Integrated 3D Analysis of Flight Image Sequences *

Sanghoon Sull and Narendra Ahuja

Beckman Institute and Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
405 North Mathews Avenue
Urbana, IL 61801, USA

Abstract. This paper is concerned with three-dimensional (3D) analysis of images showing 3D motion of an observer relative to a scene. It presents an approach to recovering 3D motion and structure parameters from multiple features present in a monocular image sequence such as points, regions, lines, texture gradient and vanishing line. For concreteness, the paper focuses on flight images of a planar, textured surface. In this paper, a linear integrated estimation method using two views is developed. Then, for robust estimation, a nonlinear integrated estimation method using multiple frames is presented. The integration of information in these diverse features is carried out using minimization of image errors. To reduce computation, a sequential-batch method is used to compute motion and structure. Performance is evaluated through simulations and experiments with a real image sequence digitized from a commercially available laserdisc of films taken from flying aircrafts.

1 Introduction

In this paper, we present an approach to recovering 3D motion and structure from multiple features present in a monocular image sequence. A key feature of the approach is an integrated use of multiple image features. These features carry the motion and structure information of interest to different degrees, and have different, often complementary, strengths and shortcomings. Thus when a given feature does not contribute significantly to the estimation process, other, more pertinent features help achieve reliable estimation. The goal is to estimate motion and structure parameters such that the estimates best explain the presence of *all* of the observed image features throughout the sequence.

The identification and analysis of the relative strengths of different features for the problem at hand is a research problem in itself. In general, the available features, and sometimes even their relative merits, depend upon the scene under consideration. In this paper we focus on the problem of an observer moving above a planar surface such as while in an aircraft which is landing or taking off.

The eight steps of our algorithm [4] are shown in Fig. 1.

* The support of Defense Advanced Research Projects Agency and the National Science Foundation under grant IRI-89-02728 is gratefully acknowledged.

Lecture Notes in Computer Science, Vol. 800
Jan-Olof Eklundh (Ed.)
Computer Vision – ECCV '94

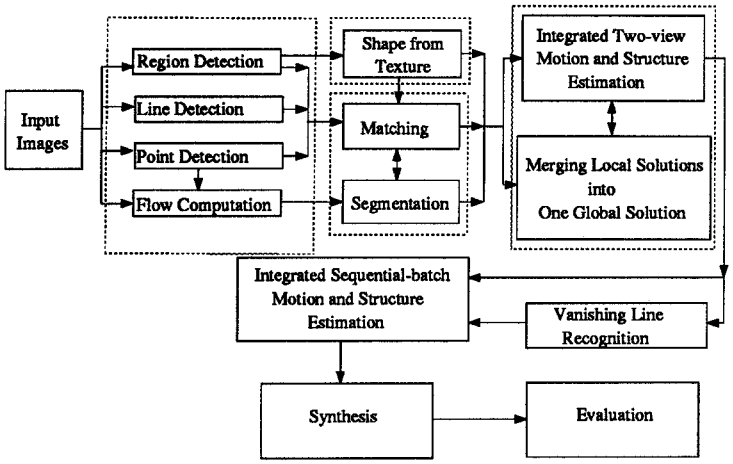

Fig. 1. A flow diagram summarizing the presented approach.

To evaluate the estimation results, we compute alignment error between estimated vanishing lines and the actual vanishing lines, compute image plane differences between observed and 3D predicted image features, and perceptually compare the synthesized sequence with the original sequence.

Sections 2 describes the mathematical formulation. The moving objects are assumed to be rigid, piecewise planar, undergoing general 3D motion, and viewed under perspective projection. Section 3 presents the experimental results.

2 Mathematical Formulation

Assume a planar surface ($aX + bY + cZ = 1$) undergoes a rigid motion described by the rotation ($\mathbf{R}$) followed by translation ($\mathbf{T}$). Let (x, y) and (x', y') be the image coordinates of a moving point on the plane at two time instants. Then, we get [1]

$$x' = \frac{a_1 x + a_2 y + a_3}{a_7 x + a_8 y + a_9}, \qquad y' = \frac{a_4 x + a_5 y + a_6}{a_7 x + a_8 y + a_9} \tag{1}$$

where $a_1, \ldots, a_9$ are expressed by using $\mathbf{R}$, $\mathbf{T}$ and $\mathbf{n}_{S,k} = [a_k, b_k, c_k]'$ (plane normal at t_k).

From Eqs. (1), we have two equations for each point correspondence:

$$x a_1 + y a_2 + a_3 - x x' a_7 - x' y a_8 - x' a_9 = 0, \tag{2}$$

$$x a_4 + y a_5 + a_6 - x y' a_7 - y y' a_8 - y' a_9 = 0. \tag{3}$$

The image error of a point correspondence i between t_k and t_{k+1} is defined as

$$E_{k,i,P} \stackrel{\text{def}}{=} \sqrt{E_{x,k,i,P}^2 + E_{y,k,i,P}^2}, \tag{4}$$

where $E_{x,k,i,P}$ and $E_{y,k,i,P}$ are defined by the errors of Eqs. (1), respectively.

For a pair (L_1, L_2) of the corresponding lines at two time instants, the 2D equations of L_1 and L_2 in the image plane are given by $A_1 x + B_1 y + C_1 = 0$

and $A_2x + B_2y + C_2 = 0$, respectively. Consider two end points $P = (x_p, y_p)$ and $Q = (x_q, y_q)$ on L_1. Let l_p and l_q be the perpendicular distances from two predicted image coordinates to L_2, respectively. Then, we have

$$l_p \stackrel{\text{def}}{=} \frac{A_2\widehat{x}_p + B_2\widehat{y}_p + C_2}{\sqrt{A_2^2 + B_2^2}}, \qquad l_q \stackrel{\text{def}}{=} \frac{A_2\widehat{x}_q + B_2\widehat{y}_q + C_2}{\sqrt{A_2^2 + B_2^2}}. \tag{5}$$

By using Eqs. (1), we have two equations for each line correspondence [4]:

$$\frac{A_2x_pa_1 + A_2y_pa_2 + A_2a_3 + B_2x_pa_4 + B_2y_pa_5 + B_2a_6 + C_2x_pa_7 + C_2y_pa_8 + C_2a_9}{\sqrt{A_2^2 + B_2^2}} = 0, \tag{6}$$

$$\frac{A_2x_qa_1 + A_2y_qa_2 + A_2a_3 + B_2x_qa_4 + B_2y_qa_5 + B_2a_6 + C_2x_qa_7 + C_2y_qa_8 + C_2a_9}{\sqrt{A_2^2 + B_2^2}} = 0. \tag{7}$$

The image error is defined as

$$E_{k,i,L} \stackrel{\text{def}}{=} \sqrt{\frac{E_{p,k,i,L}^2 + E_{q,k,i,L}^2}{2}}. \tag{8}$$

where $E_{p,k,i,L}$ and $E_{q,k,i,L}$ are defined by the residual errors of Eqs. (5), respectively.

Let M and N be the corresponding regions at two time instants. Then, we can derive the following two equations for each region correspondence [3]:

$$\frac{N_{10}}{N_{00}} - \frac{M_{10}}{M_{00}} = a_3 + (a_1 - a_9)\frac{M_{10}}{M_{00}} + a_2\frac{M_{01}}{M_{00}} - a_8\frac{M_{11}}{M_{00}} - a_7\frac{M_{20}}{M_{00}} \tag{9}$$

$$\frac{N_{01}}{N_{00}} - \frac{M_{01}}{M_{00}} = a_6 + a_4\frac{M_{10}}{M_{00}} + (a_5 - a_9)\frac{M_{01}}{M_{00}} - a_7\frac{M_{11}}{M_{00}} - a_8\frac{M_{02}}{M_{00}}, \tag{10}$$

where

$$N_{ij} \stackrel{\text{def}}{=} \int\int_N x^i y^j dxdy, \qquad M_{ij} \stackrel{\text{def}}{=} \int\int_M x^i y^j dxdy. \tag{11}$$

The image error of a region correspondence is defined as

$$E_{k,i,R} \stackrel{\text{def}}{=} \sqrt{E_{x,k,i,R}^2 + E_{y,k,i,R}^2}, \tag{12}$$

where $E_{x,k,i,R}$ and $E_{y,k,i,R}$ are defined by the residual errors of Eqs. (9) and (10), respectively.

If we consider a 3D plane given by $aX + bY + cZ = 1$, the vanishing line is expressed as $ax + by + c = 0$ in terms of the image coordinates x and y.

Linear Integrated Estimation Using Two Frames For each pair of successive frames, we first solve for the intermediate parameters $a_1, \ldots, a_9$. To solve the six equations for points, lines and regions simultaneously (Eqs. (2), (3), (6), (7), (9) and (10)), we linearly compute the eight coefficients $a_1, \ldots, a_8$ with a_9 set to 1, since a_9 can have any value. Then, we non-iteratively compute the parameters for motion and plane normal by using the existing method in [1].

Nonlinear Integrated Estimation Using Multiple Frames For robust estimation, we minimize image errors between the observed features and those corresponding to the estimates. This makes estimation a nonlinear problem.

To obtain a measure of inconsistency between the estimates and the different features used, we define the average image error between t_k and t_{k+1} as

$$\overline{E_{k_1,k_2,I}} \stackrel{\text{def}}{=} \sqrt{\sum_{k=k_1}^{k_2-1} \frac{\sum_{i=1}^{n_P(k)} E_{k,i,P}^2 + \sum_{i=1}^{n_L(k)} E_{k,i,L}^2 + \sum_{i=1}^{n_R(k)} E_{k,i,R}^2}{\lambda_{k_1,k_2,I}}}, \tag{13}$$

where $E_{k,i,P}$, $E_{k,i,L}$ and $E_{k,i,R}$ are defined in Eqs. (4), (8) and (12), and $n_P(k)$, $n_L(k)$ and $n_R(k)$ are the numbers of point, line and region correspondences for a planar patch, respectively. In the above definition, $\lambda_{k_1,k_2,I}$ represents the normalizing factor. Note that each error term has the same unit.

When an image is paired with its predecessor and successor images in the sequence, the resulting structure parameters will in general not be identical. Thus, the requirement of such *consistency of structure* must be explicitly enforced. Tracking of each feature is not necessary to enforce the structure consistency since all features are on a plane and consistency means conservation of the plane normal. Let M_k be set of motion parameters between t_k and t_{k+1}. For each window of batch size N (3 in the experiment), define the following objective function to be minimized with respect to motion (M_k) and unit plane normals ($\mathbf{n}_{S,k}$):

$$G(M_k, S_k) \stackrel{\text{def}}{=} \sum_{k=0}^{N-2} \overline{E_{0,N-1,I}} + \sum_{k=0}^{N-1} (E_{k,V}^2 + E_{k,T}^2) \tag{14}$$

where $E_{k,V}$ and $E_{k,T}$ represent the penalty terms which make plane normals stay within a certain range of the initial values computed from recognized vanishing lines and texture gradients if they are available.

This objective function combines the contributions of multiple features to the scene characteristics to be estimated. For each overlapping batch of size N, we iteratively minimize Eq. (14) with respect to M_k and $\mathbf{n}_{S,k}$. Since the number of the iteration variables is large, it is reduced by using $\mathbf{n}_{S,k+1} = \mathbf{R}_{k,k+1}\mathbf{n}_{S,k}$, Eqs. (2), (3), (6), (7), (9) and (10)). That is, given $\mathbf{n}_{S,0}$ and $\mathbf{R}_{k,k+1}$ (interframe rotations) at each iteration step, the other unit plane normals and translations are linearly computed. We can thus reduce the number of variables of the objective function from $8(N-1)$ to $2+3(N-1)$:

$$\min_{\mathbf{n}_{S,k},k=0,\ldots,N-1,\ M_k,k=0,\ldots,N-2} G(M_k, \mathbf{n}_{S,k}) = \min_{\mathbf{n}_{S,0},\mathbf{R}_{k,k+1},k=0,\ldots,N-2} G(\mathbf{n}_{S,0}, \mathbf{R}_{k,k+1}). \tag{15}$$

For monocular sequences, we have the problem of unknown scale for the estimated structure [2]. The scale factor of any two consecutive images depends on the scale factor of the first two images. Then, translations cannot be linearly computed even though $\mathbf{n}_{S,0}$ and $\mathbf{R}_{k,k+1}$ for each k are given, resulting in an increase of the parameter space of iteration. This problem is also avoided by linking multiple frames through the unit surface normals in Eq. (14).

To reduce computation, a sequential-batch method is used. Motion parameters obtained from the overlapping batches are sequentially updated.

3 Experimental Results

Average Estimation Error from Two Views Performance of the linear integrated estimation method was compared with those using single features.

For each 12 feature correspondences of points, lines and regions, a plane normal and 3D coordinates of each type of features are randomly generated at each trial. The image coordinates of the features are quantized to the nearest integer for each resolution. The relative error of a vector is defined by the Euclidean norm of the error vector divided by the Euclidean norm of the correct vector. The plane normal $\mathbf{n}_S$ is scaled to the unit vector. The error represents average errors over 50 random trials.

Figure 2 shows the average relative error of $\mathbf{n}_S$ for a given set of motion parameters ($\mathbf{n}_\omega = [0.5774, 0.5774, 0.5774]'$, $\omega = 4$ deg and $\mathbf{T} = [0.2, 0.2, 0.2]'$) for various resolutions. We obtained graphs similar to Fig. 2 for motion estimates.

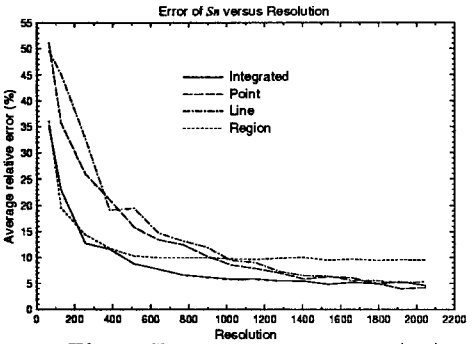

Fig. 2. Error of plane normal ($\mathbf{n}_S$).

For points and lines, the performance is similar, and the estimates becomes more accurate as the resolution becomes higher. Regions give the most reliable estimates from low to mid resolutions although the accuracy of the estimates does not improve at the higher resolutions. This is expected since the approximations are made when the region-based equations are derived. (It is not difficult to derive the exact nonlinear method for regions.) The good performance of the regions at the lower resolutions is due to the robustness of the lower order moments used in Eqs. (9) and (10) to the quantization errors of the region boundary.

From these simulations, we see that the integrated estimation gives the robust estimates at all resolutions. Note that we assumed the perfect extraction and matching of features up to the quantization errors.

Runway Sequence We derived a sequence of 34 frames from a commercially available CAV laserdisc. The digitized resolution was 640 by 480. This is a challenging sequence since the images contain partially or completely occluded vanishing lines and there is reflection of the ground on the bottom of the airplane.

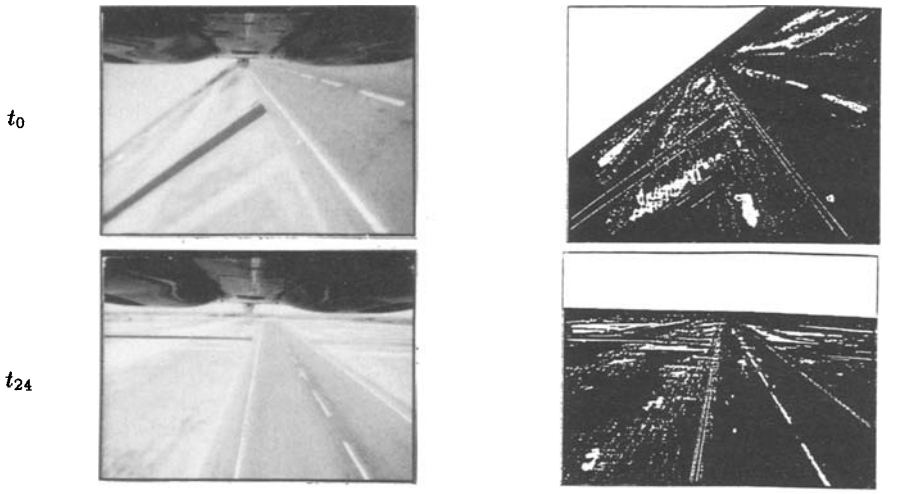

Fig. 3. Experiments with the runway sequence: (a) Input image sequence (b) Synthesized image sequence.

For this sequence, lines were important for reliable estimation since several good line features exist in the images. Estimated vanishing lines are in a good agreement with the actual vanishing lines in the image plane. The average image error $\overline{E_{0,33,I}}$ computed by Eq. (13) is 0.86 pixels. The error is less than one pixel, which is satisfactory. The original sequence and the resulting synthesized sequence are presented in Fig. 3(a) and (b). If we watch the synthesized sequence as they are played on a monitor, we perceive the same motion and structure from them in an informal viewing as in the original image sequence [2].

References

1. R.Y. Tsai and T.S. Huang, "Estimating 3D motion parameters of a rigid planar patch, II: SVD," *IEEE Trans. ASSP*, vol. 30, no. 4, pp. 525-534, Aug. 1982.
2. N. Cui, J. Weng and P. Cohen, "Extended structure and motion analysis from monocular image sequences," in *Proc. 3rd ICCV*, Osaka, Japan, pp. 222-229, 1990.
3. S. Sull and N. Ahuja, "Segmentation, matching and estimation of structure and motion of textured piecewise planar surfaces," in *Proc. IEEE Workshop on Visual Motion*, Princeton, NJ, pp. 274-279, Oct. 1991.
4. S. Sull, "Integrated 3D analysis and analysis guided synthesis," Ph.D dissertation, University of Illinois, Urbana-Champaign, 1993.

[2] A video tape showing the original image sequence and the synthesized sequences is available.