On the Computation of Motion from Sequences of Images—A Review

J. K. AGGARWAL, FELLOW, IEEE, AND N. NANDHAKUMAR, MEMBER, IEEE

Invited Paper

The present paper reviews recent developments in the computation of motion and structure of objects in a scene from a sequence of images. We highlight two distinct paradigms: i) the feature-based approach and ii) optical flow based approach. The comparative merits/demerits of these approaches are discussed. The current status of research in these areas is reviewed and future research directions are indicated.

I. Introduction

The ability to discern objects, ascertain their motion, and navigate in three-dimensional space through the use of vision is almost universal among animals. Incorporating such vision in machines is ostensibly a straightforward task given the widespread availability of microcomputers, digitizing cards, and solid-state cameras. Although it is fairly easy and inexpensive to assemble a computer vision system, it has proved surprisingly difficult to achieve a vision capability in machines, even to a limited degree. This is not to imply that we are not using all sorts of vision systems and motion detectors in a variety of applications. However, the ease with which humans detect motion and navigate around objects, and the difficulty of duplicating these capabilities in machines have recently led to major efforts by computer engineers and scientists to understand vision in man and machine. These efforts are in addition to and perhaps complement current and earlier endeavors at understanding human vision and motion by psychologists and physiologists.

Broadly speaking, there are two groups of scientists studying vision. One group is studying human/animal vision with the goal of understanding the operation of biological vision systems including their limitations and diversity. The scientists in this group include neurophysiologists, psychophysicists, and physicians. The second group of scientists includes computer scientists and engineers conducting research in computer vision with the objective of developing vision systems. Vision systems with the ability to navigate, recognize, and track objects and estimate their speed and direction are the ultimate goals of the latter research. The knowledge and results of research in neurophysiology and psychophysics have influenced the design of vision systems by engineers and scientists. At the same time, results in computer vision have provided a framework for modeling biological vision. Such cross-fertilization of ideas will continue to yield better models for biological and machine vision systems.

There is a long list of applications motivating a strong interest in sensing, interpretation, and description of motion from a sequence or a collection of images. The automatic tracking and possible ticketing of speeding vehicles on a highway is of interest to traffic engineers and law enforcement officers. The automatic recognition, tracking, and possible destruction of targets is of immense interest to the department of defense of every country. The computation, characterization, and understanding of human motion in dancing, athletics, and pilot training are important to several diverse disciplines. The analysis of scintigraphic image sequences of the human heart is of interest in assessing motility of the heart in diagnosis and supervision of patients after heart surgery. Satellite imagery provides an opportunity for interpretation and prediction of atmospheric processes through the estimation of shape and motion parameters of atmospheric disturbances for the meteorologist. The bandwidth reduction achievable through the estimation of motion allows for compression of image sequences for efficient transmission. The above examples are indicative of the diversity of applications where the computation of motion from a sequence of images is of critical importance.

This broad interest in the interpretation of motion from a sequence of images has been evident since the first workshop on motion in Philadelphia in 1979 [1]. Since that workshop, several additional meetings and special issues of various journals have contributed to the exchange of ideas and the dissemination of results. In addition, there have been several sessions on motion and related issues at meetings such as the IEEE Computer Society Computer Vision and Pattern Recognition Conference and conferences of other

Manuscript received December 18, 1987; revised March 16, 1988. The research was supported in part by a grant from the Army Research Office under Contract DAAL03-87-K-0089, and a grant from the National Science Foundation under Contract NSF/DCR-8517583.

The authors are with the Computer and Vision Research Center College of Engineering, The University of Texas at Austin, Austin, TX 78712, USA.

IEEE Log Number 8822791.

0018-9219/88/0800-0917$01.00 © 1988 IEEE

societies interested in vision. The list of workshops and special issues devoted exclusively to motion and time-varying imagery include three special issues [2]–[4], two books [5], [6], a NATO Advanced Study Institute [7], an ACM workshop [8], a European meeting on time-varying imagery [9], and a host of survery papers [10]–[15]. The extent of the breadth and depth of interest is provided by the table of contents of the book published to document the proceedings of the NATO-ASI [16]. However, this list is incomplete at best. The IEEE Computer Society workshop at Kiawah Island [17] and Second International Conference in Italy [18] are indications of the broad interest in motion at this time. The recent two-volume collection of papers in the reprint series [19] published by IEEE Computer Society includes a section on Image Sequence Analysis containing nine papers. The recent book edited by Martin and Aggarwal entitled *Motion Understanding: Robot and Human Vision* [20] gives eleven papers detailing recent developments in this area.

The above brief chronology documents the contributions from a computer vision perspective. It is not the intention of the present review to slight the earlier pioneering works of psychologists and other scientists. In particular, the kinetic depth effect demonstrated by Wallach and O'Connell [21] through the use of wire frame objects, and similar effects shown by Gibson [22] in his translucent sheet experiments, Ullman [23] in his rotating cylinders experiment, and Joahannson [24]–[28] are important contributions in the area of psychophysics of motion perception. In the same vein, the contributions of Hubel and Wiesel [29] in demonstrating the existence of specialized cortical cells tuned to the detection of motion are seminal contributions in neurophysiology. The present review, however, is only aimed at the computer vision inspired contributions to the study of motion. A more balanced review of the recent contributions in both psychophysics of vision and machine vision is found in [20].

In this paper we do not present an exhaustive compendium of recent research in the computation of motion and structure from sequences of images; instead we list some of the important work done and provide a flavor of the approaches that have been developed.

II. Methodologies for Motion Estimation

The relative motion between objects in a scene and a camera, gives rise to the apparent motion of objects in a sequence of images. This motion may be characterized by observing the apparent motion of a discrete set of features or brightness patterns in the images. The objective of the analysis of a sequence of images is the derivation of the motion of the objects in the scene through the analysis of the motion of features or brightness patterns associated with objects in the sequence of images.

Two distinct approaches have been developed for the computation of motion from image sequences. The first of these is based on extracting a set of relatively sparse, but highly discriminatory, two-dimensional features in the images corresponding to three-dimensional object features in the scence, such as corners, occluding boundaries of surfaces, and boundaries demarcating changes in surface reflectivity. Such points, lines and/or curves are extracted from each image. Inter-frame correspondence is then established between these features. Constraints are formulated based on assumptions such as rigid body motion, i.e., the 3-D distance between two features on a rigid body remains the same after object/camera motion. Such constraints usually result in a system of nonlinear equations. The observed displacement of the 2-D image features are used to solve these equations leading ultimately to the computation of motion parameters of objects in the scene.

The other approach is based on computing the optic flow or the two-dimensional field of instantaneous velocities of brightness values (gray levels) in the image plane. Instead of considering temporal changes in image brightness values in computing the optic flow field, it is possible to also consider temporal changes in values that are the result of applying various local operators such as contrast, entropy, and spatial derivatives to the image brightness values. In either case, a relatively dense flow field is estimated, usually at every pixel in the image. The optic flow is then used in conjuction with added constraints or information regarding the scene to compute the actual three-dimensional relative velocities between scene objects and camera.

A task that is closely related to the estimation of motion is the task of estimation of the structure of the imaged scene. In the case of the optic flow method, this consists of grouping pixels corresponding to distinct objects into separate regions, i.e., segmenting the optic flow map, and then computing the three-dimensional coordinates of surface points in the scence corresponding to each pixel in the image at which the flow is computed. In the case of the feature-based analysis, computing structure corresponds to forming groups of image features for each object in the scene and then computing the 3-D coordinates of each object feature associated with each image feature.

Although structure may be computed independent of motion, e.g., via stereopsis, the former process can benefit by the estimated motion. Knowledge of motion parameters for features/regions can aid segmentation of image features/regions corresponding to distinct objects. In stereopsis, knowledge of object motion can facilitate establishment of feature correspondence within a pair of stereo images, thus aiding the determination of structure. Image regions with different apparent 2-D motions can be considered to correspond to distinct objects. Psychological research has collected enough evidence to support the belief that the process of establishing correspondence and the process of estimating structure and motion are closely interwoven in the human visual mechanism. Indeed, Ullman has shown that apparent motion is a clue used by the human visual system for computing scene structure [6]. This close relationship between the estimation of structure and the estimation of motion has prompted many researchers to address both tasks as a combined problem. In this paper we discuss the combined task of computing structure and motion from image sequences.

In the following sections we discuss in greater detail the fundamental principles underlying the two distinct methodologies for computing 3-D motion from apparent motion. The basic mathematical formulations are introduced and discussed. In Section III we discuss the feature-based method for estimation of motion from a sequence of monocular images. In Section IV we discuss the optic flow method for sequences of monocular images. Section V discusses the relative merits and demerits of these two approaches. The two approaches outlined above allow for the estimation of motion without requiring that scene structure be known *a priori*. The use of stereopsis allows for the estimation of depth, i.e., the distance from the sensor to the objects. The additional information available greatly reduces the complexity of motion estimation. The variety of ways in which stereopsis can be used to facilitate the computation of motion is outlined in Section VI. Finally, Section VII concludes this paper with a few closing remarks.

III. Feature-Based Motion Estimation from Monocular Image Sequences

In this section, we discuss the feature-based approach to estimate motion from a sequence of images gathered by a single camera. A mathematical formulation is presented and variations of this formulation are discussed. The discussion focuses on the estimation of both motion and structure. No distinction is made between the situations where a) the camera is moving and imaged scene is stationary, b) camera is stationary while the imaged objects are in motion, or c) both camera and imaged objects are in motion. What is computed is the relative position and motion between the camera and the imaged scene. In the following discussion it is assumed that image features, such as points and lines, have been extracted from each image and inter-frame correspondence has already been established between the features.

We present below three approaches to feature-based analysis of monocular image sequences. The first of these is the direct formulation in which rigid body motion is assumed. In this formulation the rigidity constraint is manifest in there being single rotation and translation matrices for all observables. In the second approach rigidity is explicitly invoked with the formulation being based on preserving rigidity, e.g., preserving the angle between two intersecting 3-D lines lying on a rigid object. These two schemes use two or three views to estimate structure and motion. A third approach consists of using a long sequence of monocular images. A brief description of the salient features of each approach is presented.

A. Direct Formulations

An orthographic imaging model was used by Ullman [6], [23] to estimate the structure and motion of an object undergoing rigid motion. The position and motion of four noncoplanar points in space were recovered from three distinct orthographic projections of these points. The formulation is as follows. Let O, A, B, and C be the four points. The orthographic projection of these points in three distinct planes $\Pi 1, \Pi 2, \Pi 3$ are given and the 3-D configuration of these points is to be determined. A fixed coordinate system with origin at O is chosen. Let $\mathbf{a}, \mathbf{b}, \mathbf{c}$, be the vectors from O to A, B and C, respectively. Let each image have a coordinate system with its origin at the projection of O, and its axes along the directions $\mathbf{p}_i, \mathbf{q}_i$. Note that $\mathbf{p}_i$ and $\mathbf{q}_i$ are orthogonal unit vectors on Πi. Let the image coordinates of (A, B, C) on Πi be $(x_{ai} y_{ai}, x_{bi} y_{bi}, x_{ci} y_{ci})$, and let $\mathbf{u}_{ij}$ be the unit vector along the intersection of Πi and Πj.

The image coordinates are given by the dot products

$$x_{ai} = \mathbf{a}\cdot\mathbf{p}_i, \quad y_{ai} = \mathbf{a}\cdot\mathbf{q}_i, \quad x_{bi} = \mathbf{b}\cdot\mathbf{p}_i,$$

$$y_{bi} = \mathbf{b}\cdot\mathbf{q}_i, \quad x_{ci} = \mathbf{c}\cdot\mathbf{p}_i, \quad y_{ci} = \mathbf{c}\cdot\mathbf{q}_i.$$

The unit vector $\mathbf{u}_{ij}$ lies on Πi which is spanned by $(\mathbf{p}_i, \mathbf{q}_i)$, hence

$$\mathbf{u}_{ij} = \alpha_{ij}\mathbf{p}_i + \beta_{ij}\mathbf{q}_i, \quad \text{where} \quad \alpha_{ij}^2 + \beta_{ij}^2 = 1.$$

The unit vector $\mathbf{u}_{ij}$ also lies on Πj which is spanned by $(\mathbf{p}_j, \mathbf{q}_j)$, hence

$$\mathbf{u}_{ij} = \gamma_{ij} p_j + \delta_{ij}\mathbf{q}_j, \text{ where } \gamma_{ij}^2 + \delta_{ij}^2 = 1.$$

From the latter two equations we obtain

$$\alpha_{ij}\mathbf{p}_i + \beta_{ij}\mathbf{q}_i = \gamma_{ij}\mathbf{p}_j + \delta_{ij}\mathbf{q}_j$$

and taking the scalar product of this equation with $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ we get:

$$\alpha_{ij} x_{ai} + \beta_{ij} y_{ai} = \gamma_{ij} x_{aj} + \delta_{ij} y_{aj}$$

$$\alpha_{ij} x_{bi} + \beta_{ij} y_{bi} = \gamma_{ij} x_{bj} + \delta_{ij} y_{bj}$$

$$\alpha_{ij} x_{ci} + \beta_{ij} y_{ci} = \gamma_{ij} x_{cj} + \delta_{ij} y_{cj}.$$

These equations are linearly independent [6] and possess two solutions that are equal in magnitude but have opposite sign. Choosing one of these solutions, the vectors $\mathbf{u}_{ij}$ are determined. The distances $d1 = \|\mathbf{u}_{12} - \mathbf{u}_{13}\|$, $d2 = \|\mathbf{u}_{12} - \mathbf{u}_{23}\|$, and $d3 = \|\mathbf{u}_{13} - \mathbf{u}_{23}\|$ are then computed. When no two vectors $\mathbf{u}_{ij}$ are equal, then $di \neq 0$ and a unique triangle with sides $d1$, $d2$, and $d3$ is specified. Consider the tetrahedron formed by this triangle and the origin O, with the vertices of the triangle being placed at unit distance from the origin O. From the projections of A, B, and C on the three planes (images) a unique 3-D configuration is easily computed. In the degenerate case, i.e., when two of the $\mathbf{u}_{ij}$ are identical, straightforward trigonometric considerations provide recovery of the structure and motion of the body [23].

Although the parallel projection model is adequate in some situations it is not appropriate for most real-world applications which mandate the use of perspective projection. The use of perspective transformation substantially increases the complexity of the problem. Roach and Aggarwal [30], [31] were among the first to compute structure and motion from images via the perspective imaging transformation. A scenario consisting of a static scene and a moving camera was assumed. The goal was to investigate whether it would be possible to determine the position of the points in space and the movement (translation and rotation) of the camera.

The equations that relate the three-dimensional coordinates of a point (X, Y, Z) and its image plane coordinates (x, y) are

$$x = F\frac{a_{11}(X - X_0) + a_{12}(Y - Y_0) + a_{13}(Z - Z_0)}{a_{31}(X - X_0) + a_{32}(Y - Y_0) + a_{33}(Z - Z_0)}$$

$$y = F\frac{a_{21}(X - X_0) + a_{22}(Y - Y_0) + a_{23}(Z - Z_0)}{a_{31}(X - X_0) + a_{32}(Y - Y_0) + a_{33}(Z - Z_0)}.$$

Here F is the focal length, (X_0, Y_0, Z_0) is the projection center and $a_{11}, a_{12}, \cdots, a_{33}$ are functions of (Θ, Φ, Ψ), the orientation of the camera with respect to the global reference system.

Roach and Aggarwal showed that five points in two views are needed to recover these parameters [30], [31]. They related the number of points and the number of equations available for the solution of 3-D coordinates and motion parameters as follows: The global coordinates of each point are unknown so the five points produce 15 variables. The camera position and orientation parameters (X_0, Y_0, Z_0, Θ, Φ, and Ψ) in two views contribute another 12 variables yielding a total of 27 variables. Each 3-D point produces two projection equations per camera position thus forming a total of 20 nonlinear equations. To make the number of equations equal the number of unknowns, seven variables must be known or specified *a priori*. This is achieved by choosing the six camera parameters of the first view to be zero and setting the Z-component of one of the five points to an arbitrary positive constant to fix the scaling factor. The reason for fixing one variable as the scaling constant is that under the given camera/object constraints the information embedded in every image sequence is inherently insufficient for determining the correct scale. For example, the observed projected motion of an object moving in space can be reproduced by another object which is twice as large, twice far away from the camera, translating twice as fast, and rotating with the same speed around an axis of the same orientation as the former object. In general, the information of the absolute distance of the object from the viewer is usually lost in the image formation process. Therefore, arbitrarily setting the scale is not unreasonable in finding the solution for the structure and motion parameters.

An iterative finite difference Levenberg-Marquardt algorithm was used to solve these 18 nonlinear equations (after fixing the scale factor two of the 20 nonlinear equations have no unknown variables in them). For noise-free simulations, the methods typically converged to the correct answer within 15 seconds on a Cyber 170/50 and hence are reasonably efficient. If noise is introduced into the point positions in the image plane, a considerably overdetermined system of equations is needed to attain good accuracy of the results. Two views of 12 or even 15 points, or three views of seven or eight points are usually needed in the noisy cases.

Unlike Roach and Aggarwal [30], [31] who solved the motion parameters through a single system of equations thus creating a large search space, Nagel [32] proposed a technique which reduces the dimension of the search space through the elimination of unknown variables. The important observation made by Nagel was that the translation vector can be eliminated and the rotation matrix can be solved separately. A rotation matrix is completely specified by three parameters—namely the orientation of the rotation axis and the rotation angle around this axis. It is shown that if measurements of five points in two views are available, then three equations can be written and the three rotation parameters can be solved for separately from the translation parameters. The distance of the configuration of points from the viewer is arbitrarily fixed and the translation vector can then be determined.

Tsai and Huang [33]–[35] proposed a method to find the motion of a planar surface patch from 2-D perspective views. The algorithms consists of two steps: First, a set of eight "pure parameters" is defined. These parameters can be determined uniquely from two successive image frames by solving a set of linear equations. Then, the actual motion parameters are determined from these eight "pure parameters" by solving a sixth-order polynomial.

By exploiting the constraints of projective geometry and rigid motion, equations can be written to relate the coordinates of image points in the two frames for points on a planar surface patch $AX + BY + CZ = 1$, where A, B, and C are the structure parameters. The mapping from the (x, y) space to the (x', y') space (from one image to the next image) is given by

$$x' = \frac{a_1x + a_2y + a_3}{a_7x + a_8y + 1} \qquad y' = \frac{a_4x + a_5y + a_6}{a_7x + a_8y + 1}$$

where, a_1 through a_8 are the eight "pure parameters" and can be expressed in terms of the focal length, the structure parameters (A, B, C), and the motion parameters N_X, N_Y, N_Z, Θ, T_X, T_Y and T_Z(N specifies the rotation axis, Θ is the rotational angle, and T is the translational vector). For a particular set of pure parameters, the above equation represents a mapping from (x, y) space to (x', y') space. A set of linear equations is solved for these eight pure parameters.

After the eight pure parameters are obtained, the structure and motion parameters can be determined. Here, the Z component of the translation vector is arbitrarily chosen to fix the scale. After a series of manipulations, it is possible to get a sixth-order polynomial equation in terms of only one of the variables $T'_X = T_X/T_Z$. T'_X is solved first and then all the remaining structure and motion parameters can be easily obtained. Although potentially six real roots may result from solving a sixth-order polynomial, the authors reported that aside from a scale factor for the translation parameters, the number of real solutions never exceeded two in their simulation.

Later, Tsai and Huang [36] investigated the problem of a curved surface patch in motion. Two main results were established concerning the existence and uniqueness of the solutions. An $\boldsymbol{E}$ matrix was specified as $\boldsymbol{E} = \boldsymbol{TR}$, where $\boldsymbol{T}$ is the translation and $\boldsymbol{R}$ is the rotation. Given the image correspondences of eight object points in general positions, the $\boldsymbol{E}$ matrix can be determined uniquely by solving eight linear equations. Furthermore, the actual 3-D motion parameters can be determined uniquely given $\boldsymbol{E}$, and can be computed by taking the singular value decomposition of $\boldsymbol{E}$ without having to solve nonlinear equations. Detailed proofs of these claims are presented by the authors [36]. Although the approach results in the solution of a set of linear equations, the system is highly sensitive to noise and especially to perturbations of image coordinates. Longuet-Higgins [37], [38] worked independently to obtain results similar to those described above. He derived the $\boldsymbol{E}$ matrix and presented a method to recover $\boldsymbol{R}$ and $\boldsymbol{T}$ from $\boldsymbol{E}$ using tensor and vector analysis.

Extensions of the above approaches were proposed by several researchers [39]–[43]. One limitation of the approaches developed by Tsai and Huang [36] and Longuet-Higgins [37] is the requirement of *a priori* knowledge regarding nonzero translation. Zhuang and Haralick [39]–[41] have developed an algorithm which overcomes this limitation. Zhuang and Haralick do require that the observed object points do not lie on a specific quadratic surface passing through the origin. Faugeras, Lustman and Toscani [42] and Nagel [43] reformulated the problem in more robust manners as least-mean-squared error minimization problems.

The above approaches used 3-D points and their projections on the image planes as observables in formulating the problem. An alternative approach is to use 3-D lines and their projections as observables. When lines are used as features, two views are no longer sufficient and a minimum of three views are required. This is due to the fact that 3-D lines possess an additional degree of freedom when compared to 3-D points. In other words, one can slide a 3-D line along itself and obtain the same line. We present below an overview of some techniques that use lines as features in the estimation of structure and motion.

Yen and Huang [44], [45] have proposed an iterative method based on spherical projection and on the observation of seven line correspondences in three views for the case of general motion between views. Liu and Huang [46], [47] have used line correspondences in formulations analogous to the methods outlined above. They decompose rigid body motion into first a rotation around an axis through the origin and then a translation. For the case of pure rotation, two line correspondences over two frames are sufficient to determine the rotation matrix. The resulting nonlinear equations are solved iteratively. For the case of pure translation, five line correspondences over three frames produce a system of linear equations which can be solved to determine the translation. For the general case, Liu and Huang use six line correspondences in three frames. The rotation matrix is first determined and then the translation matrix is computed. Simulations of the iterative algorithm on synthesized data show that the approach is highly sensitive to noise and initial estimates. Moreover, estimation of the translation vector is very sensitive to errors in estimation of rotation. The algorithm has not been tested on real data.

A more robust formulation of motion estimation using line correspondences, which incorporates the effect of noise, is due to Faugeras, Lustman and Torscani [42]. An extended Kalman filtering approach is followed in solving the nonlinear equations for a "best" estimate of the motion parameters. The "best" estimate is defined to be one that minimizes an expression that involves the measurables, the unknowns, and partial derivatives of the nonlinear equation that relates the unknowns to the measurables. The measurables for each 3-D line consist of three vectors, one for each of the three image planes. Each vector corresponds to the unit normal of the plane containing the projection of the 3-D line and the center of projection for that image plane. The unknowns consist of the rotation parameters that relate the positions of the three image planes. After solving for the rotation, the translation is computed via linear equations. The structure of the object can then be computed via either a least-squares technique or via the Kalman filtering approach. Significant improvement was reported in sensitivity to noise and initial estimates.

Implicit in the above discussion was the assumption that the scene contained a single rigid object. Feature-based motion analysis has also been applied to scenes containing multiple rigid and jointed objects. Webb and Aggarwal [48] have presented a method for recovering the 3-D structure of such scenes under orthographic projection. The fixed-axis assumption is adopted to interpret images of moving objects. The fixed-axis assumption asserts that every rigid object movement consists of a translation plus a rotation about an axis which is fixed in direction for a short period of time. It is shown that, under the fixed-axis assumption, selecting any point on a rigid moving object as the origin of a coordinate system causes the other points to trace out circles in planes normal to the fixed-axis within that coordinate system. Under parallel projection, with the selected point projecting to the image origin, these circles project into ellipses. The structure of the rigid object can be recovered to within a reflection by finding the equations describing the ellipses. Furthermore, it is shown that the lengths of the long and short axes of an ellipse are functions of the position of the point in space. The position of each point in space (up to a reflection about the image plane) can then be recovered provided that the fixed axis of rotation is not parallel or perpendicular to the image plane.

A jointed object is an object made up of a number of rigid parts which cannot bend or twist. If the jointed object still moves in a way such that the fixed-axis assumption holds for each rigid part, then the motion and structure of the jointed object can be recovered. It is assumed that the rigid parts are connected by joints identified since they satisfy two sets of motion constraints. If the joints are not visible, they can be found by solving a system of linear equations. The joints can then be used to eliminate some reflections and thus the number of possible interpretations of structure is reduced. Finally, the 3-D motion of each object is reconstructed.

B. Explicit Use of Rigidity

The assumption of rigid body was implicitly used in the above formulations. We outline below a typical formulation in which the constraint of rigid body motion is explicitly invoked [49]. We discuss the case where five points in two views are used as the observables. As in the above discussion, the relative positions of the cameras are unknown, and the correspondence between points in the two views is assumed known.

The two central projection imaging systems are shown in Fig. 1. C_1 and C_2 are the centers of projection and I_1 and I_2 are the image planes. A point P_i in space with coordinates (X_i, Y_i, Z_i) in S_1 and (U_i, V_i, W_i) in S_2 is imaged as p_i on I_1 and q_i on I_2. The objective of the analysis is to derive the structure of the points and the transformation between the coordinate systems, given the image coordinates of the observed points in the two imaging coordinate systems.

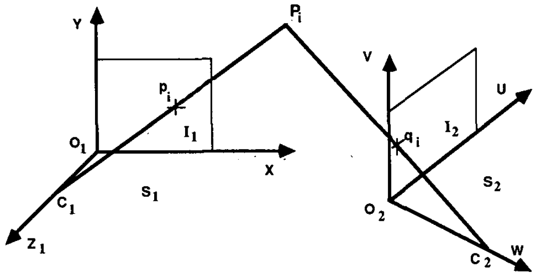

Fig. 1. Imaging geometry for the two views. P_i is the 3-D point, p_i and q_i are the images of P on the two image planes.

Because P_i is on line C_1p_i (refer to Fig. 1), there exists a real number $\lambda_i > 1$ such that

$$X_i = \lambda_i x_i, \qquad Y_i = \lambda_i y_i, \qquad Z_i = (1 - \lambda_i)f_1$$

where (x_i, y_i) are the coordinates of p_i in the I_1-image coordinate system, and f_1 is the distance from C_1 to the image plane. Similarly, P_i is on line C_2q_i and if (u_i, v_i) are the coordinates of q_i in the I_2 coordinate system then there exists $\gamma_i > 1$ such that

$$U_i = \gamma_i u_i,\ V_i = \gamma_i v_i,\ W_i = (1 - \gamma_i) f_2.$$

The squared distance between points P_i and P_j expressed in S_1 is therefore

$$d_{ij}^2(S_1) = (X_i - X_j)^2 + (Y_i - Y_j)^2 + (Z_i - Z_j)^2$$

or

$$d_{ij}^2(S_1) = (\lambda_i x_i - \lambda_j x_j)^2 + (\lambda_i y_i - \lambda_j y_j)^2 + (\lambda_i - \lambda_j)^2 f_1^2.$$

Similarly, the squared distance between P_i and P_j expressed in S_2 is

$$d_{ij}^2(S_2) = (\gamma_i u_i - \gamma_j u_j)^2 + (\gamma_i v_i - \gamma_j v_j)^2 + (\gamma_i - \gamma_j)^2 f_2^2.$$

Now, the principle of conservation of distance allows us to write (assuming, of course, identical units of measurement in S_1 and S_2)

$$d_{ij}^2(S_1) = d_{ij}^2(S_2)$$

or

$$(\lambda_i x_i - \lambda_j x_j)^2 + (\lambda_i y_i - \lambda_j y_j)^2 + (\lambda_i - \lambda_j)^2 f_1^2$$
$$= (\gamma_i u_i - \gamma_j u_j)^2 + (\gamma_i v_i - \gamma_j v_j)^2 + (\gamma_i - \gamma_j)^2 f_2^2. \quad (3.1)$$

It may be seen that each point P_i contributes two unknowns, λ_i and γ_i, and each pair of points (P_i, P_j) gives one second order equation (3.1). Therefore, 5 points yield 10 equations and 10 unknowns. Again, fixing the scale arbitrarily, we end up with a system of 10 equations in 9 unknowns. Note that each equation involves only 4 of the unknowns. Since distances between points define structure only up to a reflection in space, the solution of system (3.1) based on these distances is also subject to this uncertainty. System (3.1), although simple, is nevertheless nonlinear. Experimental results using existing iterative numerical methods do indicate, however, that the solution is well behaved [49].

When the position of the points has been computed, determining the relative position of the cameras becomes a simple matter. Indeed, take 4 noncoplanar points (from the 5 observed points in space) and call A_1 and A_2 the matrices of homogeneous coordinates of these in S_1 and S_2, respectively. Then if M is the transformation matrix (in homogeneous coordinate form) that takes S_1 onto S_2 we have

$$A_2 = A_1 M. \quad (3.2)$$

Since the 4 points are not coplanar, (3.2) can be solved for M. Now if we decompose motion M into i) a rotation through angle Θ about an axis through the origin with direction cosines n_1, n_2, n_3, followed by ii) a translation (t_1, t_2, t_3) and if it is written as

$$M = \begin{bmatrix} a_1 & a_2 & a_3 & 0 \\ a_4 & a_5 & a_6 & 0 \\ a_7 & a_8 & a_9 & 0 \\ t_1 & t_2 & t_3 & 1 \end{bmatrix}$$

then one can show that

$$\cos\Theta = (a_1 + a_5 + a_9 - 1)/2; \qquad \sin\Theta = (a_6 - a_8)/2n_1$$
$$n_1 = \sqrt{(a_1 - \cos\Theta)/(1 - \cos\Theta)}$$
$$n_2 = (a_2 + a_4)\,(1 - \cos\Theta)/2n_1$$
$$n_3 = (a_3 + a_7)\,(1 - \cos\Theta)/2n_1.$$

The algorithm has been shown to perform well on both real and synthetic data, and these results are presented in [49].

The use of lines as observables in an approach similar to the one outlined above has also been attempted by Mitiche, Seida and Aggarwal [50] who used the principle of angular invariance between 3-D lines on a rigid body undergoing motion. In their method the orientation of lines is first recovered, then the rotational component is computed, and finally, the translation is recovered. The observation of four lines in three views allows for the determination of structure and motion parameters.

The use of line correspondences has the advantage over the use of point correspondences in that extraction of lines in images is less sensitive to noise than extraction of points. Also, it is easier to match line segments between images than it is to match points.

It is possible to use both lines and points concomitantly in formulating the task. In the case of combined point and line correspondences, four points and a line in two views are sufficient to compute the structure of the scene as well as the displacement between views as described by Aggarwal and Wang [51].

The following observations may be made based on the current literature:

1) Using points or lines, or combination of points and lines for the computation of structure and motion usually gives rise to nonlinear equations.
2) The computation based upon minimum number of points or lines is usually more sensitive to noise perturbations.
3) In general, alternate formulations may give rise to different sufficiency conditions regarding minimum number of points and lines required for solving structure and motion.

C. Using Extended Sequences of Monocular Images

The approaches outlined above attempt to recover structure and motion from a limited number of views of the scene, typically 3 or 4 views. We discuss below some techniques that use long sequences of monocular images to recover structure and motion.

The first of these is the incremental approach which allows for deviations from rigid body motion. This differs from the approaches outlined above which assumed that the object being imaged undergoes rigid body motion. Psychophysical studies have shown that the human visual system can cope with less than strict rigidity [52], [26], [27]. These studies prompted Ullman to devise an algorithm that recovers the 3-D structure of viewed objects in an incremental manner using several views of an object in motion [52]. The performance of the algorithm is argued to be comparable to that of the human visual system because it possesses the following characteristics [52]:

1) At each instant there exists an estimate of the 3-D structure of the viewed object. The internal model $M(t)$ of the viewed structure at time t may be initially crude and inaccurate, and may be influenced by static sources of 3-D information.
2) The recovery process prefers rigid transformations.
3) It is able to integrate information from an extended viewing period.
4) The recovery process tolerates deviations from rigidity.
5) It eventually recovers the correct 3-D structure, or a close approximation to it.

A parallel projection system is used. $M(t)$ consists of a set of 3-D coordinates (X_i, Y_i, Z_i) where (X_i, Z_i) are the observed image plane coordinates of a point and Y_i is the depth. The estimation of structure therefore consists of finding Y_i. An initial set of values is chosen for the Y_i. Consider the situation at time t'. Let (x_i, y_i, z_i) be the new structure of the corresponding points. The task is to find y_i while minimizing deviations from rigidity. The deviation from rigidity is defined as follows. Let L_{ij} denote the distance between points i and j at time t. Let L'_{ij} denote the distance between points i and j at time t'. Under rigid motion L_{ij} should be equal to L'_{ij}. The deviation in rigidity is expressed as

$$E = \sum D_{ij}, \quad \text{where} \quad D_{ij} = \frac{(L_{ij} - L'_{ij})^2}{L_{ij}^3},$$

and the summation is for all i, j.

Two modifications to the basic scheme were explored [52]. These included using different metrics for measuring the deviation from rigidity and allowing for a correction in the initial model $M(t)$. Simulations using synthetic data were conducted. Results indicate that the model does arrive at a good approximation to the 3-D structure after several views, but does not converge to the exact solution. Also, the solution is unique upto a mirror reflection. The modification involving a flexible model quickly arrived at a good approximation with a few views but with additional views the estimated structure oscillated about the correct solution. An analysis of the convergence properties of this algorithm has also been carried out by Hildreth and Grzywacz [53]. They have also suggested a continuous formulation of the above approach wherein instantaneous velocities of the points are used instead of point positions.

Although it is argued that such a formulation is warranted when arbitrarily close frames are used, the results of Hildreth and Grzywacz indicate that local velocity information is insufficient to solve the problem, even when the object is viewed over an extended period. The major limitation of the incremental approach discussed above is that it performs well only when objects rotate about a fixed axis. In addition, orthographic projection is not generally valid. The approach does however illustrate the importance of motion in the perception of structure.

Broida and Chellappa [54] consider the case of a rigid body undergoing constant translational and rotational motion. This assumption allows for a formulation in which the number of unknown model parameters does not increase with the increase in the number of image frames. A two-dimensional object undergoing one-dimensional motion is assumed. They also assume that the object structure is known and attempt to recover the motion parameters. A Kalman filter is employed for recursive estimation of the motion parameters. The object is assumed to be transparent so that feature points are always visible and correspondence is assumed to have been established *a priori*. The unknown model parameters are represented as a vector:

$$[xc \; \dot{x}c \; zc \; \dot{z}c \; p1 \; p2 \; w]^T$$

where, (xc, zc) is the location of the center of mass of the object, $(\dot{x}c, \dot{z}c)$ is the object translational motion, $p1$ and $p2$ are unknown phase angles of moment arms $r1$ and $r2$ that connect the two feature points to the center of mass. Here $r1$ and $r2/r1$ is assumed known. The differential equation describing unforced motion is written in terms of the above vector as:

$$\dot{\mathbf{x}}(t) = [\dot{x}c \; 0 \; \dot{x}c \; 0 \; w \; w \; 0]^T$$

with arbitrary initial conditions $xc(t)$, $zc(t)$, $p1(t)$, and $p2(t)$. This system yields the following state equation:

$$\mathbf{x}(k+1) = \boldsymbol{F}(k)\,\mathbf{x}(k)$$

where

$$\mathbf{x}(k) = [xc(k) \; \dot{x}c(k) \; zc(k) \; \dot{z}c(k) \; p1(k) \; p2(k) \; w(k)]^T \quad \text{and}$$

$$\boldsymbol{F}(k) = \begin{bmatrix} 1 & \tau & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \tau & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \tau \\ 0 & 0 & 0 & 0 & 0 & 1 & \tau \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Here, τ is the time interval between successive images. The measurement model is given by

$$X1 = L[xc + r1 \cos(p1)]/[zc + r1 \sin(p1)] = h1[\mathbf{x}(k)]$$

$$X2 = L[xc + r2 \cos(p2)]/[zc + r2 \sin(p2)] = h2[\mathbf{x}(k)]$$

where $X1$ and $X2$ are the images of the two feature points and L is the focal length of the sensor. The vector representation is given by

$$\boldsymbol{X}(k) = [X1(k) \; X2(k)]^T = \boldsymbol{h}[\mathbf{x}(k)] + \boldsymbol{n}(k)$$

where $\boldsymbol{h}[x] = [h1(\mathbf{x}) \; h2(\mathbf{x})]$ and $\boldsymbol{n}(k)$ is the term corresponding to zero mean, Gaussian, spatially correlated, and temporally white noise.

The above formulation is then used to design an iterated extended linear Kalman filter that solves for the state variables—in this case the translation and rotation parameters. The performance of the algorithms on Monte Carlo simulations are discussed in [54], while extensions of this approach are presented in [55].

Weng, Huang and Ahuja [56] have proposed a method of characterizing rigid body motion from long monocular image sequences, i.e., over extended viewing periods. Their approach involves first extracting structure and motion parameters with two views of 8 points [33]–[36] and then computing the trajectory of the rotation center which is the center of mass or some fixed point of the object. They assume that the angular momentum of the object is locally constant and the object possesses an axis of symmetry. They argue that if motion is smooth and the time interval covered by the model is relatively short, then the trajectory of the rotation center can be approximated by a polynomial. The developed model is applied to subsequences of images to estimate the trajectory and predict the new locations of object points. The main characteristic of interest is the existence of precessional motion and the parameters thereof. A least-squares method is adopted to compute the parameters. The authors present a detailed analysis of the relationship between the parameters of precessional motion and discrete two-view motion. The simulations discussed, however, deal only with 3-D point sets and no testing has been conducted using real data extracted from monocular image sequences.

D. The Correspondence Problem

In the above discussions it is repeatedly assumed that correspondence was available between features extracted from one image in a sequence of images and those extracted from the next image. The task of establishing and maintaining such correspondence is, however, nontrivial. The ambiguity is aggravated by the effects of occlusion which cause features to appear or disappear and also give rise to "false" features. The development of robust techniques to solve the correspondence problem is an active area of research that is still in its infancy. We present a brief description of a few of the approaches developed. The problem of finding correspondence is common to other areas of computer vision such as stereoscopy and optic flow. Some of the techniques developed for solving the correspondence problem in these other areas can be applied to the feature-based analysis of monocular images as well, and vice versa.

Aggarwal *et al.* [57] have classified correspondence processes into two categories: those that are based on iconic models and those that are based on structural models. The former class consist of templates extracted from the first frame which are then detected in the second and subsequent frames. The second approach consists of extracting tokens with a number of attributes from the first image, and using domain constraints and structural models to match these tokens with those extracted from the second and subsequent images. The latter approach is computationally more expensive but also more robust than the former.

Sethi and Jain [58] describe a method for finding correspondence and maintaining correspondence between feature points extracted from a long sequence of monocular images. They present algorithms based on preserving the smoothness of velocity changes. The iterative optimization algorithms search for an optimum set of trajectories for feature points in a sequence of images based on constraints on the direction and magnitude of change in motion. A hypothesize and test approach is also proposed to handle occlusion. This method hypothesizes occlusion if the number of feature points detected in a frame is less than that detected in two or more preceeding or succeeding frames. Interpolating the missing point position using the preceeding two frames and testing this with the subsequent two frames verifies the existence of occlusion. Experiments with manually extracted features illustrate that the approach is able to deal with limited occlusion. The problem of automated extraction of features, however, has not been addressed by the authors.

Fang and Huang [59] have presented experimental results of motion parameter estimation using a modified version of an algorithm initially developed by Ranade and Rosenfeld [60]. The relaxation algorithm is modified by incorporating different scales to allow for large scale changes in the images (due to large translations in depth). Another relaxation technique for establishing correspondence is due to Kim and Aggarwal [61]–[63] who have applied their technique to matching features in stereo imagery as well as for matching 3-D features in depth maps. Barnard and Thompson [64] have proposed an iterative relaxation labeling technique for matching features in stereo imagery based on smoothness in change of depth. This method may be applied to matching features in two monocular images based on smoothness in spatial displacement of image features. Prager and Arbib [65] describe a technique similar to Barnard and Thompson but have included an additional temporal constraint on feature displacements. Many other approaches to matching image features can be found in recent literature, for example see [66]–[68].

In this section we discussed the feature-based extraction of motion from monocular image sequences. It was assumed that image features, such as points and lines, had been extracted from each image and inter-frame correspondence had been established between them. Three approaches to the problem were discussed: the direct formulation method where rigid body motion is implicitly used, a formulation in which rigidity is explicitly invoked, and the third approach using long sequences of monocular images.

IV. Optic Flow Based Motion Estimation

In this section we present approaches in which the instantaneous changes in brightness values in the image are analyzed to yield a dense velocity map called image flow or optic flow. The three-dimensional motion and structure parameters are then computed based on various assumptions and/or additional information. No correspondence between features in successive images is required. The optic flow techniques rely on local spatial and temporal derivatives of image brightness values. This approach, as will be evident from the following discussion, is distinct from the feature-based analysis of monocular image sequences discussed in the previous section where 1) a relatively sparse set of two-dimensional features is extracted from the images, 2) inter-frame correspondence is established between these features, 3) constraints are formulated based on assumptions such as rigid body motion, and 4) the observed displacement of the 2-D image features are used to solve these equations to produce 3-D structure and motion estimates.

The relative motion of a scene with respect to the viewer gives rise to a distribution of velocities on the image plane. This phenomenon manifests itself as temporal change in brightness values (gray levels) in the image plane. The image velocities are, in general, functions of the motion of viewed objects relative to the camera, objects' locations in 3-D space, and 3-D structure of the objects. The recovery of the 3-D motion and structure information from the sequence of monocular images can be decomposed into two steps: 1) compute image plane velocities from changes in image intensity values, and 2) use optic flow to compute 3-D motion and structure. We discuss below some basic formulations of these two problems and outline the salient features in solutions to these two tasks.

A. Computing Optic Flow

Let $g(x, y, t)$ be the image intensity at point (x, y) in the image at time t. With the assumption that the intensity is the same at time $t + \Delta t$ for the point $(x + \Delta x, y + \Delta y)$ of the image, we have

$$g(x + \Delta x, y + \Delta x, t + \Delta t) = g(x, y, t) \qquad (4.1)$$

where Δt, Δx, and Δy are small. Approximating the left-hand side by a Taylor series

$$g(x + \Delta x, y + \Delta y, t + \Delta t) = g(x, y, t) + g_x \Delta x + g_y \Delta y + g_t \Delta t + \text{higher order terms.} \qquad (4.2)$$

Ignoring the higher order terms in (4.2), using (4.1) in (4.2) and taking the limit as $\Delta t \rightarrow 0$

$$g_x u + g_y v + g_t = 0. \qquad (4.3)$$

In this equation, the partial derivatives g_x, g_y, and g_t are estimated from the image; $u = dx/dt$ and $v = dy/dt$ are the velocity components in the directions x and y, respectively, associated with the point (x, y). The collection of such velocity vectors for the entire image constitutes the optic flow for the image.

Equation (4.3) embodies two unknowns u and v, and is not sufficient by itself to specify the optical flow uniquely. It does constrain the solution. It is possible to compute optical flow for images using the optical flow constraint equation together with additional assumptions. Popular assumptions include one of the following:

a) optical flow is smooth and neighboring points have similar velocities,
b) optical flow is constant over an entire segment of the image,
c) optical flow is the result of restricted motion, for example, planar motion.

One such constraint is the smoothness constraint, i.e., motion field varies smoothly in most parts of the image [69]–[72]. Horn and Schunck [69] imposed this constraint by minimizing the error in optic flow expressed as:

$$E^2(x, y) = (\text{error in (4.3)}) + \lambda^2 \,(\text{deviation from smoothness}) = (g_x u + g_y v + g_t)^2 + \lambda^2 \{(u_x^2 + u_y^2) + (v_x^2 + v_y^2)\} \qquad (4.4)$$

where λ is a constant. The task is to find u and v so as to minimize R in the following

$$R = \iint \{(g_x u + g_y v + g_t)^2 + \lambda^2[(u_x^2 + u_y^2) + (v_x^2 + v_y^2)]\} \, dx \, dy. \qquad (4.5)$$

The integral equation may be solved by methods of calculus of variation. Differentiating (4.5) with respect to u and v and equating $\partial R/\partial u$ and $\partial R/\partial v$ to zero (for minimum error R), and writing $(u_x^2 + u_y^2) = u - u_{\text{ave}}$, and $(v_x^2 + v_y^2) = v - v_{\text{ave}}$, we get the following:

$$u = u_{\text{ave}} - g_x P/D, \qquad v = v_{\text{ave}} - g_y P/D \qquad (4.6)$$

where

$$P = (g_x u_{\text{ave}} + g_y v_{\text{ave}} + g_t), \text{ and } D = \lambda^2 + g_x^2 + g_y^2.$$

Equation (4.6) may be solved iteratively, i.e., obtain $u(t)$, $v(t)$ using $u_{\text{ave}}(t - 1)$, $v_{\text{ave}}(t - 1)$.

Horn and Schunck show that the iterative method converges when the optic flow is static, i.e., when the velocity vectors do not change with time, e.g., a sphere rotating about a stationary axis. When this condition is violated, e.g., when an object translates in front of a stationary background, there exist boundaries where local smoothness of optic flow will not hold. If the boundaries can be detected then the technique may be limited to smooth regions. Some techniques for determining such boundaries are discussed by Schunck [73].

The first-order approximation of (4.2) is unsatisfactory for edges and corners in the image [74]. First- and second-order derivatives of the Taylor series expansion of (4.2) were used by Snyder *et al.* [75] who obtained a single nonlinear equation in the two unknowns u and v. Prazdny [76] used the approach suggested by Snyder *et al.* [75] to solve the problem where only pure translation of the sensor was involved. Prazdny further assumes that the Focus of Expansion (FOE)[1] of image flow is assumed known and then solves for the magnitude of the image flow.

Yachida [77] extended Horn and Schunck's iterative method discussed above [69] for computing optic flow. The smoothness constraint considered not only a spatial neighborhood within the frame but also a temporal neighborhood, i.e., areas in the preceeding and succeeding frames.

In order to devise additional constraints to solve the image flow equation (4.3) Nagel [74], [78] has posed specific conditions on local gray value distributions and has presented an operator (gray value corner detector) that detects locations in the image that satisfy these conditions. He develops the Taylor series of (4.2) up to second-order terms. Minimizing an error functional results in a system of two nonlinear equations in u and v. These yield a closed form solution for the optic flow at the image locations detected by the corner detector. Nagel and Enkelmann [79] use these values as initial estimates in an iterative algorithm that extends the solution of the nonlinear system of equations into image areas surrounding the gray value corner. Nagel [80] has also proposed a modification of Horn and Schunck's smoothness criterion to take into consideration occluding edges. Nagel introduced a weight matrix which depends on gray level changes in such a way that smoothness requirement is retained only for the optical flow component which is perpendicular to strong gray value transitions.

Haralick and Lee [81] use (4.3) in conjunction with the requirement that the first derivatives of the gray value structure that has been displaced in the image due to object motion must remain the same. This yields three additional equations:

$$\begin{aligned} g_{xx}u + g_{xy}v + g_{xt} &= 0 \\ g_{yx}u + g_{yy}v + g_{yt} &= 0 \\ g_{tx}u + g_{ty}v + g_{tt} &= 0. \end{aligned} \qquad (4.7)$$

Equations (4.7) and (4.3) form an overdetermined system of four linear equations in u and v. Tretiak and Pastor [82]

[1] The Focus of Expansion (FOE) is defined as the intersection of the axis of camera translation with the image plane, when the intersection occurs on the positive half of the axis. When this intersection lies on the negative half of the axis of translation, it is termed the Focus of Contraction (FOC).

also independently arrived at a similar formulation. The solution of the system of equations is effected by the pseudoinverse formalism [78], [82].

Hildreth [83] has developed a scheme for computing image velocity vectors along contours formed by detecting zero-crossings of the Laplacian of Gaussian (LOG) filtered image [67]. This approach is based on Marr's theory that initial motion measurements by the human visual system are made only at locations of significant intensity changes. The two-dimensional velocity field along the contour is described by the vector function $\boldsymbol{V}(s)$, where s denotes arclength. $\boldsymbol{V}(s)$ can be decomposed into components $v^n(s)$ and $v^t(s)$ that are perpendicular and tangential, respectively, to the contour.

$$\boldsymbol{V}(s) = v^t(s)\boldsymbol{u}^t(s) + v^n(s)\boldsymbol{u}^n(s), \tag{4.8}$$

where $\boldsymbol{u}^n(s)$ and $\boldsymbol{u}^t(s)$ are unit vectors in the directions perpendicular and tangential to the curve. An orthographic projection geometry is used. Solutions to (4.8) for the simple cases of constant velocity and rigid motion in image plane are discussed. The application of a more general constraint is then discussed, i.e., the assumption that velocity varies smoothly along the contour. To measure total variation in the velocity field the following continuous functional is proposed

$$\Theta(\boldsymbol{V}) = \int \left|\frac{\partial \boldsymbol{V}}{\partial s}\right| ds. \tag{4.9}$$

This is combined with the constraint that the perpendicular component of the computed velocity field $\boldsymbol{V} \cdot \boldsymbol{u}^n$ must be close to the measured perpendicular component v^n to form the following functional

$$\Theta(\boldsymbol{V}) = \int \left[\left(\frac{\partial \boldsymbol{V}}{\partial s}\right)^2 + \left(\frac{\partial \boldsymbol{V}}{\partial s}\right)^2\right] ds + \beta \int (\boldsymbol{V} \cdot \boldsymbol{u}^n - v^n)\, ds \tag{4.10}$$

where β is a weighting factor. A discrete form of the above functional is specified

$$\Phi = \Phi_1 + \Phi_2 \tag{4.11}$$

$$\Phi_1 = \sum_{i=2}^{k} [(V_{x_i} - V_{x_{i-1}})^2 + (V_{y_i} - V_{y_{i-1}})^2] + (V_{x_1} - V_{x_n})^2 + (V_{y_1} - V_{y_n})^2 \tag{4.12}$$

$$\Phi_2 = \beta \sum_{i=1}^{k} [V_{x_i} u^n_{x_i} + V_{y_i} u^n_{y_i} - v^n_i]^2 \tag{4.13}$$

where k is the number of points in the contour. In order to find the velocities (V_{x_i}, V_{y_i}) which minimize Φ, $\partial\Phi/\partial V_{x_i}$ and $\partial\Phi/\partial V_{x_i}$ are equated to zero. This yields $2k$ linear equations which are solved via the conjugate gradient algorithm [83]. Experimental results using real data have been conducted where the initial perpendicular components of velocity were computed from the time derivative between two LOG filtered images, and the gradient along the zero-crossing contours of the first filtered image. Experiments on synthetic data show that the smoothness criterion does not guarantee accurate estimates of image flow. It is argued that the velocity field, even though incorrect, is perceptually valid.

Nagel [78] has presented a comparative analysis of the above schemes of Horn and Schunk [69], Haralick and Lee [81], Tretiak and Pastor [82], Nagel [80], and Hildreth [83] using a mathematical formalism developed by him and has shown the relationship between these approaches.

The above approaches deal with images at a single scale of resolution, i.e., the finest resolution available from the imaging sensor. Several hierarchical schemes have been developed [84]–[87]. Enkelmann [84] creates a Guassian low-pass pyramid for each image. Processing begins at a coarse level wherein the initial displacement vectors are set to zero. These vectors are projected to finer levels via bi-linear interpolation. Within each level, the velocity field is computed via Nagel's approach [80] which embodies the oriented smoothness criterion. A finite difference approach yields a large sparse system of linear equations which is solved using a multi-resolution relaxation approach. Glazer's approach [85] uses Horn and Schunk's criteria [69]. Glazer uses a Gaussian pyramid with quad-tree connectivity to propagate velocity vectors from coarse to fine levels. Glazer uses a finite difference approach and a complex multi-level relaxation approach which involves dynamic switching between levels. Anandan [87] uses a Laplacian pyramid which provides a set of bandpass filters (as opposed to the low-pass filters provided by Gaussian pyramids). A coarse to fine control strategy is also employed via an "overlapped projection scheme" that allows for multiple choices in the propagation of velocity vectors. Anandan's technique is based on establishing matches between image events in successive frames. The match criterion used is the minimization of a Gaussian weighted sum-of-squared-differences (SSD) in a 5 × 5 window and a confidence measure based on the distribution of the SSD values. A smoothness constraint similar to that of Glazer is used. The minimization problem is solved via a finite-element method that takes into consideration known discontinuities in the displacement field.

Another method, called the multi-constraint method, is emerging with promise. In this method one considers several functions $f_1, f_2, \cdots, f_n$ such that each of them satisfy the constraint equation. In particular,

$$\frac{\partial f_i}{\partial x} u + \frac{\partial f_i}{\partial_y} v + \frac{\partial f_i}{\partial t} = 0, \qquad i = 1, \cdots, n. \tag{4.14}$$

Candidate functions include directional derivatives. However, the results based upon these functions have not been promising. Other candidate functions include $g = \boldsymbol{O}(f)$ where $\boldsymbol{O}$ is an operator like the contrast, entropy, average, etc. Mitiche, Wang and Aggarwal [88] have reported preliminary success in the computation of optical flow using multi-constraint methods.

Fleet and Jepson [89] and Tsotsos *et al.* [90] have investigated the extraction of motion information using Fourier techniques. They proposed a hierarchical computational framework for early processing in the human visual system which involves the use of spatiotemporal linear filters tuned to specific frequencies corresponding to specific image velocities. A cascaded configuration of orientation specific filters followed by speed specific filters was proposed. Recently, Heeger [91] demonstrated that a family of motion-sensitive Gabor filters can be used to compute optic flow. He used 3-D (space-time) Gabor filters tuned to different spatiotemporal-frequency bands and described a method for combining the outputs of the filters to compute local velocity vectors. He has further suggested a parallel implementation and has illustrated the performance of his approach with synthetic as well as real data.

The determination of optical flow for a scene consisting of several moving objects has also been attempted. Research has focused on segmenting the optic flow into regions corresponding to distinct objects that undergo different motion. Murray and Buxton [92] use a Bayesian approach to formulate the segmentation problem. The optic flow field is modeled as spatial and temporal Markov random fields. The search for the globally optimal segmentation is performed using simulated annealing. Thompson [93] combines optical flow and contrast information in a region growing scheme that segments images into regions corresponding to surfaces moving with different velocities. Thompson *et al.* [94] detect flow boundaries using an algorithm patterned after the Marr-Hildreth zero-crossing detector. O'Rourke proposed a method to group rotating random dot patterns [95]. Fennema and Thompson extract moving regions by collecting similar optical flow vectors [96]. Adiv segmented an optic flow field using a grouping method based on a Hough voting approach [97]. Webb and Aggarwal [48] analyzed relative motion between multijointed parts of objects. More recently, Tsukune and Aggarwal [98] describe a method for extracting multiple rotational flow fields in the Hough space for orthographically projected 3-D velocity vector fields.

B. Computing Structure and 3-D Flow

Having computed optical flow, there still remains the problem of computing the motion and the structure of the object in three-dimensional space. A mathematical formulation of the basic problem is first presented. The formulation is that used by Prazdny [99], [100], Longuet-Higgins and Prazdny [101], Waxman *et al.* [102], [103], and Subbarao [104], [105], among others.

A camera centered Cartesian coordinate system (X, Y, Z) is used. The Z axis is directed along the viewing direction. The image plane is normal to the Z axis and is at unit distance from the origin. The image coordinate system (x, y) has its origin at (0, 0, 1). The x and y axes are parallel to the X and Y axes, respectively. In the perspective projection geometry, the image of a point (X, Y, Z) is formed by drawing a line from it to (0, 0, 0) which intersects the image plane at (x, y). Therefore

$$x = X/Z \qquad \text{and} \qquad y = Y/Z. \tag{4.15}$$

The camera is assumed to be in motion, with $\boldsymbol{V} = (V^X, V^Y, V^Z)$ being the translational velocity and $\boldsymbol{\Omega} = (\Omega^X, \Omega^Y, \Omega^Z)$ being the rotational velocity. The instantaneous velocity of a point $\boldsymbol{R} = (X, Y, Z)$ is given by $(\dot{X}, \dot{Y}, \dot{Z}) = -(\boldsymbol{V} + \boldsymbol{\Omega} \times \boldsymbol{R})$ as follows:

$$\begin{aligned}\dot{X} &= -V^X - \Omega^Y Z + \Omega^Z Y\\ \dot{Y} &= -V^Y - \Omega^Z X + \Omega^X Z\\ \dot{X} &= -V^Z - \Omega^X Y + \Omega^Y X.\end{aligned} \tag{4.16}$$

From this the instantaneous image velocity $(u, v) = (\dot{x}, \dot{y})$ can be written as

$$u = \left(x\frac{V^Z}{Z} - \frac{V^X}{Z}\right) + (xy\Omega^X - (1 - x^2)\,\Omega^Y + y\Omega^Z) \tag{4.17a}$$

$$v = \left(y\frac{V^Z}{Z} - \frac{V^Y}{Z}\right) + ((1 + y^2)\,\Omega^X - xy\Omega^Y - x\Omega^Z). \tag{4.17b}$$

The estimation of structure and motion is based on the key assumptions that i) the optic flow varies smoothly and ii) the surface of the object is smooth. Assumption i) allows the optic flow in a small image neighborhood around image location (x, y) to be specified by a Taylor series as:

$$\begin{aligned}u(x, y) = u_0 + u_x x + u_y y + u_{xx} x^2\\ + u_{xy} xy + u_{yy} y^2 + O_3(x, y)\end{aligned} \tag{4.18a}$$

$$\begin{aligned}v(x, y) = v_0 + v_x x + v_y y + v_{xx} x^2\\ + v_{xy} xy + v_{yy} y^2 + O_3(x, y)\end{aligned} \tag{4.18b}$$

where the partial derivatives can be computed from the optic flow. Assumption ii) allows a small surface patch $Z(X, Y)$ around the line of sight to be described as:

$$\begin{aligned}Z = Z_0 + Z_X X + Z_Y Y + \tfrac{1}{2} Z_{XX} X^2\\ + Z_{XY} XY + \tfrac{1}{2} Z_{YY} Y^2 + O_3(X, Y)\end{aligned} \tag{4.19}$$

for $Z_0 > 0$ is the distance of the surface patch along the line of sight. Substituting the relation (4.15) for Z in (4.19) in a recursive manner it is possible to further approximate the surface in terms of image plane coordinates as:

$$Z(x, y) = \frac{Z_0}{(1 - Z_X x - Z_Y y - \frac{1}{2} Z_{xx} x^2 - Z_{xy} xy - \frac{1}{2} Z_{yy} y^2 - O_3(x, y))} \tag{4.20}$$

where $Z_{xx} = Z_0 Z_{XX}$, $Z_{yy} = Z_0 Z_{YY}$, $Z_{xy} = Z_0 Z_{XY}$.

Further, the scaled translational velocities are denoted as follows:

$$V^x = \frac{V^X}{Z_0}, \quad V^y = \frac{V^Y}{Z_0}, \quad V^z = \frac{V^Z}{Z_0}, \quad \text{for } Z_0 > 0. \tag{4.21}$$

From (4.17), (4.18), (4.20), and (4.21) it is possible to derive the following relations [101]–[103], [105] assuming rigid uniform motion:

$$\begin{aligned}
u_0 &= -V^x - \Omega^Y & v_0 &= -V^y + \Omega^x\\
u_x &= -V^z + V^x Z_X & v_y &= V^z + V^y Z_Y\\
u_y &= \Omega^Z + V^x Z_Y & v_x &= -\Omega^Z + V^y Z_X\\
u_{xx} &= -2\,V^z Z_X + V^X Z_{xx} - 2\Omega^Y & u_{xy} &= -V^z Z_Y + V^x Z_{xy}\\
& & &\quad + \Omega^X\\
u_{yy} &= V^x Z_{yy} & v_{xx} &= V^y Z_{xx}\\
v_{xy} &= -V^z Z_X + V^y Z_{xy} - \Omega^Y & v_{yy} &= -2\,V^z Z_Y + V^y Z_{yy}\\
& & &\quad + 2\,\Omega^X.
\end{aligned} \tag{4.22}$$

The system of equations (4.22) relates the optic flow (u, v) and its first- and second-order spatial derivatives to the 3-D structure and motion parameters. The geometric structure for the smooth surface is specified locally by the surface slopes and curvatures, i.e., Z_X, Z_Y, Z_{xy}, Z_{xx}, and Z_{yy}. The three-dimensional motion parameters are the components of $\boldsymbol{V}$ and $\boldsymbol{\Omega}$. The system (4.22) comprises twelve nonlinear equations in eleven unknowns and is thus overdetermined. The optic flow and its derivatives are available using any of the methods outlined in the previous subsection. The overdetermined system (4.22) may, hence, be solved to yield the structure and motion parameters.

Many interesting observations may be made regarding the above equations. Note from (4.21) and (4.22) that Z_0 is not recoverable and only scaled translational velocity and curvatures may be computed. Every nonlinear term in (4.22) is a product of a structural parameter and a translational velocity component. Every curvature parameter in (4.22) is multiplied by a component of translational velocity (V^x or V^y) which is parallel to the image plane. Hence, if there is no translation parallel to the image plane, surface curvatures cannot be determined.

The nonlinear overdetermined system (4.22) may or may not yield a unique solution. Many situations give rise to dependent equations in (4.22) engendering multiple solutions. A detailed analysis of numerous cases has been presented by Subbarao [104], [105] and Waxman *et al.* [102], [103] who have derived closed form solutions for these cases. Subbarao shows that in general the solution is unique, and at most four solutions are possible in certain situations. Negahdaripour [106] also addressed the ambiguity in interpreting optic flow produced by curved surfaces in motion. He argues that the ambiguity is at most three-fold for the case of certain hyperboloids of one sheet viewed by an observer moving parallel to the image. The ambiguities inherent in interpreting noisy flow fields are discussed by Adiv [107].

An overview of some of the approaches for computing structure and motion parameters from optic flow is given below. The approaches typically involve restricting the nature of motion to be purely translatory or rotational and/or restricting the imaged surface to be planar. These assumptions significantly reduce the complexity of the system of equations (4.22).

Williams [108] considered the computation of the structure of imaged scene components for the situation where the sensor was involved in purely translatory motion. The Focus of Expansion (FOE) of image flow is assumed known and the scene is considered to consist of planar surfaces. A height and position is hypothesized for each segmented region. An image is generated for the known camera motion and compared with the actual image. Error in the hypothesized structure is computed from the difference between these two images and appropriate corrections are made to the hypothesized scene structure. This procedure is repeated until the error falls below a threshold. This approach has also been suggested for detecting the FOE.

An approach for determining scene structure from a sequence of images acquired by a translating camera is credited to Lawton [109]. In this method, features are extracted from each image. Several directions of camera motion are hypothesized. Each corresponds to a unique FOE or FOC. Image feature displacements are computed for each motion and compared with actual displacements. The motion corresponding to minimum error in feature displacements is chosen to be the best estimate. Scene structure is computed in units of relative depth, i.e., ratio of depth to change in depth. The technique allows for the segmentation of objects at different depth.

Rieger and Lawton [110] have devised a method for determining the instantaneous axis of translation for a camera undergoing general motion. Their method is based on the observation of Longuet-Higgins and Prazdny [101] that two surface points which lie on the same ray of projection but at different depths will have image velocities that differ only by the difference in the translational components of their 3-D velocity. Difference vectors are computed at optic flow discontinuities and the intersection of these difference vectors are estimated via an optimization technique similar to that used in [109]. The translational axis is specified by this procedure and the computation of camera rotation and translation is simplified.

Prazdny [111] proposed an approach in which the velocity field is decomposed into rotational and translational components. The rotational motion is hypothesized and the FOE is identified for the resultant translational field. An error function of three parameters is used to evaluate the estimated motion. Minimization of the error yields the best estimate. The algorithm has been tested on data generated by simulated planar surfaces in motion.

Bruss and Horn [112] and Horn [71] discuss the formulation of an iterative least-mean-squared error approach to the estimation of 3-D motion from optical flow. They make no *a priori* assumptions about the motion. They derive a system of seven equations, three of which are linear in V^x, V^y, and V^z, and four which are solved via a numerical method. No experimental results, however, have been shown. Horn and Weldon [113] have proposed methods for computing purely translational or purely rotational 3-D motion directly from brightness gradients without computing optical flow. They employ only first derivatives of the image gray levels, and analyticity of the surface is not required. Negahdaripour and Horn [114] discuss the recovery of motion of a camera relative to a planar surface. They also do not compute optic flow, and use instead the spatial and temporal derivatives of brightness values directly. They present iterative schemes for solving nine non-linear equations based on a least-squares formulation, and also present a closed form solution.

Chou and Kanatani [115] use a scheme in which object motion is initially hypothesized and iteratively refined. They extract features from the images obtained before and after motion. They do not require that feature correspondence be established *a priori*. They transform the first set of features and evaluate the discrepancy between the estimated feature positions and the true feature positions (in the second image) after motion. Assuming infinitesimal motion, they relate the discrepancy to optic flow parameters. They use a numerical least-squares technique to solve the linear constraints for a better estimate of the motion. This process is repeated until the estimated motion produces feature positions that are sufficiently close to the true ones obtained in the second image after motion.

In this section we have presented the optic flow approach for the estimation of motion parameters from a sequence of monocular images. We discussed the basic formulation of the problem and outlined some of the recently developed techniques for computing the optic flow. The above discussion included the problem of inferring 3-D structure and motion from optic flow and overviews of some of the solutions to this problem.

V. Comparing Optic Flow and Feature-Based Methods

In the preceeding sections we discussed two distinct approaches for the estimation of motion from monocular image sequences, i.e., feature-based analysis and optical flow methods. In this section we compare the two approaches and discuss some of the advantages and disadvantages associated with each of the methods.

Feature-based approaches require that correspondence be established between a sparse set of features extracted from one image with those extracted from the next image in the sequence. Although several methods have been discussed for extracting and establishing feature correspondence, the task is difficult and only partial solutions suitable for simplistic situations have been developed. In general, the process is complicated by occlusion which may cause features to be hidden, false features to be generated and hidden features to reappear. Much more work needs to be done in this area before the advent of one or more general techniques that can be reliably applied to real imagery. In comparison, the optic flow approach, in general, does not require any feature correspondence to be established.

The computation of the optical flow as well as the interpretation of motion and structure from optic flow requires the evaluation of first and second partial derivatives of image brightness values and also of the optic flow. Real images are, in general, noisy. The evaluation of derivatives is a noise enhancing operation. The higher the order, the more noise sensitive is the derivative. Hence, even in cases where closed form solutions for the 3-D structure and motion exist, the optical flow techniques do not produce usable results because of the sensitivity to noise [71]. Also, there are discontinuities in the optical flow depending upon occlusion, and these regions must be detected reliably otherwise violations of the continuity assumption will have adverse and global effects on the estimate of optical flow.

In contrast to the method of global minimization, another approach depends upon solving a set of constraints in a small neighborhood. However, the local and global methods rely on similar assumptions of smoothness of optical flow field. The common weakness of both methods is the inaccurate estimates at points where the flow changes sharply or is discontinuous. The global method propagates the errors across the entire image, while the neighborhood size limits the propagation in local methods. Schunck [70] and Kearney *et al.* [116], [117] address these difficulties in detail. Kearney *et al.* present a detailed analysis of the sources of errors in local optimization techniques for computing optical flow [116]. They identify three main sources of error:

1) Poor estimation of brightness gradients in highly textured image regions. The problem is especially severe for temporal gradients in moving regions.
2) Variations in optic flow across the image violate the assumption of locally constant flow. Significant error arises at discontinuities in the flow field.
3) Insufficient local variation in the orientation of the brightness gradient which causes error propagation in the ill-conditioned system.

Sensitivity to noise is also a problem with the feature-based techniques though to a lesser degree. The techniques reported in the literature have all been only marginally tolerant to noise. One method of decreasing the sensitivity to noise has been to use more than the required minimum number of features in an iterative least-squares technique. Although this usually has a smoothing effect, it can cause additional complications. For example, if all the additional points chosen are coplanar, then all that has been achieved is a significant increase in the computation time and probable instability of the solution. The establishment of correspondence also becomes computationally expensive.

Recently, Verri and Poggio [118] argued that the optic flow does not correspond to the 2-D velocity field unless very special conditions are satisfied. They argue against the use of optic flow for quantitative estimates of 3-D motion. They apply the theory of stability of dynamical systems to the optic flow formulation and conclude that the optic flow may provide stable qualitative information such as the Focus of Expansion and motion discontinuities.

When numerical techniques are used for the solution of structure and motion using either approach one must consider the many caveats involved in such a solution. A discussion of these caveats would be inappropriate in this paper and the reader is directed to the literature in numerical analysis for possible pitfalls and remedies.

Much attention has been devoted recently by the computer vision community to the use of regularization techniques in many vision tasks including both feature-based formulations and the optic flow approach for motion and structure estimation [119]–[123]. This technique is used to reformulate certain ill-posed problems into well-posed problems. The ill-posed problems are those for which either 1) the solution exists but is not unique, or 2) the solution does not depend continuously on the input data. Regularization is typically formulated as an error minimization and involves a stabilizing functional that is applied to the input data and perhaps an additional smoothing parameter. Due to the seemingly infinite choice of possible stabilizing functions and smoothness parameters it is difficult to specify a best regularizing algorithm for an application.

VI. Computing Motion from a Sequence of Stereo Images

The technique described in the previous sections determine the motion and structure of an object given a sequence of monocular images of the scene. It was seen that in both the feature-based methods as well as in the optic flow techniques, the solutions for structure and motion remain ambiguous with respect to absolute value of distance between the camera and the scene. In other words, structure and motion parameters are unique only up to a scaling factor. The use of stereoscopy can provide this additional parameter to uniquely determine depth and hence absolute values for the structure and motion parameters.

The fusion of stereo and motion may be effected with different objectives in mind. Stereoscopic processing may be used to aid motion recovery, or conversely, motion analysis may be used to help establish feature correspondence in stereo image pairs. The fusion of these two processing modules in human and other biological visual systems has been detected via neurobiological and psychophysiological investigation [124], [125]. Recent research in both the feature-based and optic flow based approaches has addressed the fusion of stereoscopic analysis and motion estimation. We outline the salient features of such effort.

A. Feature Based Analysis

The overall analysis consists of the following steps: i) From the sequences of stereo images, the depth map for each stereo pair is determined, ii) the correspondence between three-dimensional features in successive depth maps is established, and iii) the motion of the objects is computed based upon the matched features. This formulation of motion analysis based on sequences of stereo images has several advantages and disadvantages which are briefly discussed below.

Kim and Aggarwal discuss the estimation of motion parameters from a sequence of depth maps extracted from stereo images [63]. The depth map for each stereo pair is computed using an edge-based stereo algorithm. 3-D features (consisting of lines and points) are extracted from each depth map. These features are matched between successive depth maps using a two pass relaxation process [61], [62]. In the process of extraction, search and matching, the search space is limited to the area of the motion in the image by an image differencing technique.

In general, correspondences between two 3-D lines extracted from one depth map and those from another may be used to determine the motion of a rigid object, assuming that the motion is small. Here, a three-dimensional line is specified by a three-dimensional direction and a point on the line. The same method can be used for three-dimensional point correspondences since two points determine a line. In general, three point correspondences, or one line correspondence and one point correspondence are sufficient to determine the three-dimensional motion parameters of a moving object. In the former case, the three points should not be collinear, and in the latter case, the point should not lie on the same line. A system of linear equations is derived and the solution is straightforward. A system based upon these observations has been implemented to derive the structure and the displacement of the objects between the views. In this study the motion of simple toy objects was estimated with excellent results [63].

Although it is theoretically quite easy to estimate the motion parameters given the correspondence between two sets of 3-D points, practical considerations complicate the implementation of the system. In stereo imagery, the range values estimated are subject to a great deal of uncertainty due primarily to quantization of disparity. More robust formulations of the problem of motion estimation using sequences of stereo images have been proposed [126]–[128]. One approach has been to estimate motion parameters via a system of linear equations using 3 points in each depth map [126]. Several sets of 3 points are chosen from the large number of available points and the motion parameters are computed for each set. For each set of computed motion parameters, all available points in the first depth map are subjected to the estimated motion. The discrepancy between the points in the second depth map and transformed points from the first depth map is computed via a simple distance measure. The set of estimated motion parameters that yields the lowest error is chosen. Although the solution of the system of linear equations is easy, the estimation of large sets of motion parameters and especially the search for the best set of motion parameters is computationally intensive.

An alternative approach has been to use a least-mean-squared error analysis [127], [129]. The underlying principle here is again the invariance of distance between points on an object subjected to rigid motion. The formulation is analogous to the approach followed by Magee and Aggarwal [130], [131] for determining motion parameters from sequences of range images. While the direct method of solution is adopted in [130], [131], a two-part iterative approach is adopted in [127]. The displacement between the centroids of two sets of registered 3-D points is used to determine the translation vector. The rotation matrix is decomposed into three factors corresponding to rotations about the z, x, and y axes. Each of these is individually solved for while the other two are fixed. This is repeated in a cyclic manner until a least mean squared error criterion is satisfied. The advantage of the above decomposition is that the 3-D estimation problem reduces to a set of 2-D problems which are more tractable.

The above approaches consider the determination of structure and motion as separate issues. Hence, if structure is first computed (as is usually the case for stereo imagery) then errors accrued due to quantization of disparity will continue to plague the estimation of motion. To alleviate this problem a new approach has been developed by Kiang, Chou and Aggarwal [132] based on iterative refinement of both structure and motion estimates. The approach is based on a 1-D model for triangulation error in stereoscopy. The strategy for modifying structure and motion estimates is based on the structural relationship between the corresponding uncertainty polyhedra in successive depth maps. Experimental results using synthetic as well as real data demonstrate significant improvement in the estimation of both structure and motion when compared to the conventional techniques based on reducing least-mean-squared error in motion alone.

Aloimonos and Rigoutsos [133] have developed a scheme for computing 3-D motion parameters from a sequence of stereo imagery which does not require *a priori* establishment of correspondences. The features extracted from the left and right images are assumed to lie on a planar surface $Z = pX + qY + c$. Perspective imaging geometry is assumed. The image planes are parallel to the $X - Y$ plane. The parameters p, q, and c are acquired by solving a set of linear equations in which the coefficient of each of the unknowns consists of a function of a sum of the image coordinates. The solution of the linear equation provides the structure of the scene. Applying this process before and after the planar surface undergoes motion allows for the estimation of the motion parameters. The method developed was not as robust as was expected and was modified by including a third camera. The performance of the algorithm in presence of noise is described in [133].

Another technique for estimating 3-D motion parameters from two 3-D point sets without establishing correspondence has been presented by Lin *et al.* [134]. The algorithm is based on the property that a function and its Fourier transform must experience the same rotation. The translation is first determined from the displacement of the centroid. Two functions are defined on the feature set. A correlation between the Fourier transforms of these functions is determined. The rotation axis and angle are computed based on this procedure. Some simulation results have been presented [134].

The above techniques are representative of the approaches wherein stereopsis aids the recovery of motion. There exist many reports in recent literature discussing the use of motion in recovering structure, e.g., Jenkin [135] used instantaneous velocities at feature points to aid the establishment of stereo correspondence, Nevatia [136], Mutch [137], Xu *et al.* [138], and Jain [139], among others, used known motion parameters to simulate stereo. We feel that although this approach is related to the estimation of motion, it is a separate field in itself. Hence, we do not pursue any further the discussion of the use of known motion to aid stereopsis, and we limit our discussion to the use of stereoscopy for estimation of motion.

B. Multiple Optic Flow Fields

In Section IV we discussed the interpretation of optic flow fields obtained from a sequence of monocular images. Another approach has been to compute multiple optic flow fields from different views, to establish correspondence between them and reconstruct 3-D velocity vector fields.

Mitiche [140] assumes that optic flow is computed for each view in a stereoscopic imaging system for which the stereoscopy parameters are known. He further assumes that correspondence between points in the two images are available which allows for the estimation of depth. Mitiche shows that given this information it is posssible to compute the 3-D motion parameters in a straightforward manner. Waxman and Sinha [141] have used a similar approach. In addition, they have filtered the optic flow field to minimize the effects of noise. Nagel [142] has also attempted such stereo-motion fusion techniques and has devised an approach based on the minimization of an error function. Tsukune and Aggarwal [98] have used this approach for reconstructing 3-D velocity fields for a scene containing multiple objects in motion.

Richards [143] demonstrated that the relative rate of change of disparity (ratio between temporal rate of change of disparity and disparity) due to object/camera motion is a useful aid in establishing feature correspondence within a pair of stero images. Waxman and Duncan [144] used the ratio between relative flow and disparity to aid the establishment of stereo correspondence. The relative flow is defined to be the difference between the optic flow at a point in the left image and that at the corresponding point in the right image. Waxman and Duncan show that their ratio is identical to the one devised by Richards [143].

VII. Conclusion

In this paper we have reviewed recently developed techniques for estimating structure and motion from sequences of monocular and stereoscopic images. We discussed two distinct approaches: feature-based analysis and optic flow techniques. We described some of the different mathematical formulations that have been developed for each of these tasks. A comparison of the feature-based and optic flow methods was then presented in which the relative merits and demerits of both approaches were discussed. An overview of the fusion of stereoscopy and motion analysis, especially for aiding the estimation of motion, was presented.

The optic flow approach consists of computing the two-dimensional field of instantaneous velocities of brightness values (gray levels) in the image plane. Instead of considering temporal changes in image brightness values in computing the optic flow field, it is possible to also consider temporal changes in values that are the result of applying various local operators such as contrast, entropy, and spatial derivatives to the image brightness values. In either case, a relatively dense flow field is estimated, usually at every pixel in the image. The optic flow is then used in conjunction with added constraints or information regarding the scene to compute the actual three-dimensional relative velocities between scene objects and camera.

The feature-based approach is based on extracting a set of relatively sparse but highly discriminatory set of two-dimensional features in the images corresponding to three-dimensional object features in the scene such as corners, occluding boundaries of surfaces, and boundaries demarcating changes in surface reflectivity. Such points, lines and/or curves are extracted from each image. Inter-frame correspondence is then established between these features. Constraints are formulated based on assumptions such as rigid body motion, e.g., the 3-D distance between two features on a rigid body remains the same after object/camera motion. Such constraints usually result in a system of nonlinear equations. The observed displacement of the 2-D image features are used to solve these equations leading ultimately to the computation of motion parameters of objects in the scene.

In the feature-based approach, the main problems encountered are seen to be: 1) establishing and maintaining correspondence between the image plane features, 2) robust formulation of the problem which is usually based on the assumption that the viewed object undergoes rigid motion, and 3) developing appropriate iterative algorithms which are stable and accurate. The optic flow based approach suffers from a different set of drawbacks, i.e., 1) it is highly noise sensitive due to its dependence on spatio-temporal gradients, 2) it requires that motion be smooth and small thus requiring a high rate of image acquisition, and 3) it requires that motion vary continuously over the image. Both approaches also are affected by object occlusion and choice of initial/boundary conditions. The use of sequences of stereoscopic images provides three-dimensional points and lines which somewhat simplify the problem of estimating motion.

A great deal of future research effort is warranted to overcome the obstacles mentioned above. The significant contributions made by various researchers in this area during the recent past is to be noted and this trend may be expected to continue in the future. Two workshops, one in Europe [145] and one in the USA [146] are planned in the near future to engender progress in this challenging area.

Acknowledgment

The authors wish to thank the reviewers, the editor Dr. H. Li, and Dr. W. Martin for their comments, which helped in improving this manuscript.

References

[1] J. K. Aggarwal and N. I. Badler, eds., *Abstracts for the Workshop on Computer Analysis of Time-Varying Imagery*, University of Pennsylvania, Moore School of Electrical Engineering, Philadelphia, PA, Apr. 1979.

[2] ——, "Special issue on motion and time-varying imagery," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, no. 6, Nov. 1980.

[3] W. E. Snyder, (Guest Ed.), "Special issue on computer analysis of time-varying images," *Computer*, vol. 14, no. 8, Aug. 1981.

[4] J. K. Aggarwal, (Guest Ed.), "Special issue on motion and time varying imagery," *Computer Vision Graphics and Image Processing*, vol. 21, nos. 1 and 2, Jan., Feb. 1983.

[5] T. S. Huang, Ed., *Image Sequence Analysis.* New York, NY: Springer-Verlag, 1981.

[6] S. Ullman, *The Interpretation of Visual Motion.* Cambridge, MA: MIT Press, 1979.

[7] NATO Advanced Study Institute on Image Sequence Processing and Dynamic Scene Analysis, *Advance Abstracts of Invited and Contributory Papers*, Braunlage, West Germany, June 21–July 2, 1982.

[8] Siggraph/Siggart Interdisciplinary Workshop on Motion: Representation and Perception, Toronto, Canada, April 4–6, 1983, and *Computer Graphics*, vol. 18, no. 1, Jan. 1984.

[9] International Workshop on Time-Varying Image Processing and Moving Object Recognition, Florence, Italy, May 1982.

[10] W. N. Martin and J. K. Aggarwal, "Dynamic scene analysis: A survey," *Computer Graphics and Image Processing*, vol. 7, pp. 356–374, 1978.

[11] H.-H. Nagel, "Analysis techniques for image sequences," in *Proc. IJCPR-78*, (Kyoto, Japan), pp. 186–211, Nov. 1978.

[12] J. K. Aggarwal and W. N. Martin, "Dynamic scene analysis," in *Image Sequence Processing and Dynamic Scene Analysis*, T. S. Huang, Ed. New York, NY: Springer-Verlag, 1983, pp. 40–74.

[13] J. K. Aggarwal, "Three-dimensional description of objects and dynamic scene analysis," in *Digital Image Analysis*, S. Levialdi, Ed. New York, NY: Pitman Books, Ltd., 1984, pp. 29–46.

[14] H.-H. Nagel, "What can we learn from applications?," in *Image Sequence Analysis*, T. S. Huang, Ed. New York, NY: Springer-Verlag, 1981, pp. 19–228.

[15] ——, "Overview on image sequence analysis," in *Image Sequence Processing and Dynamic Scene Analysis*, T. S. Huang, Ed. New York, NY: Springer-Verlag, 1983, pp. 2–39.

[16] T. S. Huang, Ed., *Image Sequence Processing and Dynamic Scene Analysis.* Berlin, West Germany: Springer-Verlag, Proceedings of NATO Advanced Study Institute at Braunlage, West Germany, 1983.

[17] IEEE Computer Society Workshop on Motion: Representation and Analysis, Kiawah Island, SC, May 1986.

[18] The 2nd International Workshop on Time-Varying Image Processing and Moving Object Recognition, Florence, Italy, September 1986.

[19] R. Chellappa and A. A. Sawchuk, Eds., *Digital Image Processing and Analysis: Volume 2: Digital Image Analysis.* New York, NY: IEEE Computer Society Press, 1985.

[20] W. Martin and J. K. Aggarwal, Eds., *Motion Understanding: Robot and Human Vision.* Norwell, MA: Kluwer Academic Publishers, 1988.

[21] H. Wallach and D. N. O'Connell, "The kinetic depth effect," *J. Exp. Psychol.*, vol. 45, pp. 205–217, 1953.

[22] E. J. Gibson, J. J. Gibson, O. W. Smith, and H. Flock, "Motion parallax as a determinant of perceived depth," *J. Exp. Psychol.*, vol. 8, pp. 40–51, 1959.

[23] S. Ullman, "The interpretation of structure from motion," in *Proc. R. Soc. London, B203*, pp. 405–426, 1979.

[24] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Perception and Psychophysics*, vol. 14, pp. 201–211, 1973.

[25] ——, "Spatio-temporal differentiation and integration in visual motion perception," *Psych. Res.*, vol. 38, pp. 379–383, 1976.

[26] ——, "Visual event perception," in *Handbook of Sensory Physiology*, R. Held, H. W. Leibowitz, and H.-L. Teuber, Eds. Berlin, West Germany: Springer-Verlag, 1978.

[27] G. Jansson and G. Johansson, "Visual perception of bending motion," *Perception*, vol. 2, pp. 321–326, 1973.

[28] G. Johansson, "Visual motion perception," *Sci. Amer.*, vol. 232, pp. 76–88, 1975.

[29] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat," *J. Neurophysiol.*, vol. 28, pp. 229–289, 1965.

[30] J. W. Roach and J. K. Aggarwal, "Computer tracking of objects moving in space," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-1, no. 2, pp. 127–135, Apr. 1979.

[31] ——, "Determining the movement of objects from a sequence of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, no. 6, pp. 554–562, Nov. 1980.

[32] H. H. Nagel, "Representation of moving rigid objects based on visual observations," *Computer*, pp. 29–39, Aug. 1981.

[33] R. Y. Tsai and T. S. Huang, "Estimating 3-D motion parameters of a rigid planar patch, I," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-29, no. 6, pp. 1147–1152, Dec. 1981.

[34] R. Y. Tsai, T. S. Huang, and W. L. Zhu, "Estimating three-dimensional motion parameters of a rigid planar patch, II: Singular value decomposition," *IEEE Trans. Acoust. Speech Signal Processing*, vol. ASSP-30, pp. 525–534, Aug. 1982.

[35] T. S. Huang and R. Y. Tsai, "Image sequence analysis: Motion Estimation," in *Image Sequence Processing and Dynamic Scene Analysis*, T. S. Huang, Ed. New York, NY: Springer-Verlag, 1981.

[36] R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surface," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, no. 1, pp. 13–26, Jan. 1984.

[37] H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, vol. 293, pp. 133–135, Sept. 1981.

[38] ——, "The reconstruction of a scene from two projections-configurations that defeat the 8-point algorithm," in *Proceedings of the First Conf. on Artificial Intelligence Application*, (Denver, CO), pp. 395–397, Dec. 5–7, 1984.

[39] X. Zhuang, T. S. Huang, and R. M. Haralick, "Two-view motion analysis: A unified algorithm," *J. Opt. Soc. Amer.*, vol. 3, no. 9, pp. 1492–1500, Sept. 1986.

[40] X. Zhuang and R. M. Haralick, "Two view motion analysis—Theory and algorithm," in *Proc. ICASSP*, Mar. 1985.

[41] ——, "Two view motion analysis," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 686–690, June 1985.

[42] O. D. Faugeras, F. Lustman, and G. Toscani, "Motion and structure from motion from point and line matches," in *Proc. 1st Int. Conf. Computer Vision*, (London, England), pp. 25–34, June 1987.

[43] H.-H. Hagel, "Image sequences—Ten (octal) years—From phenomenology towards a theoretical foundation," in *Proc. Int. Conf. of Pattern Recognition*, pp. 1174–1185, Oct. 1986.

[44] B. L. Yen and T. S. Huang, "Determining 3-D motion and structure of a rigid body using straight line correspondences," in *Image Sequence Processing and Dynamic Scene Analysis*, T. S. Huang, Ed. New York, NY: Springer Verlag, 1983.

[45] ——, "Determining 3-D motion/structure of a rigid body over 3 frames using straight line correspondences," in *Proceedings of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, (Washington, DC), pp. 267–272, June 19–23, 1983.

[46] Y. C. Liu and T. S. Huang, "Estimation of rigid body motion using straight line correspondences," in *Proc. IEEE Computer Society Workshop on Motion: Representation and Analysis*, pp. 47–52, May 1986.

[47] ——, "Estimation of rigid body motion using straight line correspondences: Further results," in *Proc. Int. Conf. of Pattern Recognition*, pp. 306–309, Oct. 1986.

[48] J. A. Webb and J. K. Aggarwal, "Structure and motion of rigid and jointed objects," *Artificial Intelligence*, No. 19, pp. 107–130, 1982.

[49] A. Mitiche, S. Seida, and J. K. Aggarwal, "Determining posi-

tion and displacement in space from images," in *Proceedings of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition,* (San Francisco, CA), pp. 504-509, June 19-23, 1985.

[50] ——, "Line-based computation of structure and motion using angular invariance," in *Proc. IEEE Computer Society Workshop on Motion*, pp. 175-180, May 1986.

[51] J. K. Aggarwal and Y. F. Wang, "Analysis of a sequence of images using point and line correspondences," in *Proceedings of the 1987 IEEE International Conf. on Robotics and Automation*, (Raleigh, NC), pp. 1275-1280, Mar. 31-Apr. 3, 1987.

[52] S. Ullman, "Maximizing rigidity: The incremental recovery of 3D structure from rigid and non-rigid motion," *Perception*, vol. 13, pp. 255-274, 1984.

[53] E. C. Hildreth and N. M. Grzywacz, "The incremental recovery of structure from motion: Position vs. velocity based formulations," in *Proc. IEEE Computer Society Workshop on Motion*, pp. 137-143, May 1986.

[54] T. J. Broida and R. Chellappa, "Estimation of object motion parameters from noisy images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 1, pp. 90-99, Jan. 1986.

[55] ——, "Kinematics and structure of a rigid object from a sequence of noisy images," in *Proc. IEEE Computer Society Workshop on Motion*, pp. 95-100, May 1986.

[56] J. Weng, T. S. Huang, and N. Ahuja, "3-D motion estimation, understanding and prediction from noisy image sequences," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 3, pp. 370-389, May 1987.

[57] J. K. Aggarwal, L. S. Davis, and W. N. Martin, "Correspondence processes in dynamic scene analysis," *Proc. IEEE*, vol. 69, no. 6, pp. 562-572, May 1981.

[58] S. K. Sethi and R. Jain, "Finding trajectories of feature points in a monocular image sequence," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 1, pp. 56-73, Jan. 1987.

[59] J.-Q. Fang and T. S. Huang, "Some experiments on estimating the 3-D motion parameters of a rigid body from two consecutive image frames," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 545-554, Sept. 1984.

[60] S. Ranade and A. Rosenfeld, "Point pattern matching by relaxation," *Pattern Recognition*, vol. 12, pp. 269-275, 1980.

[61] Y. C. Kim and J. K. Aggarwal, "Finding range from stereo images," in *Proceeding IEEE Conf. Computer Vision and Pattern Recognition*, (San Francisco, CA), pp. 289-294, June 1985.

[62] Y. C. Kim, "Structure and motion of objects from stereo images," Ph.D. Dissertation, Department of Electrical and Computer Engineering, The University of Texas at Austin, May 1986.

[63] ——, "Determining object motion in a sequence of stereo images," *IEEE J. Robotics Automat.*, vol. RA-3, no. 6, pp. 599-614, Dec. 1987.

[64] S. T. Barnard and W. B. Thompson, "Disparity analysis of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, pp. 333-340, July 1980.

[65] J. M. Prager and M. A. Arbib, "Computing the optic flow: The MATCH algorithm and prediction," *Computer Vision, Graphics and Image Processing*, vol. 24, pp. 271-304, 1983.

[66] M. Jenkin and J. K. Tsotsos, "Applying temporal constraints to the dynamic stereo problem," *Computer Vision, Graphics and Image Processing*, vol. 33, pp. 16-32, 1986.

[67] D. Marr, *Vision. New York, NY: Freeman, 1982.*

[68] L. Dreschler and H.-H. Nagel, "Volumetric model and 3-D trajectory of a moving car derived from Monocular TV frame sequence of a street scene," *Computer Graphics and Image Processing*, vol. 20, pp. 199-228, 1982.

[69] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.

[70] B. G. Schunck, "Image flow: Fundamentals and future research," in *Proc. of IEEE Conf. on Pattern Recognition and Image Processing*, pp. 560-571, 1985.

[71] B. K. P. Horn, *Robot Vision*. Cambridge, MA: MIT Press, 1986.

[72] D. H. Ballard and O. A. Kimball, "Rigid body motion from depth and optical flow," *Computer Vision, Graphics and Image Processing*, vol. 22, pp. 95-115, 1983.

[73] B. G. Schunck, "Image flow: Fundamentals and algorithms," in *Motion Understanding: Robot and Human Vision*, W. N. Martin and J. K. Aggarwal, Eds. Norwell, MA: Kluwer Academic Publishers, 1988.

[74] H.-H. Nagel, "Displacement vectors derived from second-order intensity variations in image sequences," *Computer Vision, Graphics and Image Processing*, vol. 21, pp. 85-117, 1983.

[75] W. E. Snyder, S. A. Rajala, and G. Hirzinger, "Image modeling, the continuity assumption and tracking," in *Proc. Int. Conf. of Pattern Recognition*, pp. 1111-1114, 1980.

[76] K. Prazdny, "A simple method for recovering relative depth map in the case of a translating sensor," in *Proc. of Int. Joint Conf. on Artificial Intelligence*, pp. 698-699, 1981.

[77] M. Yachida, "Determining velocity map by 3-D iterative estimator," in *Proc. of Int. Joint Conf. on Artificial Intelligence*, pp. 716-718, 1981.

[78] H.-H. Nagel, "On the estimation of optical flow: Relations between different approaches and some new results," *Artificial Intelligence*, vol. 33, pp. 299-324, 1987.

[79] H.-H. Nagel and W. Enkelmann, "Investigation of second-order gray value variations to estimate corner point displacements," in *Proc. Int. Conf. of Pattern Recognition*, (Munich, W. Germany), pp. 768-773, 1982.

[80] H.-H. Nagel, "Constraints for the estimation of displacement vector fields from image sequences," in *Proc. of Int. Joint Conf. on Artificial Intelligence*, pp. 945-951, 1983.

[81] R. M. Haralick and J. S. Lee, "The facet approach to optic flow," in *Proceedings of Image Understanding Workshop*, (Arlington, VA), pp. 84-93, 1983.

[82] O. Tretiak and L. Pastor, "Velocity estimation from image sequences with second order differential operators," *Proc. Int. Conf. of Pattern Recognition*, pp. 16-19, 1984.

[83] E. C. Hildreth, "Computations underlying the measurement of visual motion," *Artificial Intelligence*, vol. 23, pp. 309-354, 1984.

[84] W. Enkelmann, "Investigations of multigrid algorithms from the estimation of optical flow fields in image sequences," in *Proc. IEEE Computer Society Workshop on Motion: Representation and Analysis*, pp. 81-87, May 1986.

[85] F. Glazer, "Hierarchical motion detection," Ph.D. Dissertation, COINS Department, University of Massachusetts, Amherst, MA, Feb. 1987.

[86] P. Anandan, "A unified perspective on computational techniques for the measurement of visual motion," in *Proc. 1st Int. Conf. Computer Vision*, (London, England), pp. 219-230, June 1987.

[87] ——, "Computing dense displacement fields with confidence measures in scenes containing occlusion," in *Proc. DARPA Image Understanding Workshop*, (New Orleans, LA), pp. 236-246, 1984.

[88] A. Mitiche, Y. F. Wang, and J. K. Aggarwal, "Experiments in computing optical flow with the gradient-based, multiconstraint method," *Pattern Recognition*, vol. 20, no. 2, pp. 173-179, 1987.

[89] D. J. Fleet and A. D. Jepson, "Velocity extraction without form interpretation," in *Proceedings of the Third Workshop on Computer Vision: Representation and Control*, (Bellaire, MI), pp. 179-185, Oct. 1985.

[90] J. K. Tsotsos, D. J. Fleet, and A. D. Jepson, "Towards a theory of motion understanding in man and machine," in *Motion Understanding: Robot and Human Vision*, W. N. Martin and J. K. Aggarwal, Eds. Norwell, MA: Kluwer Academic Publishers, 1988.

[91] D. J. Heeger, "Optical flow from spatiotemporal filters," in *Proc. 1st Int. Conf. Computer Vision*, (London, England), pp. 181-190, June 1987.

[92] D. W. Murray and B. F. Buxton, "Scene segmentation from visual motion using global optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 2, pp. 220-228, Mar. 1987.

[93] W. B. Thompson, "Combining motion and contrast for segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, pp. 543-549, 1980.

[94] W. B. Thompson, K. M. Mutch, and V. A. Berzins, "Dynamic occlusion analysis in optical flow fields," *IEEE Trans. Pattern*

Anal. Machine Intell., vol. PAMI-7, no. 4, pp. 374-383, July 1985.

[95] J. O'Rourke, "Motion detection using Hough techniques," in *Pattern Recognition and Image Processing Conference*, (Dallas, TX), pp. 82-87, Aug. 3-4, 1981.

[96] C. L. Fennema and W. B. Thompson, "Velocity determination in scenes containing several moving objects," *Computer Graphics and Image Processing*, vol. 9, pp. 301-305, Apr. 1979.

[97] G. Adiv, "Determining three-dimensional motion and structure from optical flow generated by several moving objects," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 7, no. 4), pp. 384-401, 1985.

[98] H. Tsukune and J. K. Aggarwal, "Analyzing orthographic projection of multiple 3-D velocity vector fields in optical flow," *Computer Vison, Graphics and Image Processing*, in press.

[99] K. Prazdny, "Motion and structure from optical flow," in *Proc. of Int. Joint Conf. on Artificial Intelligence*, pp. 702-704, 1979.

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

[101] H. C. Longuet-Higgins and K. Prazdny, "The interpretation of a moving retinal image," *Proc. Roy. Soc. London*, vol. B208, pp. 385-397, 1980.

[102] A. M. Waxman, B. Kamgar-Parsi, and M. Subbarao, "Closed form solutions to image flow equations," in *Proceedings of the First Conf. on Artificial Intelligence*, (Denver, CO), pp. 12-23, 1984.

[103] A. M. Waxman and K. Wohn, "Image flow theory: A framework for 3-D inference from time-varying imagery," in *Advances in Computer Vision*, C. Brown, Ed. Hillside, NJ: Erlbaum Publishers, 1987.

[104] M. Subbarao, "Solution and uniqueness of image flow equations for rigid curved surfaces in motion," in *Proc. 1st Int. Conf. Computer vision*, (London, England), pp. 687-692, June 1987.

[105] —, "Interpretation of image motion fields: Rigid curved surfaces in motion," Tech. Rep. CAR-TR-199, Center for Automation Research, Univ. of Maryland, College Park, MD, April. 1986.

[106] S. Negahdaripour and B. K. P. Horn, "Determining 3-D motion of planar objects from image brightness measurements," in *Proceedings of the International Joint Conf. on Artificial Intelligence*, (Los Angeles, CA), pp. 898-901, Aug. 18-23, 1985.

[107] G. Adiv, "Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field," in *Proc. of IEEE Computer Vision and Pattern Recognition Conf.*, (San Francisco, CA), pp. 70-77, June 1985.

[108] T. D. Williams, "Depth from camera motion in a real world scene," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, no. 6, pp. 511-516, Nov. 1980.

[109] D. T. Lawton, "Processing translational motion sequences," *Computer Vision, Graphics and Image Processing*, vol. 22, pp. 116-144, 1983.

[110] J. H. Rieger and D. T. Lawton, "Determining the instantaneous axis of translation from optic flow generated by arbitrary sensor motion," in *Proc. ACM Interdisc. Workshop Motion*, (Toronto, Ont., Canada), pp. 33-41, 1983.

[111] K. Prazdny, "Determining the instantaneous direction of motion from optical flow generated by a curvilinearly moving observer," *Computer Graphics and Image Processing*, vol. 17, pp. 238-248, 1981.

[112] A. R. Bruss and B. K. P. Horn, "Passive navigation," *Computer Vision, Graphics and Image Processing*, vol. 21, no. 1, pp. 3-20, Jan. 1983.

[113] B. K. P. Horn and E. J. Weldon, "Computationally efficient methods for recovering translational motion," in *Proceedings of the First International Conf. on Computer Vison*, (London, England), pp. 2-11, June 8-11, 1987.

[114] S. Negahdaripour and B. K. P. Horn, "Direct passive navigation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 1, pp. 168-176, Jan. 1987.

[115] T.-C. Chou and K. Kanatani, "Recovering 3-D rigid motions without correspondence," in *Proc. 1st Int. Conf. Computer Vision*, (London, England), pp. 534-538, June 1987.

[116] J. K. Kearney, W. B. Thompson, and D. L. Boley, "Optical flow estimation: An error analysis of gradient-based methods with local optimization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 2, pp. 229-244, Mar. 1987.

[117] J. K. Kearney, "Gradient-based estimation of optical flow," Ph.D. Dissertation, University of Minnesota, 1983.

[118] A. Verri and T. Poggio, "Against quantitative optical flow," in *Proc. 1st Int. Conf. Computer Vision*, (London, England), pp. 171-180, June 1987.

[119] Y. Yasumoto and G. Medioni, "Robust estimation of three-dimensional motion parameters from a sequence of image frames using regularization," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 4, pp. 464-471, July 1986.

[120] T. E. Boult, "What is regular in regularization?," in Proc. 1st Int. Conf. Computer Vision, (London, England), pp. 457-462, June 1987.

[121] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, no. 6035, pp. 314-319, 1985.

[122] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 4, pp. 413-424, July 1986.

[123] A. Rougee, B. C. Levy, and A. S. Willsky, "Reconstruction of two-dimensional velocity fields as a linear estimation problem," in *Proc. 1st Int. Conf. Computer Vision,* (London, England), pp. 646-650, June 1987.

[124] G. F. Poggio and T. Poggio, "The analysis of stereopsis," *Ann. Rev. Neurosci.*, vol. 7, pp. 379-412, 1984.

[125] D. Regan and K. I. Beverly, "Binocular and monocular stimuli for motion in depth: Changing-disparity and changing-size feed the same motion-in-depth stage," *Vision Research*, vol. 19, pp. 1331-1342, 1979.

[126] T. S. Huang and S. D. Blostein, "Robust algorithms for motion estimation based on two sequential stereo image pairs," in *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition*, (San Francisco, CA), pp. 518-523, June 19-23, 1985.

[127] Z. C. Lin, T. S. Huang, S. D. Blostein, H. Lee, and E. A. Margerum, "Motion estimation from 3-D point sets with and without correspondences," in *Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition*, (Miami Beach, FL), pp. 194-201, June 22-26, 1986.

[128] Z. Lin, H. Lee, and T. S. Huang, "Finding 3-D point correspondences in motion estimation," in *Proc. 8th Int. Conf. Pattern Recognition*, (Paris, France), pp. 303-305, Oct. 1986.

[129] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least squares fitting of two 3-D point sets," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 5, pp. 698-700, Sept. 1987.

[130] M. J. Magee and J. K. Aggarwal, "Determining motion parameters using intensity guided range sensing," in *Proceedings of the 7th International Conf. on Pattern Recognition*, (Montreal, Canada), pp. 538-541, 1984.

[131] J. K. Aggarwal and M. J. Magee, "Determining motion parameters using intensity guided range sensing," *Pattern Recognition*, vol. 19, no. 2, pp. 169-180, 1986.

[132] S. M. Kiang, R. J. Chou, and J. K. Aggarwal, "Triangulation errors in stereo algorithms," in *IEEE Computer Vision Workshop,* (Miami Beach, FL), pp. 72-78, Dec. 1-2, 1987.

[133] J. Aloimonos and I. Rigoutsos, "Determining the 3-D motion of a rigid planar patch without correspondence, under perspective projection," in *Proc. IEEE Computer Society Workshop on Motion: Representation and Analysis*, pp. 167-174, May 1986.

[134] Z.-Ch. Lin, H. Lee, and T. S. Huang, "A frequency domain algorithm for determining motion of a rigid object from range data without correspondences," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 194-201, 1986.

[135] M. R. M. Jenkin, "The stereopsis of time-varying images," Tech. Rep. RBCV-TR-84-3, Dept. of Computer Science, Univ. of Toronto, Ontario, Canada, Sept. 1984.

[136] R. Nevatia, "Depth Measurement by motion stereo," *Computer Graphics and Image Processing*, vol. 5, pp. 203-214, 1976.

[137] K. M. Mutch, "Determining object translation information using stereoscopic motion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 6, pp. 751-755, Nov. 1986.

[138] G. Xu, S. Tsuji, and M. Asada, "A motion stereo method based on coarse-to-fine control strategy," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 2, pp. 332-336, Mar. 1987.

[139] R. Jain, S. L. Bartlett, and N. O'Brien, "Motion stereo using ego-motion complex logarithmic mapping," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, no. 3, pp. 356-369, May 1987.

[140] A. Mitiche, "On combining stereopsis and kineopsis for space perception," in *Proceedings of the First Conf. on Artificial Intelligence*, (Denver, CO), pp. 156-160, Dec. 1984.

[141] A. M. Waxman and S. Sinha, "Dynamic stereo: Passive ranging to moving objects from relative image flows," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 4, pp. 406-412, July 1986.

[142] H.-H. Nagel, "Dynamic stereo vision in a robot feedback loop based on the evaluation of multiple interconnected displacement vector fields," in *Proceedings of the International Symp. on Robotics Research*, pp. 200-206, 1985.

[143] W. Richards, "Structure from stereo and motion," *J. Opt. Soc. Amer.*, vol. 2, pp. 343-349, Feb. 1985.

[144] A. M. Waxman and J. H. Duncan, "Binocular image flows: Steps toward stereo-motion fusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 6, pp. 715-729, Nov. 1986.

[145] Third International Workshop on Time-Varying Processing and Moving Object Recognition, Florence, Italy, May 29-31, 1989.

[146] IEEE Computer Society Workshop on Visual Motion, Irvine, CA, Mar. 20-22, 1989.

J. K. Aggarwal (Fellow, IEEE) received the B.S. degree in mathematics and physics from the University of Bombay, in 1956, the B.Eng. degree from the University of Liverpool, England, in 1960, and the M.S. and Ph.D. degrees from the University of Illinois, Urbana, in 1961 and 1964, respectively.

He joined The University of Texas in 1964 as an Assistant Professor and has since held positions as Associate Professor (1968) and Professor (1972). Currently, he is the John J. McKetta Energy Professor of Electrical and Computer Engineering and Computer Sciences at The University of Texas at Austin. Further, he was a Visiting Assistant Professor at Brown University, Providence, RI (1968), and a Visiting Associate Professor at the University of California, Berkeley, during 1969-1970. He has published numerous technical papers and several books, *Notes on Nonlinear Systems* (1972), *Nonlinear Systems: Stability Analysis* (1977), *Computer Methods in Image Analysis* (1977), *Digital Signal Processing* (1979), and *Deconvolution of Seismic Data* (1982). His current research interests are image processing, computer vision, and parallel processing of images. He was co-Editor of the special issues on Digital Filtering and Image Processing (IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, March 1975) and on Motion and Time Varying Imagery (IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, November 1980) and editor of the two volume special issue of Computer Vision, Graphics and Image Processing on Motion (CVGIP, January and February 1983). Currently he is an Associate Editor of the journals *Pattern Recognition, Image and Vision Computing*, and *Computer Vision, Graphics and Image Processing* and IEEE EXPERT. Further, he is a member of the Editorial board of IEEE Press and the Editor of the IEEE Selected Reprint Series. He was the General Chairman for the IEEE Computer Society Conference on Pattern Recognition and Image Processing, Dallas, TX, 1981, and was the Program Chairman for the First Conference on Artificial Intelligence Applications sponsored by IEEE Computer Society and AAAI held in Denver, CO, 1984. In June 1987, he began his term as the Chairman of the Pattern Analysis Machine Intelligence Technical Committee of the IEEE Computer Society.

Dr. Aggarwal is an active member of IEEE, IEEE Computer Society, ACM, AAAI, The International Society for Optical Engineering, Pattern Recognition Society, and Eta Kappa Nu.

N. Nandhakumar (Member, IEEE) received the B.E. (Hons) degree in electronics and communication engineering from the P.S.G. College of Technology, University of Madras, India, in 1981, the M.S.E. degree in computer, information and control engineering from the University of Michigan, Ann Arbor, in 1983, and the Ph.D. degree in electrical engineering from The University of Texas at Austin in 1987.

He is currently a Research Associate at the Computer and Vision Research Center at The University of Texas at Austin and conducts research in the integration of diverse sensing modalities for computer vision. He is an active member of the IEEE, and as a student he held offices in the student chapters of the IEEE and the IE.