

# RIGID AND NONRIGID MOTION ANALYSIS: ROBUST RECOVERY OF 3-D STRUCTURE AND MOTION

*Hiroyuki Morikawa and Hiroshi Harashima*

Department of Electrical Engineering  
The University of Tokyo

7-3-1, Hongo, Bunkyo-ku, Tokyo 113 JAPAN

## ABSTRACT

This paper presents an incremental approach to understanding 3D structure and motion of rigid as well as nonrigid objects from image sequences. To be applicable to natural imagery like TV signals, the recovery process should meet two requirements: robustness to noise and ability of coping with deformable objects. These requirements can be met by the use of smoothness-of-motion constraint. The smoothness-of-motion constraint provides a framework for temporal integration of motion information over a longer sequence, rather than a set of image pairs. Based on the nature of motion, i.e. motion smoothness, 3D structure and motion information is successively and smoothly updated so as to agree with the observed transformation in the image. A model of rigid and nonrigid motion is introduced, and the smoothness-of-motion constraint is formulated as a stabilizing function. Some preliminary results are also given.

## 1. INTRODUCTION

In this paper we consider the problem of computing the structure and motion of objects in a scene from a sequence of images. The problem of reconstructing the shapes and motion has been studied by computer scientists with the objective of developing vision systems for many applications, such as robot vision, surveillance systems, object tracking, autonomous vehicle navigation, and computer graphics[1]. In addition, this has recently attracted some attention in image coding field, e.g., 3-D structure extraction coding[2] and model-based coding[3].

In studying the computation of structure from motion, one immediately faces the problem that the recovery of structure is underconstrained, ill-posed problem: there are infinitely many 3-D structures consistent with a given pattern of motion in the changing 2-D image. This problem cannot be solved without some assumptions about the world. Computational studies of the recovery of structure from motion establish that the rigidity is a sufficiently powerful constraint for imposing uniqueness upon the 3-D interpretation. The rigidity assumption allows recovery of the structure of objects, under certain condition, in 2 or 3 views[4].

These theoretical studies have given rise to algorithms for the recovery of 3-D information of rigid objects. Thus, to date almost all research on recovering structure and motion has been concerned with rigid objects using rigidity assumption, and attempt to recover 3-D information from a limited number of views of the scene, typically 2 or 3 views. However, the rigidity assumption and 2-3 view approach can be sensitive to noise[5], and clearly is inappropriate when dealing with deformable objects.

To be applicable to natural images like TV signals, the process of recovering structure and motion should meet the two requirements: robustness to noise and ability of coping with nonrigid objects.

Regarding the robustness, one way to combat the effect of noise would be the temporal integration of motion information. Significant smoothing can be achieved by the use of a larger number of images in the sequence. This conclusion is supported in recent computational studies. This approach is sometimes referred to as trajectory-based approach[6].

Regarding the nonrigid objects, we must weaken the strict rigidity assumption to allow for more general motion. There are some researches done to treat the restricted classes of nonrigid motion, such as global deformations such as shear, bending, and divergence. General nonrigid motion problem was recently formulated. Ullman[7] attempts to recover the structure of a moving object by assuming a minimal change in the rigidity of this object between frames. The results show that the algorithm is able to recover the approximate 3-D structure. The scheme, however, show the depth reversal phenomenon, and often wobbles somewhat around the correct solution to the 3-D structures. Subbarao[8] seeks to recover structure of nonrigid objects on a surface-patch-by-surface-patch basis. In this approach the problem may be sensitive to noise because of restricting oneself to using only local measurements.

In this paper we present an alternative approach, i.e., smoothness-of-motion approach, to recover the 3-D structure and motion of rigid as well as nonrigid objects from a longer sequence. It is shown that the use of the smoothness-of-motion would be more flexible and general approach than strict rigidity assumption.

## 2. REPRESENTATION OF RIGID AND NONRIGID MOTION

We are interested in estimating three-dimensional structure and motion parameters of rigid and deformable bodies from image sequences. Generally, the motion of a three-dimensional solid is a mathematical function  $F$  which explicitly modifies the global coordinates of points in space

$$\mathbf{x}_i(t+1) = F(\mathbf{x}_i(t)) \quad (1)$$

where  $\mathbf{x}_i(t) = (x_i(t), y_i(t), z_i(t))^T$  represents the coordinates of the  $i$ th point in the object at time  $t$ , and  $\mathbf{x}_i(t+1) = (x_i(t+1), y_i(t+1), z_i(t+1))^T$  is the coordinates of the same point after motion at time  $t+1$ .

According to Helmholtz's fundamental theorem of kinematics, the most general motion of a sufficiently small element of a deformable body can be represented as the sum of a rotation, a uniform deformation, and a translation[10]. Thus, if we divide the body into small parts which undergo a uniform deformation, the motion of each one part can be described by a linear affine model mathematically,

$$\mathbf{x}_i(t+1) = (\mathbf{R}_{t+1} + \mathbf{S}_{t+1})\mathbf{x}_i(t) + \mathbf{T}(t+1) \quad (2)$$

where  $\mathbf{R}_{t+1}$  is the rotation matrix from time  $t$  to  $t+1$ ,  $\mathbf{S}_{t+1}$  is the linear deformation matrix from time  $t$  to  $t+1$ , and  $\mathbf{T}(t+1)$  is the displacement

vector of the object between time  $t$  and  $t+1$  due to translation motion. The sum of the matrices  $R_{t+1}$  and  $S_{t+1}$  is sometimes referred to as generalized motion parameters.

Clearly, this uniform deformation model, or a linear affine model, is a suitable representation only when we assume a very small element of a body, and it is inadequate to represent a global or nonuniform deformation that occur often in nature. In answer to this problem we use a more general and flexible model:

$$x_i(t+1) = R_{t+1} \cdot D_{t+1}^i [x_i(t)] + T(t+1) \quad (3)$$

where  $D_{t+1}^i$  is the transformation function of deformation of  $i$ th point from time  $t$  to  $t+1$ . This can be described schematically by

$$\text{Trans}(\text{Rot}(\text{Deform}(x))). \quad (4)$$

This representation is highly intuitive and easily visualized, since this is equivalent to view the object motion as any translation or rotation (rigid motion) is performed after deformation. Also various deformation, e.g., bending, twisting, tapering, cavity deformations, can be incorporated into the transformation function  $D_{t+1}^i$ , widening the application of the representation. For example, the transformation function  $D_{t+1}^i$ , of tapering deformation along axis  $z$  is

$$D_{t+1}^i: (x_i(t), y_i(t), z_i(t))^T \rightarrow (f_x(z)x_i(t), f_y(z)y_i(t), z_i(t))^T \quad (5)$$

where  $f_x$  and  $f_y$  are the tapering functions in the  $x$ - and  $y$ -axes of the object centered coordinate system.

### 3. COMPUTING 3-D STRUCTURE AND MOTION

This section first discuss the smoothness-of-motion constraint used in our approach, rather than strict rigidity assumption. Then a nonrigid motion model based on the smoothness-of-motion constraint is introduced, and a formulation will be presented that uses the smoothness-of-motion constraint as a stabilizing function.

#### 3.1 Smoothness of Motion

In general, the moving objects exhibit a smooth motion due to inertia and elasticity, i.e., the motion parameters between consecutive frames are correlated. If a frame sequence is acquired at a rate such that no dramatic changes take place between frames, then observed changes in the motion will be gradual for most physical objects. Thus, the *smoothness-of-motion* is a very reasonable assumption for the analysis of 3-D dynamic scene. If unusual case such as a collision occurs, then some high-level process may be required to analyze the motion after collision.

The advantage of the smoothness-of-motion approach is that it provides a framework for integrating time content information over a larger number of image frames. That is, this approach to the recovery of structure will allow successive refinement of the estimated structure of objects as more frames are observed. We believe that the use of a longer sequence would meet the two requirements: robustness to noise and ability of coping with nonrigid objects. The multiframe approach helps to combat the errors due to noises: significant smoothing can be achieved by the use of a larger number of images in the sequence. In addition, since a longer sequence gives more constraints than a 2-3 view approach, we may admit nonrigid objects in our analysis by relaxing the rigidity assumption. The smoothness-of-motion approach to the recovery of structure would be more suitable for natural scenes.

Perceptual studies also indicate that the integration of motion measurements over time is required to reach an accurate perception of rigid as well as nonrigid objects, and that the noise sensitivity of the system improves with an increase in the number of frames[10]. Thus, the recovery of structure from motion is not an all-or-none process. For a short viewing times, objects sometimes appear flatter than the true structure of the moving objects. These properties of the human visual

system are qualitatively consistent with the behavior of the incremental multiframe approach based on the smoothness-of-motion constraint. We believe that the smoothness-of-motion constraint is more general and can free ourselves from the strict rigidity assumption.

We exploit the smoothness-of-motion assumption for computing 3-D structure and motion of rigid and nonrigid objects. Based on this constraint about the nature of motion, 3-D structure and motion information is successively and smoothly updated so as to agree with the observed transformation in the image. In other words, update is made by resisting changes in structure and motion as much as possible, and as rigid as possible. Consequently, we consider the objects undergoing a rigid transformation combined with some nonrigid distortions, i.e., the deviations from rigidity are not so strong. This consideration is comparable to perceptual studies suggesting that the visual system can cope with less than strict rigidity, but cannot cope with completely unstructured nonrigid objects such as an amoeba. This tolerance for deviations from rigidity allows the recovery process to be applicable to various environments, and also implies that the recovery process has a certain immunity to noise.

#### 3.2 Formulation

Let  $x = (x, y, z)$  represent a spatial point coordinate, and let  $u = (u, v)$  represent a corresponding image plane coordinate. The configurations of object-coordinates and image-coordinates are chosen such that  $u, v$  axes coincide with  $x, y$  axes, and the  $z$ -axis is aligned with the optical axis. The formulation in this section assumes orthographic projection as imaging model. The low-level problem is not addressed, and image coordinates of  $n$  object match points are assumed to be available.

Based on the smoothness-of-motion constraint, we introduce the model for the rotation  $R_{t+1}$ , translation  $T(t+1)$ , and deformation  $D_{t+1}^i$  in the rigid and nonrigid motion representation (3), (4).

If we assume small rotation angles between frames, the rotation matrix  $R_{t+1}$  can be approximated by

$$R_{t+1} = R[\mathbf{w}(t+1)] = \begin{pmatrix} 1 & -w_z(t+1) & w_y(t+1) \\ w_z(t+1) & 1 & -w_x(t+1) \\ -w_y(t+1) & w_x(t+1) & 1 \end{pmatrix} \quad (6)$$

where  $\mathbf{w}(t+1) = (w_x(t+1), w_y(t+1), w_z(t+1))^T$ , and  $w_x(t+1), w_y(t+1), w_z(t+1)$  denote the rotation angle around the  $x, y, z$  axis, respectively, between time  $t$  and  $t+1$ . Considering the smoothness-of-motion, the rotation  $\mathbf{w}(t+1)$  currently observed depends on previous rotation  $\mathbf{w}(t)$ . Thus we introduce the term  $\Delta\mathbf{w}(t) = (\Delta w_x(t), \Delta w_y(t), \Delta w_z(t))^T$  representing the changes in rotation between time  $t$  and  $t+1$ . That is:

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \Delta\mathbf{w}(t). \quad (7)$$

Similarly, the translation  $T(t+1)$  can be expressed by

$$T(t+1) = T(t) + \Delta T(t) \quad (8)$$

where  $T(t+1) = (T_x(t+1), T_y(t+1), T_z(t+1))^T$  is a translation vector, and  $\Delta T(t) = (\Delta T_x(t), \Delta T_y(t), \Delta T_z(t))^T$  is the changes in translation between time  $t$  and  $t+1$ .

One way to describe deformation is the movement of each point. But to describe such completely unstructured motion leads the recovery process to be badly underconstrained. To transform the recovery process to a overconstrained problem, one must invoke some simplifications to nonrigid motion such as articulated motion, or bending, tapering, pinching deformation. Here we consider only the deformation along the optical axis.

$$D_{t+1}^i: (x_i(t), y_i(t), z_i(t))^T \rightarrow (x_i(t), y_i(t), z_i(t) + \Delta z_i(t))^T \quad (9)$$

where  $\Delta z_i(t)$  represent deformation of the  $i$ th point at time  $t$  along the  $z$ -axis, i.e., along the optical axis in the orthographic projection, and may be considered as a measure of the deviation from rigidity. Although this deformation model represent restricted types of nonrigid motion, empirical studies suggest that this restricted model is able to cope with

general nonrigid motion.

The deviation terms  $\Delta w(t)$ ,  $\Delta T(t)$ ,  $\Delta z_i(t)$  are introduced in the modelling of motion to represent the smoothness-of-motion constraint explicitly. In particular, the terms  $\Delta w(t)$ ,  $\Delta T(t)$  can be regarded as corresponding to the inertia of the objects, and the term  $\Delta z_i(t)$  to the elasticity of the objects.

Assuming a orthographic projection as an imaging model  $h$ , defined by

$$h : (x, y, z)^T \rightarrow (u, v)^T, \quad (10)$$

and  $u_i(t) = (u_i(t), v_i(t))^T$  and  $u_i(t+1) = (u_i(t+1), v_i(t+1))^T$  are the image coordinates corresponding to the points  $x_i(t)$ ,  $x_i(t+1)$ , respectively, then

$$u_i(t) = h[x_i(t)] = (x_i(t), y_i(t))^T \quad (11)$$

$$u_i(t+1) = h[x_i(t+1)] = h[R_{t+1} \cdot D_{t+1}^T [x_i(t)] + T^*(t) + \Delta T^*(t)] \quad (12)$$

where  $T^*(t) = h[T(t)] = (T_x(t), T_y(t))^T$ , and  $\Delta T^*(t) = h[\Delta T(t)] = (\Delta T_x(t), \Delta T_y(t))^T$ .

Using the model described above 3-D structure and motion information is computed. Our objective is to estimate 3-D information at time  $t+1$ , i.e.,  $w(t+1)$ ,  $T^*(t+1)$ ,  $z(t+1) = (z_1(t+1), z_2(t+1), \dots, z_n(t+1))^T$ , given 3-D information at time  $t$ , i.e.,  $w(t)$ ,  $T^*(t)$ , and  $z(t)$ , and the positions of the moving points at time  $t$  and  $t+1$ , i.e.,  $u_i(t)$  and  $u_i(t+1)$ . From the equations (7), (8) and (9), this requires the computation of the unknown deviation values  $\Delta w(t)$ ,  $\Delta T^*(t)$ , and  $\Delta z(t) = (\Delta z_1(t), \Delta z_2(t), \dots, \Delta z_n(t))^T$ .

In the lack of static information about the 3-D structure and motion, the initial 3-D information at time  $t=0$  are all zero, i.e., no depth and no motion are assumed. As each view of the moving objects appears, 3-D information at time  $t$  is updated so as to agree with the new frame. The update is made such that changes in 3-D information is as smooth as possible, and hence as rigid as possible. In other words, new 3-D information is obtained by the minimal change of the previous 3-D information that is sufficient to account for the new frame.

To derive equations (11) and (12), we have been making two assumptions: that the objects being observed exhibit motion represented by (3), and the images of the objects are noise free. Thus, to reduce errors introduced by these assumptions, we employ a least-squares approach which minimizes the deviation between the input frame and that predicted from the estimated 3-D information. This approach is adopted because of its robustness.

In addition, the *smoothness-of-motion* constraint means that the deviation terms  $\Delta w(t)$ ,  $\Delta T^*(t)$ , and  $\Delta z(t)$  should be small. Hence, we introduce the function  $\|P\|^2$  as a measure of the smoothness, and formulate the smoothness-of-motion constraint as a functional  $\|P\|^2$  to be minimized.

$$\|P\|^2 = \alpha \|\Delta w(t)\|^2 + \beta \|\Delta T^*(t)\|^2 + \gamma \|\Delta z(t)\|^2 \quad (13)$$

where  $\alpha, \beta, \gamma$  are scale parameters, and  $\|\cdot\|$  is a  $L_2$  norm.

Thus, to determine the new 3-D information, we choose the function  $E$  to be minimized as a sum of two terms: the first one is the difference between the predicted and input frame measured in the least square sense, and the second term is the cost function corresponding to the smoothness-of-motion constraint shown in (13).

$$E(\Delta w(t), \Delta T^*(t), \Delta z(t)) = \sum_{i=1}^n (u_i(t+1) - h[x_i(t+1)])^2 + \|P\|^2 \quad (14)$$

After the deviation terms  $\Delta w(t)$ ,  $\Delta T^*(t)$ , and  $\Delta z(t)$  have been determined with the minimization of the functional  $E$ , new 3-D information at time  $t+1$ , i.e.,  $w(t+1)$ ,  $T^*(t+1)$ ,  $z(t+1)$ , can easily be derived from the equations (3), (6)-(9). A new frame is then registered,

and the process described above repeats itself.

Note that the algorithm for the recovery of 3-D structure and motion can be formulated within the framework of regularization theory. We can consider the functional  $\|P\|^2$  as a stabilizing functional in regularization to restrict admissible solutions to space of smooth functions.

## 4. SIMULATIONS

In this section we illustrate some results of applying the recovery algorithm to both rigid and nonrigid objects. The evaluation functional  $E$  in (14) relates the nonlinearity of deviation terms  $\Delta w(t)$ ,  $\Delta T^*(t)$ , and  $\Delta z(t)$ . In general this equation can be solved by using nonlinear least squares such as Levenberg-Marquardt Method.

Because the motion smoothness constrains the magnitude of deviations, we assume the deviation terms higher than second order to be small or infinitesimal in the minimization of the evaluation functional  $E$ . This assumption eliminated thorny issues such as convergence or initial value specification that typically plagued most nonlinear optimization problems. Our implementation results demonstrated that the linear approximation affect little the integrity of the optimized solution.

In all the examples presented here, the scale parameters in (3) are  $\alpha=1$ ,  $\beta=0.01$ ,  $\gamma=0.01$ . The value of the parameters  $\alpha$  is designed to be almost  $10^2$  or  $10^3$  times larger than  $\beta$  and  $\gamma$ , since  $\Delta w(t)$  is radian while  $\Delta T^*(t)$  and  $\Delta z(t)$  are pixel value. Furthermore, the scale parameters  $\alpha, \beta, \gamma$  are chosen from the compromise among the speed of the convergence, robustness to noises, and ability of coping with nonrigid objects. The empirical studies, however, show that the sensitivity of the parameter selection to estimation results is not so strong.

1) *rigid motion* : Following Ullman[7], we generate synthetic objects containing six points: the vertices of the solid outlined pentagon, and a sixth point at the origin. The objects are three-dimensional, not merely planar. The solid line of Fig.1 illustrates the projection of this object on the  $x$ - $z$  plane. The input to the recovery process consisted of the projection of six points of the object on the  $x$ - $y$  plane. The dashed line in figure shows the estimated structure.

At frame 1 no depth is assumed, and estimated structure is flat (Fig.1(a)). The object is then rotated around the  $y$  axis. The rotation angle between  $n$  frame and  $n+1$  frame is given by

$$w_y = 2.0 + \sin(2\pi n/30) \text{ (deg)}. \quad (15)$$

Fig.1 illustrates the behavior of the recovery process. The estimated structure is almost similar to the correct structure at the frame 41, and the 3-D motion parameters are also very well estimated.

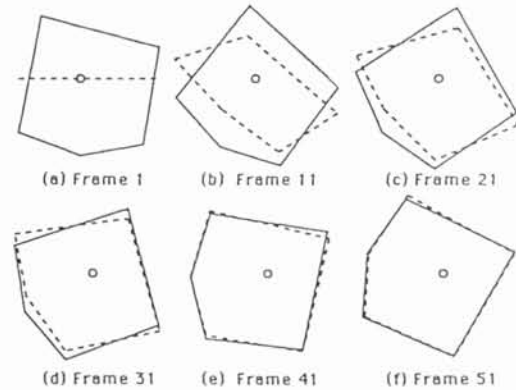


Fig.1. The recovery of a 6-point rigid object ( $x$ - $z$  plane:  $x$ -horizontal axis,  $y$ -vertical axis). The estimated structure (dashed line) is compared to the correct structure (solid line). The rotation angle of the object is a function of frame number.

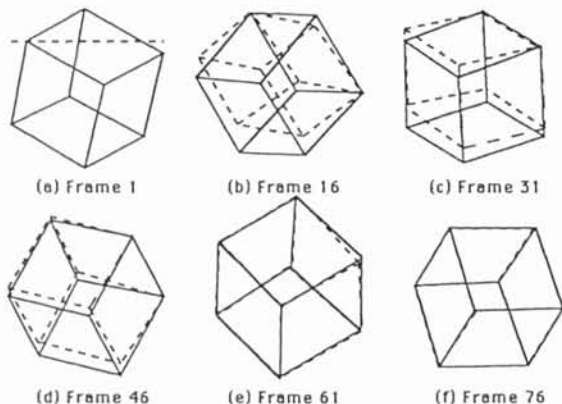


Fig.2. The recovery of a wire-frame cube ( $x-z$  plane). The rotation angle of the object is a function of frame number.

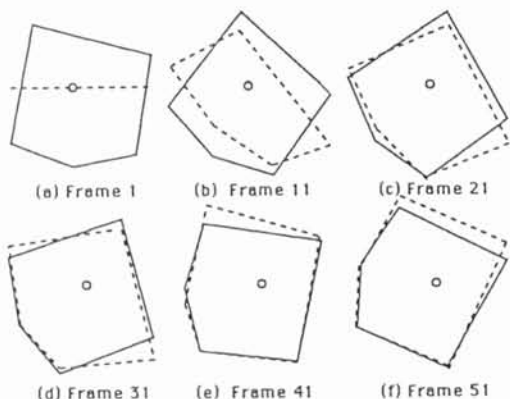


Fig.3. The recovery of a 6-point rigid object ( $x-z$  plane). The rotation angle of the object is a function of frame number. The random noise (range from -3 pixels to +3 pixels) is added.

Empirical studies show that as long as the motion is constant or smoothly changed as (15), qualitatively similar results to these shown in this paper (Fig.1-Fig.4) are obtained. It can also be seen that the rate of convergence and quality of the solution gradually deteriorates with larger angular displacement than 10 degrees. The reason for the deterioration is the disagreement with the assumption of the small rotation angle in equation (6).

Fig.2 illustrates the recovery process of wire-frame cube, with feature points on its vertices. The object is transparent, as would be the case with a true wire-frame cube. It can be seen that the scheme recovers the correct 3-D structure as 6 point case.

In order to understand the effects of noise, the random noise is added to the points of the objects. The range of the random noise is from -3 pixels to 3 pixels. This noise level can be considered to be very high, because the maximum displacement of the points is about 3.5 pixels. Even in this noisy case, the scheme can still recover the 3-D structure successfully as illustrated in Fig.3.

2) *nonrigid motion* : Fig.4 shows the example of recovering nonrigid motion from a sequence of images. The object at frame 1 is identical in shape to the object examined in Fig.1. In this case nonrigid transformation is added to the rotation of the object. The result shows that the scheme copes successfully with such nonrigid motion as well as rigid motion. This tolerance for the deviations from rigidity is also implied by the robustness to noise in the rigid motion case.

Empirical studies suggest that the scheme works well with noise and nonrigid motion, and agrees with the principle of the graceful degradation[11].

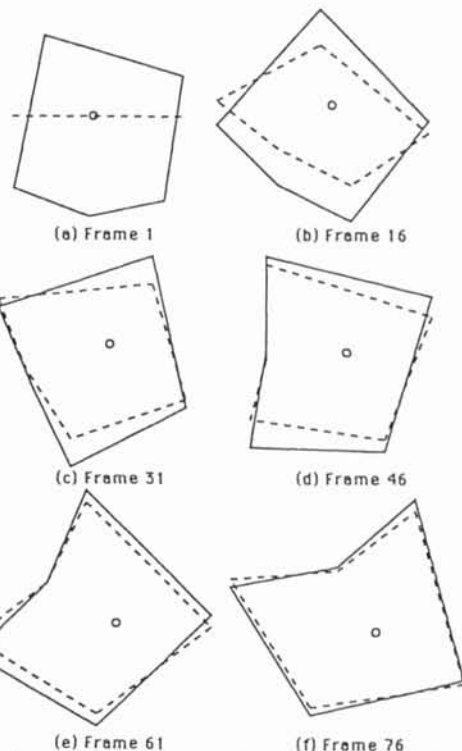


Fig.4. The recovery of a 6-point nonrigid object ( $x-z$  plane). The object is rotated by 2 degrees at a frame.

## 5. CONCLUSION

We have described the utility of an incremental approach for recovering 3-D structure and motion of rigid as well as nonrigid objects from a sequence of images. The basic idea is to successively estimate the 3-D information by using the constraint about the nature of motion, i.e., smoothness-of-motion. The use of a longer sequence can meet the two requirements: robustness to noise and ability of coping with nonrigid objects. A focus of our current work is the mathematical analysis of the scheme.

## REFERENCES

- [1] B.K.P.Horn, *Robot Vision*. Cambridge, MA and New York, NY: M.I.T Press, and McGraw-Hill, 1986.
- [2] H.Morikawa and H.Harashima, "3-D structure extraction coding of image sequences," in *Proc. Int. IEEE Conf. Acoust. Speech, Signal Process.*, M4.4, pp.1969-1972, Albuquerque, NM, April 1990.
- [3] K.Aizawa, H.Harashima and T.Saito, "Model-Based Analysis Synthesis Image Coding System for a person's face," *Signal Processing: Image Communication*, Vol.1, pp.139-152, Oct. 1989.
- [4] S.Ullman, *The Interpretation of Visual Motion*. Cambridge, MA: M.I.T Press, 1979.
- [5] Y.Yasumoto and G.Medioni, "Robust estimation of three-dimensional motion parameters from a sequence of image frames using regularization," *IEEE Trans. Pattern Ana. Machine Intell.*, vol.PAMI-8, no.4, pp.464-471, July 1986.
- [6] I.K.Sethi and R.Jain, "Finding trajectories of feature points in a monocular image sequence," *IEEE Trans. Pattern Ana. Machine Intell.*, vol.PAMI-9, no.1, pp.56-73, Jan. 1987.
- [7] S.Ullman, "Maximizing rigidity: the incremental recovery of 3-D structure from rigid and nonrigid motion," *Perception*, vol.13, pp.255-274, 1984.
- [8] M.Subbarao, "Interpretation of image flow : A spatio-temporal approach," *IEEE Trans. Pattern Ana. Machine Intell.*, vol.PAMI-11, no.3, pp.266-278, March 1989.
- [9] A.Sommerfeld, *Mechanics of Deformable Bodies*. New York: Academic Press, 1964.
- [10] E.C.Hildreth, N.M.Grzywacz, E.H.Adelson and V.K.Inada, "The perceptual buildup of three-dimensional structure from motion," A.I. Memo No.1141, Artificial Intelligence Lab., MA Inst. Technol., 1989.
- [11] D.Marr, *Vision*. San Francisco, CA: Freeman, 1982.