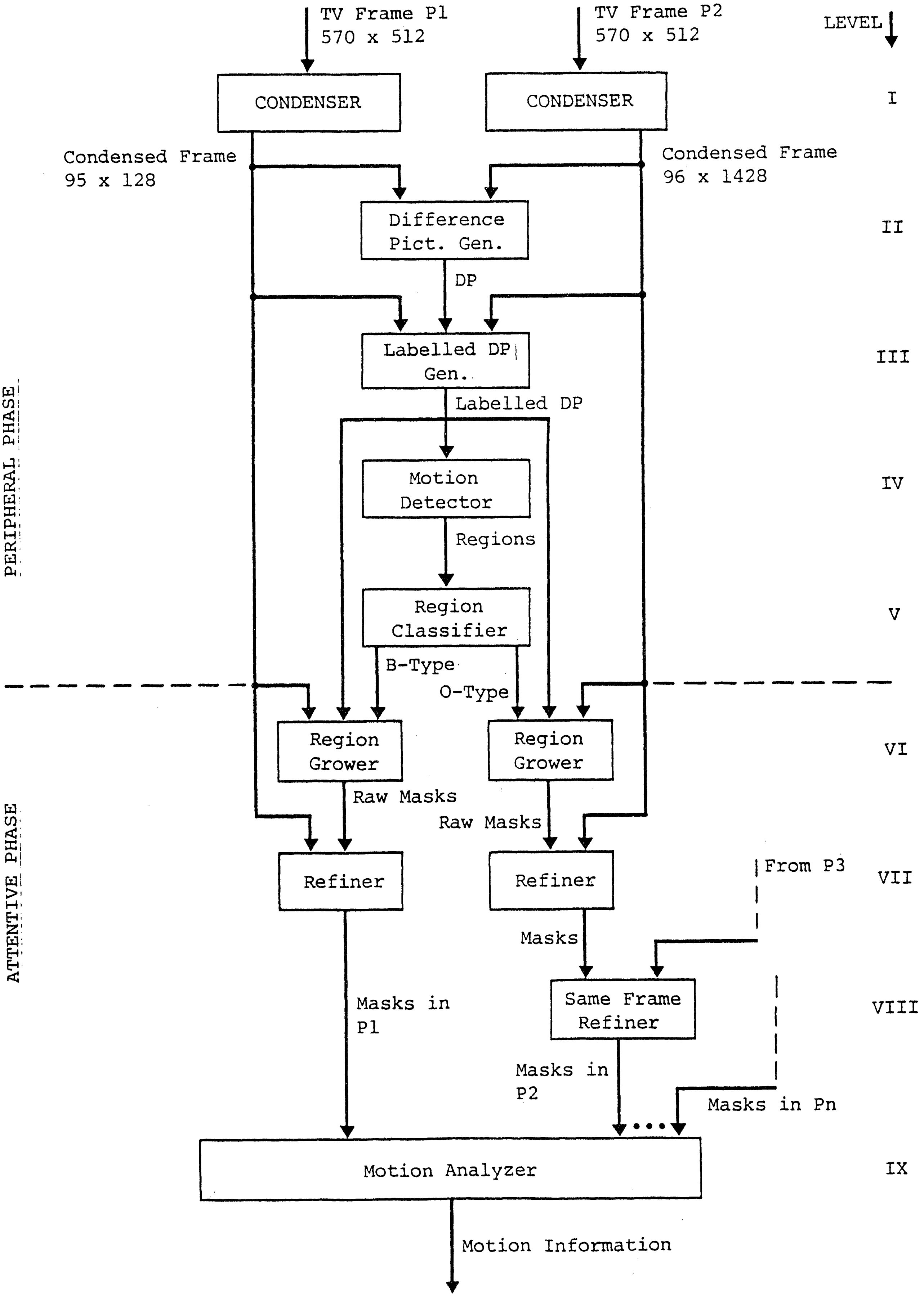
An algorithm was developed to combine the properties of regions in the ADP and the DP for the extraction of the images of the moving objects in the reference and the current frames after the objects are displaced from their projections in the reference frame. As discussed in [JaN79, Jai81b] this approach may be used to construct the reference frame containing only stationary components of the scene by replacing the part of the reference frame corresponding to the moving object by the background after the object images have been recovered. Since such a reference frame gives the images of moving objects in the difference picture, this may play an important role in the subsequent analysis, particularly in object recognition and motion understanding.

This algorithm has been used in many different real world TV frame sequences and with generally good results [Jai81a, DrN81]. A limitation of this algorithm, however, is that the object masks can not be recovered before the object is completely displaced from its projection in the reference frame. The limitation results in the failure of this algorithm in presence of the running occlusion. In many applications, even without running occlusion, it may be required to extract images of moving objects earlier.

Jain, Martin, and Aggarwal [JMA79] exploit the properties of a difference picture for the extraction of the images of moving objects. By comparing successive frames of the sequence, difference pictures are obtained and classified into o_grow and b_grow regions. Starting with the regions in the difference pictures many domain-independent properties of motion are employed for growing a region. The region growing approach depends on the type of region also. The masks thus obtained are further refined by comparing masks of the object in a frame obtained from two different pair of frames. The outline of the approach is shown in Figure 6. It was shown that this region growing approach results in good masks for the moving objects from two or three frames. A possible hardware implementaion of this approach for real time segmentation is discussed in [AgJ82]. A similar approach for segmentation is proposed in [YMA82] using some connectivity properties of difference pictures. Tang et al [TSR82] propose that the segmentation may be performed by finding difference region and then studying their frame-to-frame changes.

Figure 6

This figure gives the information flow in the algorithm for segmentation by Jain, Martin, and Aggarwal. Starting with the difference pictures, several refinements are applied to the regions to obtain the masks for the moving objects.


TV Frame P1
570 x 512
TV Frame P2
570 x 512
LEVEL
CONDENSER
CONDENSER
I
Condensed Frame
95 x 128
Condensed Frame
96 x 1428
Difference
Pict. Gen.
II
DP
Labelled DP
Gen.
III
Labelled DP
PERIPHERAL PHASE
Motion
Detector
IV
Regions
Region
Classifier
V
B-Type
O-Type
Region
Grower
Region
Grower
VI
Raw Masks
Raw Masks
ATTENTIVE PHASE
Refiner
Refiner
From P3
VII
Masks
Same Frame
Refiner
VIII
Masks in
P1
Masks in
P2
Masks in Pn
Motion Analyzer
IX
Motion Information

Recently, Jain [Jai83b, Jai83d] showed that using ADPs it is possible to segment a scene with very little computation. He defined absolute, positive and negative difference and accumulative difference pictures as following:

$$DP_{12}(x,y)=1 \ \text{if} \ |F(x,y,1)-F(x,y,2)| \ > \ T$$

$$=0 \ \ otherwise$$

$$PDP_{12}(x,y)=1 \ \text{if} \ F(x,y,1)-F(x,y,2)>T$$

$$=0 \ \ otherwise$$

$$NDP_{12}(x,y)=1 \ \text{if} \ F(x,y,1)-F(x,y,2)<-T$$

$$=0 \ \ otherwise$$

and

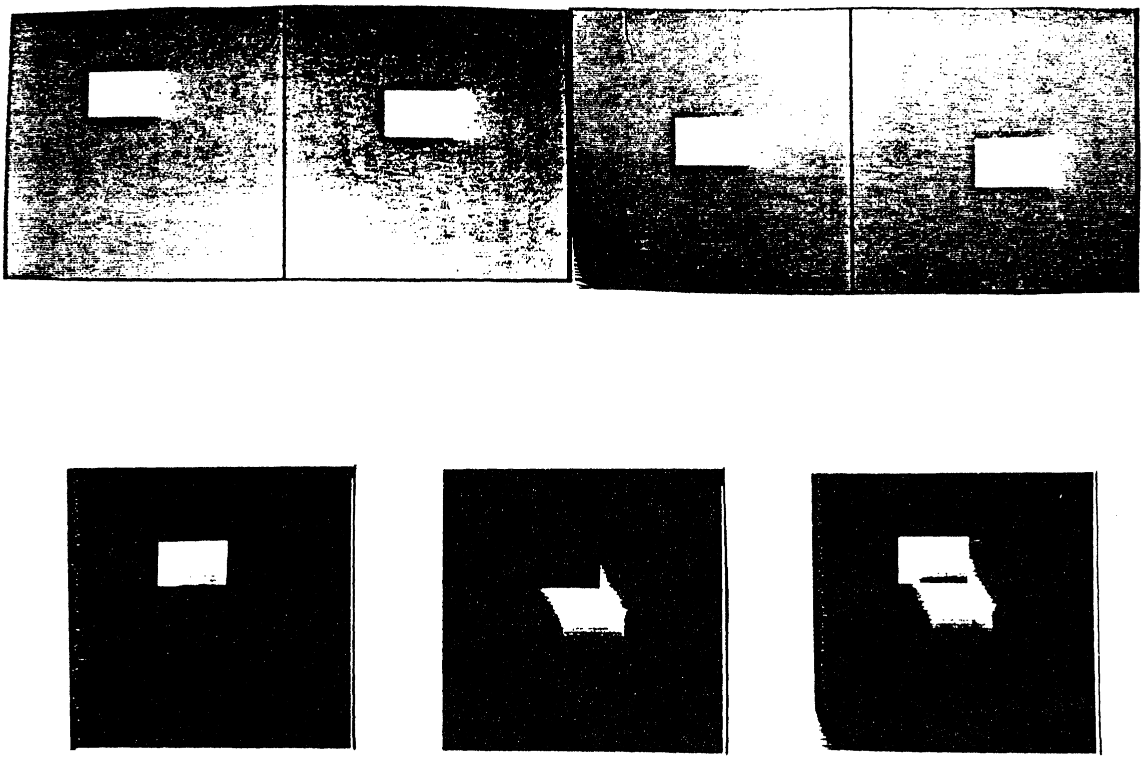
$$AADP_n(x,y)=AADP_{n-1}(x,y)+DP_{1n}(x,y)$$

$$PADP_n(x,y)=PADP_{n-1}(x,y)+PDP_{1n}(x,y)$$

$$NADP_n(x,y)=NADP_{n-1}(x,y)+NDP_{1n}(x,y)$$

Figure 7

Figure 7a shows frame 1, 5, 10, and 15 of a scene containing a moving object. The intensity coded positive, negative, and absolute ADPs are shown in figures 7b, c, and d, respectively.

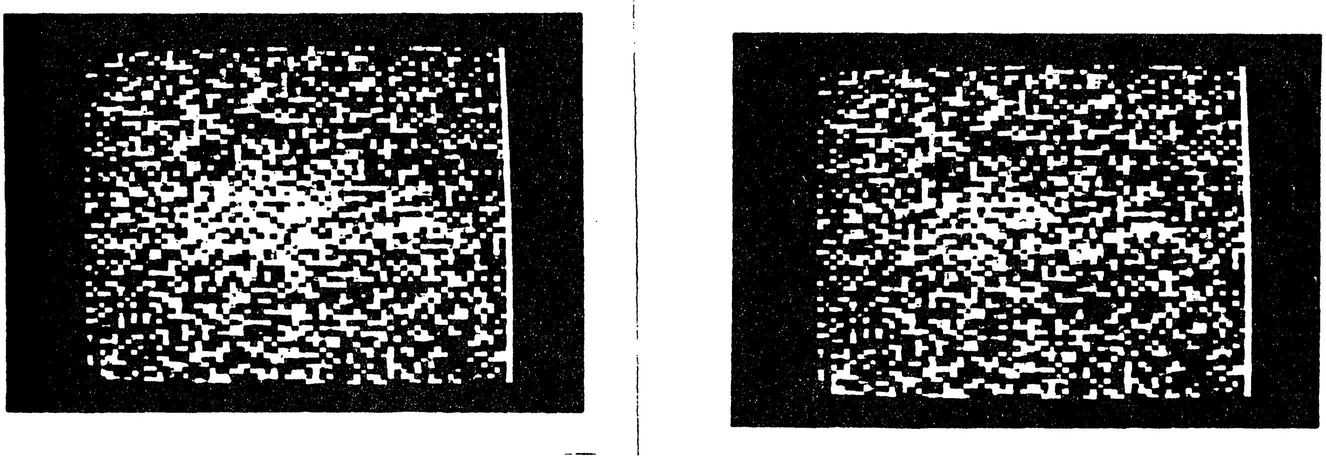


It is shown that in one of the ADPs (either PADP or NADP) the region due to the motion of the object continues growing even after the object has been completely displaced from its projection in the reference frame while in the other ADP it stops growing. The entries continue increasing in value in the area covered by the projection in the first frame. The accumulative difference pictures for a synthetic scene are shown in Figure 7. To obtain the mask of an object in the reference frame a test to check whether a region is still growing is required. The mask in the current frame may be obtained by considering the accumulative difference picture of the other kind. This method was applied in [Jai83b]. This approach, however, in its simplest form has the same limitation as the original approach based on the accumulative difference pictures proposed in [JaN79], namely, it can extract images of moving objects only after the object has completely displaced from its projection in the reference frame. It appears, however, that by computing properties in difference and accumulative difference picture, it is possible to segment images in complex situations, such as in running occlusion, also [Jai83d]. To avoid running occlusions from disruption of the segmentation process, the segmentation process should not wait until the object image is completely displaced from its position of the reference frame. In the system reported in [Jai83d], the region in the accumulative difference pictures are monitored to find the regions in place of the reference frame position of the object. It is shown that a simple test on the rate of increase of regions and on the presence of the *stale* entries allows the determination of the regions that are finally going to mature and result in the mask of the objects in the reference frames. The early determination of the reference frame position of the objects and hence extraction of masks for the objects allows necessary action to prevent the running occlusion.

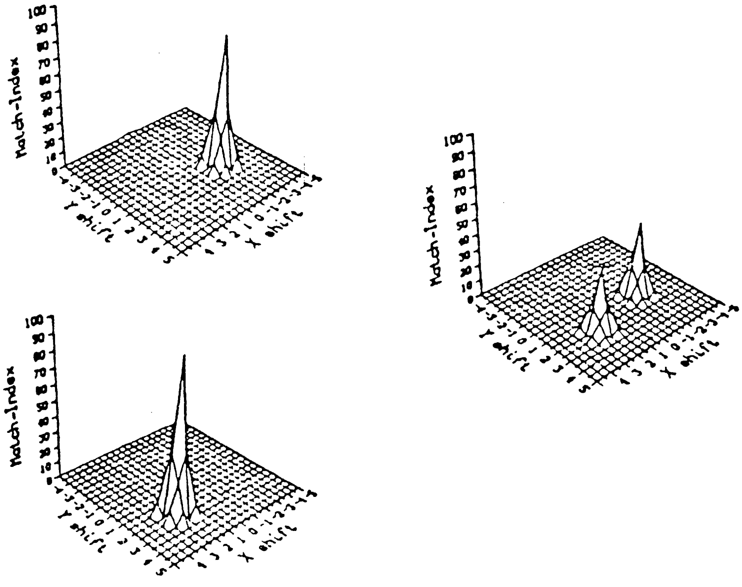
An approach for the segmentation of scenes containing textured objects and textured backgroud was proposed by Jayaramamurthy and Jain [JaJ82, JaJ83]. This approach also starts with the difference regions, then uses a shift-match and Hough transform to detect moving objects and their displacement component. The basic idea in this approach is to assume some displacement component for the objects and to verify the hypothesis using the structure of the intensities in the neighborhood of the points. A very attractive feature of the approach is the possibility of recovering objects images and the motion characteristics in presence of occlusion in textured scenes. In Figure 8 we show a textured sequence in which it is difficult to find what is happening. The algorithm successfully extracts the masks of the moving images and determines their motion characteristics.

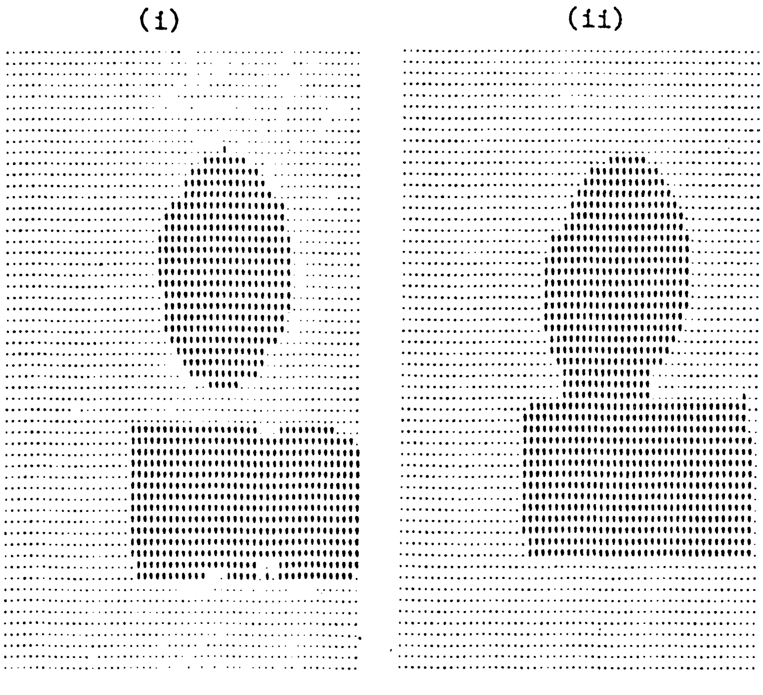
Figure 8

In Figure 8a and 8b two frames of a scene containing textured objects against textured backgrounds are shown. It is difficult to detect the objects and their motion. The approach presented in [JaJ82, JaJ83] gives the direction of motion and the images of the moving objects. The direction of motion is indicated by the accumulator array for the displacement component, shown in the Figure 8c, and the image of the moving objects are shown in the Figure 8d.



(a)


Match-Index
Y shift
X shift
Match-Index
Y shift
X shift
Match-Index
Y shift
X shift


(i)
(ii)

3.1.2. Using Velocities

Fennema and Thompson [FeT79] exploit the relationship, (see also [LiM75, CaR76]), between the spatial and temporal gradients of intensities to segment images. Using the relation

$$\frac{dF(x,y,t)}{dt} = -(F_x.u + F_y.v)$$

they show that if the direction of the velocity and intensity gradients at a point are ϑ_v and ϑ_g , respectively, and the velocity over a surface is constant, then values of v_g plotted against ϑ_g will lie along a cosine curve. Assuming sufficient variations in ϑ_g for the surface, the relationship between v_g and ϑ_g will uniquely define v. They use a clustering approach to determine the dominant velocities in the scene for the classification of each point.

Note that this approach is based on two assumptions: the linear relationship between the temporal and spatial gradient, and constant velocity of the object. To satisfy the first assumption, they require blurring of the images. The constant velocity component requirement means that this approach will not work in the presence of rotation. The results for the real world scenes containing translating objects are encouraging.

Thompson [Tho80] suggests that the velocity information be combined with the intensity information for segmentation. Starting with the velocity information one may extract regions of uniform velocity and then use region growing based on intensities for the extraction of the images of the moving objects. It appears that rotation will still cause problems for this approach. Velocity values at different points of a rotating object will be different in a frame sequence and may not form a good cluster for starting region growing based on intensity values. Occlusion of moving object will also give difficult time to the algorithm based on this approach.

Potter [Pot74] used template matching for segmentation of a dynamic scene. His approach is very sensitive to noise and may have difficulties even with laboratory scenes.

3.2. Moving Camera

If the camera is moving then every point in the image, with the exception of pathological case of points that are also moving with the same velocity, has non-zero relative velocity. The velocity of points depends on their distance from the camera and their own velocity. The difference picture based approaches may be extended for the segmentation of the scene, but if the aim is to extract the images of moving objects then more information will be required to decide whether the motion component at a point is due to its depth alone or is a combination of components due to it's depth and motion. The gradient based approach will also require additional information.

If the direction of the motion of the camera is known then the FOE with respect to the stationary component of the scene can be easily computed since the FOE will have coordintaes

$$x_f = \frac{dx}{dz}$$

and

$$y_f = \frac{dy}{dz}$$

in the image plane. As discussed in a later section, the velocity vectors for all stationary points of the scene project on the image plane so that they will intersect at the FOE. Jain [Jai82, Jai83b] proposed the Ego-Motion Polar transform of an image such that a frame $F(x,y,t)$ is transformed to $E(r,\vartheta,t)$ using

$$E(r,\vartheta,t)=F(x,y,t)$$

where

$$r = \sqrt{(x - x_f)^2 + (y - y_f)^2}$$

and

$$\vartheta = Tan^{-1}(\frac{y - y_f}{x - x_f})$$

In the EMP space all stationary points show displacement along the ϑ axis, the points that belong to moving objects have a displacement component along the r axis also. Thus by using any technique for the stationary camera case, discussed above, we may determine the displacement component in the EMP space and segment a scene into its stationary and non-stationary components. The results of the experiments reported in [Jai83b] with real scenes are very encouraging. In Figure 9 we show 3 frames of a sequence acquired using a moving camera; the results of the segmentation are shown in Figure 10.

Figure 9

Three frames of a scene acquired using a moving camera are shown in Figure 9a, b, and c.

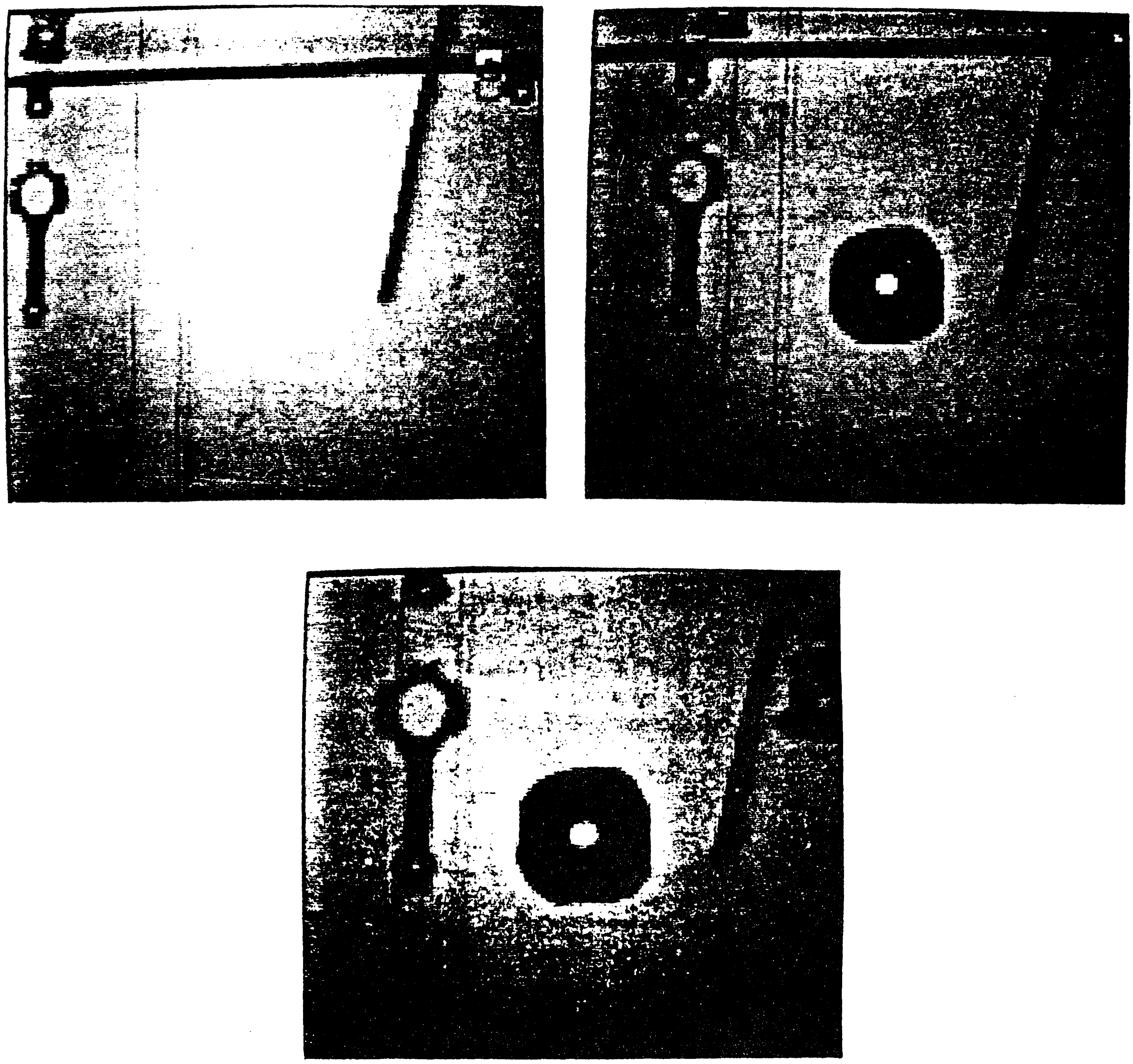
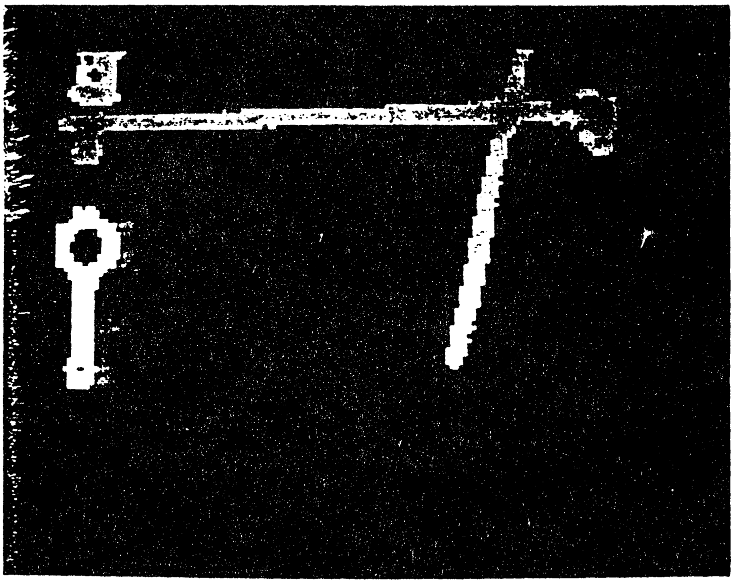


Figure 10

The segmentation of the frames shown in Figure 9 as obtained by [Jai83b] is shown here. The moving objects are brighter.



It appears that more information about moving as well as stationary objects may be extracted using a complex logarithmic mapping rather than the simple polar mapping about the FOE. This mapping will be discussed in a later section.

3.3. Edge-Based Methods

If moving edges are detected, then images of the moving objects in the case of the SCMO can be obtained easily. The segmentation of the static and dynamic component of the scenes may require techniques similar to that discussed in a previous section for the MCMO. No efforts have been made to extract masks of the moving objects from moving edges obtained using a moving edge detector such as [HaJ82, HaJ83]. The results of the moving edge detector for different real world scenes indicate the feasibility of the approach. A problem in this endeavor is likely to arise due to missing edges. The edge detectors fail on the points for which the motion is along the spatial gradient. The boundaries of an object obtained using a moving edge detector may, therefore, require some technique for filling such gaps in the boundary of the object.

4. Optical Flow

Gibson [Gib66, Gib79] proposed the concept of optical flow in his theory of ecological optics. Optical flow is the distribution of velocity at each point of the image relative to the observer. It has been shown that optical flow carries valuable information for the analysis of dynamic scenes [Lee80, Clo80, Pra80, LoP80, Pra82]. Some methods have been proposed for the extraction of the information assuming that optical flow is available. Techniques developed for the computation of optical flow do not, however, result in the optical flow of the quality required by the information recovery processes. In this section we first discuss current methods for the computation of the optical flow and then consider the intrinsic information and its recovery.

4.1. Computation of the Optical Flow

The optical flow is determined by computing the velocity vector at each pixel of the image. Several schemes have been devised for obtaining optical flow from two or more frames of a sequence. These schemes can be classified in two general categories: feature based and gradient based. Interestingly, though optical flow is more useful for a moving camera, in most of the proposed methods for the computation of the optical flow the results are shown for scenes containing moving objects acquired by a stationary camera. When the camera is stationary, most points in a frame have zero velocity, since usually a very small subset of a scene is in motion. The main application of the optical flow is in the case of a moving camera; it is implicitly assumed that the techniques developed for the stationary camera scenes will be extensible to the mobile camera too.

4.1.1. Feature Based Methods

These methods compute optical flow in the form of disparity vectors by first selecting some features in frames and then using matching to solve the correspondence problem. Barnard and Thompson demonstrated that disparities may be computed using relaxation [BaT79]. This approach was discussed in section 2.2. As discussed there, the problem of selection of features and establishing correspondence is not an easy one. Moreover, this method gives velocity vectors only at sparse points. Nagel [Nag82b] and Nagel and Enkelmann [NaE82] have recently proposed a method based on the second order intensity variation to compute velocity vectors for corner points and then use relaxation to obtain vectors for the missing points. He obtains the displacement vector at a point by minimizing the function

$$MD=\sum[F(x,y,t_1)-F(x-\delta_x,y-\delta_y,t_2)]^2$$

where δ_x and δ_y are displacement components in the x and y directions, respectively.

He developed a formalism for iterative refinement of a displacement estimate for image locations around a corner. The displacement estimate obtained at the corner may then be used to compute the estimates in the neighborhood of the corner point. This approach, thus, starts at the corners and propogates the displacement field over the image. Hill and Jain [HiJ83] use intensity profiles in conjunction with a bilinear-recursive hill climbing approach to obtain the velocity vectors to sub-pixel precision and then used relaxation across several frames to refine the vectors obtained over a sequence. No effort was made to propogate the vectors to the neighborhood. The behavior of relaxation algorithms for the propagation of velocity vectors in images is not yet understood and hence efforts for propagation are usually sensitive to noise and occluding boundaries in images.

No experience is yet available for scenes obtained using a moving camera for the computation of optical flow using the feature based approach. It appears that the difficult problem in this case would be that of extrapolating, or filling, the disparity vectors to obtain the optical flow. Since most approaches obtain disparity vectors only to the resolution determined by the feature locations, the extrapolation may result in noisy optical flow.

4.1.2. Gradient Based Methods

The methods in this class exploit the relationship between the spatial and temporal gradients of intensity at a point in a frame. Limb and Murphy[LiM75] and Cafforiao and Rocca [CaR76] use the relationship to compute velocity at points for reduction of bandwidth for transmitting TV images. As discussed earlier, Fennema and Thompson [FeT79] applied this to the segmentation of images using the velocity of points. These researchers use the relation:

$$F_x u + F_y v + F_t = 0$$

where $u = \frac{dx}{dt}$ and $v = \frac{dy}{dt}$ and F_x, F_y, and F_t are the partial derivatives of image intensity with respect to x, y, and t, respectively. This equation can be written in the following form:

$$(F_x, F_y) \cdot (u, v) = -F_t$$

Fennema and Thompson [FeT79] compute

$$V_g = -\frac{F_t}{\sqrt{F_x^2 + F_y^2}}$$

where V_g is the component of the velocity in the direction of the intensity gradient in an image. They assumed that surfaces have uniform velocity and computed the velocity component using Hough transform. Their method will not work in the presence of rotation due to the assumption of uniform velocity. Moreover, this approach is good for segmentation of images, as it gives velocity for an object rather than velocity at every point.

Horn and Schunck [HoS81] assume that a velocity field varies smoothly everywhere in an image. They developed an iterative approach for the computation of the optical flow using two or more frames. The following iterative equations were used for the computation of the optical flow:

$$u = u_{av} - f_x \frac{P}{D}$$

$$v = v_{av} - f_y \frac{P}{D}$$

where

$$P = f_x u_{av} + f_y v_{av} + f_t$$

$$D = \lambda^2 + f_x^2 + f_y^2$$

In the above equations f_x, f_y, f_t, and λ represent the spatial gradients in the x and y directions, the temporal gradient, and a multiplier, respectively. For the case of two frames the method was iterated on the same frames many times, in case of more frames each iteration used a new frame. They demonstrate their method with synthetic frame sequences. The smoothness constraint is not satisfied at the boundaries of objects because the objects may be at different depths. If the objects are moving then also the constraint will be violated. The abrupt changes in the velocity field at the boundaries cause problems in this approach. Schunck and Horn [ScH81] discuss some heuristics to solve these problems. Our experience in applying this approach to real scenes [Bor82] showed that the optical flow computed using this approach is very noisy.

An important fact about gradient based methods is that they assume linear variation of the intensities [HLS80] in images and compute the velocity at a point using this assumption. Typically, it is expected that such an assumption will be satisfied at the edges in images and hence velocity can be computed at the edge points. Fennema and Thompson [FeT79] blur their images to satisfy this condition; Horn and Schunck [HoS81] use relaxation to fill in the regions of uniform intensity. Some efforts have been made to combine the feature matching with the gradient approaches for the computation of the optical flow [Gla81,Yac81]. Nagel [Nag82a, Nag82b] argues that the linear model of intensity variation at the edges is too simplistic; he proposes a quadratic model for the computation of the velocity. The intensity at a point in an image must be represented by

$$I(x,y)=I(x_o,y_o)+I_x(x-x_o)+I_y(y-y_o)$$

$$+\frac{1}{2}I_{xx}(x-x_o)^2+I_{xy}(x-x_o,y-y_o)+\frac{1}{2}I_{yy}(y-y_o)$$

where

$$I_{xx}=\frac{\partial^2 I}{\partial x^2} \quad I_{xy}=\frac{\partial^2 I}{\partial x\,\partial y} \quad I_{yy}=\frac{\partial^2 I}{\partial y^2}$$

The results obtained, using this approximation, for real world frame sequences are encouraging.

4.2. Information in Optical Flow

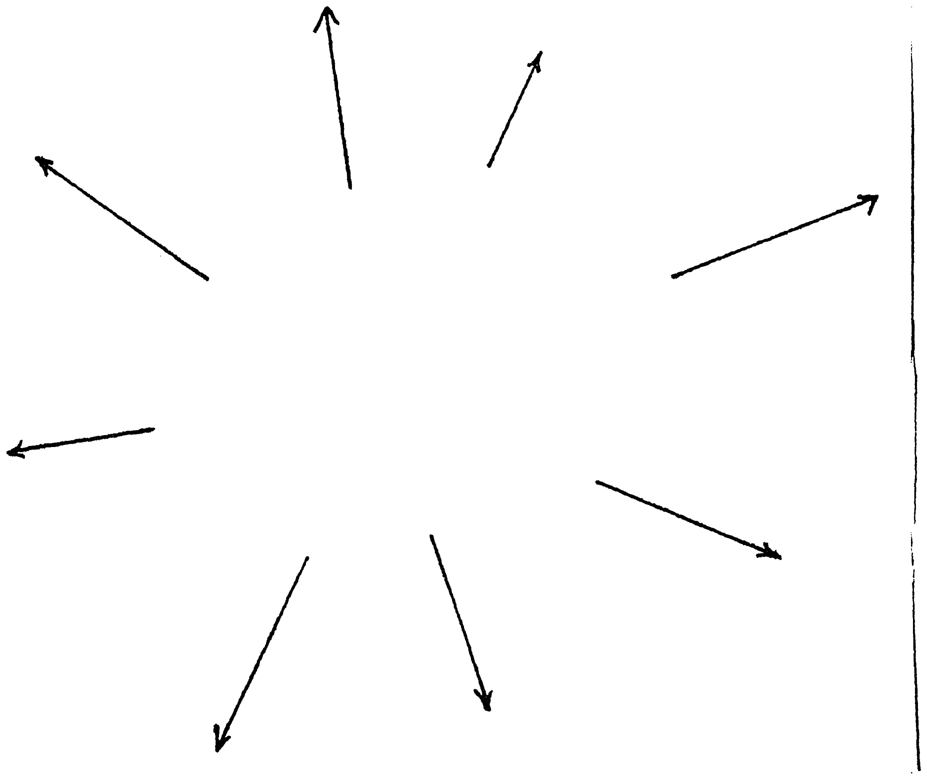
Assuming that somehow high quality optical flow has been computed, researchers have studied what kind of information is available in extractable form in the flow field. Clocksin [Clo80] and Prazdny [Pra80, LoP80, Pra81] addressed the computational aspect of this problem. They assume that the environment contains rigid stationary surfaces at known depths and that the observer, the camera, locomotes through this world. The optical flow can be derived from the known structure. They then show how to invert the process giving the structure of the environment from the computed optical flow field.

Clocksin [Clo80] argues that areas with smooth velocity gradients relate .to a single surface and can give information about the structure of the surface. The areas with large gradients give information on occlusion and boundaries because only two different objects at different depths, can move at different speeds relative to the camera. Using an observer-based coordinate system, he derives a relationship to recover the surface orientation from the smooth velocity gradients. The orientation is specified with respect to the direction of the motion of the observer.

Longuet-Higgins and Prazdny [LoP79] allow for rotational motion of the observer. They used the fact that in this case the velocity field is the vector sum of a translational component and a rotational component. The translational component is directed towards a point, called the Focus of Expansion (FOE, for the approaching motion of the observer) or the focus of contraction (for the receding motion of the observer), in the image. This is shown in Figure 11. This point is the intersection of the direction of motion and the image plane. The structure of surfaces can be recovered from the first and second spatial derivatives of the translational component. The rotational component is fully determined by the angular velocity.

Figure 11

The velocity vectors due to the stationary components of the scene, for a translating observer, meet at the FOE.



The importance of the FOE for the recovery of the structure from the translational component of optical flow encouraged several researchers to develop methods for the determination of the FOE. Jain [Jai83a] proposes a method, using a geometric argument, based on the minimization of the disparities of tokens in individual frames. The FOE in an image is the point at which the function

$$M(\hat{O})=\left[\sum_{k=1}^{k=n} L(\hat{O}P^{1})-\sum_{k=1}^{k=n} L(\hat{O}P)\right|$$

is extremized. In the above equation $L(.)$ is the Euclidean distance between two points, P^1 and P are the tokens in the second and first frames, $\hat{O}$ is a point in the search space which may be considered a tentative choice for the FOE, and n is the number of tokens in frames. The attractive feature of this method is the fact that no correspondence is required. Another attractive feature of this approach is that the search space is gradual, allowing application of the irrevocable search methods such as gradient techiques. Figure 13 shows the function for an approaching observer for the two frames shown in Figure 12. Emerging and disappearing tokens may create problems for this approach. Efforts to apply this to real scenes have not been encouraging [JeJ83a]. Lawton [Law81, Law82] combines the determination of the FOE and the extraction of tokens and shows that some success can be achieved in real scenes. Prager [Pra79] has emphasized the importance of the known value of the FOE.

Figure 12

A composite frame showing the locations of the tokens in two frames. The features for different frames are shown marked x and o.

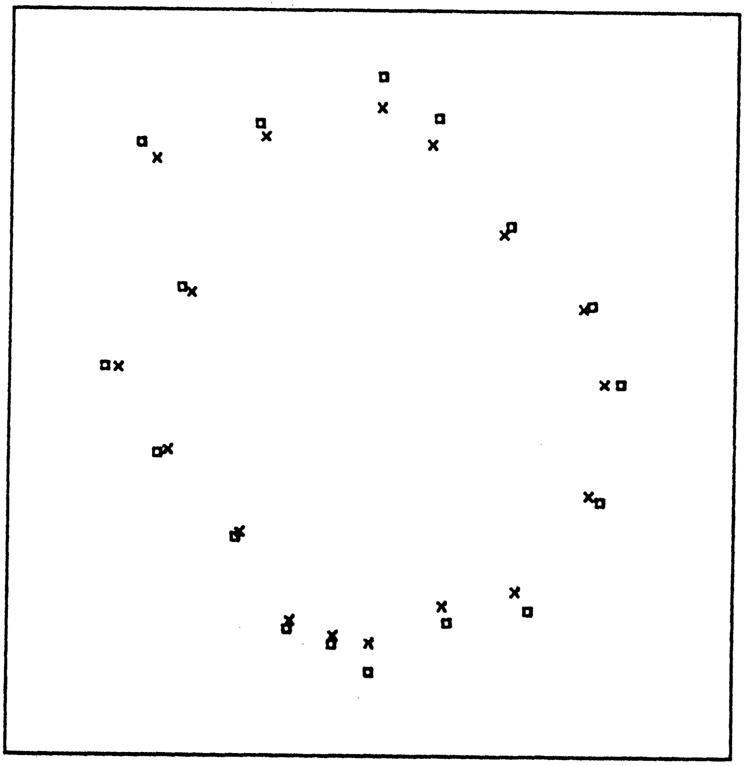
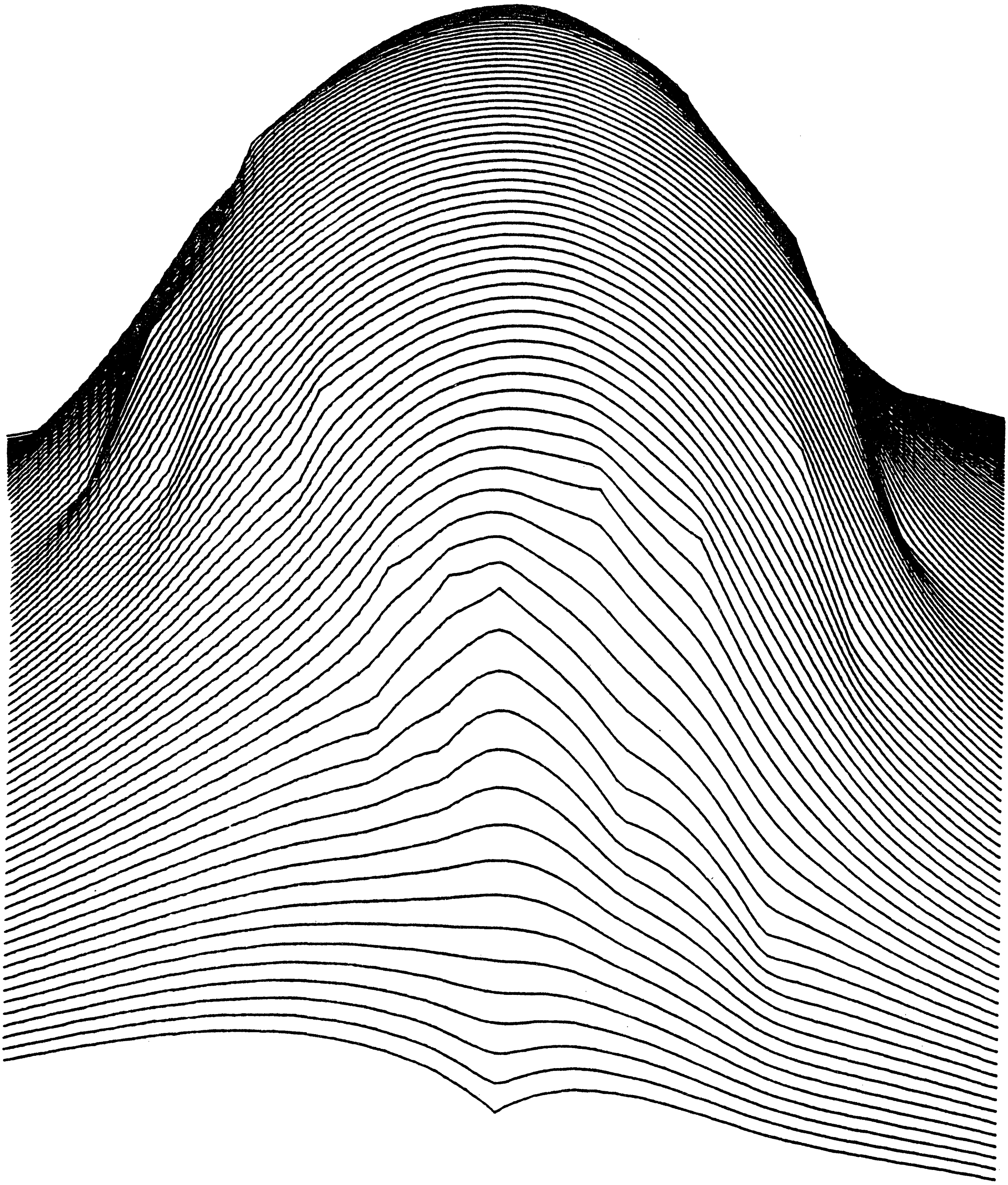


Figure 13

The function for computing the FOE for the frames shown in the figure 12. Note that the function is gradual and hence gradient techniques may be used for the determination of the FOE.



Jerian and Jain [JeJ83a] point out that after the FOE is determined using a direct method, it should be used to solve correspondence problem. The next step is to recompute the FOE using the matching thus obtained. It is expected that this type of approach may result in better correspondence and a more precise location of the FOE.

If the FOE is correctly determined then it may be used for the computation of the translational component of the optical flow. Since all flow vectors meet at the FOE, the direction of the flow vector is determined, only magnitude remains to be computed. Thus, the two dimensional problem of the computation of the optical flow is reduced to a one dimensional problem. This fact has been specified by many researchers, but not applied, possibly due to the uncertainty in the location of the FOE in real scenes using the proposed approaches.

By detecting edges in optical flow field, one may detect depth edges in frames. Thompson et. al. [TMB82] propose an approach for edge detection in the flow field. The flow field is computed in hierarchical structures using

5. Recovering 3-Dimensional Information

Most systems for the determination of motion parameters were concerned with the parameters in the image plane, i.e. in 2-D space. The interpretation of 2-D displacements in terms of 3-D motion is much more complicated due to the fact that the picture formation process results in a 2-D projection of the 3-D events leading to a loss of information. The recovery of the 3-D motion parameters and the 3-D structure of the objects has attracted much attention from researchers in the last few years. Ullman [Ull79] argues that the ***rigidity*** assumption about an object may help in the recovery of the structure. The rigidity assumption states that any set of elements undergoing a 2-D transformation which has a unique interpretation as a rigid body moving in space should be so interpreted. The research in human perception [Tod82, Joh76] suggests that the human visual system exploits this fact.

The research in the recovery of 3-D structure from image sequences may be divided in two general classes. Some researchers have been concerned with the problem of recovery of the structure and motion using a minimal number of points in a minimal number of frames. Recently, the trajectory based recovery has attracted some attention.

5.1. Recovering Structure Using Tokens

Suppose that we apply an interest operator to consecutive frames of a sequence and extract some interesting points or tokens, such as corners, then using some method discussed in an earlier section, manage to succeed in solving the correspondence problem. Ullman shows that if token correspondences have been established, it is possible to recover the 3-D location of four noncoplanar points from their 3 orthogonal projections. This gives an implicit 3-D structure of the object. He shows that if the points are noncoplanar then a unique structure can be recovered; in the case of coplanar points, the structure, to a reflection, can be recovered. Note that these results are obtained for the orthographic or parallel projection. For the case of perspective projections, two views of five non coplanar points are required. The equations for the recovery are nonlinear and require iterative methods for the solution. Williams uses a heuristic method [Wil80] for the recovery. Roach and Aggarwal [RoA80] present results of their efforts to solve the resulting nonlinear equations using standard techniques available under IMSL. In their studies they observed that in the presence of noise, the minimal solution does not give correct results; significant overdetermination (2 views of 15 points, or 3 views of 8 points) may be required.

Nagel and Neumann [NaN81] give a compact representation of the nonlinear equations required for the specification of 3-D rotations of rigid objects. They show that Ullman's polar equations are a special case of their solution. They derive the following compact relation for the recovery of structure:

$$(X_{p21}\ X\ X_{p22}D')X'_{p11}(X_{pm1}\ X\ X_{pm2}D'_{.})DX'_{p12}=$$

$$(X_{p21}\ X\ X_{p22}D')DX'_{p12}(X_{pm1}\ X\ X_{pm2}D')X'_{p11}$$

where X_{pmn} denotes a vector representing image plane coordinates in the form $[x_p, y_p, f]$ where f is the focal length of the imaging device, of m-th object point in n-th frame, and D is the rotation matrix. For $m=3,4,5$ this equation gives a set of three equations for the determination of 3 rotation parameters.

Tsai and Huang [TsH81,THZ82] introduce eight *pure parameters* for the case of a rigid planar patch undergoing general 3-D motion. Using the Lie group theory of transformations, they show that for two given successive views the solution is unique. From computational point of view, a very attractive feature of their approach is that the parameters can be computed by solving a set of linear equations. They demonstrate that though theoretically 6 solutions are possible, practically the maximum number of solutions is two. In [THZ82] it is shown that the actual motion parameters can be estimated by computing the singular value decomposition of a 3 X 3 matrix consisting of the eight pure parameters and that the number of solutions is either one or two. This approach has been extended to rigid objects with curved surfaces in [TsH82], where it is shown that seven points in two frames are required to uniquely estimate the 3-D motion parameters. The points should not be traversed by two planes with one plane containing the origin, nor by a cone containing the origin. For the estimation of 3-D translation by solving a set of linear equations, derived using a set of 8 points in two frames, *essential parameters* were introduced.

By avoiding uncertain and time consuming iterative methods required for the solution of non linear equations, one can hope to recover the 3-D motion in realistic situations in real time. An effort [JeJ83b] to apply the approach using the IMSL package to recover the motion parameters in a real scene shows, however, that the proposed approach is very sensitive to the location of points. It was observed that a very high precision in the location of the tokens may be required to obtain reliable results.

The feature based methods for the recovery of the structure or for the estimation of the motion parameters require two difficult steps before they are applied: the precise location of the tokens or points, and the correspondence. If we apply interest operators based on small neighborhoods, then the number of tokens extracted in a real image is very large making correspondence a difficult problem. The operators based on a large neighborhood and higher order greylevel characteristics do result in a reasonable number of tokens, reducing the complexity of the correspondence, but their location may not be precise. Even if the location is obtained at the resolution of the pixel, it appears that results obtained using the methods discussed above may not be reliable.

5.2. Trajectory Based Methods

All above methods depend on a set of points in two or three frames. If a token is traced over several frames, by solving correspondences, then one obtains the 2-D trajectory of the point. It appears that the efforts to recover the structure and motion in 3-D from the trajectories may be more reliable than those based on a set of features in a few frames. A trajectory may be interpolated by using curve fitting techniques to obtain a better resolution in the 2-D path. Moreover, experiments by Dreschler and Nagel [DrH81] and Hill and Jain [HiJ83] show that the correspondence problem may be simplified by considering more than two frames and extending relaxation across frames.

Webb and Aggarwal [Web81, WeA82b] propose the use of several monocular views for recovering the 3-D structure of moving rigid and jointed objects. They assume that a general motion of objects may be considered, over at least a short interval, as a rotation about a fixed axis and a translation. The fixed axis assumption allows recovery of the structure, under parallel projection, for even two points. If a point on the object is fixed then the other points trace out circles in planes normal to the axis of rotation. These circles are projected as ellipses under parallel projection. The structure of points can be recovered to within a reflection by finding the parameters of the ellipse.

Jointed objects are composed of two or more rigid objects. Webb and Aggarwal [WeA82b] present an algorithm, assuming that the feature selection and the correspondence problems have been solved and that the fixed axis assumption is satisfied for each rigid component, for identifying jointed components and for recovering their structure. The most attractive feature of this approach is that the recovery requires fitting an ellipse and hence the recovery method is not as sensitive as the methods based on 2 or 3 frames.

Recently, Sethi and Jain [SeJ83] have extended the trajectory based approach using perspective projections. They have shown that rotation of a point about a fixed axis results in an ellipse in the image plane. For the case of rotation about an axis parallel to the X or Y axis the 3-D coordinates can be recovered, to a scale, using one point only. Their work is inspired by some experiments of Todd [Tod82], who shows that humans appear to be interpreting rigid and non-rigid rotations based on trajectories. Sethi and Jain [SeJ83] shows that if a point is rotating about an axis parallel to the Y axis that passes through a point z_0 on the Z axis, then the equation of the ellipse can be written in the general form:

$$\frac{(x_p-h)^2}{a^2}+\frac{(y_p-k)^2}{b^2}=1$$

The following relations can be used to recover the structure from the parameters of the ellipse:

$$y=\frac{2b^2}{a^2k}(1+z_o)$$

$$x=y\frac{x_p}{y_p}$$

and the radius of the rotation is

$$r=\frac{b}{k}(1+z_o)$$

It was shown further that if the axis of rotation passes through a point x_0, then the major and minor axes of the ellipse are oriented differently. By using transformations given by

$$\hat{x}_p=x_p-\frac{y_p}{y}x_o$$

and

$$\hat{y}_p=y_p$$

it is possible to transform the ellipse to the standard form and use the standard equations to recover the structure. The experiments with the synthetic data are encouraging.

The case of an arbitrary axis was not considered for the recovery of structure in [SeJ83]. Since the parameters of an ellipse can be obtained using only 5 points, it is possible to recover the structure from 5 frames. Note, however, that this method considers rotation about a fixed axis. It will be interesting to consider trajectories by relaxing the fixed axis assumption and see whether structure may be recovered.

6. Motion Understanding

The analysis of a frame sequence may result in the extraction of the images of moving objects, their 3-D structure, and their frame-to-frame displacements. In many applications the aim may be to name moving objects and to describe the events taking place in the scene. The recognition of the object may be performed using an image or the 3-D structure of the object. It appears that the motion characteristics of objects may also help in the recognition. Different objects have different kinds of motion: humans walk, cars run, aeroplanes fly. The motion characteristics of objects are difficult to obtain directly from the frame-to-frame displacement. Much of the recognition and analysis of things by humans does not require fine detail and precise analysis of small details and parts. Details are only brought into play when the object or the process is the focus of attention.

An object with complex motion in each of its several parts can be abstracted to a simple moving block undergoing rigid motion. The abstracted motion can be simple translation or rotation, whereas the real motion may be very complex. At the next level of analysis one may try to analyze the motion of parts of the object; the knowledge of the motion of the abstraction of the object may help in the analysis of the motion of individual components of the object. It appears that a correct approach is to *somehow* compensate for the known motion of a higher level abstraction of the object to get detailed motion of the lower level parts. Jerian and Jain present their approach for such a system in [JeJ83b].

6.1. Motion Representation

Asada, Yachida, and Tsuji [AYT81] present some interesting experiments in understanding motion of blocks in 3-D space. Blocks can be in complex configurations and may be undergoing coincidental motion, such as a block rotating around a joint attached to another moving block. Assuming orthogonal projection and exploiting knowledge about the blocks-world they solved the correspondence problem. The most interesting and the novel part in their system is the hierarchical repesentation for articulated motion. The complex rotation of objects is interpreted as the combination of simple rotations of individual blocks. Suppose that there are three blocks, A, B, and C. A is rotating about some fixed axis, B is connected to A and is rotating about some other axis, and C is connected to B and is also rotating about yet another axis. It is possible to recognize the motion of A; but motion of B and C becomes complex. They show that a motion comprising a translation and rotation between two frames may be represented as a rotation. They developed methods for obtaining axis of rotation and the rotation angle to describe an arbitrary motion. If the orientations of an axis as determined from several frames forms a close cluster then the object is assumed to be undergoing simple motion. In the above situation first the motion for A is determined; for B and C there is no close cluster. After the motion of A is determined, the motion of B is obtained by transforming its motion to the coordinate system fixed to the object A. The motion of C can be determined similarly by transforming its motion to a coordinate system fixed to B. The experiments show that this approach is successful in the analysis of complex motion. Using projections on the Gaussian sphere, this approach may be extended for perspective cases [AYT82].

A method for the representation of the motion of objects in images, acquired using a moving camera, is proposed by Jain[Jai82, Jai83b, Jai83c]. This approach is based on the fact that all velocity vectors corresponding to the stationary objects in a scene acquired using a translating observer intersect at the FOE. Using the FOE as the origin, we may convert a frame to a second frame in which the abcissa is r and the ordinate is ϑ. Using this transformation, as discussed in an earlier section, it is possible to segment a dynamic scene into moving and stationary objects. What is more interesting, is that by using complex logarithmic mapping (CLM) about the FOE some interesting properties may be obtained in the EMP space. Let us define

$$z = x + iy$$

$$w = u + iv$$

where

$$w = log(z)$$

Thus

$$u = log(r) \qquad v = \vartheta = Tan^{-1}\frac{y}{x}$$

Using this transformation it can be shown that

$$\frac{du}{dZ} = -\frac{1}{Z}$$

and

$$\frac{dv}{dZ} = 0$$

The above result states that if the observer is moving then for a stationary point the horizontal displacement depends only on its depth and the vertical component is zero. This fact appears to be very useful in not only the segmentation of a dynamic scene into moving and stationary components, but also in the determination of the depth of points using the known motion of the observer. Schwartz [Sch80a, Sch80b, Sch81, Sch82] and Cavanaugh [Cav78, Cav81] have studied this mapping in the context of biological vision systems. Schwartz has found that the retino-striate mapping can be approximated using a complex log function. This mapping is responsible for size, rotation and projection invariance in these systems. Cavanaugh argues that the mapping is justified only in limited cases. Jain [Jai83c] showed that some of the limitations, with respect to the projection invariance, may be removed if the mapping is obtained with respect to the FOE. The complex EMP space allows an observer-centered representation of the sequence by considering the ego-motion of the observer. This representation may play an important role in MCMO dynamic scenes.

6.2. Motion Understanding

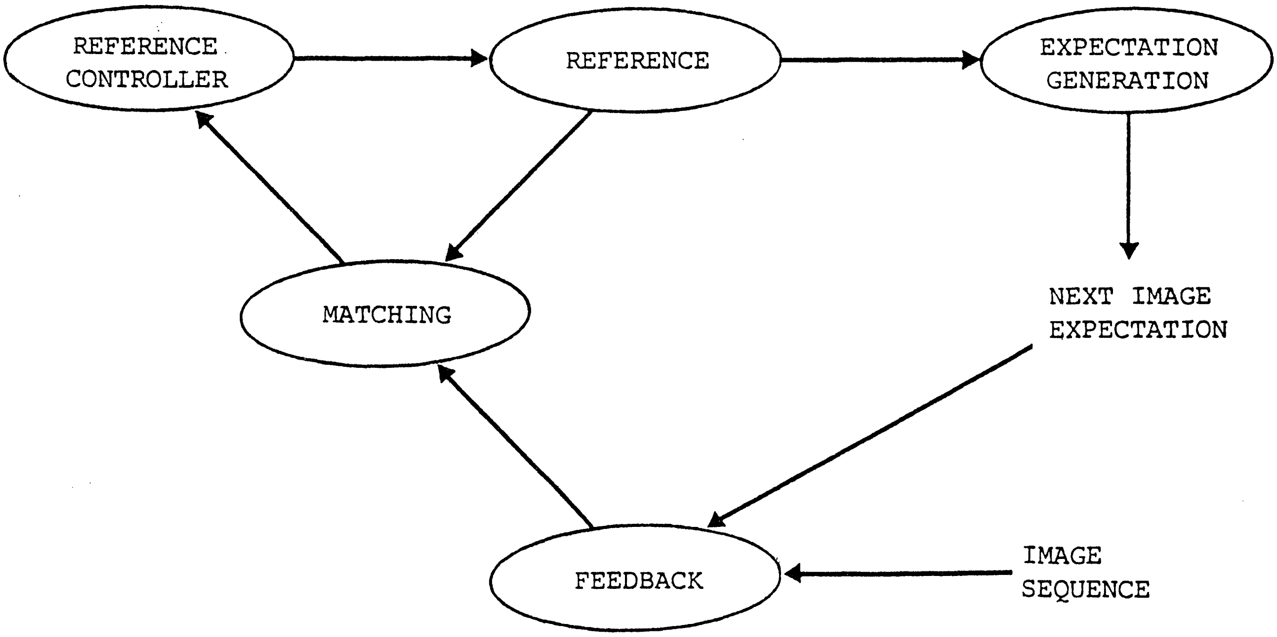
The motion exhibited over a sequence of frames may be described using motion verbs. To describe the motion of an object in a frame sequence using a motion verb requires abstraction and recognition of motion concepts. Badler[Bad75] and Tsotsos[Tso77, Tso80] have developed methods for the representation of motion verbs and for the recognition of the motion in terms of the predefined verbs.

Badler [Bad75] was the first attempt to describe the events taking place in a frame sequence using natural language. He outlined a methodology for describing the events using motion verbs, adverbs, and directionals. The motion concepts are defined using lower level location, orientation, and spatial relation changes. It is assumed that lower level processes will be able to give such information and then the system will work mainly in bottom-up mode for recognizing the motion events taking place in the sequence. He developed a methodology for the representation of the motion concepts and for the recognition of the events in terms of the motion concepts from the low level data. No effort was made to consider a real vision system and hence synthetic data was used as the input to the system.

Tsotsos [Tso80] suggests the use of feedback and knowledge for motion understanding. He developed a motion understanding system for the left ventricular wall motion. By using feedback at different levels it is expected to overcome the problems due to poor quality images and the inherent ambiguity in the recognition of motion concepts based on low level processes. His scheme for the knowledge based analysis of a frame sequence using feedback is shown in Figure 14. In this system, the objects in the first frame are identified and classified using static scene analysis techniques or manually. The analysis of motion of objects in the later frames is guided by expectations. Based on the previous motion of the objects, a new location in the next frame is predicted and verified. The changes detected using this approach are considered to be due to motion and are represented using low level vision constructs, such as axes, area, and arclength. Note that if a new object enters in the field of view in a later frame, the system will not be able to analyze its motion.

Figure 14

The knowledge based feedback model for the analysis of dynamic scenes as proposed by Tsotso in [Tso80].



The low level description of the motion is called the *essential trace* of the object. The next higher level of the representation is called the *essential kineses* and is obtained by combining pairs of adjacent essential traces. Four types of essential kineses are:

1. location changes,

2. length changes,

3. area changes,

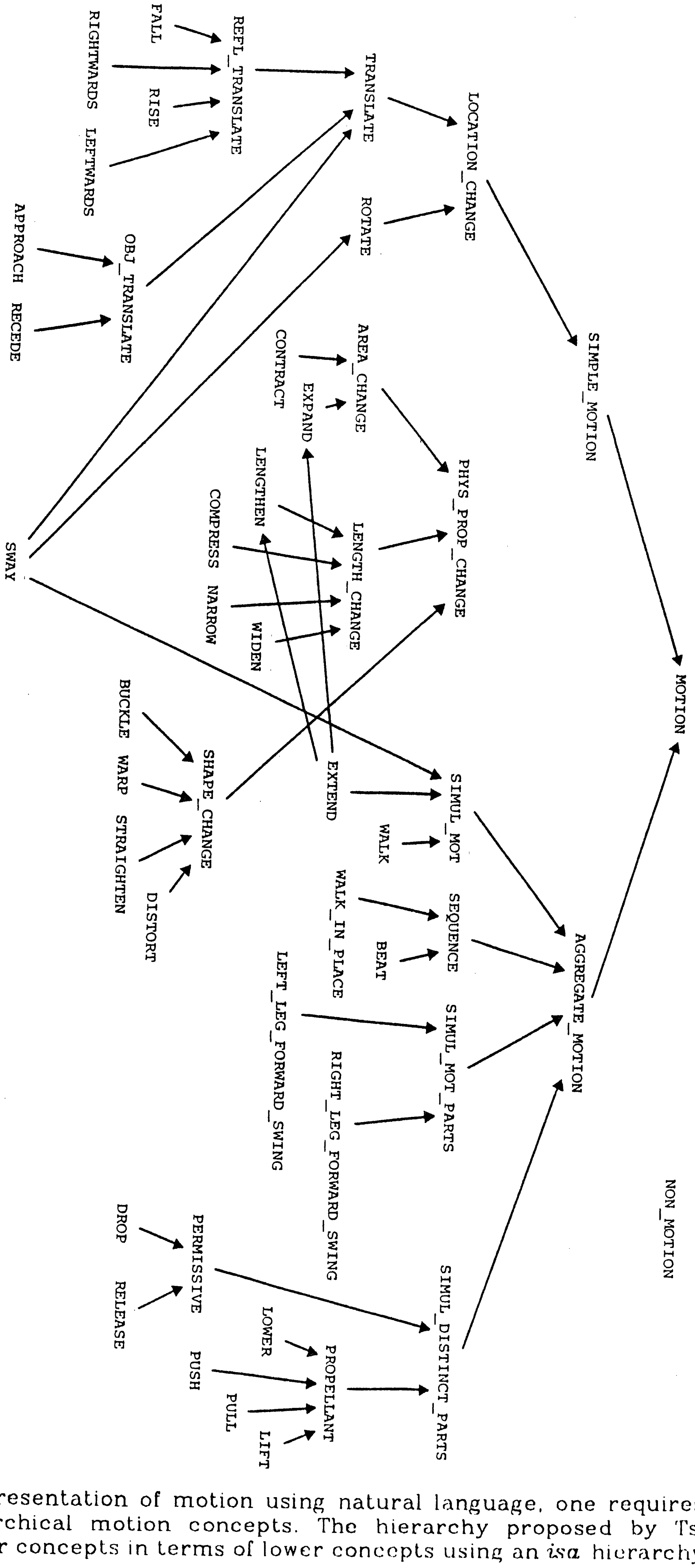


Figure 15

For the representation of motion using natural language, one requires methods to represent hierarchical motion concepts. The hierarchy proposed by Tsotso [Tso80] represents higher concepts in terms of lower concepts using an *isa* hierarchy.

4. shape changes.

These primitive kineses are used as a representation for relating quantitative changes to qualitative ones and are also used to match against the hypotheses which are active for a particular object's motion.

The objects and the concepts in this system are represented using frames. The motion concepts are organized hierarchically and are shown in Figure 15. This hierarchy is an extension of work done by Badler [Bad75]. The details of the representation of objects and other concepts used by Tsotsos are beyond the scope of this paper. It should be pointed out here that Tsotsos considered more aspects of a motion understanding system than considered by any other system. This system starts at the image level and uses *markers* to simplify the task of low level processing and analyzes the left ventricular wall motion and finally describes the motion in high level concepts related to the domain. A project for the natural language description of traffic scenes is described in [Neu82]. This project extends the methodologies developed by Tsotsos.

7. Discussion

The last few years have seen significant advances in the field of dynamic scene analysis. The early systems were strongly influenced by static scene analysis; most researchers now try to exploit the extra dimension offered by a frame sequence. Better techniques are available for change detection at pixel (or super pixel) level, a good understanding of methods for the extraction of interesting points, such as corners, has been achieved. Methods for the segmentation of dynamic scenes of all types perform reasonably well, though they are still not of the level required by higher level processes. Recovery of 3-D information has received significant attention leading to a good knowledge about the theoretical aspects, their performance in real scenes requires better location of features. Methods for the determination of the FOE for the recovery of information from optical flow show promise, at least in limited domains. The problem of how to represent motion is receiving increasing attention, both at the low-level and at the level of the description of motion concepts using motion verbs.

In the quest for solving problems in different phases of dynamic scene analysis, the techniques used are being influenced by different fields: such as differential geometry, natural language understanding, psychology, neuro-physiology, and photogrammetry. Many mathematical techniques are being applied to different aspects of the analysis leading to, generally, a better understanding of the problem and hence the development of better tools for the extraction of the information.

In many cases the assumptions made are too severe and result in techniques having applicability for only a very constrained domain. In some cases the problems arise due to the assumption that the input to the process, from a lower level process, will be perfect; and unfortunately the lower level process produces output, if at all, that is far from acceptable. A good example of this is the recovery of information from optical flow. Feature based recovery techniques are also very sensitive to the location of the tokens and fail to give acceptable results for the real images. By studying the sensitivity of the proposed approaches one could make stronger statements about their applicability.

We feel that the principle of least commitment and opportunistic procrastination [JaH82] have not been exploited to the desirable extent. Bnowing that the output of a process is going to be inherently imperfect, one should refrain from coming to faulty conclusions. Moreover, dynamic scenes allow the freedom to procrastinate, but not to sleep, until an appropriate time comes. As suggested in [JaH82] we may consider a dynamic scene analysis system in three phases and assume that the outputs of various processes may not be perfect. We should study the interaction of various processes and develop techniques that will be more tolerant to the noise coming from other processes. It appears that the paradigm of distributed problem solving [ErL75, MSC82] may help the integration of the imprecise information obtained from different sources.

Acknowledgment

The presentation and clarity of this paper was significantly enhanced by the comments and suggestions of Susan Haynes and Nancy O'Brien. I am thankful to them.

REFERENCES

[AgD75]
Aggarwal, J.K. and R.O. Duda, "Computer analysis of moving polygonal images," *IEEE Trans. on Computers*, vol. C-24, No.10, Oct. 1975, 966-976.

[AgJ82]
Agrawal, D.P. and Jain, R., "A multiprocessor system for dynamic scene analysis" *IEEE Trans. on Computer*, vol. C-31, pp.952-962, Oct. 1982.

[ALR75]
Arking,A.A., R.C. Lo, and A. Rosenfeld, "An evaluation of Fourier transform techniques for cloud motion estimation," TR-351, Dept. of Computer Science, University of Maryland, Jan. 1975.

[AYT81]
Asada, M., M. Yachida, and S. Tsuji, "Understanding of three-dimensional motions in block world", *Tech. Report no. 81-01*, Dept. Control Engin., Osaka University, 1981.

[AYT82]
Asada, M., M. Yachida, and S. Tsuji, "Representation of motions in time-varying imagery" *Proc. ICPR 82*, pp.306-309, 1982.

[BaB82]
Ballard, D.H. and C. M. Brown, *Computer Vision*, Prentice Hall, 1982.

[Bad75]
Badler, N.I., *Temporal Scene Analysis: Conceptual descriptions of object movements*, TR-80, Dept. of Computer Science, University of Toronto, 1975.

[BaT79]
Barnard, S.T. and W.B. Thompson, "Disparity analysis of images," *IEEE Trans. on PAMI*, vol. PAMI-2, 1980, pp. 333-340.

[Bor82]
Bornaee, Z., "Computation of optical flow", *M.S. Thesis*, Dept. of Computer Science, Wayne State University, 1982.

[CaR76]
Cafforio, C. and F. Rocca, "Methods for measuring small displacements of television images," *IEEE Trans. on Info. Theory*, vol. IT-22, pp. 573-579, Sept. 1976.

[Cav78]
Cavanagh, P., "Size and position invariance in the visual system," *Perception*, vol. 7, pp.167-177, 1978.

[Cav81]
Cavanagh, P., "Size invariance: reply to Schwartz," *Perception*, col. 10, pp.469-474, 1981.

[ChJ75]
Chien, R.T. and V.C. Jones, "Acquisition of moving objects and hand eye coordination," *Proc. 4th IJCAI, 1975, pp. 737-741.*

[ChW77]
Chien, Y.T. and M.O. Ward, "Representation and detection of position and shape change in time varying images," *Proc. IEEE Workshop on Picture Data Description and Management*, April 1977, pp.220-225.

[Clo80]
Clocksin, W.F., "Perception of surface slant and edge labels from optical flow: A computational approach," *Perception*, vol. 9, 1980, pp.253-269.

[DrN81]
Dreschler, L. and H.-H. Nagel, "Volumetric model and 3-D trajectory of a moving car derived from monocular TV-frame sequences of a street scene," *Proceedings of IJCAI*, 1981, pp.692-697.

[ErL75]
Erman, L and V. Lesser, "A multilevel organization for problem solving using many diverse cooperating sources of knowledge," *Proc. 4th IJCAI*, 1975, pp. 483-490.

[FeT79]
Fennema, C. L. and W. B. Thompson, "Velocity determination in scenes containing several moving objects," *Computer Graphics and Image Proc.*, vol 9, pp. 301-315, Apr. 1979.

[GrJ83]
Grosky, W.I. and R. Jain, "Region matching in pyramids for dynamic scene analysis", *in Multi-resolution image processing*, Ed. A. Rosenfeld, 1983.

[Gib66]
Gibson, J. J., *The senses considered as perceptual systems*, Boston: Houghton Mifflin, 1966.

[Gib79]
Gibson, J.J., *The ecological approach to visual perception*, Houghton Mifflen, Boston, 1979.

[GGF80]
Gilbert, A.L., M.K. Giles, G.M. Flachs, R.B. Rogers and Y. Hsun," "A real-time video tracking system using image processing," *IEEE Trans PAMI, vol. PAMI-2, 1980, pp.47-56.*

[Gla81]
Glazer, F., "Computing optical flow," *Proceedings IJCAI-81*, 1981, pp. 644-647.

[HiJ83]
Hill, Richard G. and R. Jain, "On determining optical flow", *Technical Report, Department of Computer Science, Wayne State University*, 1983.

[Hay82]
Haynes, Susan, "Detection of Moving Edges", *M.S. Thesis*, Dept. of Computer Science, Wayne State University, Detroit, 1982.

[HaJ82]
Haynes, S. and R. Jain, "Time varying edge detection", *Proc. ICPR*, Munich, pp. 754-756, 1982.

[HaJ83]
Haynes, S. and Jain, R., "Time varying edge detector," *Computer Graphics and IMage Processing*, (In Press) 1982.

[Hil82]
Hildreth, E.C., "The integration of motion information along contours," *Proc. of Computer Vision Workshop*, 1982, pp.83-91.

[HLS80]
Hirzinger, G., Landzettel and W.E. Snyder, "Automated TV tracking of moving objects," *Proc. IJCPR 1980, pp.1255-1261.*

[HNR82]
Hsu, Y.Z., H.-H. Nagel, and G. Rekers, "New likelihood test methods for change detection in image sequences", IFI-HH-M 104/82, University of Hamburg, Hambureg, West Germany, Nov. 1982.

[HoS81]
Horn, B. K. P. and B. G. Schunck, "Determining optical flow," *Proc. DARPA Image*

Understanding Workshop, pp. 144-156, Apr. 1981.

[Jai81a]
Jain, R., "Extraction of motion information from peripheral processes," *IEEE Trans on PAMI*, vol. PAMI-3, 1981, pp. 489-503.

[Jai81b]
Jain, R., "Dynamic scene analysis using pixel based processes," *IEEE Computer*, Aug 1981, pp.12-18.

[Jai82]
Jain, R., "Segmentation of moving observer frame sequences," *Pattern Recognition Letters*, vol. 1, pp. 115-120, 1982.

[Jai83a]
Jain, R., "Direct computation of the focus of expansion," *IEEE Trans. on PAMI*, vol. PAMI-5, pp.58-64, 1983.

[Jai83b]
Jain, R. "Segmentation of frame sequences obtained by a moving observer," *General Motors Research Publication*, Nov. 1982.

[Jai83c]
Jain, R., "Complex Logarithmic Mapping and the Focus of Expansion", *Proc. of Workshop on Motion: Representation and Control*, April 4-6, Toronto, 1983.

[Jai83d]
Jain, R. "On difference and accumulative difference pictures in dynamic scene analysis", *GMR publication*, 1983.

[JaJ82]
Jayaramamurthy, S. N. and R. Jain, "An approach for segmentation of textured dynamic scenes", *Proc. ICPR*, Munich, pp. 925-930, 1982.

[JaH82]
Jain, R. and S. Haynes, "Imprecision in computer vision," *IEEE Computer*, Aug 1982.

[JaJ83]
Jayaramamurthy, S.N. and Jain, R., "Segmentation of textured dynamic scenes" *Computer Graphics and Image Processing*, (in press) 1982.

[JaN79]
Jain, R., and H. H. Nagel, "On the analysis of accumulative difference pictures from image sequences of real world scenes," *IEEE Trans. PAMI*, vol. PAMI-1, pp. 206-214, Apr. 1979.

[JMA79]
Jain, R., W. N. Martin, and J. K Aggarwal, "Segmentation through the detection of changes due to motion," *Computer Graphics and Image Proc.*, vol. 11, pp. 13-34, Sept. 1979.

[JMN77]
Jain, R., D. Militzer, and H.-H. Nagel, "Separation of stationary from nonstationary components of a scene," *Proc. IJCAI*, 1977, pp.612-618.

[JeJ83a]
Jerian, C. P. and R. Jain, "Determining motion parameters for scenes with translation and rotation", *Proc. Workshop on Motion: Representation and Control*, April 4-6, Toronto, 1983.

[JeJ83b]
Jerian, C. P. and R. Jain, "On recovering 3-D information in dynamic scenes", *Technical report*, 1983.

[Joh76]
Johansson, G., "Spatio-temporal differentiation and integration in visual motion

perception", *Psych. Research*, vol. 38, pp.379-383, 1976.

[KoK80]
Korn, A. and R. Korris, "Motion analysis in natural scenes picked up by a moving sensor," *Proc. IJCPR, 1980, pp. 1251-1254.*

[Law81]
Lawton, D.T., "Optic flow field structure and processing image motion," *Proceedings IJCAI-81*, 1981, pp. 700-704.

[Law82]
Lawton, D. T., "Motion analysis via local translational processes," *Proc. Computer Vision Workshop, 1982, pp. 59-72.*

[Lee80]
Lee, D.N., "The optic flow field: The foundation of vision," *Phil. Trans. Royal Society of London*, vol. B290, 1980, pp. 169-179.

[LeG82]
Lenz, R and A. Gerhard, "Image sequence coding using scene analysis and spatio-temporal interpolation", *Proc. of NATO advanced studies institute on Image Sequence Processing* and Dynamic Scene Analysis, 1982.

[LeY82]
Legters, G.R. and T.Y.Young, "A mathematical model for computer image tracking", *IEEE Trans. on PAMI*, vol. PAMI-4, pp.583-594, 1982.

[LiM75]
Limb, J. O., and J. A. Murphy, "Estimating the velocity of moving images in television signals," *Computer Graphics and Image Proc.*, vol. 4, pp. 311-327, Dec. 1975.

[LNY82]
Levine, M.D., P. B. Noble, and Y.M. Youssef, "Understanding the dynamic behavior of a moving cell", *Proc. ICPR 82*, pp. 316-319, 1982.

[LoP80]
Longuet-Higgins, C. and K. Prazdny, "The interpretation of a moving retinal image", *Proc. Royal Society London*, vol. B-208, pp.385-397, 1980.

[MaA78]
Martin, W. N. and J. K. Aggarwal, "Dynamic Scene Analysis", *Computer Graphics and Image Processing*, vol. 7, pp. 356-374, 1978.

[MaA79]
Martin, W.N. and J.K. Aggarwal, "Computer analysis of dynamic scenes containing curvilinear objects," *Pattern Recognition*, vol. 11, 1979, pp.169-179.

[MaU79]
Marr, D. and S. Ullman, "Directional selectivity and its use in early visual processing," *MIT AI-Memo 524*, June 1979.

[Mor81]
Moravec, H.P., "Rover visual obstacle avoidance", *Proc. IJCAI 81*, pp. 785-790, 1981.

[MSC82]
McArthur, D., R. Steeb and S. Cammarata, "A framework for distributed problem solving," *Proc. AAAI*, 1982, pp.181-184.

[Nag78a]
Nagel, H.-H., "Formation of an image concept by analysis of systematic variations in the optically perceptible environment," *Computer Graphics and Image Processing*, vol 7, 1978, pp. 149-194.

[Nag78b]
Nagel, H.-H., "Analysis techniques for image sequences", *Proc. IJCPR*, pp. 186-211, 1978.

[Nag82a]
Nagel, H.-H., "On change detection and displacement vector estimation in image sequences," *Pattern Recognition Letters*, vol.1, pp.55-59, 1982.

[Nag82b]
Nagel, H.-H., "Displacement vectors derived from second order intensity variations in image sequences," Report IFI-HH-M-97/82, Fachbereich Informatik, Universitaet Hamburg, 1982.

[Nag82c]
Nagel, H.-H., "Overview on image sequence analysis", *Mitteilung No. 100*, Fachbereich Informatik, Universitaet Hamburg, Aug. 1982.

[NaE82]
Nagel, H.-H. and W. Enkelmann, "Investigation of second order greyvalue variations to estimate corner point displacements", *Proc. ICPR*, pp. 768-773, 1983.

[NaN81]
Nagel, H.-H. and B. Neumann, "On the derivation of 3-D rigid point configuration from image sequences," *Proc. IJCAI, 1981.*

[NaR82]
Nagel, H.-H. and G. Rekers, "Moving object masks based on an improved likelihood test", *Proc. ICPR*, pp. 1140-1142, 1982.

[Neu82]
Neumann, B., "Towards natural language description of real-world image sequences", *TR. No. IfI-HH-M-101/82* Fachbereich Informatik, Universitat Hamburg, Nov. 1982.

[Neu80]
Neumann, B., "Exploiting image formation knowledge for motion analysis," *IEEE Trans. PAMI*, PAMI-2, 1980, pp. 550-554.

[NeR79]
Netravali, A. N. and J. D. Robbins, "Motion compensated television coding: part 1," *Bell Sys. Tech. Journal*, vol. 58, pp. 631-670, March 1979.

[NeR80]
Netravali, A. N. and J. D. Robbins, "Motion-compensated coding: Some new results", *BSTJ*, vol. 59, pp. 1735-1745, 1980.

[NeS79]
Netravali, A. N. and J. A. Stuller, "Motion compensated transform coding", *PRIP 79*, pp. 561-567, 1979.

[OHO73]
Onoe, R. M., N. Hammano, and K. Ohba, "Computer analysis of traffic flow observed by subtractive tevevision", *Computer Graphics and Image Processing*, pp. 377-399, 1973.

[Pot74]
Potter, J.L., "Motion as a cue to segmentation," *Milwaukee symposium on Automatic Control*, 1974, pp.100-104.

[Pra79]
Prager, J.M., "Segmentation of static and dynamic scenes," *COINS TR-79-7*, University of Massachusetts at Amherst, May 1979.

[Pra80]
Prazdny, K., "Egomotion and relative depth map from optical flow," *Biological Cybernetics*, vol. 36, 1980, pp. 87-102.

[Pra81]
Prazdny, K., "A simple method for recovering relative depth map in the case of a translating sensor", *Proc. IJCAI 81*, pp. 698-699,1981.

[Pra82]
Prazdny, K., "Computing motions of planar surfaces from spatio-temporal changes in image brightness," *Proc. IEEE Conf. Pattern Recog. and Image Processing*, June 14-17, 1982, Las Vegas, Nevada, pp. 256-258.

[PrR77]
Price, K. and R. Reddy, "Change detection and analysis in multispectral analysis", *Proc. IJCAI*, Cambridge, pp.619-625, 1977.

[ReJ83]
Rheaume, D. P. and R. Jain, "A visual tracking system", *General Motors Research Publication*, Feb. 1983.

[RoA79]
Roach, J.W. and J.K. Aggarwal, "Computer tracking of objects moving in space," *IEEE Trans PAMI*, vol PAMI-1, No.2, April 1979, pp. 127-135.

[RoA80]
Roach, J.W. and J.K. Aggarwal, "Determining the movement of objects from a sequence of images," *IEEE Trans. on PAMI*, vol. PAMI-2, No.6, Nov. 1980, pp. 554-562.

[Sch79]
Schalkoff, R. J., "Algorithms for a real-time automatic video tracking system," PhD dissertation, Univ. of Virginia, Charlottesville, Va., May 1979.

[ScM81]
Schalkoff, R. J., and E. S. McVey, "A model and tracking algorithm for a class of video targets," *IEEE Trans. PAMI*, vol. PAMI-4, pp. 2-10.

[SCH82]
Schalkoff, R. J., "Dynamic imagery via distributed parameters systems: Characteristics, discontinuities, weak solution and shocks", *PRIP 82*, pp. 119-125, 1982.

[Sch80a]
Schwartz, E. L., "Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to coding," *Vision Research*, 20, 1980, pp.645-669.

[Sch80b]
Schwartz, E.L., "A quantitative model of the functional architecture of human striate cortex with application to visual illusion and cortical texture analysis", *Biological Cybernetics*, vol.37, pp.63-76, 1980.

[Sch81]
Schwartz, E. L., "Cortical anatomy, size invariance, and spatial frequency analysis," *Perception*, vol.10, pp.455-468, 1981.

[Sch82]
Schwartz, E.L., "Columnar architecture and computational anatomy in primate visual cortex: Segmentation and feature extraction via spatial frequency coded difference mapping," *Biological Cybernetics*, vol. 42, pp.157-168, 1982.

[ScH81]
Schunck, B.G. and B.K.P. Horn, "Constraints on Optical flow computations," *Proc. IEEE Conf. on PRIP*, 1981, pp.205-210.

[SeJ83]
Sethi, I. K. and R. Jain, "Determining three dimensional structure of rotating objects using a sequence of monocular views", *Technical Report, Dept. of Computer Science, Wayne State University, Detroit*, 1983.

[SNR80]
Stuller, J. A., A. N. Netravali, and J. D. Robbins, "Interframe television coding using gain and displacement compensation", *BSTJ*, vol. 59, pp.1227-1240, 1980.

[ThB81]
Thompson, W. B. and S. T. Barnard, "Lower-level estimation and interpretation of visual motion," *Computer*, vol. 14, pp. 20-28. Aug. 1981.

[Tho80]
Thompson, W. B., "Combining contrast and motion for segmentation", *IEEE Trans. on PAMI*, PAMI-2, 1980.

[TMB82]
Thompson, W.B., K.M. Mutch, and V. Berzins, "Edge detection in optical flow fields," *Proc. AAAI*, 1982, pp.

[Tod82]
Todd, J. T., "Visual information about rigid and nonrigid motion: A geometric analysis", *J. Experimental Psychology: Human Perception and Performance*, vol. *, pp.238-252, 1982.

[TsH81]
Tsai, R. Y. and T. S. Huang, "Estimaing three-dimensional motion parameters of a rigid planar patch", *Proc of PRIP*, pp. 94-97, 1981.

[TsH82]
Tsai, R. Y. and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters or rigid objects with curved surfaces," *Proc. IEEE Conf. Pattern Recog. and Image Processing*, 1982, pp. 112-118.

[THZ82]
Tsai, R. Y., T. S. Huang, and W.L. Zhu, "Estimating three-dimensional motion parameters of a rigid planar patch, II: Singular value decomposition", *IEEE Trans. Acoustics, Speech and SIgnal Processing*, vol. ASSP-30, pp.525-534.

[TSR82]
Tang, I, W. E. Snyder, and S. A. Rajala, "Extraction of moving objects in dynamic scenes", *Proc. ICPR*, Munich, pp. 1143-1146, 1982.

[Tso77]
Tsotsos, J.K., "Some notes on motion understanding," *Proc. of IJCAI*, 1977, p611.

[Tso80]
Tsotsos, J.K., *A framework for visual motion understanding*, TR114, Dept. of Computer Science, University of Toronto, 1980.

[TsY79]
Tsuji, S., M. Osada, and M. Yachida, "Tracking and segmentation of moving objects in dynamic line images," *IEEE Trans. on PAMI*, PAMI-2, 1980, pp.516-522.

[WaC80]
Ward, M.O. and Y.T. Chien, "Analysis of time-varying imagery through the representation of position and shape changes," *Proc. IJCPR*, 1980, pp.1236-1238.

[WeA82]
Webb, J.A. and J.K.Aggarwal, "Structure from motion of rigid and jointed objects", *Artificial Intelligence*, vol. 19, pp.107-130, 1982.

[Web81]
Webb, J., *Shape and structure from motion of objects*, Ph. D. Thesis, University of Texas, Dec. 1981.

[Wil80]
Williams, T.D., "Depth from motion in real world scenes," *IEEE Trans. PAMI*, PAMI-2, 1980,pp.511-516.

[Ull79]
Ullman, S., *The interpretation of visual motion* Cambridge, Mass, MIT Press, 1979

[Yac81]
Yachida, M., "Determining velocity map by 3-D iterative estimation," *Proceedings IJCAI-81* 1081, pp.716-718.

[YAT78]
Yachida, M., Asada, M., and S. Tsuji, "Automatic motion analysis system of moving objects from the records of natural processes", *Proc. IJCPR*, Kyoto, pp.726-730, 1978.

[YMA80]
Yalamanchili, S., W.N. Martin, and J. K. Aggarwal, "Differencing operations for the segmentation of moving objects in dynamic scenes", *Proc. of 5 ICPR*, pp. 1239-1242, 1980.

[ZuH81]
Zucker, S.W. and R.A. Hummel, "A three-dimensional edge detector," *IEEE Trans. PAMI*, PAMI-3, 1981, pp. 324-330.