A CONCEPT FOR AUTOMATED 3-D MEASUREMENT OF OBJECTS IN MOTION

Hartmut Schaefer and Shunji Murai
Institute of Industrial Science, University of Tokyo
7-22, Roppongi, Minatoku, Tokyo
Japan

Commission V

Abstract :

A conceptual solution for automated 3-D measurements of objects in motion is presented by the authors, based on the method of "Shape-from Motion", developed in the field of computer vision. For accurate measurements camera calibration, effective image preprocessing, and eventually available knowledge of object and motion have to be introduced into the process. Some parts of the concept have been realized. Therefore, an on-line system for verifiable experiments has been developed.

1. Introduction :

In the perception of space, motion plays besides stereo and texture a very important role in the human visual system /1/. For instance, suppose the observer moves relative to a scene with objects at different distances, then each object appears as shifted by different amounts (motion parallax), depending on the distance observer-object.

From a sequence of time-ordered images, taken by a single camera, velocity vectors of prominent feature points can be estimated and then related to the motion and structure of objects in space. The motion is described, basically, in terms of "optical flow" which is a field of 2-D displacement vectors per time on the image, these vectors being the projection of 3-D velocities of points moving in space.

Applications for analysis of time-ordered images, also called "motion-analysis" or "image sequence analysis", cover a broad range of fields including computer graphics, industry, medicine, robotics, remote sensing etc. with tasks such as /4/:

- target tracking
- 3-D shape from motion
- change detection
- behavior studies of objects

Note that in motion analysis, only relative motion can be recorded. In our system, the camera is stationary and the object moves. We assume that the frames are close enough together in time and the displacement vectors became small relative to object size. However, one of the basic assumption in motion analysis is the rigidity of objects and a continuous smooth motion.

For accurate measurement we have to introduce camera calibration, image preprocessing, and knowledge of object and its motion into the process. We discuss here a solution, on a conceptual level, for automated 3-D measurement of objects in motion. Until now, only few studies in verifying the motion experiments with accuracy checks have been presented /20/, since many of the researchers in computer vision are more interested in quantitative scene description and understanding (e.g. autonomous visual navigation of robots) than in qualitative 3-D reconstruction of objects in motion. Therefore, an 'on-line' system for verifiable experiments has been developed. Some early results can be demonstrated on real images.

2. Processing Flow in Motion Analysis :

The flow in motion analysis as it is proposed by the authors is shown in Fig.1.

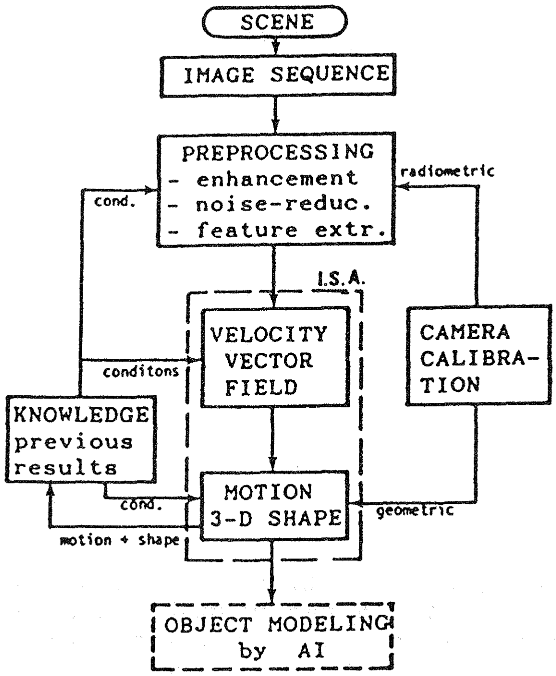

Fig.1 : Basic process in motion analysis

An electronic camera takes a sequence of images of the scene which can be processed either directly in on-line frame by frame, or after the sequence is taken and stored in a special video memory.

In the preprocessing step noise reduction, enhancement, segmentation, and feature extraction are the main issues. In the 'Image Sequence Analysis' step, the features of consecutive images are matched and related to optical flow, form which the relative motion parameters and the 3-D shape can be estimated. Preliminary results from each consecutive image pair can be used to guide later steps in the form of constraints given into the process.

Camera calibration is essential for accurate measurements. Geometric and radiometric calibration results as well have to be considered in the 'Motion Estimation' and 'Preprocessing' steps, respectively.

Finally, motion and shape must be converted into an interpretable and understandable form for the user, for instance by techniques of artificial intelligence (AI).

3. Conceptual Solution in Motion Analysis :

3.1 Camera and Preprocessing :

In motion analysis, solid state video cameras seem to be most suitable (e.g. CCD-, MOS-camera). Although their resolution is rather low, they are capable of working in 'on-line' together with an image processing system. Most of the CCD-sensor systems are designed for standard video signals (e.g. the American/Japanese NTSC-standard). The NTSC- frame rate is 30 frames/sec. Simple CCD-cameras are not equipped with a shutter and their use in motion analysis is restricted only to slow moving objects. Recently, CCD-standard video cameras with high-speed shutter (up to 1/5600 sec) and "slow-motion" ability are available with a resolution of 754 * 488 sensors per chip, which are able to freeze objects in high speed motion.

Row digitized CCD-video images are usually prone to noise due to A/D conversion, quantisation, and dark current in the sensor. The noise cleaning process, according to radiometric calibration results, depends on each CCD-camera type and cannot be given in general /14/.

In our case, we first subtracted a fixed pattern noise from each image obtained by closed lens and stored in the image memory. By this procedure almost all of the visible systematic noise could be removed. Then we applied a 3*3 smoothing filter to the image together with a gain of 2:1. In both steps hardware functions of the image processing system were used. Special averaging techniques by temporal image filtering may be used but they are very time consuming /3/.

However, we are not able to remove all of the noise. Therefore, the algorithm in the further motion analysis process should be noise robust.

3.2 Finding Optical Flow :

The several methods proposed in computer vision to compute optical flow fall into two broad classes : feature based methods and gradient based methods.

In the feature based method, prominent features are extracted from each image and tracked by a matching method from frame to frame. The problem here, however, is that only a sparsely set of points can be obtained, which are not dense enough to be useful for tasks such as 3-D interpretation and change detection.

Gradient techniques rely on the fact that spatial and temporal intensity changes are not independent of each other and satisfy the relation /1,4,6/:

$$-I_t = I_x V_x + I_y V_y \tag{1}$$

with

$I(x,y,t)$	=	differentiable image function
I_x, I_y	=	components of spatial intensity gradient
I_t	=	gray value difference between consecutive frames at point (x,y)
V_x, V_y	=	components of velocity vector

The disadvantage here is that the single equation does not allow one to determine both velocity components without additional constraints, and a linear model such as Eq.(1) is often too simplistic around edges or corners /5/.

Therefore, we propose a two-step approach combining both methods /6,7/. First, extracting and tracking prominent features and second, propagating their velocity estimates to a large number of image points along edges based on Eq.(1).

For feature tracking, corner points are most appropriate since they are invariant to rotations. For a corner detector we used the interest-operator of MORAVEC /8/, applied independently to each image.

Point patterns of consecutive images are then matched by relaxation method as described by BERNARD,THOMPSON /9/ and RANADE,ROSENFELD /10/. This method makes use of the consistency property of disparity which relies upon the fact that disparity varies continuously across unoccluded surfaces and discontinuously only at occluding edges. It is used as a measure of how well a particular match conforms to near-by matches. False matches, based on similarity alone, can be avoided, which is particularly true in motion analysis where objects can rotate and simple correlation techniques would fail.

The consistency constraint can even be extended to the optical flow, since the optical flow changes continuously from frame to frame due to smooth motion.

3.3 Motion and Structure form Optical Flow :

If only one object is present in each image, motion and shape estimation from optical flow is trivial. If many objects are present and expected, the object candidates must be classified by segmentation techniques with the aid of optical flow /11/. Then, for each object candidate one can determine its motion parameters and 3-D structure from their optical flow region, respectively.

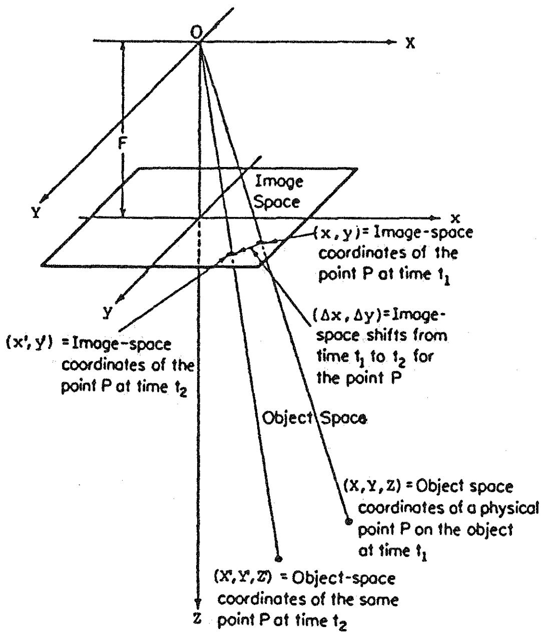

Fig.2: Basic geometry in 3-D motion estimation

Any 3-D motion of a rigid body can be decomposed into a rotation and a translation /2/.

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{bmatrix} \qquad (2)$$

Introducing image coordinates using perspective rotation

$$x = F * X/Z \qquad y = F * Y/Z \qquad (3)$$

and solving for Z yields :

$$Z = \frac{F\,\Delta X - x'\,\Delta Z}{x'(r_{31}x + r_{32}y + r_{33}F) - (r_{11}x + r_{12}y + r_{13}F)} = \frac{F\,\Delta X - y'\,\Delta Z}{y'(r_{31}x + r_{32}y + r_{33}F) - (r_{21}x + r_{22}y + r_{23}F)} \qquad (4)$$

Eq.(4) is non-linear in 5 independent unknowns $\hat{\Delta x}, \hat{\Delta y}, \omega, \phi, \kappa$, i.e the translation vector T can only be determined within a scale factor. Therefore, we want to find only the unit translation vector :

$$\hat{T} = (\hat{\Delta X}, \hat{\Delta Y}, \hat{\Delta Z}) = (\Delta X, \Delta Y, \Delta Z)^{-1/2}\ T \qquad (5)$$

Five points are necessary to solve Eq.(4). Thus the motion estimation problem is equivalent to the 'relative-orientation' problem in photogrammetry, and in fact Eq.(4) expresses the coplanarity condition.

The nonlinear motion equation (4) can be solved by any of the standard methods for relative orientation, known in analytical photogrammetry. All these methods, however, need a good initial guess which is very hard to obtain for the initial match in motion analysis. Therefore, LONGUET-HIGGINS /12/ and HUANG /2/ proposed a linear algorithm and called it "8-point solution" because of its 8 unknowns. This algorithm was implemented in our system and successfully tested with simulated data.

As in relative orientation, the solution is indeterminable if the moving object points and the origin lie on a 'critical surface' or close to it (e.g. cylinder, cone). Hence, tracking a concave object could become critical and should be avoided. The object coordinates are also indeterminable if T = 0 (pure rotation), but we can observe its rotation parameters /2/.

Finally, the object coordinates can be determined from Eq.(4) and (3) within a scale factor. The models ,obtained from consecutive images, can then be grouped together and referred to one unique reference system, for instance by the 'independent model method' of phototriangulation.

3.4 Camera Calibration :

The calibration of a CCD-camera can be separated into two groups : a testfield independent part and a testfield dependent part.

The testfield independent part comprises:

(i.) basic sensor check
(ii.) radiometric calibration

This part can be carried out independently from applications and in our system, we assume an already radiometric calibrated camera, as mentioned in 3.1.

The testfield dependent part contains determination of:

(i.) interior orientation parameters
- principal distance F
- principal point xp,yp
- transformation parameter between comparator coordinates (u,v) and image coordinates (x,y)

(ii.) lens distortion parameters

Due to very good stability in the inner orientation of CCD-cameras and a fixed focus throughout our experiments, which was assumed, the interior orientation parameters can be regarded as constant and the calibration process will be done only at the beginning of each motion experiment.

For the functional calibration model, we choose a modified DLT-approach. Their basic equations are /13/ :

$$u = \frac{l_{11}X + l_{12}Y + l_{13}Z + l_{14}}{l_{31}X + l_{32}Y + l_{33}Z + l_{34}} = \frac{L_1 * X + l_{14}}{L_3 * X + l_{34}}$$

$$v = \frac{l_{21}X + l_{22}Y + l_{23}Z + l_{24}}{l_{31}X + l_{32}Y + l_{33}Z + l_{34}} = \frac{L_2 * X + l_{24}}{L_3 * X + l_{34}} \qquad (6)$$

In the DLT-approach, these equations have been solved for the 11 unknowns l_{ij} by setting $l_{34} = 1$, but then the interior parameters will depend on the choice of the reference system. Therefore, FAUGERAS, TOSCANI /15/ suggested the constraint $\|L_3\|^2 = 1$ which left the orientation parameters invariant to rigid motion of the reference system. Under this assumption, we can find a set of observation equation from Eq.(6) :

$$r = B\ X_9 + C\ X_3 \qquad (7)$$

with the condition

$$X_3^T\ X_3 = 1 \qquad (8)$$

where $X_9 = [L_1, l_{14}, L_2, l_{24}, l_{34}]$, $X_3 = [L_3]$, and r is the residual error vector. Using the method of Lagrange multipliers, differentiating to X_9 and X_3, and setting the derivatives to zero, yield:

$$X_9 = -(B^T B)^{-1} B^T C\ X_3 \qquad (9a)$$

$$(C^T C - C^T C(B^T B)^{-1} B^T C)\ X_3 = \lambda\ X_3 \qquad (9b)$$

or

$$D \qquad X_3 = \lambda\ X_3$$

The solution is obtained by taking the eigenvector X_3 of D corresponding to the smallest eigenvalue, this yields X_9 by Eq.(9a) /15/.
This process has to be iterated until a stable solution is obtained. The external and internal orientation parameters can then be recovered uniquely from X_9 and X_3 solutions.

Taking additional parameters for lens distortion into consideration seems to be questionable because of

(i.) low resolution due to coarse CCD pixel structure

(ii.) systematic errors in the interior orientation have little effect to the determination of small displacement vectors, particularly true when only a single camera is used /16/.

3.5 Knowledge of Scene and Motion and Object Modeling :

The algorithm in motion analysis are relatively sensitive to image noise. Good accuracy depends on considerable over-determination due to increasing object points and views.
A stable solution may be obtained if we make use of some knowledge about the object and its expected motion. Therefore, several authors proposed stable and unique solutions for objects with planar faces /17,18/. Other authors suggest a kind of parallel algorithm, coupling computation of optical flow and motion parameters. However, an initial guess of object shape or depth is necessary /19/.

In our solution, we assume that no a-priori information either of object shape nor motion is given. But while the motion analysis proceeds from frame to frame, preliminary results from earlier steps can be used to guide the later processing (e.g. the appearance of objects in the next frame can be predicted from former motion data), and later results can be used to update or correct results from earlier steps. Such processing methods are known as "Feedback-Control" or "Heterachical-Control" schemes in computer vision /21/.

Although object or scene modeling is necessary to complete the entire process in motion analysis, it will be discussed only very briefly, since it is not the goal in our experimental system.
Therefore, a hierarchical model of object description with increasing level of representational completeness will be given next, as suggested by NAGEL /22/.

(i.) 3-D geometric description of a single object
- rigid 3-D point model
- rigid 3-D wire frame model
- rigid 3-D surface model
- rigid 3-D volume model

(ii.) 3-D configuration of independently movable objects

(iii.) 3-D configuration of objects surrounded by environment (foreground, background)

4. Experiments :

4.1 The System Configuration :

The basic structure of the system, as it is set up in the laboratory, is shown in Fig.3.

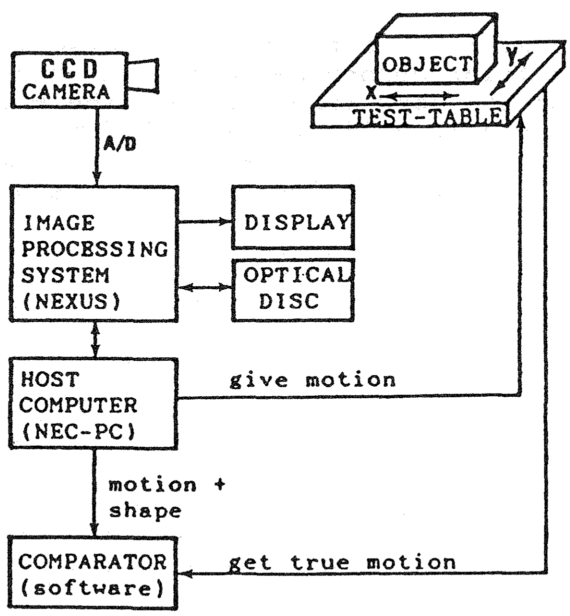

Fig.3: Experimental system for verifiable experiments in motion analysis

As a CCD-camera, we used a simple low resolution camera, SONY XC-37 with 384*491 pixel size, for checking the system and early experiments. For later experiments a camera with high-speed shutter and higher resolution will be used. The camera is connected via a real-time A/D converter to an image processing system, NEXUS KASHIWAGI RESEARCH Corp. with a 4*512*480 8bit frame memory. Images can be stored on two optical discs ,NATIONAL DU-15 with 1 Gbyte per disc. The system is controlled by a 16bit micro computer,NEC PC 9801vm.

For verifying 3-D motion a special test-table was designed. The test-table is driven by a high precision pulse motor for translation and equipped with a micrometer for rotation around one coordinate axis:

- maximal range for X- and Y translation = 230 mm ± 10 µm
- maximal angle for rotation = 5° 30' ± 5 "

The test-table is controlled by the host computer. The estimated motion parameter, obtained from the image sequence analysis, will be compared with the true motion, grabbed from the test-table, in the comparator.

4.2 Results :

The first step in motion analysis is to find an optical flow field from consecutive images. Some parts of the suggested concept of finding optical flow have been implemented and tested. First, selecting prominent feature points by MORAVEC interest operator and second, matching these sets of points be relaxation method.

The algorithm will be demonstrated on two different image pairs:

(i.) a pure translation, shifted by 40 mm (Fig.4a,b)
(ii.) a pure rotation, rotated by 2°.75 (Fig.4c,d)

As a simple rigid body, we chose a 'Rubik's Cube' which was placed on the test-table and their images were taken by a CCD-Camera and digitized into a 512*480 8bit form.

The interest operator was applied to a coarse 4 times reduced image in order to avoid false corner detection in busy textured regions (e.g. reflectance on the dark lines between the squares). The detected points were then used to guide the search for corner points in the finer steps - 2 times reduced image and original image. This approach worked quite well, except in low contrast regions where only few points could be detected (see Fig. 5a,b upper right corner of the cube) even if an contrast dependent threshold was chosen. Sometimes the interest operator responds to non-smooth edges which lead to fails corner points (Fig. 5c,d).

In the second step the feature points of the consecutive images were then matched by relaxation method. An initial set of possible matches is constructed by pairing each candidate point from image 1 with every candidate point from image 2 within some range of maximum detectable disparity. Each pair is then labeled by a weight. As initial weight we used the normalized cross-correlation value multiplied by 100. These initial weights, depending only on similarity, are then iteratively updated by the consistency property (see 3.2). After some iteration pairs with weights greater then some threshold are then considered as matched. In our experiments only 4 iteration were necessary and a threshold of 0.001.

In the pure translation case the matching was of no problem and good results were obtained. In the pure rotation case, however, one mismatch can be seen near the upper right corner of the cube (Fig.4c,d) since the corners are not corresponding points but very close to each other. Unreliable matches can be seen on points which are no corner points but edge points due to false corner detection.

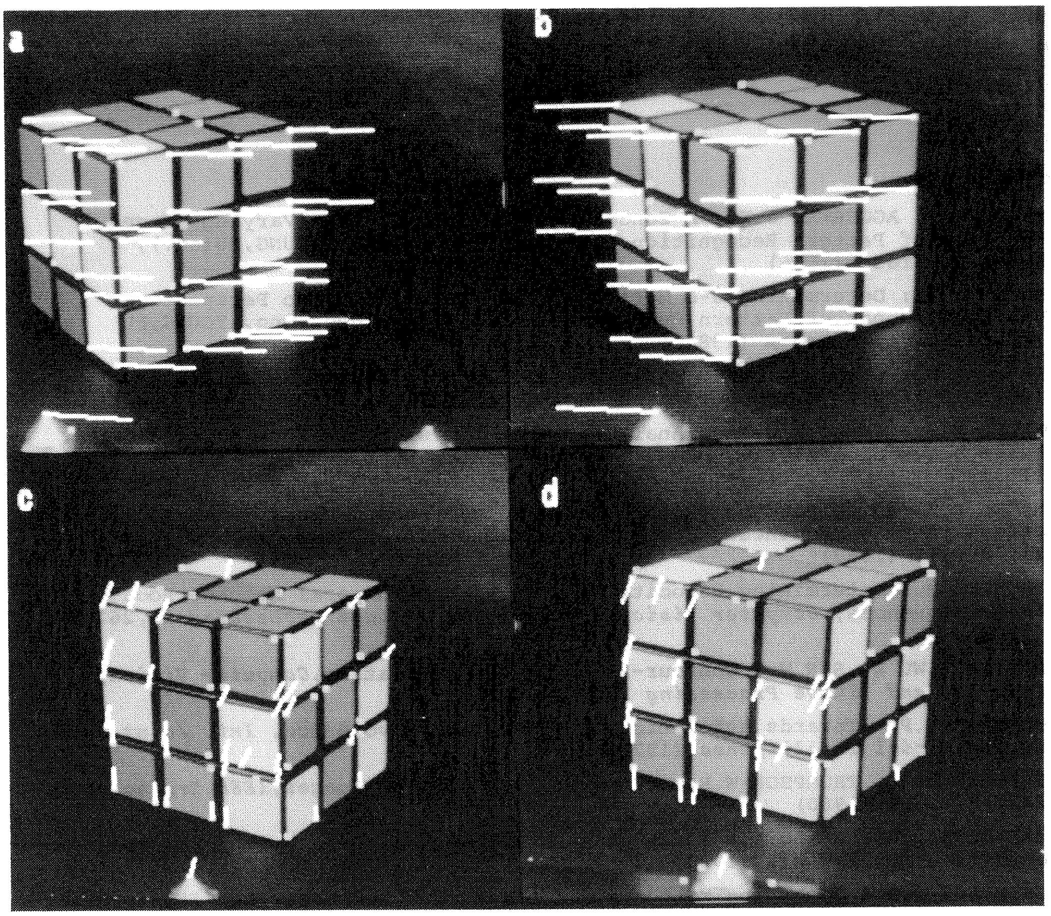

Fig.4: Two sequences with detected corner points and its displacement vectors; a) translation time t1, b) translation time t2, c) rotation time t1, d) rotation time t2.

5. Summary and Remarks :

A conceptual solution for automated 3-D measurements of objects in motion from a sequence of images has been presented by the authors. It consists of four parts; (1) estimate a velocity flow field from consecutive images, (2) determine the motion parameter and 3-D shape of each object moving in space, (3) calibration of the camera, and (4) introducing knowledge of object and its motion into the process obtained in earlier steps.

In the current investigation step (1) has been partly realized for detecting and matching of prominent feature points. The MORAVEC interest operator shows some weakness in detecting corner points in low contrast regions. A refined interest operator with a more sophisticated gray value model, as proposed by FOERSTNER /23/, is recommended. Good results in point pattern matching were obtained by relaxation method.

The motion analysis as proposed in this paper is, however, restricted to rigid body motion only. Motion and 3-D shape analysis from nonrigid bodies, like animals, human bodies and organs in medicine or 3-D deformation measurements in civil engineering require a combination of stereo method (e.g. two cameras instead of one) and shape-from-motion method.

One important problem, which is not considered in this paper is, how to handle the big amount of image data. Therefore, further studies and special techniques are required.

REFERENCES

1. MITCHI A., AGGERWAL J.K.: A Conceptual Analysis of Time-Varying Images, Handbook of Pattern Recognition and Image Processing, YOUNG,FU(ed.), Academic Press (1986)
2. HUANG T.S.: Determining 3-D Motion and Structure from Two Perspective Views, Handbook of Pattern Recognition and Image Processing, YOUNG,FU (ed.), Academic Press (1986)
3. HUANG T.S.: Image Sequence Enhancement, Image Sequence Analysis,HUANG (ed.), Springer Verlag Berlin and New York (1981)
4. NAGEL H.H.: Image Sequence Analysis: What can we learn from applications, Image Sequence Analysis, HUANG (ed.), Springer Verlag Berlin and New York (1981)
5. NAGEL H.H.: Displacement Vectors Derived from Second-order Intensity Variations in Imag Sequences, *Computer Vision, Graphics, and Image Processing* 21, 85-117 (1983)
6. YACHIDA M.: Determining Velocity Maps by Spatio-Temporal Neighborhoods from Image Sequences, *Computer Vision, Graphics, and Image Processing* 21,262-279 (1983)
7. DAVIS L.,WU Z.,SUN H.: Contour-Based Motion Estimation,*Computer Vision, Grapics, and Image Processing* 23, 313-326 (1983)
8. MORAVEC H.P.: Towards Automatic Visual Obstacle Avoidance, *Int. Joint Conf. of Artificial Intelligence* (1977)
9. BERNARD S.T., THOMPSON W.B.: Disparity Analysis of Images,*IEEE Transaction*, PAMI-2, No.4 (1980)
10. RANADE S., ROSENFELD A.: Point Pattern Matching by Relaxation, *Pattern Recognition* 12, 269-275 (1980)
11. THOMPSON W.B.: Combining Motion and Contrast for Segmentation, *IEEE Trans.*, PAMI-2, No.6 (1980)
12. LONGUET-HIGGINS H.C.:A Computer Algorithm for Reconstructing a Scene from Two Projections, *Nature(London)* 293, 133-135 (1981)
13. ABDEL-AZIS Y.I.,KARARA H.M.: Photogrammetric Potentials of Non-Metric Cameras, Dept. of Civil Eng., University of Illinois, Urbana (1974)
14. GUELCH E.: Calibration of CCD Video Camera, Progress in Imaging Sensors, *Proc. ISPRS Symposium, 1-5.Sept. 1986, Stuttgart (FRG)*,391-398 (1986)
15. FAUGERAS O.D., TOSCANI G.: Camera Calibration for 3-D Computer Vision, *IEEE Proc. Inter. Workshop on Industrial Applications of Machine Vision and Machine Intelligence, Febr. 2-5,1987, Tokyo,* 240-247 (1987)
16. WIJK M.C., ZIEMANN H.: The Use of Non-Metric Cameras in Monitoring High Speed Processes, *Photogrammetry Engineering and Remote Sensing,* Vol.42, No.1 (1976)
17. KANATANI K., CHOU T.: Tracing Finite Motions Without Correspondance, *IEEE Proc. Inter. Workshop on Industrial Applications of Machine Vision and Machine Intelligence, Febr. 2-5,1987, Tokyo,* 118-123 (1987)
18. TSAI R.Y., HUANG T.S: Estimating 3-D Motion Parameters of a Rigid Planar Patch, *IEEE Trans.* ASSP-29, No.6, 1147-1152 (1981)
19. BALLARD D.H, KIMBALL O.A.: Rigid Body Motion from Depth and Optical Flow, *Computer Vision, Graphics, and Image Processing* 22, 95-115 (1983)
20. FANG J., HUANG T.S.: Some Experiments on Estimating the 3-D Motion Parameters of a Rigid Motion from Two Consecutive Image Frames, *IEEE Trans.* PAMI-6, No.5, 545-554 (1984)
21. SHIRAI Y.: Three Dimensional Computer Vision, Springer Verlag (1987)
22. NAGEL H.-H.: Spatio- Temporal Modeling Based on Image Sequences, *Proc. of the Inter. Symp. on Image Processing and it Applications,* Jan 18-21,1984, Tokyo, M. ONOE (ed.), (1984)
23. LUHMANN T., ALTROGGE G.: Interest Operator for Image Matching, *Proc. of the Symp. From Analytical to Digital, ISPRS Comm. III,* August 19-22,1986, Rovaniemi, (1986)