Estimation of motion and structure of planar surfaces from a sequence of monocular images

Sanghoon Sull and Narendra Ahuja *
Beckman Institute and Coordinated Science Laboratory
University of Illinois, Urbana, IL 61801

Abstract

We present an algorithm which estimates 10 parameters for motion and structure of a rigid planar patch given point correspondences in a monocular image sequence under perspective projection. The rotational velocity is assumed constant and the rotation center arbitrary. The algorithm mainly consists of two steps. First, the 3D space of $(\omega_x, \omega_y, \omega_z)$ is searched exhaustively and for each $(\omega_x, \omega_y, \omega_z)$ we compute linearly all the other parameters with the value of an objective function. Some of the $(\omega_x, \omega_y, \omega_z)$ and the corresponding structure values are used as the initial guesses in the second step. The objective function is iteratively minimized with respect to five variables for rotation and structure. The solution corresponding to the global minimum is used to obtain least squares estimates of the remaining unknowns, for translation and rotation center. We have experimentally found that the objective function converges well so that we do not have to search the 3D space densely. Results are presented for three image sequences, two simulated and one real.

1 Introduction

The motion algorithms available for two perspective views[1] suffer from quantization errors especially when the moving object is small, and located near the optical center, and the motion is difficult. To reduce this sensitivity, additional constraints are used such as restrictions on motion and structure, and the use of many frames. We are concerned with one such approach.

Since the problem of interpreting image sequence under perspective projection becomes nonlinear, it is necessary to have a good initial guess. The estimates from linear two-view algorithms may not be sufficiently reliable to be used as initial guess to this nonlinear problem. Several algorithms for image sequence have been developed under the assumption of constant motion. Since the iteration is over a large number of unknowns, it is difficult to ensure that the global solution is reached. It may not be clear how to optimally use arbitrary numbers of features and frames. The assumption of orthographic projection, whenever valid, is helpful [2].

*The support of Air Force Office of Scientific Research under grant AFOSR90-0061 is gratefully acknowledged.

2 Motion model

We are given K images of a rigid plane under perspective projection and assuming constant rotational and translational velocity. The initial plane equation in 3D is $aX + bY + cZ = 1$ with $a^2 + b^2 + c^2 = 1$ as a scale factor. So, we can parametrize the unit surface normal $\vec{n}_S = (a, b, c)'$ with two angles ϕ and ψ.

Let $\vec{Q}_0$ be the 3D coordinates of the rotation center at t=0. $\mathbf{R}$ is the rotation of the object about the axis $\vec{n}_\omega = [n_x, n_y, n_z]'$ by ω during one time unit. Then, $\mathbf{R}^k$ is the rotation between t=0 and t=k. And, we define $[\omega_x, \omega_y, \omega_z]' \stackrel{\text{def}}{=} [n_x\omega, n_y\omega, n_z\omega]'$.

Then, the 3D motion equation of a point p at t=k becomes

$$\vec{X}_p^k = \vec{T}_k + \mathbf{R}^k \vec{X}_p^0 \tag{1}$$

where

$$\vec{T}_k = (\mathbf{I} - \mathbf{R}^k)\vec{Q}_0 + k\vec{t} \tag{2}$$

$$= [T_x(k), T_y(k), T_z(k)]'. \tag{3}$$

Here, $\vec{t}$ represents the translation. To resolve the ambiguity regarding the location of the rotation center along the rotation axis, $\vec{Q}_0$ is described by two coordinates α and β to describe $\vec{Q}_0$.

$$\vec{Q}_0 = \begin{bmatrix} 1 - \frac{n_x^2}{1+n_z} & \frac{-n_x n_y}{1+n_z} \\ \frac{-n_x n_y}{1+n_z} & 1 - \frac{n_y^2}{1+n_z} \\ -n_x & -n_y \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{4}$$

x_p^k and y_p^k represent the image coordinates of the point p at t=k.

The resulting objective function becomes

$$G_s = \sum_{k=1}^{K} \sum_{p=1}^{N} (x_p^k Z_{pk} - X_{pk})^2 + (y_p^k Z_{pk} - Y_{pk})^2 \tag{5}$$

where $(() \stackrel{\text{def}}{=} (k))$

$X_{pk} = (r_{11}() + aT_x())x_p^0 + (r_{12}() + bT_x())y_p^0 + (r_{13}() + cT_x())$
$Y_{pk} = (r_{21}() + aT_y())x_p^0 + (r_{22}() + bT_y())y_p^0 + (r_{23}() + cT_y())$
$Z_{pk} = (r_{31}() + aT_z())x_p^0 + (r_{32}() + bT_z())y_p^0 + (r_{33}() + cT_z()).$

We derive a relationship between $\mathbf{R}^k$ and $\vec{n}_S$ so that a good estimate of $\vec{n}_S$ can be linearly computed given $\mathbf{R}$. Hence, the effective search space is reduced to have a dimension of 3 and it becomes practical to search the space exhaustively to find the global minimum. An estimate of $\mathbf{A}^k$ between time 0 and k can be obtained

from 4 or more point correspondences using a least squares method where $\mathbf{A}^k$ is the matrix $\mathbf{A}$ defined in [1] between t=0 and t=k. Then, we have

$$\mathbf{U}_k \vec{n}_S = 0 \tag{6}$$

where

$$\mathbf{U}_k = \begin{bmatrix} (\vec{r}_1(k) \times \vec{a}_1(k))' \\ (\vec{r}_2(k) \times \vec{a}_2(k))' \\ (\vec{r}_3(k) \times \vec{a}_3(k))' \\ (\vec{r}_2(k) \times \vec{a}_1(k) + \vec{r}_1(k) \times \vec{a}_2(k))' \\ (\vec{r}_2(k) \times \vec{a}_3(k) + \vec{r}_3(k) \times \vec{a}_2(k))' \\ (\vec{r}_3(k) \times \vec{a}_1(k) + \vec{r}_1(k) \times \vec{a}_3(k))' \end{bmatrix} \tag{7}$$

$\vec{r}_1(k)$,$\vec{r}_2(k)$,$\vec{r}_3(k)$ and $\vec{a}_1(k)$,$\vec{a}_2(k)$,$\vec{a}_3(k)$ are the row vectors of $\mathbf{R}^k$ and $\mathbf{A}^k$, respectively. We can prove the rank of $\mathbf{U}_k$ is two if $\vec{T}_k$ is not zero. For K frames together, we get

$$\mathbf{U}_K \vec{n}_S = 0 \tag{8}$$

where $\mathbf{U}_K$ is 6K × 3 matrix.

Now, we summarize the algorithm.

ALGORITHM

0) Compute $\mathbf{A}^k$ between t=0 and k for $k = 1, \cdots, K$ from 4 or more point correspondences using a least squares method.

1) For each quantized set of values of ω_x, ω_y and ω_z, compute the smallest eigenvalue and corresponding eigenvector $\vec{n}_S$ of $\mathbf{U}'_K\mathbf{U}_K$ where $\mathbf{U}_K$ is defined in Eq.(8). If the smallest eigenvalue is less than the predetermined threshold, compute G_s (See Eq.(5)).

2) From Step 1), choose the set of values of ω_x, ω_y and ω_z with ϕ and ψ from the corresponding eigenvector $\vec{n}_S$ which yield the predetermined number of the smallest values of G_s. Using those candidates as initial guesses, minimize iteratively G_s w.r.t. 5 variables of ω_x, ω_y, ω_z, ϕ and ψ. Note that given ω_x, ω_y, ω_z, ϕ and ψ, we can compute linearly the other variables such as $\vec{t}$ for translation and α and β for the initial rotation center.

3 Experimental results

The coordinates are rounded off to integers for experiment 1 and 2. Units for translation and rotation angle are in meters and radians, respectively. Focal length was 35mm and the resolution was 512 by 512 in all cases. The field of view of the camera was 8.4 degrees. For Step 1) of our algorithm, ω_x, ω_y and ω_z are quantized with the resolution of 4°. We also applied the linear two-view algorithm of [1] between frame 0 and frame k, but the resulting values were not good.

3.1 Experiment 1.

16 frames and six feature points were used. The coordinates at t=0 are (230,300), (250,350), (300,230), (280,250), (290,230) and (300,350). Note that the points are located near the image center.

Actual values:

$\vec{n}_w = [0.5774, 0.5774, 0.5774]'$ and $\omega = 0.03$

$\vec{n}_S = [0.65, 0.3, 0.7]'$, $\vec{t} = [-0.005, 0.005, 0.02]'$

$\vec{Q}_0 = [-0.4767, -0.4767, 0.9533]'$

Estimated values:

$\vec{n}_w = [0.5905, 0.5646, 0.5766]'$ and $\omega = 0.0297$

$\vec{n}_S = [0.6484, 0.3166, 0.6924]'$

$\vec{t} = [-0.0051, 0.0049, 0.0215]'$

$\vec{Q}_0 = [-0.5000, -0.4638, 0.9663]'$

3.2 Experiment 2.

30 frames and six feature points were used. The coordinates at time=0 are (130,90), (250,150), (200,40), (280,50), (290,140) and (250,150).

Actual values:

$\omega = 0$, $\vec{n}_S = [0, 0, 1]'$, $\vec{t} = [0, 0.004, 0.01]'$

Estimated values:

$\vec{n}_w = [-0.9914, -0.1311, 0.0046]'$ and $\omega = -0.0007$

$\vec{n}_S = [-0.0035, 0.0138, 0.9999]'$

$\vec{t} = [-0.0001, 0.0048, 0.0096]'$

$\vec{Q}_0 = [0.0920, -0.6990, -0.0933]'$

3.3 Experiment 3.

Feature points were extracted [2], and nine of these were chosen from the larger arm of the robot as shown in Fig.1. Ten frames were used and no attempt was made to calibrate the camera. For $\vec{Q}_0$, we do not have the exact value, but it appears to be reasonable.

Actual values:

$\vec{n}_w = [-0.0214, 0.618, 0.786]'$ and $\omega = -0.0655$

$\vec{n}_S = [-0.0214, 0.618, 0.786]'$, $\vec{t} = [0, 0, 0]'$

Estimated values:

$\vec{n}_w = [0.0075, 0.6461, 0.7632]'$ and $\omega = -0.0661$

$\vec{n}_S = [0.0025, 0.6419, 0.7668]'$

$\vec{t} = [0.00008, -0.00001, 0.00007]'$

$\vec{Q}_0 = [-0.0010, -0.6990, 0.5917]'$

References

[1] R.Y. Tsai and T.S. Huang, "Estimating 3D Motion Par.of a Rigid Planar Patch, II: SVD", *IEEE Trans. ASSP*, pp.525-534, August, 1982

[2] C. H. Debrunner, *Structure and Motion from Long Image Sequences*. Ph.D thesis, UIUC, 1990

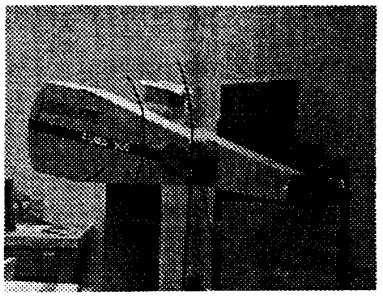

Figure 1: The initial frame superposed by trajectories