MVA'94 IAPR Workshop on Machine Vision Applications Dec. 13-15, 1994, Kawasaki

A statistically unbiased method for computing the rotation angles from image sequences

Han Wang Stan Z Li E K Teoh
School of Electrical & Electronic Engineering
Nanyang Technological University, Singapore 2263
e-mail: hw@ntuix.ntu.ac.sg phone: (+65)799-1253 fax: (+65)791-2687

ABSTRACT

Given a set of points correspondence from two views, there exists infinite number of solutions for interpreting the angular motion–this is called the bas-relief ambiguity. Solving this problem relies on either prior knowledge of the motion or further projection such as 3 views. It is shown that the unique solution exists from only one pair of correspondence with the prior knowledge of known rotation axis. This paper presents two new algorithms that compute the rotation angle from two and three views using least squares approaches: in the case of 2-view, we compute the rotation angle uniquely, an improvement over the Harris algorithm which produces dual solutions; in the 3-view algorithm, we compute the rotation axis first using the numerical approach and then recover the rotation angle using the 2-view algorithm. A robust method for dealing with the outliers (false matches) is also reported.

1 Introduction

A problem in computer vision is to recover the motion from two dimensional projections, this is called structure-from-motion (SFM). Rigid motion is decomposed into two parts, the translation and the rotation. Of the two, the rotational motion is relatively difficult to compute. Ambiguity exists when the interpretation of the rotation axis is deep (far away) or shallow (close). This is called the bas-relief ambiguity[2]. The solution to the rotation requires additional information of either *a priori* knowledge of the direction of the rotation axis.

Harris[1] showed a 2-view algorithm that can compute the rotational motion from a quadratic equation. This algorithm gave dual solutions and requires an additional phase to determine which one is the wanted solution.

This paper is organised as follows. A least squares approach is derived after the study of the uniqueness of rotation from a single pair of correspondence. In the case of three views, the Newton's method is used to compute the rotation axis and then the rotation angle is recovered using the least square approach. Experiments using real image sequences are presented. Lastly, a robust M-estimation (maximum likelihood type estimation) method is presented to reject outliers when there are false matches.

2 The 2-View Algorithm

We discuss in this section algorithm that recovers rotation angle from 2 views with the constraint of known rotation axis given that the correspondence from two images is provided correctly. Assuming orthographic projection (also known as parallel projection), the image plane shares coordinates with the world X and Y coordinates. This is a special case of the perspective projection when the focal length is at infinity.

Let $P_i = (X_i, Y_i, Z_i)$ denote a point on the object and this point is projected onto the image plane at $p_i = (x_i, y_i)$, hence $(x_i, y_i) = (X_i, Y_i)$. This implies that the distance (Z_i) from the object to the image plane will not have an effect to the motion. This assumption holds for weak perspective when the object is placed a distance very much larger than its own size.

Further, we denote the motion by a rotation, specified by R, followed by a translation, specified by $\mathbf{t} = (t_x, t_y, t_z)$, hence,

$$\begin{pmatrix} X_i' \\ Y_i' \\ Z_i' \end{pmatrix} = R \begin{pmatrix} X_i \\ Y_i \\ Z_i \end{pmatrix} + \mathbf{t} \qquad (1)$$

where $(X_i', Y_i', Z_i')^\top$ is the corresponding point of $(X_i, Y_i, Z_i)^\top$ after motion and

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}.$$

Equation (1) are in fact three linear equations of which the first two are used since the depth is not recoverable. Replacing $(X'_i, Y'_i)^\top$ and $(X_i, Y_i)^\top$ with $(x'_i, y'_i)^\top$ and $(y_i, y_i)^\top$, we have

$$x'_i = r_{11}x_i + r_{12}y_i + r_{13}Z_i + t_x, \tag{2}$$

and

$$y'_i = r_{21}x_i + r_{22}y_i + r_{23}Z_i + t_y \tag{3}$$

The translational components t_x and t_y can be dropped if we do the following manipulation: find a point (x_1, y_1) on the object, and set the coordinate origin to this point; after the motion, this point becomes (x'_1, y'_1) and we move it to the origin again. Making use of the identities of orthogonal matrix:

$$\begin{aligned} r_{11}r_{23} - r_{21}r_{13} &= -r_{32} \\ r_{12}r_{23} - r_{22}r_{13} &= r_{31}, \end{aligned} \tag{4}$$

we now have

$$r_{23}x' - r_{13}y' + r_{32}x - r_{31}y = 0, \tag{5}$$

This equation shows that under orthographic projection, there is an infinite number of solutions from 2 views regardless the number of correspondences (Huang and Lee [2]. Introducing the constraint of known axis, we show that a unique solution exists by using only one pair of correspondence.

Let $\ell = (\ell_x, \ell_y, \ell_z)^\top$ denote the unit vector of the rotation axis, the rotation matrix R of angle θ around ℓ can be represented by

$$R = \begin{pmatrix} C + \ell_x^2 V & \ell_x\ell_y V - \ell_z S & \ell_x\ell_z V + \ell_y S \\ \ell_y\ell_x V + \ell_z S & C + \ell_y^2 V & \ell_y\ell_z V - \ell_x S \\ \ell_z\ell_x V - \ell_y S & \ell_z\ell_y V + \ell_x S & C + \ell_z^2 V \end{pmatrix} \tag{6}$$

where $C = \cos\theta, S = \sin\theta$, and $V = 1 - \cos\theta$ (see [3] for details). Substitute the elements from equation (6) into equation (5) and rearrange it, we obtain,

$$a(1 - \cos\theta) + b\ \sin\theta = 0, \tag{7}$$

where

$$\begin{aligned} a &= \ell_z(\ell_y(x + x') - \ell_x(y + y')) \\ b &= \ell_x(x - x') + \ell_y(y - y'). \end{aligned} \tag{8}$$

Since $\tan\frac{\theta}{2} = \frac{1-\cos\theta}{\sin\theta}$, equation (7) evolves to

$$(a\ \tan\frac{\theta}{2} + b)\ \sin\theta = 0. \tag{9}$$

In equation (9), $\sin\theta \neq 0$ since the assumption of known axis effectively implies that $\theta \neq 0$. Hence,

$$\theta = 2\arctan(\frac{-b}{a}). \tag{10}$$

The above equation establishes the relation of the rotation angle from two views using only one correspondence with the *priori* knowledge of rotation axis.

Subject to noise, equation (9) will not be valid, hence an estimation method is required to find the solution where the squared error (given below) can be kept minimal with a large number of observations (correspondences),

$$\mu(\theta) = \sum_{i=1}^{n}\{A_i \tan(\theta/2) + B_i\}^2 \tag{11}$$

We can indeed find a unique solution[8] that is

$$\theta_0 = 2\arctan(-\frac{\sum A_i B_i}{\sum A_i^2}) \tag{12}$$

where

$$\begin{aligned} A_i &= \ell_z(\ell_y(x_i + x'_i) - \ell_x(y_i + y'_i)) \\ B_i &= \ell_x(x_i - x'_i) + \ell_y(y_i - y'_i). \end{aligned} \tag{13}$$

3 The 3-View Algorithm

It is shown in this section that the rotation axis can be recovered from three frames. We assume that the object rotates the same amount of angles between frames. This assumption is valid when the object has sufficient inertia or the sampling interval is reasonably small. Let the double prime denote the coordinates from the third frame. From (12) we have

$$\tan\frac{\theta}{2} = -\frac{\sum A_i B_i}{\sum A_i^2} = -\frac{\sum A'_i B'_i}{\sum (A'_i)^2} \tag{14}$$

where

$$\begin{aligned} A'_i &= \ell_z(\ell_y(x'_i + x''_i) - \ell_x(y'_i + y''_i)) \\ B'_i &= \ell_x(x'_i - x''_i) + \ell_y(y'_i - y''_i). \end{aligned}$$

Rearrange (14), we have

$$\sum(A_i^2)\sum(A'_iB'_i) - \sum(A'_i)^2\sum(A_iB_i) = 0. \tag{15}$$

Extract ℓ_z from A_i and A'_i, resulting in $A_i = \mathcal{A}_i\ell_z$ and $A'_i = \mathcal{A}'_i\ell_z$; substitute them into (15), canceling ℓ_z, it gives

$$\sum(\mathcal{A}_i^2)\cdot\sum(\mathcal{A}'_iB'_i) - \sum(\mathcal{A}'_i)^2\cdot\sum(\mathcal{A}_iB_i) = 0. \tag{16}$$

Notice that $\mathcal{A}_i$, $\mathcal{A}'_i$, B_i and B'_i consist of only two variables of ℓ_x and ℓ_y. The correspondence between frame one and frame three results in twice the rotation angle, hence

$$\tan\theta = -\frac{\sum A_i'' B_i''}{\sum (A_i'')^2}. \tag{17}$$

Using the identities of $\tan\theta = \frac{2\tan\frac{\theta}{2}}{1-\tan^2\frac{\theta}{2}}$ and $\ell_z^2 = 1 - \ell_x^2 - \ell_y^2$, it is analogous to derive from (17) the following equation

$$\begin{aligned} &(1-\ell_x^2-\ell_y^2)[(\sum \mathcal{A}_i^2)^2 \sum \mathcal{A}_i'' B_i'' - \\ &2\sum \mathcal{A}_i^2 \sum (\mathcal{A}_i'')^2 \sum \mathcal{A}_i B_i] - \\ &(\sum \mathcal{A}_i B_i)^2 \sum \mathcal{A}_i'' B_i'' \quad = \quad 0 \end{aligned} \tag{18}$$

Let $f_1(\ell_x, \ell_y)$ and $f_2(\ell_x, \ell_y)$ denote the left hand side of (16) and (18), hence

$$\mathbf{f} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} = \mathbf{0}, \tag{19}$$

which is a nonlinear equation of order 8 and the 2-dimensional Newton's method converges quickly,

$$\begin{pmatrix} \ell_x \\ \ell_y \end{pmatrix}^{(k+1)} = \begin{pmatrix} \ell_x \\ \ell_y \end{pmatrix}^{(k)} - \mathbf{J}^{-1}\mathbf{f}^{(k)} \tag{20}$$

where $\mathbf{J}$ is the Jacobian. ℓ_z can be computed using $\ell_z = \pm\sqrt{1-\ell_x^2-\ell_y^2}$. Ullman[6] and Huang[2] indicated that point correspondences from 3 views allow two solutions to motion and one is the reflection of the other with respect to the image plane.

Having recovered the rotation axis, the rotation angle can be computed efficiently using the least square approach since the summation terms in equation (12) have already been computed.

4 Experiments

Experiments were conducted using a sequence of images with the objects rotating on a turntable. The camera looked down on the turntable with an angle of 30 degrees. Feature points were detected using the Wang-Brady corner detector [7]. Figure 1 was obtained from using equation (12) and the results is indeed smooth and robust.

Figure 2 shows the results of the 3-view algorithm for recovering of the rotation axis.

5 Dealing with Outliers

Previously, we have assumed that there are no mistakes in finding corresponding matches. This as-

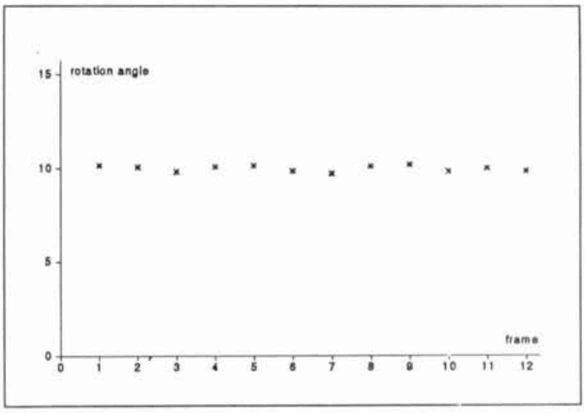

Figure 1: *Rotations recovered from the 2-view algorithm.*

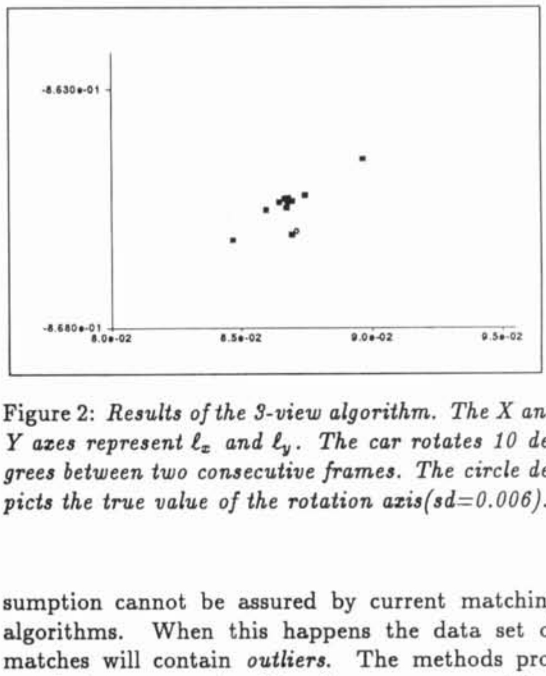

Figure 2: *Results of the 3-view algorithm. The X and Y axes represent ℓ_x and ℓ_y. The car rotates 10 degrees between two consecutive frames. The circle depicts the true value of the rotation axis(sd=0.006).*

sumption cannot be assured by current matching algorithms. When this happens the data set of matches will contain *outliers*. The methods proposed above can get arbitrarily wrong even with a small percentage of outliers. We use an M-estimator from robust statistics methods [5] for solving this problem.

The idea is to introduce an adaptive interaction (weight) function h into equation (11) in the following way. Consider the individual error terms

$$\eta_i = A_i \tan(\theta/2) + B_i.$$

We want to give a larger weight $h(\eta_i)$ to the term containing A_i and B_i if η_i is relatively small so that the i-th match is likely an "inlier"; or a smaller weight otherwise. In Tukey's biweight [5], the function h is defined as

$$h(\eta_i) = \begin{cases} \left(1-\left(\frac{\eta_i}{cS}\right)^2\right)^2 & \text{if } \left|\frac{\eta_i}{cS}\right| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

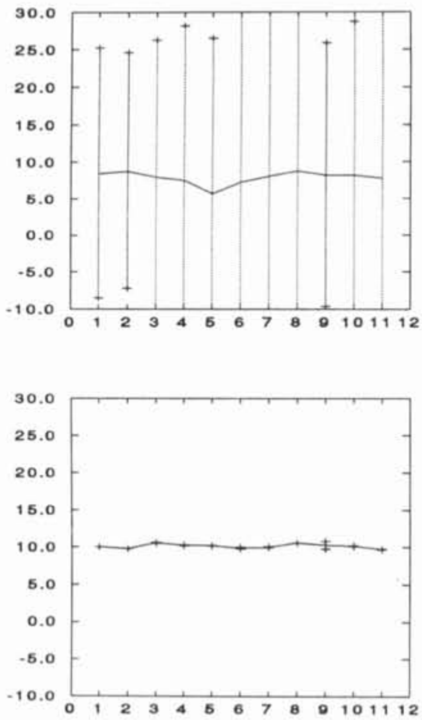

Figure 3: Rotation angles computed from corresp-ndence data containing 20% of outliers using the the 2-view algorithm (upper) and the M-estimator (lower). The abscissa represents frame number and the ordinate rotation angle. The solid lines represent the mean angle and the vertical dashed lines indicate the standard deviation computed for each frame.

where S is an estimate of spread and c a constant. A choice may be

$$S = \text{median}\{|\eta_i - \text{median}\{\eta_i\}|\}$$

with $c = 1.4826$. The M-estimate for our problem is defined as the solution of the following fixed-point equation:

$$\theta = 2\arctan\left(-\frac{\sum_{i=1}^{n} h_i A_i B_i}{\sum_{i=1}^{n} h_i A_i^2}\right) \tag{22}$$

Our experiments show that this scheme is reliable when outliers are present (see Figure 3).

6 Conclusion

The 2-view algorithm computes reliably the rotation angle and the requirement for camera intrinsics is relaxed due to the assumption of parallel projection and the object is placed at a distance from the viewer. In many applications, the rotation axis is provided or its direction is known within certain range. Experiments show that accurate measurements of the rotation axis is not necessary. In fact, the pose of the camera in the above experiments was estimated manually. No calibration was carried out!

The 3-view algorithm requires an estimation of the direction of the rotation axis for its initial value when the numerical method is employed. With given initial values, the algorithm converges fast and accurately. A new algorithm has been developed [4] in a least square approach which utilises a one dimensional search to locate the minimal energy.

The two and three view algorithms require robust correspondence among frames. In the presence of false matches, the algorithm works less reliably. The robust M-estimation method has the added advantage of rejecting outliers and the results are very promising.

Acknowledgments—Authors thank Eric Sung, Chan Kap Luk and Professor D Mital for their valuable comments, discussion and kind support.

References

[1] C Harris. Resolution of the bas-relief ambiguity in structure-from-motion under orthographic projection. In *Proc. of the British Machine Vision Conference*, pages 67–72, Oxford, UK, September 1990.

[2] T S Huang and C H Lee. Motion and structure from orthographic projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11(5):536–540, May 1989.

[3] K Kanatani. *Geometric Computation for machine vision.* Oxford Press, 1993.

[4] E Sung and H Wang. Computation of rotation from perspective views. In *Proc. ICARCV'94: 3nd International Conference on Automation, Robotics, and Computer Vision*, Singapore, Nov. 1994.

[5] J. W. Tukey. *Explortary Data Analysis.* Addison-Wesley, Reading, MA, 1977.

[6] S. Ullman. *The interpretation of visual motion.* MIT Press, New Jersey, 1979.

[7] Han Wang and M Brady. A practical solution to corner detection. In *Proc. of ICIP'94: IEEE Int. Conf. on Image Processing*, Austin, 1994.

[8] Han Wang and Stan Z Li. Solving the Bas-relief ambiguity. In *Proc. of ICIP'94: IEEE Int. Conf. on Image Processing*, Austin, 1994.