

# Structure from Motion Using the Ground Plane Constraint

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

Department of Computer Science, University of Reading  
Reading, Berkshire RG6 2AY, ENGLAND

**Abstract.** This paper concerns the interactive construction of geometric models of objects from image sequences. We show that when the objects are constrained to move on the ground plane, a simple direct SFM algorithm is possible, which is vastly superior to conventional methods. The proposed algorithm is non-iterative, and in general requires a minimum of three points in two frames. Experimental comparisons with other methods are presented in the paper. It is shown to be greatly superior to general linear SFM algorithms not only in computational cost but also in accuracy and noise robustness. It provides a practical method for modelling moving objects from monocular monochromatic image sequences.

## 1. Introduction

The work described here was carried out as part of the ESPRIT II project P2152 (VIEWS - Visual Inspection and Evaluation of Wide-area Scenes). It is concerned with semi-automatic methods to construct geometric object models using monocular monochromatic image sequences. A common feature in the images used in the VIEWS project is that the movement of objects satisfies the *ground plane constraint* [1, 2], i.e., the objects 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. In this WCS, an object can only translate along the X- and Y-axis, and rotate about the Z-axis, leaving 3 degrees of freedom of motion.

We show in this paper that, in order to make the most effective use of the ground plane constraint in structure from motion (SFM), it is necessary to formulate structure (and motion) constraint equations in the WCS. This allows us to derive simple yet robust SFM algorithms. The paper is organised as follows. We first discuss the use of the ground plane constraint to simplify the constraint equations on the relative depths (i.e., the structure) of the given rigid points. We then describe simple robust methods for solving the constraint equations to recover the structure and motion parameters. Experimental studies of both a conventional 6 degrees of freedom SFM algorithm [3] and the proposed 3 degrees of freedom algorithm are then reported.

## 2. Constraint Equations

We assume a pinhole camera model with perspective projection as shown in Fig.1. Under this

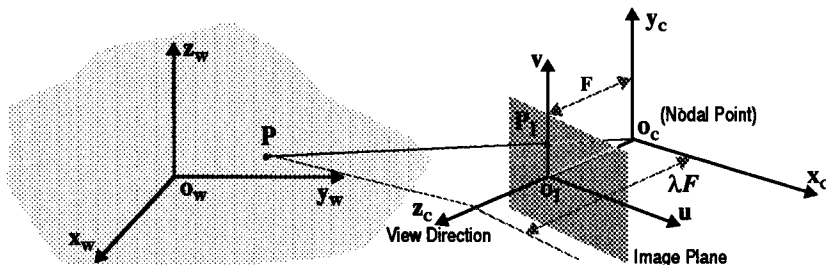


Figure 1. Coordinate Systems and Imaging Geometry

imaging model, the squared distance  $d_{mn}^2$  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

$$d_{mn}^2 = (\lambda_m U_m - \lambda_n U_n)^2 + (\lambda_m V_m - \lambda_n V_n)^2 + (\lambda_m W_m - \lambda_n W_n)^2 \quad (1)$$

where  $\lambda_m$  and  $\lambda_n$  are the depths (scales) of  $P_m$  and  $P_n$  respectively ( $\lambda F = z_c$ ), and  $U, V$  and  $W$  are terms computable from known camera parameters and image coordinates [1, 2]. A similar equation can be written for the point pair in a subsequent frame. Using primed notation to indicate the new frame, we have

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

It is shown in [1, 2] that by using the distance invariance property of the rigidity constraint [4, 5] and the height invariance property of the ground plane constraint, we can obtain from (1) and (2) the following second-order polynomial equation:

$$A_m \lambda_m^2 + B_{mn} \lambda_m \lambda_n + A_n \lambda_n^2 = 0 \quad (3)$$

where  $A_m, B_{mn}$  and  $A_n$  are terms computable from  $U, V$  and  $W$  [1, 2]. (3) is the basic constraint on the relative depths of two points of a rigid object whose movement satisfies the ground plane constraint. For  $N$  such points  $P_1, P_2, \dots, P_N$ , there are  $N(N-1)/2$  different point pairs thus  $N(N-1)/2$  constraint equations of type (3):

$$A_m \lambda_m^2 + B_{mn} \lambda_m \lambda_n + A_n \lambda_n^2 = 0, \quad n, m \in \{1, 2, \dots, N\}; \quad n > m \quad (4)$$

which can be solved for the  $N$  unknown depths  $\lambda_n, \quad n = 1, 2, \dots, N$ .

### 3. Estimation of Structure and Motion Parameters

Since the equations in (4) are homogeneous in the  $N$  unknown depths, the point depths can only be determined up to a global scale factor as is the case in all SFM algorithms [6], and we can set the depth of an arbitrary point (the *reference point*) to be an arbitrary value. For instance, we can set the depth ( $\lambda_1$ ) of the first point  $P_1$  to be 1, then the  $N-1$  constraint equations associated with  $P_1$  in (4) become  $N-1$  quadratic equations each of which specifies the depth of a single additional point:

$$A_m \lambda_m^2 + B_{1m} \lambda_m + A_1 = 0, \quad m \in \{2, 3, \dots, N\} \quad (5)$$

The correct root may usually be determined by imposing the physical constraint on the depths  $\lambda_m > 0, m = 2, 3, \dots, N$ . In situations where both roots are positive, we need to use the constraint between the point  $P_m$  and one additional point [1], i.e., in general, at most three points in two frames are required to solve the constrained SFM uniquely.

We take each of the given points in turn as the reference point, and repeat the above procedure, to obtain  $N$  sets of depths for a set of  $N$  points:

$$\{(\lambda_1^n, \lambda_2^n, \dots, \lambda_i^n, \dots, \lambda_N^n) : n = 1, 2, \dots, N\} \quad (6)$$

where  $\lambda_i^n = 1$  for  $i = n$ , and the superscript  $n$  indicates the depths computed under reference point  $P_n$ . The depths of each set in (6) are normalised with respect to the depth of the same single point (say  $P_1$ ) in the corresponding set to get  $N$  normalized sets of depths as

$$\{ (\tilde{\lambda}_1^n, \tilde{\lambda}_2^n, \dots, \tilde{\lambda}_i^n, \dots, \tilde{\lambda}_N^n) : n = 1, 2, \dots, N \} \quad (7)$$

where  $\tilde{\lambda}_1^n = 1, n = 1, 2, \dots, N$ . A unique solution for the depths of the  $N$  points is obtained by computing

$$\lambda_m = \mathbf{median} \{ \tilde{\lambda}_m^n, n = 1, 2, \dots, N \}, m = 1, 2, \dots, N \quad (8)$$

Equation (8) is justified by the fact that all sets of normalized depth scales in (7) describe the same relative structure of the given  $N$  points.

Other approaches have also been explored to solve (4) but are not included here because of space limitation. Once the depths are known, the computation of point coordinates in the WCS, and the estimation of the three motion parameters (i.e., the translations  $T_x$  and  $T_y$  along the X- and Y-axis, and the rotation angle about the Z-axis in the WCS) are straightforward. Details may be found in [1, 2].

#### 4. Experimental Results

We have compared the performance of the new algorithm with that of a recent linear SFM algorithm proposed by Weng *et al.* [3]. For convenience, we call the proposed algorithm the TSB algorithm, and the algorithm in [3] the WHA algorithm in the subsequent discussions.

Using synthetic image 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. Comprehensive testing has been carried out [1, 2]. Numerous results show that in general the TSB algorithm performs much better than the WHA algorithm especially under high noise conditions.

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 [3] to use the *standard image error (SIE)*  $\Delta\epsilon$  defined as follows [3]

$$\sqrt{\sum_{i=1}^N \frac{d_i^2 + d_i'^2}{2N}} \quad (9)$$

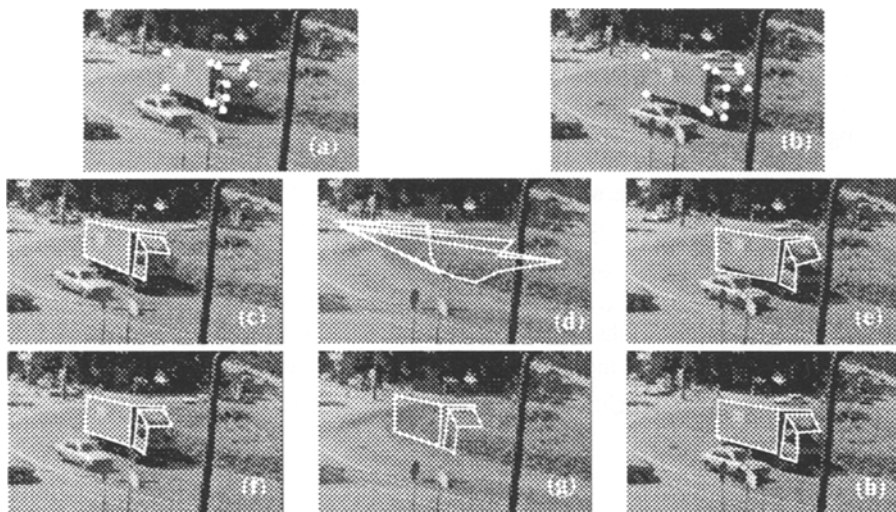
where  $N$  is the number of points,  $d_i$  and  $d_i'$  the distances in the two images between the projection of the reconstructed 3D point  $i$  and its observed positions in the images. The SIEs of the two algorithms applied to five different moving objects are listed in Table I. In these terms,

**Table 1**

**SIE (in pixels) of Algorithms TSB and WHA Under Real Image Sequences**

Moving Object	Standard Image Error		Error Ratio (WHA / TSB)
	Algorithm WHA	Algorithm TSB	
Lorry	0.160	0.00134	119
Estate1	0.455	0.00144	316
Estate2	0.532	0.00147	362
Saloon1	0.514	0.00177	290
Saloon2	1.188	0.00171	695

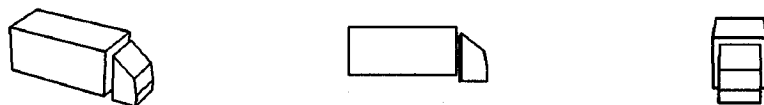
the TSB algorithm performs several hundred times better than the WHA algorithm. It is argued in [1, 2] that the SIE gives a poor measure of performance. In addition to the observed error in the two frames, the performance should be analysed by a qualitative assessment of the projected wire-frame model of all given points from other views. For example, Fig.2 shows two images of



**Figure 2.** For detailed captions, see text.

a moving lorry and an intermediate view of the partial wire-frame lorry model recovered by the two algorithms. Fig.2(a) shows 12 lorry points in the first frame and Fig.2(b) the same points in the second frame. The inter-frame point correspondences were known. The recovered point coordinates are converted into a partial wire-frame lorry model simply by connecting appropriate point pairs. The model recovered by the WHA algorithm appears to fit the two original images reasonably well as illustrated in Fig.2(c) and (e), as indicated by the small standard image error (0.160) given in Table 1. We can now perturb the recovered model slightly (as if the lorry had undergone a small motion between (c) and (e)). The outcome (Fig.2(d)) is far from expectation, and the recovered model is clearly not lorry-shaped. In contrast, the partial wire-frame lorry model recovered by our algorithm proves to be accurate and consistent as can be seen in Fig.2(f)-(h). Other image pairs from the lorry sequence were used, and the results obtained were similar. The disastrous performance of the WHA algorithm shown in the above example is attributed to several causes. Firstly, the linear constraint used by [3] is not the same as the rigidity constraint [8], and as a consequence there are non-rigid motions that would satisfy the linear constraint (see Fig.2(d)). Secondly, no explicit use of the ground plane constraint is made in [3]. Finally, the number of points used in the given example is small.

To further assess the performance of the TSB algorithm, the WCS coordinates of the 12 lorry points recovered by the algorithm from the two images Fig.2(a) and (b) were converted by means of an interactive models-from-motion tool [7] into a full polyhedral lorry model, which is displayed in Fig.3 under three different viewpoints. The completed model has been matched



**Figure 3.** Three different views of the lorry model recovered by the new algorithm

against the lorry image sequence and tracked automatically using the methods reported in [9] as illustrated in Fig.4. The match is very good.

We have also investigated the sensitivity of the proposed algorithm to systematic errors such as errors in rotational camera parameters. It was found that such errors have small effects



**Figure 4.** Matching between the recovered lorry model and four lorry images

on the estimation of the rotation angle, and moderate impact on that of the translational parameters. Detailed results cannot be reported here due to space limitation.

## 5. Discussion

In the real world, the movement of many objects (e.g., cars, objects on conveyor belts, etc.) is constrained in that they 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 this physical motion constraints (the ground plane constraint). 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 usually requires two points in two frames. The algorithm has been shown to be greatly superior to existing linear SFM algorithms in accuracy and robustness, especially under high noise conditions and when there are only a small number of corresponding points. The recovered 3-D coordinates of object points from outdoor images enable us to construct 3-D geometric object models which match the 2-D image data with good accuracy.

## References

- [1] 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.
- [2] T. N. Tan, G. D. Sullivan, and K. D. Baker, Structure from Constrained Motion Using Point Correspondences, *Proc. of British Machine Vision Conf.*, 24-26 September 1991, Glasgow, Scotland, Springer-Verlag, 1991, pp.301-309.
- [3] 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.
- [4] 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.
- [5] S. Ullman, *The Interpretation of Visual Motion*, MIT Press, 1979.
- [6] 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.
- [7] 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.
- [8] D. J. Heeger and A. Jepson, Simple Method for Computing 3D Motion and Depth, *Proc. of IEEE 3rd Inter. Conf. on Computer Vision*, December 4-7, 1990, Osaka, Japan, pp.96-100.
- [9] A. D. Worrall, G. D. Sullivan, and K. D. Baker, Model-based Tracking, *Proc. of British Machine Vision Conf.*, 24-26 September 1991, Glasgow, Scotland, Springer-Verlag, 1991, pp.310-318.