調査論文

이동 물체 인식을 위한 Optic Flow

正會員 李 鍾 秀*

Optic Flow for Motion Vision; Survey

Jong Soo LEE* *Regular Members*

요 약 Optic flow는 이동물체표면의 3차원 속도를 영상판(image plane)에 투영시킨 2차원 속도이다. 본 논문은 이동물체의 연속영상으로부터 optic flow를 구하는 기술들을 조사분석하고 정해진 optic flow로부터 물체의 인식및 3차원 속도를 결정하는 기술들을 논하였다. 연속영상으로부터 구해진 optic flow는 영상압축기술인 영상간 부호 (inter-frame image coding)에 해당되며 컴퓨터비전 시스템(computer vision system)에서 이동물체 인식에 사용된다.

ABSTRACT Optic flow is 2D velocity projected on the image plane of 3D velocity of a moving surface element. In this paper, we survey techniques computing optic flows from an image time sequence of moving objects and techniques determining 3D velocities and surface structures of the moving objects from the optic flows determined.

1. Introduction

When objects in the environment are illuminated, light is reflected from surfaces of the objects. The reflected light forms a densely structured optic array at a point of observation. The optic array may be thought of as a bundle of narrow cones of light with their apices at the point of observation and their bases at distinct surface elements. Each surface element may produce a distinct light cone from a different intensity and spectral composition of the light at each surface element [LEE (1980)].

For moving objects or the observation point in motion, a light cone from a surface element moves with some velocity causing change of the optic array continuously over time, giving rise to an optic flow field. Formally, the optic flow field is a field of 2D velocities projected on the image plane of 3D velocities of surface elements as shown in Fig 1. For convenience and realism, the optic flow field is described in terms of and extracted from the changing pattern of light incident on an image plane that intercepts the time-varying optic array.

Motion has been conceptualized classically as an inference from analyzing each static 2D brightness intensity pattern of image time sequence [HELMHOLTZ (1925)]. However, motion and structure can be perceived directly from the optic flow field as one psychological theory indicates [GIBSON (1950abc)]. In that theory, motion and change are basic for a vision system while a static intensity pattern is a rare thing.

In the following, we discuss techniques com-

* 한국전자통신연구소, 접속처리연구실
Communications processing section, Electronics and Telecommunications Research Institute
論文番號 : 86-01 (接受 1986. 2. 20)

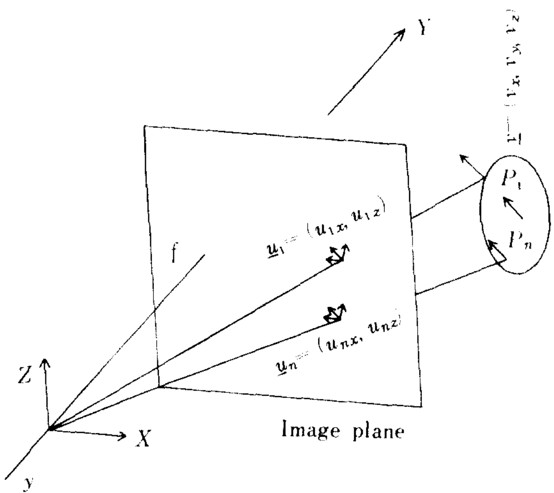

Fig. 1 Geometry of optic flow vector $\underline{u}$.

puting optic flow directly in section 2 and techniques recovering 3D motions and depth from optic flow in section 3. In section 4 we consider techniques determining corresponding points from which optic flows can be easily computed and in section 5 we show techniques determining 3D motion and depth from corresponding points. We conclude in section 6.

2. Computing Optice Flow

To compute the optic flow field directly from a time sequence of image frames, numerous techniques have been proposed. A technique can be assigned to one of two categories depending upon the requirement of a matching scheme. The so called 'non-matching' approaches, which do not require the correspondence process, determine two components of a flow vector based on the temporal and spatial brightness variation of the image and motion model. When a surface models an image brightness, the brightness variation at an image point can be represented by the surface gradient at the point. Thus, the 'non-matching' approaches are sometimes called gradient based techniques [ULLMAN (1981)].

In their experimental studies of moving television images, LIMB and MURPHY (1975) use frame brightness difference for the temporal brightness variation and consecutive row pixel brightness difference for the spatial brightness variation in estimating the horizontal component of a flow vector. These brightness differences deserve consideration in a mathematical model of image and motion for the optic flow computation [CAFFORIO and ROCCA (1976, 79), FENNEMA and THOMPSON (1979), THOMPSON and BARNARD (1981), HORN and SCHUNK (1981)].

A time-varying image can be represented mathematically by I (x,z,t) where x and z are the image plane coordinates and t is the time coordniate. Suppose that the image is displaced a distance δx in the x-direction and a distance δz in the z-direction in a time δt. Then the intensity at image point $(x+\delta x, z+\delta z)$ at time $t+\delta t$ can be represented using the Taylor series expansion as

$$
\begin{aligned}
I(x+\delta x,\, z+\delta z,\, t+\delta t) \\
= I(x, z, t) + \delta x \frac{\partial I}{\partial x} + \delta z \frac{\partial I}{\partial z} + \delta t \frac{\partial I}{\partial t} \\
+ \frac{\delta x^2}{2} \frac{\partial^2 I}{\partial x^2} + \frac{\delta x \delta z}{2} \frac{\partial^2 I}{\partial x \partial z} \\
+ \frac{\delta z^2}{2} \frac{\partial^2 I}{\partial z^2} + \frac{\delta z \delta t}{2} \frac{\partial^2 I}{\partial z \partial t} \qquad (1) \\
+ \frac{\delta t^2}{2} \frac{\partial^2 I}{\partial t^2} + \frac{\delta t \delta x}{2} \frac{\partial^2 I}{\partial t \partial x}
\end{aligned}
$$

where the partial derivatives are computed at the image point (x, z) at the time t either from a differentiable image model or using a finite difference method.

For the optic flow computation, some techniques consider only the linear and constant terms in (1) while others consider the second or higher terms as well. In the following two subsections, we will first discuss techniques using the linear and constant terms in (1) and then techniques requiring higher terms.

2-1. Linear Model of Image and Motion

Neglecting the second and higher order terms as the error measurement in (1), we have

$$I(x+\delta x,\ z+\delta z,\ t+\delta t) = I(x,z,t) + \delta_x \frac{\partial I}{\partial x} + \delta_z \frac{\partial I}{\partial z} + \delta_t \frac{\partial I}{\partial t} \tag{2}$$

Assuming that the brightness remains the same after the motion, we can set $I(x+\delta x,\ z+\delta z,\ t+\delta t) = I(x, z, t)$. Thus, we have

$$\frac{\delta x}{\delta t}\frac{\partial I}{\partial x} + \frac{\delta z}{\delta t}\frac{\partial I}{\partial z} + \frac{\partial I}{\partial t} = 0 \tag{3}$$

Setting $\delta t \rightarrow 0$, we have a linear relationship between the two optic flow components u and v as

$$u\frac{\partial I}{\partial x} + v\frac{\partial I}{\partial z} + \frac{\partial I}{\partial t} = 0 \tag{4}$$

where $u = \lim_{\delta t \to 0} \frac{\delta x}{\delta t}$ and $v = \lim_{\delta t \to 0} \frac{\delta z}{\delta t}$

This relationship is called the motion constraint or the optic flow equation [FENNEMA and THOMPSON (1979), THOMPSON and BARNARD (1981), HORN and SCHUNK (1981)].

Using the optic flow equation and assuming optic flow constancy over a moving image area, a flow vector can be computed [FENNEMA and THOMPSON (1979), THOMPSON and BARNARD (1981), HORN and SCHUNK (1981)]. CAFFORIO and ROCCA (1976) take the statistical analysis on the optic flow equation and estimate motion of multiple objects using a clustering scheme. CAFFORIO and ROCCA (1979) investigate the noise effect on the optic flow equation and consider a hardware implementation of motion estimation. FENNEMA and THOMPSON (1979) represent the optic flow equation in a polar coordinate system and determine the representative flow vector in a clustering scheme. THOMPSON and BARNARD (1981) derive a number of optic flow equations from the moving image area and determine the flow vector by solving the equations using the pseudo inverse method. They compare the pseudo inverse method to the clustering method. Further discussion about these two methods is in [DUBOIS et al. (1981), NAGEL (1981a)]. DAVIS et al. (1983) use local optic flow component constancy which follows from the local 2D object motion assumption. From this constraint as well as the optic flow equation, they propagate, using geometry, the optic flow along a line segment with two given flow vectors at both end points.

Since it cannot be said generally which pixels move with the same velocity, HORN and SCHUNK (1981) use the global constraint that the estimated optic flow field should vary smoothly with the image plane coordinates. They combine the smoothness constraint and the motion constraint by a weighted sum of their error measurements and apply a minimization technique to the summed error to derive an iterative formula. The global smoothness constraint, however, causes unrealistic optic flows on stationary image areas or on the background across the occluding boundaries of moving areas.

To cope with this difficulty, YACHIDA (1983) assumes optic flows at prominent feature points like corners such as at the points A, B, ..., G in Fig 2 and propagates the flow vectors smoo-

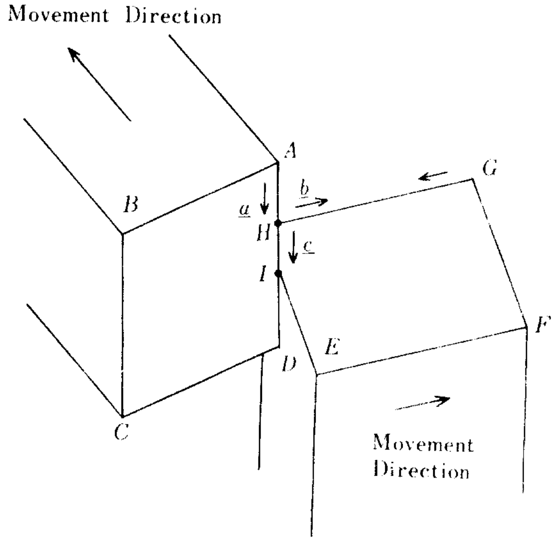

Fig. 2 Propagation of optic flow vectors along the edges.

thly along the connecting edge points between the corners. He computes the smooth flow vectors by the same equations derived by HORN and SCHUNK (1981) and their error measurements for termination of the propagation process. He mentions a difficulty in propagating the optic flow at the intersection (e.g. the point H in Fig 2) of an occluding and an occluded boundary. He solves the problem by terminating the propagation process when the error of newly estimated flow vector is equal to or larger than the previously estimated one. In Fig 2, the propagation of flow vectors to b direction is terminated by the propagation from d direction at a point between the points H and G.

Some researchers use additional constraints for the optic flow computation, but these are not absolutely necessary. Some researchers [NETRAVALI and ROBBINS (1979, 80), PAQUIN and DUBIOS (1983)], use the motion constraint only, with frame-to-frame prediction, to derive an iterative estimator by minimizing an error arising from the prediction. They simplify the estimator to use the optic flow (displacement vector) for encoding TV signals in real time. Using the optic flow equation as the linear measurement, the displacement vector is estimatated as a random process by applying Kalman filter theory [STULLER and KRISHNAMURPHY (1983)].

Instead of imposing constraints on optic flows, we can derive additional constraints by differentiating the linear model of image and motion. Specifically, we can differentiate the linear model of (2) with respect to x, producing

$$0 = \delta x \frac{\partial^2 I}{\partial x^2} + \delta z \frac{\partial^2 I}{\partial x \partial z} + \delta t \frac{\partial^2 I}{\partial x \partial t} \tag{5}$$

Dividing both sides of (5) by δt and setting $G = \partial I / \partial x$, we have, for $\delta t \rightarrow 0$,

$$u \frac{\partial G}{\partial x} + v \frac{\partial G}{\partial z} + \frac{\partial G}{\partial t} = 0 \tag{6}$$

Thus we derive another constraint (6) in the same form as the optic flow equation for the two optic flow components u and v. Although we use the x-directional derivative for G, any directional derivative can be used.

WOHN et al. (1983) use the gradient directional derivative and some other differential image features (e.g. curvature and moment) for G. However, feature estimation requiring high-order spatial derivatives is not reliable since the finite difference operation for the derivatives tends to amplify the noise. Thus, their estimated optic flows are often inaccurate. Such flow vectors are enhanced by a smoothing scheme which they devised with local 2½ D rigid motion to smooth those flow vectors which do not cross occluding boundaries.

Let P and Q be two optic flow image points where the optic flow vector is given by v and v', respectively. In the 2½ D rigid motion, they can compute the rotational velocity ω_q of a point Q with respect to a point P by

$$\omega_q = \frac{\triangle r \times \triangle v}{||\triangle r||^2} \tag{7}$$

and they define the dilation D_q by

$$D_q = \frac{\triangle r \cdot \triangle v}{||\triangle r||^2} \tag{8}$$

where

$$\triangle v = v' - v$$
$$\triangle r = \underline{Q} - \underline{P}$$

Thus for each pixel in a neighborhood, they can compute rotational velocity and dilation with respect to its center pixel position. For the center pixel, they average the computed values. To smooth a flow vector at a point P, they choose its neighbor P′ which has the minimum variation of computed rotational velocities and dilations in its neighborhood and update the flow vector v_P at P by

$$v = v' - \triangle r \times \omega' \tag{9}$$

where ω' is the computed rotational velocity at the point P′. In Fig 3, the flow vector at P is recomputed with respect to P′ because the 3x3 neighborhood of P′ has minimum variation of rotational velocities and dilations.

Similar multiple constraints are derived by TRETIAK and PASTOR (1984). They derive two independent equations directly from the motion constraint by differentiating the optic flow equation with respect to the image plane coordinates. The same two equations, however, can be derived from the linear model of image and motion.

Recently SCHUNK (1984) proved that the

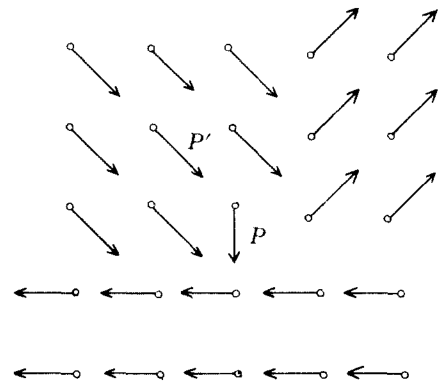

Fig. 3 Updating Optic Flow Vector at P with respect to Optic Flow Vector at P'.

motion constraint is valid at the image points of intensity discontinuities at an occluding boundary and pointed out that rotational motion or perspective projection may cause a change in the brightness pattern of an image sequence due to the forshortening effects.

2-2. Nonlinear Model of Image and Motion

The linear model of intensity variation is too simple to represent edges and corners [SNYDER et al. (1980), NAGEL (1983)]. Thus, SNYDER et al. (1980) represent (1) up to the second order at the time $t + \delta_t$ as

$$\begin{aligned} I(x+\delta x, z+\delta_z, t+\delta t) &= I(x, z, t+\delta t) + \delta x \frac{\partial I}{\partial x} + \delta_z \frac{\partial I}{\partial z} \\ &+ \frac{\delta x^2}{2}\frac{\partial^2 I}{\partial x^2} + \frac{\delta x \delta z}{2}\frac{\partial^2 I}{\partial x \partial z} + \frac{\delta z^2}{2}\frac{\partial^2 I}{\partial z^2} \end{aligned} \tag{10}$$

where the partial derivatives are computed at the image plane coordinate (x, z) at the time Setting $(x+\delta x, z+\delta z, t+\delta t) = I(x, z, t)$.

(brightness constancy after displacement), they derive

$$I(x, z, t) - I(x, z, t+\delta t) = \delta x \frac{\partial I}{\partial x} + \delta z \frac{\partial I}{\partial z} + \frac{\delta x^2}{2}\frac{\partial^2 I}{\partial x^2} + \frac{\delta x \delta z}{2}\frac{\partial^2 I}{\partial x \partial z} + \frac{\delta z^2}{2}\frac{\partial^2 I}{\partial z^2} \quad (11)$$

For each pixel, they derive one quadratic equation (11) in the two unknown displacement vector components δx and δz. Thus, with the assumption of displacement vector constancy over a moving image area (i.e. $I(x, z, t) = I(x, z, t+\delta t)$) they can determine the displacement vector using clustering.

NAGEL (1983) fits the image model of (10) over a local neighborhood such that the partial derivatives are substituted for the fitting coefficients minimizing the sum of weighted squared differences between the measurement and the image model. Using the determined fitting coefficients and the proper transformation of the local coordinate system, he derives two coupled nonlinear equations for the two displacement vector components from (11) by minimizing the sum of the squared differences between the measurement and the approximating equation.

For the corner points defined by DRESCHLER and NAGEL (1982) or KITCHEN and ROSENFELD (1980), which are shown to be equivalent in [NAGEL (1983)], the two nonlinear equations can be further simplified into closed forms. This method is implemented and applied to a TV frame sequence and the results, which show sparse displacement vectors computed, are reported in [NAGEL and ENKELMANN (1982)]. To overcome the sparseness, NAGEL (1983) combines the above minimization problem with a global constraint called the 'oriented smoothness' constraint to estimate displacements for image area with minor intensity variation. This constraint imposes smooth variation of displacements only in the direction perpendicular to the gradient direction, perpendicular to the gradient direction allowing significant displacement change in other directions. Since the intensity gradient direction usually occurs across an occluding boundary, the above constraint allows significant displacement change across such a boundary. NAGEL and ENKELMANN (1984) implement the scheme and report preliminary results.

A higher order Taylor expansion than those in (1) has been used for the optic flow computation by HARALICK and LEE (1983). They fit a polynomial function to intensities in a local 3D neighborhood consisting of two image coordinates and one coordinate for time. They determine optic flow from the fit by locating the point which matches up with the relative origin.

3. From Optic Flow to 3D Motion and Depth

Early systems of moving image analysis concentrated on the extraction of 2D motion information from an image sequence. LEESE et al. (1970) and ENDLICH et al. (1971) use a matching scheme to automatically determine the cloud movement on the image plane from satellite images of clouds. AGGARWAL and DUDA (1975) analyze moving polygonal images to determine motion of polygone and to segment the images. With the help of a moving object model, CHOW and AGGARWAL (1977) analyze planar curvilinear motion. POTTER (1975) uses the 2D motion information from the 'cross-shaped' template for image segmentation.

From the psychophysical and biological study of a vision system perceiving three dimensional objects, researchers began to concentrate on the 3D scene analysis containing moving objects. Some researchers identify moving elements (tokens) from changing brightness pattern to get motion and structural information of a moving object. From the brightness difference of two image frames, LILLESTRAND (1972), ULSTAD (1973), LIMB and MURPHY (1975) obtain motion information and JAIN et. al. (1977, 79), JAIN and NAGEL (1979ab), JAIN (1981), YALAMANCHILI et al. (1982) extract structural information of moving objects as well as their motion information. This differencing technique is further developed for image segmentation and motion information extraction [JAIN et al. (1979), JAIN and NAGEL (1979a, b), JAIN (1981), YALAMANCHILI et al. (1982). This differencing technique is referred to as a peripheral process in the survey by MARTIN and AGGARWAL (1978) where the motion understanding is described on three levels, peripheral, attentive, and cognitive.

Moving object boundaries can be extracted directly from an image time sequence. HAYNES and JAIN (1983) define the time-varying 'edginess' as the product of the temporal brightness change and the static 'edginess' of the Sobel edge detector. They detect the moving object boundaries by thresholding the time-varying 'edginess'. SHAH and JAIN (1984) detect the time-varying corners in a similar way.

THOMPSON (1980) segments images containing multiple moving objects based on both the motion and the brightness. He assumes brightness motion constancy at image points from the same object and uses the motion constraint described in the previous section for the Hough transform to determine dominant motions. He segments images by merging regions having the same motion and similar brightness. Motion constancy is not realistic in general, especially when the motion involves some rotation.

In a restricted scene domain consisting of only vertical and horizontal surfaces, WILLIAMS (1980) searches over distances from the camera to the vertical surfaces and heights of horizontal surfaces which result in the best image model of one frame from the other.

Another approach which GIBSON (1950 abc) pioneered derives 3D motion and structure information from the optic flow field obtained. GORDON (1965) shows an optic flow field of the ground plane observed by a moving observer and discusses the motion and depth perception from the optic flow field.

Optic flow field from a pure translational motion contains information about moving object surface structure as well as 3D motion. NAKAYAMA and LOOMIS (1974) show that relative depth can be determined from optic flow of an observer in translational motion. CLOCKSIN (1980) derives a mathematical formula for the surface orientation and the edge discrimination from optic flows generated by a pure translational motion. RIEGER and LAWTON (1983) employ an optimization procedure on thresholded difference vectors at discontinuities of the optic flow field to determine the instantaneous axis of translational motion.

Many researchers study optic flow field from general motion. KOENDERINK and VAN DOORN (1976) discuss the surface structure from a differential geometric analysis of optic flow field. PRAZDNY (1980, 81) studies the curvilinear motion of an observer and computes the egomotion parameters from at least five optic flow points. The computation involves solving a system of nonlinear (third order polynomial) equations for the rotational components. He also computes the depth map from the motion parameters determined. LONGUET-HIGGINS and PRAZDNY (1980) assume the first and second order spatial derivatives of the optic flow field are observable and compute the surface orientation and relative motion at each retinal point. They also show that the dilation, shear, and vorticity components of the first derivative can be used for the above 3D information.

BRUSS and HORN (1983) measure the discrepancy between the observed optic flow and the one expected by the computed motion parameters and use a least square technique to determine the motion parameters minimizing the discrepancy. In the case of pure translational and rotational motion, they derive equations in closed form for the motion parameters from the use of a proper error norm. For the general motion, they come up with a nonlinear system for the motion parameters.

Given depth and optic flow at a point, BALLARD and KIMBALL (1983) compute 3D optic flow at the point and use it for the determination of motion parameters in decomposed parameter spaces. They assume a small external force on moving objects and compute the acceleration from 3D flows.

4. Determining Corresponding Points

Optic flow can be computed without difficulty from the corresponding points in successive image frames. However, establishing the correspondence is generally recognized as a quite difficult task from the consideration of various factors [AGGARWAL et al. (1981)]. Marr and co-workers [GRIMSON and MARR (1979), MARR and HILDRETH (1980), MARR and POGGIO (1979) study the correspondence problem in human stereo vision and develop a computational theory of human vision. ESKENAZI and CUNNINGHAM (1978) and MORAVEC (1979ab) derive a set of corresponding points from two cameras in a stereo arrangement. However, stereo vision requires extremely high resolution of the cameras to get an acceptable 3D description. AGGARWAL et al. (1981) consider two general approaches to the correspondence problem in their review; one based on iconic or picture-like models of an image pattern and the other based on the structure (identifiable tokens) present in the image pattern. In the next subsection, we will discuss matching techniques using a template or an iconic model followed by token matching techniques.

4-1. Template Matching

The first approach usually takes a subimage (a template) containing a moving object from one frame and then matches it against the subsequent frame based on the chosen iconic representation. The iconic representation can be a portion of the image itself, a segmented binary image, or the edge image. Similarity (Dissimiliarity) is measured at each possible match and the match having the maximum (minimum) cross-correlation (absolute difference sum) is considered to be the best match. An equivalent process can be carried out through the use of the Fourier Transform of the iconic representation [LEESE et al. (1970)].

This search process usually requires much computation. To speed up the search process, several schemes have been introduced. BARNEA and SILVERMAN (1972) introduce a class of fast algorithms for image registration which speed up the search by essentially doing no calculation for similarity measure on positions when errors exceed the best match so far. VANDERBURG and ROSENFELD (1977) employ a two-stage template matching scheme in which, at first, a subtemplate is selected and used to determine positions which, result in a (dis) similarity measure above (below) a specified threshold. Next, the remainder of the template is evaluated at those chosen positions for the best match. Two difficulties in the two-stage matching scheme are in selecting the subtemplate and determining a proper threshold value of (dis) similarity measure. HALL et al. (1980) discuss the selection of the best subset for matching purposes in the sense of the minimum correlation length. GOSHTASBY et al. (1984) analytically derive the threshold value from the template size and false dismissal probability. Another scheme for the fast search is the coarse-fine matching scheme [GOSHTASBY et al. (1984)] which uses a similar strategy.

To establish point correspondence in multiple views at the specified positions, TSAI (1983) presents two methods called the joint moment method and the window variance method which result in a sharper similarity measure than the usual two frame cross-correlation. These methods, however, usually require more computation.

The techniques described up to now work only for image translation. Motion which produce a scaling or rotation of images can introduce much more computation [ALTMANN and REITBOCK (1984)].

4-2. Token Matching

Instead of matching picture-like models, one can identify a set of structures in one image frame and then search the corresponding structures in the subsequent image frame. Several researchers consider the object boundary structure for the match. POTTER (1975) generates a 'cross-shaped template' at a point by calculating the distance from the point to the closest object boundary in the row and column directions, and searches the corresponding point, in the next frame, at which the same size of the template is computed. MARTIN and AGGARWAL (1978) segment boundaries detected from both image frames into primitives such as straight lines and circular arcs and match a primitive of one frame against those of the other frame based on the length and curvature of the primitives. Using the generalized Hough Transform of an object boundary in one image frame, one can find the instance of that boundary in the subsequent frame [BALLARD (1981)].

Locating distinctive features present in images and imposing constraints on them, one can establish the correspondence of the feature points. ENDLICH et al. (1971) locate the brightness center by an ISODATA technique in successive image frames and establish the correspondence between sets of the brightness centers by assuming constant brightness motion. RENEDE and ROSENFELD (1980) apply a relaxation algorithm to establish the correspondence between sets of points selected by hand from two image frames with the assumption of the global brightness displacement constancy. BARNARD and THOMPSON (1980) locate the feature points in an image pair by applying the Moravec interest operator [MORAVEC (1977)] and assign an initial match probability to each pair of feature points which lie within some specified distance of each other allowing for the possibility that the feature may not exist in the second image. They refine the probability for each pairing at a feature point iteratively by applying a relaxation algorithm such that the probability for a pairing is updated by the supporting evidence of similar pairings in the local neighborhood. DRESCHLER and NAGEL (1982) make some modifications in updating the probability by considering both evidence of the support and contradiction in the local neighborhood. At each point, the pairing assigned the highest probability is selected for the correspondence.

ULLMAN (1979) introduces a simple cost function to compare possible matches between two dot patterns and establishes the correspondence by minimizing the cost function. He shows that minimizing the cost function is optimal if the dots in the pattern move independently. PRAGER and ARBIB (1983) assume several feature types on dot patterns and they use a feature metric between feature types to define a distance between any two position-feature pairs. Using this distance, they define an attraction function for a nearest neighbor. With the help of the attraction function and with the assumption of local consistency of displacement, they can establish the correspondence.

Some other structures present in images can be used. AYALA et al. (1982) segment image frames and establish correspondence between sets of segments based on the similarity of the segment feature values. They use features such as segment position, size, intensity, and area ratio. Since incorrect image segmentation from noise and occlusion introduces incorrect feature values, ROACH and AGGARWAL (1979) suggest using multiple methods arranged in a hierarchy. JACOBUS et al. (1980) encode boundary, region, and surface information in a graph form using a primitive called the half chunk and use a graph matching scheme to establish the correspondence.

The problem with most of these techniques is that they must employ some combinatorial computation to establish the match. This kind of computation is very expensive. Furthermore, these techniques require preprocessors to extract proper structures from image frames.

5. From Corresponding Points to 3D Motion and Depth

To determine 3D motion and structure of moving objects, a set of corresponding points can be used. ROACH and AGGARWAL (1980) require two known perspective views of five non-coplanar points in space to set up the problem and come up with 18 nonlinear equations in 18 unknowns. They solve the system of nonlinear equations iteratively with an initial guess for each unknown by using a modified finite difference Lavenberg-Marquetdt algorithm [BROWN and DENNIS (1972), LAVENBERG (1944)]. From experiments with noisy views of 3D points, they find that they obtain more accurate results with considerable overdetermination (two views of 12 or 15 points, three views of 7 or 8 points).

When the surface structure of a moving object is known, the motion estimation problem can be formulated differently. For a known planar surface, TSAI and HUANG (1981) define eight pure parameters which characterize the planar surface and can be uniquely determined by a linear method given two perspective views of four or more points. They derive a sixth order polynomial of one variable whose coefficients are expressed by the pure parameters and determine the motion parameters by solving the polynomial. Instead of solving the sixth order polynomial, the motion parameters can be determined by computing the singular value decomposition of a 3x3 matrix containing the pure parameters TSAI et al. (1982). For a known curved surface, TSAI and HUANG (1984) show that two perspective views of seven points on the curved surface are sufficient to determine the 3D motion parameters. YEN and HUANG (1983) project image points centrally on the unit sphere at the origin and use the projected points in correspondence to determine the 3D motion parameters and their 3D positions. They analyze three types of motions; pure translation, pure rotation, and general motion, and interpret the uniqueness of the motion solution by a simple geometry of the points on the sphere. Unfortunately, we can not find any experimental results, especially of noisy views, in their paper.

NAGEL (1981b) uses two coordinate systems; one attached to the camera and the other to a moving object. He derives a relationship between known image point coordinates and unknown object point coordinates and sets up a minimization problem from the relationship and sets of image points in correspondence. DRESCHLER and NAGEL (1982) and NAGEL (1983) apply this technique to a sequence of feature (corner point) images obtained from a street scene containing a moving taxicab and approximate the cab by a polyhedra.

MARTIN and AGGARWAL (1983) assume orthograhic projection of images and develop a scheme describing a 3D object from multiple views of its occluding object boundary. They obtain the object boundaries from the thresholded brightness image. Their 3D description is a bounding volume approximation (to the actual object) represented by a volume segment data structure. AGGARWAL (1983) reviews techniques of deriving 3D descriptions from image sequences.

6. Conclusion

Two categories of techniques computing optic flows have been discussed. The one which does not require the correspondence process computes an optic flow vector based on the temporal and spatial brightness variation of the image and motion model. The other one determine the optic flow from the corresponding points in successive image frames. However, establishing the correspondence is quite a difficult task dealing with massive amounts of image data.

Techniques extracting information (3D velocity and surface structure) from either optic flows or corresponding points usually set up and solve a minimization problem which is formulated for the motion or structure information using the relationship derived.

References

AGGARWAL and DUDA (1975) DUDA
Computer Analysis of Moving Polygonal Images, J.K. Aggarwal and R.O. Duda IEEE Trans. Computers, vol. C-24, Oct. 1975, pp. 966-976.

AGGARWAL et al. (1981)
Correspondence Processes in Dynamic Scene Analysis, J.K. Aggarwal, L.S. Davis, and W.N. Martin Proceedings of the IEEE. vol. 69, no. 5, May 1981, pp. 562-572.

AGGARWAL (1983)
Three Dimensional Description of Objects and Dynamic Scene Analysis, J.K. Aggarwal TR-83-1-20, The University of Texas at Austin, Laboratory for Image and Signal Analysis, January 1983.

ALTMANN and REITBOCK (1984)
A Fast Correlation Method for Scale-and Translation-Invariant Pattern Recognition, D. J. Altmann and H. J. P. Reitbock IEEE on PAMI, vol. PAMI-6 no. 1, Jan., 1984, pp. 46-57.

AYALA et al. (1982)
Moving Target Tracking Using Symbolic Registration I.L. Ayala, D.A. Orton, J.B. Larson, and D.F. Elliott IEEE Trans. on PAMI, vol. PAMI 4, no. 5, Sep., 1982, pp. 515-520.

BALLARD (1981a)
Generalizing the Hough Tr.nsform to detect Arbitrary Shape D.H. Ballard Pattern Recognition, vol. 13 no. 2, 1981, pp. 111-122

BALLARD (1981b)
On Shapes, D.H. Ballard, Proc., 7th IJCAI, Vancouver, B.C., Aug., 1981.

BALLARD and KIMBALL (1983)
Rigid Body Motion from Depth and optic Flow, D.H. Ballard and O.A. Kimball, Computer Vision, Graphics, and Image Processing, vol. 22, 1983, pp. 95-115

BARNARD and THOMPSON (1980)
Disparity Analysis of Images, S.T. Barnard and W. B. Thompson IEEE Trans. on PAMI, vol. PAMI-2, no. 4, July 1980, pp. 333-340.

BARNEA and SILVERMAN (1972)
A Class of Algorithm For Fast Image Registration, D. I. Barnea and H.H. Silverman IEEE Transactions On Computers, vol. C-21, no. 2, 1972. pp. 179-186.

BROWN and DENNIS (1972)
Derivative Free Analogues of the Lavenberg-Marquardt and Gauss Algorithms for Nonlinear Least Squares, K. M. Brown and J. E. Dennis, Numer. Math., vol. 18, 1972, pp. 289-297.

BRUSS and HORN (1983)
Passive Navigation, A.R. Bruss and B.K.P. Horn Computer Vision, Graphics, and Image Processing, vol. 21, 1983, pp. 3-20.

CAFFORIO and ROCCA (1976)
Methods for Measuring Small Displacements of Television Images, C.Cafforio and F. Rocca, IEEE Trans. Information Theory, vol. 22, no. 5, Sept. 1976, pp. 573-579.

CAFFORIO and ROCCA (1979)
Tracking Moving Objects in TV Images C.Cafforio

and F. Rocca Signal Process. 1, 1979, pp. 133-140

CHOW and AGGARWAL (1977)
Computer Analysis of Planar Curvilinear Moving Images, W. K. Chow and J. K. Aggarwal, IEEE Trans. Computers, vol. C-26, Feb. 1977, pp. 179-185.

CLOCKSIN (1980)
Perception of Surface Slant and Edge Labels from Optic Flow, W. F. Clocksin Perception vol. 9, 1980, pp. 253-269.

DAVIS et al. (1983)
Contour-Based Motion Estimation, L. S. Davis, Z. Wu, and H. Sun Computer Vision, Graphics, and Image Processing, vol 23, pp. 313-326.

DRESCHLER and NAGEL (1982)
Volumetric Model and 3D-Trajectory of a Moving Car Derived from Monocular TV-Frame Sequences of a Street Scene, L. Dreschler and H.-H. Nagel, Computer Graphics and Image Processing, vol. 20(3) 1982, pp. 199-228.

DUBOIS et al. (1981)
Image Sequence Coding, E. Dubois, B. Prasada, and M. S. Sabri, Image Sequence Analysis, T. S. Huang edt., pp. 229-287, Springer-Verlag, Berlin/Heidelberg /Newyork, 1981.

ENDLICH et al. (1971)
Use of a Pattern Recognition Technique for determining Cloud Motions from Sequences of Satellite Photographs, R. M. Endlich, D. E. Wolf, D. J. Hall, and A. E. Brain J. Appl. Meteorol., vol. 10, Feb. 1971, pp. 105-117.

ESKENAZI and CUNNINGHAM (1978)
Real Time-Tracking of Moving Objects in TV Images R. Eskenazi and R. T. Cunningham IEEE Workshop Pattern Recognition and Artificial Intelligence, Princeton, N. J., Apr., 1978, pp. 4-6.

FENNEMA and THOMPSON (1979)
Velocity Determination in Scenes Containing Several Moving Objects, C. L. Fennema and W. B. Thompson Computer Graphics and Image Processing, vol. 9, Apr., 1979, pp. 301-315.

GIBSON (1950a)
The Perception of the Visual World. J. J. Gibson Houghton Mifflin, Boston, 1950.

GIBSON (1950b)
The Senses Considered as Perceptual System, J. J. Gibson Houghton Mifflin, Boston, 1950.

GIBSON (1950c)
The Ecological Approach to Visual Perception J. J. Gibson. Houghton Mifflin, Boston. 1950.

GORDON (1965)
Static and Dynamic Visual Fields in Human Space Perception D. A. Gordon, Journal of the Optic Society of America vol. 55 no. 10, Oct., 1965, pp. 1296-1303.

GOSHTASBY et al. (1984)
A Two-Stage Cross Correlation Approach to Template Matching, A. Goshtasby, S. H. Gage, and J. F. Bartholic IEEE Trans. on PAMI-6 no. 3, May 1984

GRIMSON and MARR (1979)
A Computer Implementation of a Theory of Human Stereo Vision, W. E. L Grimson and D. Marr, Proc. Image Understanding Workshop, (Baumann, L. S. etd.) Palo Alto, Calif, Apr., 1979, pp. 41-47.

HALL et al. (1980)
The Selection of Critical Subsets for Signal, Image, and Scene Matching, E. L Hall D. L. Davies, and M. E. Casey IEEE Trans. on PAMI, vol. PAMI-2 no. 4, July 1980, pp. 313-322.

HARALICK and LEE (1983)
Facet Approach to Optic Flow, R. M. Haralicj and J. S. Lee DARPA Image Understanding Workshop. ARlington VA, June 1983.

HAYNES and JAIN (1983)
Detection of Moving Edges S. M. Haynes and R. Jain Computer Vision, Graphics, and Image Processing, vol. 21, 1983, pp. 345-367

HELMHOLTZ (1925)
The Treatise on Physiological Optics, H. von Helmholtz J. P. C. Southhall edt., New York Dover, 1925.

HORN and SCHUNK (1981)
Determining Optic Flow, B. K. P. Horn and B. G. Schunk, Artificial Intelligence, no. 17, 1981, pp. 185-203.

JACOUBS et al. (1980)
Motion Detection and Analysis of Matching Graphs of Intermediate-Level Primitives, C. J. Jacoubs, R. T. Chien, and J. M. Selander IEEE Trans. on PAMI, vol. PAMI-2, no. 6, Nov., 1980, pp. 495-510.

JAIN et al. (1977)
Separating Nonstationary from Stationary Scene Components in a Sequence of Real World TV Images, R. Jain, D. Militzer, and H.-H. Nagel, Proc. IJCAI 5, Cambridge, MA., 1977, pp. 612-618.

JAIN et al. (1979)
Segmentation through the Detection of Changes due to Motion, R. Jain W. N. Martin, and J. K. Aggarwal, Computer Graphics and Image Processing, vol. 11, 1979, pp. 13-34.

JAIN and NAGEL (1979a)
On a Motion Analysis Process for Image Sequences from Real World Scenes. R. Jain and H.-H. Nagel,

IEEE Trans. PAMI, vol. PAMI-1 no. 2, Apr., 1979, pp. 206-214.

JAIN and NAGEL(1979b)
On the Analysis of Accumulative Difference Pictures from Image Sequences of Real World Scene, R. Jain and H. -H. Nagel IEEE Trans. on PAMI, vol. PAMI-1 no. 2, Apr., 1979.

JAIN(1981)
Extraction of Motion Information from Peripheral Processes, R. Jain IEEE Trans. on PAMI, vol. PAMI-3 no. 5, Sep., 1981, pp. 489-503.

KITCHEN and ROSENFELD(1980)
Grey-Level Corner Detection, L. Kitchen and A. Rosenfeld, TR-887, Computer Science Center, University of Maryland, College Park, 1980.

KOENDERINK and van DOORN(1976)
Local Structure of Movement Parallax of the plane, J. J. Koenderink and A. J. van Doorn, Journal of the Optical Society of America, vol. 66, 1976, pp. 717-723.

LAVENBERG(1944)
A Method for the Solution of Certain Non-linear Problems in Least Squares, K. Lavenberg, Quart. Appl. Math., vol. 2, 1944, pp. 164-168.

LEE(1980)
The Optic Flow Field: The Foundation of Vision, D. N. Lee, Phil. Trans. Soc. Lond. B290, pp. 169-179, 1980.

LEESE et al. (1970)
The determination of Cloud Pattern Motions from Geosynchronous Satellite Image Data, J. A. Leese, C. S. Novak, and V. R. Taylor Pattern Recognition, vol. 2 no. 4. Dec., 1970, pp. 279-292.

LILLESTRAND(1972)
Techniques for Change Detection, R. L. Lillestrand IEEE Trans. Computers, vol. C-21, 1972, pp. 654-659

LIMB and MURPHY(1975)
Estimating the Velocity of Moving Images in Television Signals, J. O. Limb and J. A. Murphy Computer Graphics and Image Processing vol 4, pp. 311-327, 1975.

LONGUET-HIGGINS and PRAZDNY (1980)
The Interpretation of a Moving Retinal Image, H. C. Longuet-Higgins and K. Prazdny Proc. R. Soc. Lond. B vol. 208, 1980, pp. 385-397.

MARR and POGGIO(1979)
A Computational Theory of Human Stereo Vision, D. Marr and T. Poggio, Proc. Soc. London Ser. B vol. 204, 1979, pp. 301-308.

MARR and HILDRETH(1980)
Theory of Edge Detection, D. Marr and E. Hildreth Proc. Roy. Soc. London Ser. B vol. 207, 1980, pp. 187-217.

MARTIN and AGGARWAL(1978a)
Survey, Dynamic Analysis, W. N. Martin and J. K. Aggarwal Computer Graphics and Image Processing, vol. 7 no. 3, June 1978.

MARTIN and AGGARWAL(1978b)
Computer Analysis of Dynamic Scenes Containing Curvilinear Figures, W. N. Martin and J.K. Aggarwal, Pattern Recognition, vol. 11, Dec., 1978, pp. 169-178.

MARTIN and AGGARWAL (1983)
Vlumetric Description of Objects from Multiple Views, W. N. Martin and J. K. Aggarwal, IEEE PAMI, vol. PAMI-5, no. 2, 1983, pp. 150-158.

MORAVEC(1977)
Towards Automatic Visual Obstacle Avoidance, H. P. Moravec, International Joint Conference on Artificial Intelligenec, 1977, pp. 584.

MORAVEC(1979a)
Visual Mapping by a Robot Rover, H. P. Moravec, IJCAI, Tokyo, Japan, 1979, pp. 598-600

MORAVEC(1979b)
Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover H. P. Moravec Ph. D Dissertation, Stanford University, Stanford, Calif.

NAGEL(1981a)
Image Sequence Analysis: What can we learn from Applications ?, H. -H. Nagel, Image Sequence Analysis, T. S. Huang edt., SpringerVerlag, Berlin / Heidelberg / Newyork, 1981, pp. 19-228.

NAGEL(1981b)
Representation of Moving Rigid Objects Based on Visual Observations, H. -H. Nagel, IEEE on Computers, vol. C-14, Aug., 1981, pp. 29-39.

NAGEL(1981c)
On the Derivation of 3D Rigid Point Configurations from Image Sequences, H. -H. Nagel, Pattern Recognition and Image Processing, August, 1981.

NAGEL and ENKELMANN(1982)
Investigation of Second Order Grey Value Variations to Estimate Corner Point Displacements, H. -H. Nagel and W. Enkelmann, International Conference on Pattern Recognition, 1982, pp. 768-773.

NAGEL(1983a)
Displacement Vectors Derived from Second-Order Intensity Variations in Image Sequences, H. -H. Nagel, Computer Vision, Graphics, and Image Processing,

vol. 21, 1983, pp. 85-117.

NAGEL(1983b)
Constraints for the Estimation of Displacement Vector Fields from Image Sequences, H.-H. Nagel. Proc. ICJAI-83, 1983, pp. 945-951.

NAGEL and ENKELMANN(1984)
Towards the Estimation of Displacement Vector Fields by 'Oriented Smoothness' Constraints. H.-H. Nagel and W. Enkelmann International Conference on Pattern Recognition 7, Montreal Canada, 1984, pp. 6-8

NAKAYAMA and LOOMIS(1974)
Optic Velocity Patterns, Velocity Sensitive Neurons, and Space Perception: a Hypothesis K. Nakayma and J. M. Loomis Perception vol. 3, 1974, pp. 63-80

NETRAVALI and ROBBINS (1979)
Motion-Compensated Television Encoding: Part I. A. N. Netravali and J.P. Robbins. Bell System Tech. J. vol. 58, 1979, pp. 631-670.

NETRAVALI and ROBBINS (1980)
Motion-Compensated Coding: Some New Results, A.N. Netravali and J.P. Robbins. Bell System Tech. J vol. 59, 1980, pp. 1735-1745

PAQUIN and DUBIOS(1983)
A Spatio-Temporal Gradient Method for Estimating the Displacement Field in Time-Varying Imagery, R. Paquin and E. Dubios. Computer Vision, Graphics, and Image Processing, vol. 21, 1983, pp. 205-221

POTTER(1975)
Velocity as a Cue to Segmentation, J.L. Potter. IEEE Trans. Systems, Man, and Cybernetics, vol. 5, May 1975, pp. 390-394

PRAGER and ARBIB(1983)
Computing the Optic Flow: The Match Algorithm and Prediction. Pragger and Arbib Computer Vision, Graphics, and Image Processing, vol. 24, 1983, pp. 276-304.

PRAZDNY(1980)
Egomotion and Relative Depth Map from Optic Flow K. Prazdny, Biological Cybernetics, vol. 36, 1980, pp. 87-102.

PRAZDNY(1981)
Determining the Instantaneous Direction of Motion from Optic Flow Generated by Curvilinear Moving Observer, K. Prazdny, Computer Graphics and Image Processing, vol. 17, pp. 238-248, 1981

RENADE and ROSENFELD(1980)
Point Pattern Matching by Relaxation, S. Renade and A. Rosenfeld, Pattern Recognition, vol. 12, 1980, pp. 269-275.

RIEGER and LAWTON(1983)
Determining the Instantaneous Axis of Translation forom Optic Flow Generated by Arbitrary Sensor Motion J.M. Rieger and D.T. Lawton ACM, 1983.

ROACH and AGGARWAL(1979)
Computer Tracking of Objects Moving in Space, J.W. Roach and J. K. Aggarwal. IEEE Trans. on PAMI, vol. PAMI-1 no. 2, Apr., 1979.

ROACH and AGGARWAL(1980)
Determining the Movement of Objects from a Sequence of Images, J.W. Roach and J. K. Aggarwal, IEEE Trans. on PAMI, vol. PAMI-2 no. 6, Nov., 1980, pp. 554-562.

ROSENFELD and VANDERBURG(1977)
Coarse Fine Template Matching, A. Rosenfeld and G. J. Vanderburg. IEEE Trans. on Systems, Man, and Cybernetics, vol. 7 no. 2, Feb 1977, pp. 104-107.

SCHUNK(1984)
The Motion Constraint Equation for Optic Flow, B.G. Schunk International Conference on Pattern Recognition 7, Montreal Canada, 1984, pp. 20-22.

SHAH and JAIN(1984)
Detecting Time-Varying Corners, M.A. Shah and R. Jain, Computer Vision, Graphics, and Image Processing, vol. 28, 1984, pp. 345-355.

SNYDER et al. (1980)
Image Modelling, the Continuity Assumption and Tracking, W.E. Snyder, S.A. Rajala, and G. Hirzinger, International Conference on Pattern Recognition, 1980, pp. 1111-1111

STULLER and KRISHNAMURPHY (1983)
Kalman Filter Formulation of Low-Level Television Image Motion Estimation, J. Stuller and G. Krishnamurphy, Computer Vision, Graphics, and Image Processing, vol. 21, 1983, pp. 169-204 aaa

THOMPSON(1980)
Combining Motion and Contrast for Segmentation, W. B. Thompon, IEEE Trans. on PAMI, vol. PAMI-2 no. 6, Nov. 1980, pp 543-549

THOMPSON and BARNARD(1981)
Low-Level Estimation and Interpretation of Visual Motion, W.B. Thompson and S.T. Barnard, IEEE Computer, vol. 14, no. 8, Aug., 1981, pp20-28

TRETIAK and PASTOR(1984
Velocity Estimation from Image Sequence with Second Order Differential operator O. Tretiak and L. Pastor. International Conference on Pattern Recognition 7, Montreal Canada, 1984, pp. 16-19

TSAI and HUANG(1981)
Estimating Three-Dimensional Motion Parameters

of a Rigid Planar Patch I, R. Y. Tsai and T. S. Huang, IEEE Trans. on ASSP, vol. ASSP-29, Dec. 1981, pp. 1147 - 1152

TSAI et al. (1982)
Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch II, R. Y. Tsai, T. S. Huang, and W. Zhu. IEEE Trans. on ASSP, vol. ASSP - 30 no. 4, Aug., 1982, pp. 525 - 534

TSAI and HUANG (1984)
Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces, R.Y. Tsai and T.S. Huang, IEEE on PAMI, vol. PAMI - 6, no. 1, Jan., 1984, pp. 13 - 27

TSAI (1983)
Multiframe Image Point Matching and Surface, Reconstruction, IEEE on PAMI, vol. PAMI - 5, no. 2, Mar., 1983, pp. 159 - 174

ULLMAN (1979)
The Interpretation of Visual Motion, S. Ullman MIT Press, Cambridge, Mass., 1979

ULLMAN (1981)
Analysis of Visual Motion by Biological and Computer System, S. Ullman, Computer, vol. 14, no. 8, Aug., 1981, pp. 57 - 69

ULSTAD (1973)
An Algorithm for Estimating Small Scale Difference between Two Digital Images, M. S. Ulstad, Pattern Recognition vol. 5, 1973, pp. 323 - 333

VANDERBURG and ROSENFELD (1977)
Two Stage Template Matching, G. J. Vanderburg and A. Rosenfeld, IEEE Trans. Computers, vol. C - 26, Apr., 1977, pp. 384 - 393

WILLIAMS (1980)
Depth from Camera Motion in a Real World Scene, T. D. Williams, IEEE Trans. on PAMI, vol. PAMI - 2 no. 6, Nov., 1980, pp. 511 - 516

WOHN et al. (1983)
Motion Estimation Based on Multiple Local Constraints and Nonlinear Smoothing, K. Wohn, L. S. Davis, and P. Thrift, Pattern Recognition, vol. 16 no. 6, 1983, pp. 563 - 70

YACHIDA (1983)
Determining Velocity Map by Spatio-Temporal Nwighborhoods from Image Sequence, M. Yachida Computer Vision, Graphics, and Image Processing, vol. 21, 1983, pp. 262 - 279

YALAMANCHILI et al. (1982)
Extraction of Moving Object Descriptions via Differencing, S. Yalamanchili, W. N. Martin, and J. K. Aggarwal Computer Graphics and Image Processs - ing, vol. 18 1982, pp. 188 - 201

YEN and HUANG (1983)
Determining 3 - D Motion and Structure of a Rigid Body Using the Spherical Projection, B. L. Yen and T. S. Huang, Computer Vision, Graphics, and Image Processing, vol. 21, 1983, pp. 21 - 32

1973年 : 서울大學校 電氣工學科 卒業
1982年 : 美國 Virginia Polytech Institute 電氣工學科 (M. S.)
1985年 : 美國 Virginia Polytech Institute 電氣工學科 (ph. D.)
1986年 : 電子通信研究所 接觸處理研究室 先任研究員