Structure from Constrained Motion Using Point Correspondences

T. N. Tan, G. D. Sullivan and K. D. Baker

Intelligent Systems Group
Department of Computer Science
University of Reading, ENGLAND

ABSTRACT

Mobile objects are frequently confined to move on a ground surface which, locally at least, is approximately flat. Existing structure-from-motion (SFM) algorithms make no use of real-world motion constraints of this kind. In this paper we present a simple novel SFM algorithm whereby the above motion constraints can be conveniently incorporated into the constraint equations on the relative depths of rigid points. The simplicity and homogeneity of the constraint equations allow simple and robust direct solutions of the relative depths. The proposed algorithm is non-iterative, and in general requires a minimum of three points in two frames. It is superior to general linear SFM algorithms not only in computational cost but also in accuracy and noise robustness.

1. Introduction

Motion analysis in general and structure from motion (SFM) in particular have been areas of intense research in the last decade. Numerous approaches have been reported in the literature. For a review see [1]. Problems associated with these approaches, such as high computational complexity, high noise sensitivity, and inefficient use of available information, are enumerated in [2].

We show in this paper that, by working in appropriate coordinate systems and by making use of *a priori* knowledge about scene geometry and real-world constraints on object movements, simple yet robust SFM algorithms can be developed. The motivation of this work comes from the desire to build 3-D geometric object models from monocular monochromatic image sequences [3] in the context of the ESPRIT II project P2152 (VIEWS - Visual Inspection and Evaluation of Wide-area Scenes). The goal of the project is to build a generic vision system for the automatic monitoring and surveillance of land vehicle and airport traffic in known scenes. Fig.1(a) and (b) show images from two traffic scenarios currently used in the project. A common feature in the two scenarios is that the objects (either cars or aeroplanes) move on a ground surface which, locally at least, is approximately flat. We approximate the flat ground surface by the X-Y plane of a world coordinate system (WCS), whose Z-axis points upwards. The movement of the objects in this WCS is constrained in that they can only rotate about and cannot translate along the Z-axis (assuming, of course, that the vertical movement due to suspension etc. is negligible). Constrained motion of this kind is common: mobile objects frequently move on the ground, and objects conveyed by conveyor belts move on well-defined planes.

Existing SFM algorithms are usually defined, and constraint equations are formulated, in a camera-centred coordinate system. They are thus unable to make

Figure 1. Two traffic scenarios used in the ESPRIT II project P2152 (VIEWS)

effective use of the ground plane motion constraints described above since the constraints in the camera coordinate system, although computable, are difficult to incorporate into the constraint equations.

In this paper we describe a simple SFM algorithm based on the distance invariance property of the rigidity constraint [5], which makes effective use of the real-world motion constraints by formulating constraint equations in the world coordinate system. The resultant constraint equations are second-order polynomial equations, each of which involves only two unknowns. The simplicity and homogeneity of these equations allow simple and robust direct solutions. In the subsequent discussions, we make the following common assumptions: 1) Camera parameters (intrinsic and rotational) are known; 2) Feature points have been identified, and inter-frame feature correspondences have been established; and 3) Camera is static, and object motion is rigid.

The paper is organised as follows. In the next section, we concentrate on how to make use of known real-world motion constraints to simplify the constraint equations on the relative depths (i.e., the structure) of the given rigid points. In Section 3 we then describe simple robust methods for solving the constraint equations to recover the relative depths. The estimation of motion parameters from the recovered structure is discussed in Section 4. Experimental evaluation and comparison results of the proposed algorithm under both synthetic and real image data are summarized in Section 5. The paper is concluded in Section 6.

2. Constraint Equations

We assume a pinhole camera model with perspective projection as shown in Fig.2. Under this imaging model, the CCS (Camera Coordinate System) coordinates P_c and the WCS

Figure 2. Coordinate Systems and Imaging Geometry

(World Coordinate System) coordinates P_w of an image point $P_I = (u, v)$ are easily shown to be given by [4]

$$P_c = (\lambda u, \lambda v, \lambda F); \quad P_w = (\lambda U + j, \lambda V + k, \lambda W + l) \quad (1)$$

where F is the camera focal length, λ a positive depth scale, (j, k, l) the WCS coordinates of the origin of the CCS, and U, V and W are defined by

$$U = au + dv + gF$$

$$V = bu + ev + hF$$

$$W = cu + fv + iF$$

where a, b, c, d, e, f, g, h and i are the rotational camera parameters [4]. Thus the squared distance measured in the WCS between two points P_m with image coordinates (u_m, v_m) and P_n with image coordinates (u_n, v_n) is given by

$$d_{mn}^2 = (\lambda_m U_m - \lambda_n U_n)^2 + (\lambda_m V_m - \lambda_n V_n)^2 + (\lambda_m W_m - \lambda_n W_n)^2 \quad (2)$$

Similar equations can be written for the same points in a frame after motion. For example, the squared distance, using primed notations to indicate the new frame, is given by

$$d'_{mn}{}^2 = (\lambda'_m U'_m - \lambda'_n U'_n)^2 + (\lambda'_m V'_m - \lambda'_n V'_n)^2 + (\lambda'_m W'_m - \lambda'_n W'_n)^2 \quad (3)$$

The distance invariance property of rigid motions requires

$$d_{mn}^2 = d'_{mn}{}^2 \quad (4)$$

This basic constraint equation relates the depth scales (structure parameters) $(\lambda_m, \lambda_n, \lambda'_m, \lambda'_n)$ to image coordinates [5]. If objects are confined to move on a flat ground surface which, by definition, is the X-Y plane of the WCS, then the Z coordinates of P_m and P_n are invariant to motion, i.e., $\lambda_m W_m + l = \lambda'_m W'_m + l$ and $\lambda_n W_n + l = \lambda'_n W'_n + l$, or equivalently

$$\lambda'_p = \frac{W_p}{W'_p} \lambda_p, \quad p = m, n \quad (5)$$

From (2) to (5), we obtain

$$A_m \lambda_m^2 + B_{mn} \lambda_m \lambda_n + A_n \lambda_n^2 = 0 \quad (6)$$

where A_m, B_{mn} and A_n are terms computable from U, V and W [4]. (6) is a second-order polynomial equation, and is the basic constraint on the relative depths of two rigid points which move parallel to the ground surface. For N such points $P_1, P_2, \dots, P_N$,

there are $\frac{N(N-1)}{2}$ different point pairs thus $\frac{N(N-1)}{2}$ constraint equations of type (6):

$$A_m \lambda_m^2 + B_{mn} \lambda_m \lambda_n + A_n \lambda_n^2 = 0, \quad n, m \in \{1, 2, \dots, N\}; \quad n > m \quad (7)$$

These equations are homogeneous in the N unknown depths $\lambda_m, m = 1, 2, \dots, N$, which implies that the depths can be determined only up to a global scale factor [1]. We therefore set one of the unknowns arbitrarily to 1, leaving $N-1$ unknowns to be

determined. It thus appears that one needs only a minimum of two points in two frames to solve the constrained SFM. This is in general not the case, and at least three points in two frames are usually required as we discuss in the next section.

3. Estimation of Structure Parameters

If the depths are solved from (7), other structure parameters such as the CCS and WCS coordinates can be easily computed using (1). Hence we concentrate on how to find a correct set of depths. Several methods are possible.

3.1 Method A

An obvious choice is to employ the standard nonlinear least squares technique as used in many existing SFM algorithms [1]. However, such a technique usually requires a good initial guess, and apart from the inherent high computational complexity, it may diverge or converge to a wrong solution.

3.2 Method B

If we fix one of the depths to determine the global scale, then a simple direct solution for the depths stands out immediately. We set $\lambda_1 = 1$, then the $N - 1$ constraint equations associated with P_1 of (7) become $N - 1$ quadratic equations each of which specifies the depth of a single point:

$$A_m \lambda_m^2 + B_{1m} \lambda_m + A_1 = 0, \quad m \in \{2, 3, \dots, N\} \quad (8)$$

These equations are clearly independent of each other except under certain degenerate cases (e.g., coplanar points). Roots of the equations are given by the standard formula:

$$\lambda_m = \frac{-B_{1m} \pm \sqrt{B_{1m}^2 - 4A_m A_1}}{2A_m}, \quad m \in \{2, 3, \dots, N\} \quad (9)$$

The correct root may be determined by imposing the physical constraint on the depths $\lambda_m > 0$, $m = 1, 2, \dots, N$. But under situations where both roots are positive, we need to use the constraint between the point P_m and at least one additional point [4], i.e., at least three points are required to solve the constrained SFM uniquely.

3.3 Method C

A problem with Method B is that the depths of all other points are obtained by satisfying only the constraint equations between these points and the point (P_1) with the arbitrarily fixed depth (let us call such a point the *reference point*). Thus any noise or error (e.g. matching error, perturbation in image coordinates) associated with the reference point (P_1) will degrade the accuracy of the recovered depths of all other points. This is highly undesirable. In fact, without any *a priori* knowledge, there is no reason to choose any particular reference point. All points should be equally favoured as the reference point, and all available constraints should be imposed. This may be achieved by taking each of the given points in turn as the reference point, and solving the depths of all other points

by Method B described above. At the end of this process, we obtain N sets of depths for a set of N points:

$$\{(\lambda_1^n, \lambda_2^n, \dots, \lambda_i^n, \dots, \lambda_N^n) : n = 1, 2, \dots, N\} \quad (10)$$

where $\lambda_n^n = 1$, $n = 1, 2, \dots, N$, and the superscript n indicates the depths computed under reference point P_n . Let us normalize the depths of each set in (10) with respect to (say) the first depth of the corresponding set, and denote the N normalized sets of depths as

$$\{(\tilde{\lambda}_1^n, \tilde{\lambda}_2^n, \dots, \tilde{\lambda}_i^n, \dots, \tilde{\lambda}_N^n) : n = 1, 2, \dots, N\} \quad (11)$$

where $\tilde{\lambda}_1^n = 1$, $n = 1, 2, \dots, N$. Then we define the unique solution for the depths of the N points as

$$\lambda_m = \text{median} \{ \tilde{\lambda}_m^n, n = 1, 2, \dots, N \}, m = 1, 2, \dots, N \quad (12)$$

(12) is justified by the fact that all sets of normalized depth scales in (11) describe the same relative structure of the given N points. Other methods for finding a correct set of depths from the N un-normalized sets of depths are possible, but (11) and (12) perhaps specify the simplest.

3.4 Method D

Another possible solution to (7) is to combine Method A with Method B or C, that is to use the results of Method B or C as the initial guess to bootstrap the iteration process of Method A.

After unprimed depth scales are obtained, primed depth scales (i.e., depth scales associated with the second frame) can be calculated by (5), and the CCS and WCS coordinates by (1). For both computational and accuracy considerations, we have only implemented Method C.

4. Estimation of Motion Parameters

The motion we consider has three degrees of freedom: the rotation angle θ about the Z-axis, and translations T_x and T_y along the X- and Y-axis respectively. The pre- and post-motion WCS coordinates $P_{wm} = (x_m, y_m, z_m)$ and $P'_{wm} = (x'_m, y'_m, z'_m)$ of point P_m are related by

$$\begin{aligned} x_m \cos \theta - y_m \sin \theta + T_x &= x'_m \\ x_m \sin \theta + y_m \cos \theta + T_y &= y'_m \\ z_m &= z'_m \end{aligned} \quad (13)$$

where θ is measured anti-clockwise. To determine the rotation angle from known WCS coordinates, let the X-Y plane projections of a point pair P_{wm} and P_{wn} in the first frame, and P'_{wm} and P'_{wn} in the second frame be Q_m, Q_n and Q'_m, Q'_n , then we compute θ by (see Fig.3)

Figure 3. Computation of Rotation Angle

$$\theta^* = \frac{2}{N(N-1)} \sum_{n=m+1}^N \sum_{m=1}^N \cos^{-1} \left(\frac{(\mathbf{Q}_m \mathbf{Q}_n) \cdot (\mathbf{Q}'_m \mathbf{Q}'_n)}{|\mathbf{Q}_m \mathbf{Q}_n| |\mathbf{Q}'_m \mathbf{Q}'_n|} \right) \quad (14)$$

where we take the average of the angles between all possible point pairs to combat noise. The angle given above only represents the magnitude of the true rotation angle θ^* as the sign of the angle, i.e., the direction of rotation, is lost in the $\cos^{-1}$ operation. The sense of θ^* can be easily determined based on (13). After the rotation angle is obtained, the translations can be calculated directly from (13). Again to control noise, we take the averages of the results computed from each point, i.e.,

$$\begin{aligned} T_x^* &= \frac{1}{N} \sum_{m=1}^N (x'_m - x_m \cos \theta^* + y_m \sin \theta^*) \\ T_y^* &= \frac{1}{N} \sum_{m=1}^N (y'_m - x_m \sin \theta^* - y_m \cos \theta^*) \end{aligned} \quad (15)$$

5. Experimental Results

The proposed algorithm has been applied to both synthetic and real image data. We compare the performance of the proposed algorithm with that of a recently published SFM algorithm [6] in order to make a quantitative appraisal of our algorithm. [6] is chosen because it is the most recent version of the well-known linear SFM algorithms [7-9], and has been shown to be superior to the original linear algorithms [7,8] in performance [6].

Using synthetic data, Monte Carlo simulations were conducted to investigate the noise sensitivity of, and the influence of the number of point correspondences on the two algorithms. A set of six known motions were simulated under a typical camera configuration. True image coordinates were shifted by a random number which is uniformly distributed over $(-\Delta E, \Delta E)$. Motion parameters computed by the two algorithms were compared with the ground truth, and the mean and standard deviation of the absolute difference in each parameter were calculated. Comprehensive testing has been carried out [4]. Here we give an example. Table 1 shows the results obtained under $\Delta E = 0.5$ pixels and using 15 points. From the experimental results, the following general observations can be made [4]:

1. The motion parameter errors in [6] increase much more rapidly with noise than those in the proposed algorithm.
2. The motion parameters computed by the proposed algorithm are in general much

Table 1

Noise Performance Comparison Between Proposed Algorithm and [6]

Ground Truth	Proposed Algorithm		Algorithm [6]	
	error mean	error std. dev.	error mean	error std. dev.
theta	-4.56	0.66	1.56	0.72
Tx	1000.0	196.46	30621.18	244240.38
Ty	2000.0	185.96	13354.1	42783.53
theta	0.0	1.59	1.08	0.34
Tx	1234.0	202.88	12162.87	76293.94
Ty	567.8	189.82	11505.89	105845.36
theta	4.56	0.62	1.09	1.34
Tx	1000.0	166.02	11214.69	142727.91
Ty	2000.0	243.62	6362.82	49235.69
theta	12.34	0.55	0.95	1.11
Tx	0.0	179.8	1637.23	10349.74
Ty	0.0	305.68	2297.8	15872.71
theta	23.45	0.61	0.82	0.98
Tx	-2500.5	152.39	824.84	1147.89
Ty	234.5	424.31	1052.34	2334.57
theta	123.4	0.27	0.31	0.33
Tx	0.0	142.19	706.87	716.0
Ty	2345.6	277.19	802.45	873.09

more accurate than those by [6] especially under high noise conditions.

3. The performance of the proposed algorithm is consistently improved by using more point correspondences.
4. Under low noise conditions, the use of more points generally improves the performance of [6]. The effectiveness of using more points to reduce errors in [6] is significant when the number of points is increased from 10 to 15, but is marginal when further points (e.g., from 15 to 20) are employed. This observation agrees with the experimental results presented in [6] (see Fig.8 and Fig.9 of [6]).
5. Under high noise conditions ($\Delta E > 1.0$ pixel), the use of additional points further degrades the performance of [6].

With real image sequences, the assessment of the accuracy of the recovered structure is not straightforward as the ground truth is usually unknown. Proposals have been made [6] to use the discrepancy between the image of the reconstructed 3-D structure and the known point correspondences as a measure of accuracy. We argue that such a measure is inappropriate as it does not reflect the accuracy of the recovered relative structure since multiplying the depths of the feature points by different factors changes the relative structure of the points but does not affect the measure of this kind. We thus propose a new assessment procedure based on model matching [10]. We have applied the two algorithms to a number of real image sequences [4]. An example is given in Fig.4 which shows 12 points identified on a moving lorry, and the corresponding displacement field. The heights of the 12 points recovered by the two algorithms are tabulated in Table 2, where the height of P_1 is assumed to be 1000 mm in order to fix the global scale. Agreements (at least partially) between our perception of the image and the heights recovered by the proposed algorithm is evident in Table 2, whereas the heights obtained by [6] appear widely wrong. This is probably due to the high noise sensitivity of [6], and its making no use of the ground plane motion constraint. To further assess the accuracy of the relative structure of the 12 points recovered by our algorithm, the recovered WCS

Figure 4. The moving lorry and its displacement field

Table 2		
Recovered Heights (in mm) of Lorry Points by [6] and Proposed Algorithm		
Name of Point	Algorithm [6]	Proposed Algorithm
P12	5359.7	2323.8
P11	6182.4	3289.6
P10	4115.3	919.1
P9	5087.4	846.9
P8	5901.6	1451.5
P7	6474.9	2312.6
P6	7127.7	3201.3
P5	9136.1	3407.3
P4	7113.3	3388.7
P3	6190.6	3546.8
P2	5910.4	1304.0
P1	1010.2	1000.0

coordinates of these points are converted into a simple polyhedral model [3], which is displayed in Fig.5 under three different viewpoints. The model is then matched against

Figure 5. Three different views of the recovered lorry model

the lorry image as illustrated in Fig.6. The matching is very good. The accuracy of the

Figure 6. Matching between the recovered model and four lorry images

model, hence the accuracy of the recovered structure of the 12 points, is evident in Fig.4

to Fig.6. Results of this kind could not be obtained by [6].

6. Conclusions

In the real world, the movement of many objects (e.g., cars, objects on conveyor belts, etc.) is constrained in that they can only move on a fixed plane or surface (e.g., the ground). A new SFM algorithm has been presented in this paper which, by formulating motion constraint equations in the world coordinate system, makes effective use of the physical motion constraints of this kind. The algorithm is computationally simple and gives a unique and closed-form solution to the motion and structure parameters of rigid 3-D points. It is non-iterative, and in general requires a minimum of three points in two frames. The algorithm has been shown to be superior to the existing linear SFM algorithms in accuracy and robustness, especially under high noise conditions and real image data.

The algorithm in its present form can be extended and improved in many directions. For instance, longer image sequences can be used, and the recovered structures can be integrated over time by means of (say) extended Kalman filtering. Meanwhile, the effect of camera parameter errors on the reliability of the proposed algorithm needs to be studied.

References

- [1] J. K. Aggarwal and N. Nandhakumar, On the Computation of Motion from Sequences of Images - A Review, *Proc. of IEEE*, vol.76, no.8, 1988, pp.917-935.
- [2] H. Shariat and K. E. Price, Motion Estimation with More Than Two Frames, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.12, 1990, pp.417-434.
- [3] T. N. Tan, G. D. Sullivan, and K. D. Baker, 3-D Models from Motion (MFM) - an application support tool, *ESPRIT II P2152 project report*, RU-03-WP.T411-02, University of Reading, June 1991.
- [4] T. N. Tan, G. D. Sullivan, and K. D. Baker, Structure from Constrained Motion, *ESPRIT II P2152 project report*, RU-03-WP.T411-01, University of Reading, March 1991.
- [5] A. Mitiche and J. K. Aggarwal, A Computational Analysis of Time-Varying Images, in *Handbook of Pattern Recognition and Image Processing*, T.Y. Young and K. S. Fu, Eds. New York: Academic Press, 1986.
- [6] J. Y. Weng, T. S. Huang, and N. Ahuja, Motion and Structure from Two Perspective Views: Algorithms, Error Analysis, and Error Estimation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.11, no.5, 1989, pp.451-477.
- [7] R. Y. Tsai and T. S. Huang, Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surface, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.6, 1984, pp.13-26.
- [8] H. C. Longuet-Higgins, A Computer Algorithm for Reconstructing a Scene from Two Projections, *Nature*, vol.293, 1981, pp.133-135.
- [9] X. Zhuang, T. S. Huang, and R. M. Haralick, Two-View Motion Analysis: A Unified Algorithm, *J. Opt. Soc. Amer.*, vol.3, no.9, 1986, pp.1492-1500.
- [10] K. S. Brisdon, *Hypothesis Verification Using Iconic Matching*, PhD thesis, University of Reading, 1990.