Ego-Motion and 3D Recovery

Shape from motion algorithms: a comparative analysis of scaled orthography and perspective

Boubakeur Boufama[1] and Daphna Weinshall[2] & Mike Werman[2]

[1] Lifia–Inria,
46 Av. Félix Viallet,
Grenoble Cedex, France
[2] Institute of Computer Science
The Hebrew University
91904 Jerusalem, Israel

Abstract. This paper describes a comparative study of reconstruction algorithms from sequences of images, comparing algorithms which make the weak perspective assumption (also called scaled orthography or para perspective) to algorithms assuming perspective projection. The weak perspective assumption is usually employed to simplify the computation. Using three sequences of real images, taken under conditions corresponding to small, medium, and large fields of view, we compare two algorithms that compute invariant shape from motion; one assumes scaled orthography and one assumes perspective projection.
The paper describes also a joint algorithm, which combines the simplicity of weak perspective algorithm with the accuracy of the perspective one.

1 Introduction

In real images, objects are projected to the image via perspective projection (where, using the pinhole approximation, a space point (x, y, z) is projected to the image coordinates $(\frac{x}{z}, \frac{y}{z})$). When the maximal distance h (h is the radius of the object) is small relative to the depth of the object z, it proves computationally convenient to neglect high order terms in $\frac{h}{z}$. This approximation is called scaled orthography or weak perspective. This approximation is equivalent to assuming that the object undergoes rotation, translation and scaling in $3D$ space, followed by an orthographic projection to the image (where a space point (x, y, z) is projected to the image coordinates (x, y)).

Algorithms for the computation of $3D$ shape of objects can be divided into those that assume weak perspective for computational convenience, and those that take into account the real perspective projection. Which type of algorithm should be used depends on the given sequence of images. When only two frames are given, and when the motion is sufficiently large so that small differences in the image are observable, perspective algorithms should be used. When perspective effects are not reliably observable (i.e., for a small field of view), and when a longer sequence of images is given, weak perspective algorithms should be used. The latter tend to be more robust and are also easier to implement. On the other hand the precision of the reconstruction is lower (for a large field of view).

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

In this paper we study the relative advantages and scope of each of these approaches. In Section 3 we analyze the performance of two invariant shape algorithms (reviewed in Section 2), one which assumes scaled orthography and one which assumes perspective projection. We use three real sequences of images, with small, medium and large fields of view. We show that the weak perspective algorithm performs slightly **better** with a small field of view, slightly worse with a medium field of view, and significantly **worse** with a large field of view.

Finally, we describe in Section 4 a joint algorithm, which combines the simplicity and convergence of the weak perspective algorithm with the accuracy of the perspective algorithm.

2 The 3*D* reconstruction algorithms used in this paper

We applied two reconstruction methods to the same points and the same number of images. We give here a brief description of the two methods.

2.1 The perspective projection algorithm

This algorithm was originally described in [3]. It computes the 3*D* structure of points from a sequence of images using 5 known points as a relative frame.

Consider $v \geq 2$ images of a scene composed of p points (P_i). Each point P_i is represented by a column vector of its homogeneous coordinates $(x_i, y_i, z_i, t_i)^T$. In image j, the point P_i is projected to the point p_{ij}, represented by a column vector of its image coordinates $(U_{ij}, V_{ij})^T$. Let M_j denote the 3×4 projection matrix of the j-th image, we have:

$$U_{ij} = \frac{m_{11}^{(j)} x_i + m_{12}^{(j)} y_i + m_{13}^{(j)} z_i + m_{14}^{(j)} t_i}{m_{31}^{(j)} x_i + m_{32}^{(j)} y_i + m_{33}^{(j)} z_i + m_{34}^{(j)} t_i}, \quad V_{ij} = \frac{m_{21}^{(j)} x_i + m_{22}^{(j)} y_i + m_{23}^{(j)} z_i + m_{24}^{(j)} t_i}{m_{31}^{(j)} x_i + m_{32}^{(j)} y_i + m_{33}^{(j)} z_i + m_{34}^{(j)} t_i} \tag{1}$$

Since we have p points and v images, we have $2 \times p \times v$ equations. The number of unknowns is $11 \times v$ for the matrices, plus $3 \times p$ for the points. Therefore if $2 \times p \times v \geq 11 \times v + 3 \times p$, we have a redundant set of non-linear equations. The problem is then formulated as a conditional parameter estimation problem.

2.2 The weak perspective algorithm

This linear method was described in [5]. It computes the Euclidean structure of points from a sequence of images assuming weak perspective.

Given 4 points, it computes the Gramian of the 4 points by solving a linear system of equations. The inverse Gramian gives the complete Euclidean-invariant (metric) structure of the 4 points.

Given more than 4 points, the algorithm proceeds as follows:

- select from the data 4 points as a basis (optionally using QR factorization to maximize the independence of the selected points);
- compute the affine structure of all points by solving a linear system;

- compute the Euclidean structure of the 4 basis points by solving a linear system;
- obtain the Euclidean structure of all the points if necessary (this can be done by multiplying a vector of affine coordinates by the root of the inverse Gramian of the basis points).

3 Experimental results with real data

In order to study the behavior of the two algorithms described in Section 2, we tested three comparative reconstructions for the following three cases:

Small field of view: an object of radius 12cm at about 140cm from the camera (Fig 1a); in this case the weak perspective approximation is not only appropriate, it actually increases the accuracy of the reconstruction.

Medium field of view: an object of radius 15cm at about 60cm from the camera (Fig 1b); in this case the weak perspective approximation requires the negligence of terms which are not very small (≈ 0.25), and therefore the weak perspective algorithm leads to less accurate results.

Large field of view: objects spanning 60cm at about 60cm from the camera (Fig 1c); in this case the weak perspective approximation neglects higher order terms of $O(1)$, and the perspective algorithm is needed.

For each of the three cases, we used a sequence of 5 images taken with the same CCD camera of 12.5mm focal length. Corners were first automatically extracted using a method similar to [1], and then automatically tracked over the sequence. For simplicity we only used points which appeared in all the images. Furthermore, for comparison we measured all the $3D$ coordinates of the points with a ruler. Recall, that the perspective algorithm we use here assumes a priori knowledge of the exact $3D$ coordinates of 5 object points. Thus the comparisons reported below are meaningful only when the weak perspective reconstruction turns out to be more accurate. Following is the error analysis for each case:

Case 1: small field of view. In this case we expect weak perspective algorithm to give good results. Table 1 gives the measured coordinates and errors of both methods. Mean errors are also given. As can be seen from the table, the weak perspective approximation gives better results for this case. The conditions here make the perspective projection more sensitive to image noise, and computing epipolar geometry for example, will be numerically unstable [2].

Case 2: medium field of view. Here we expect the weak perspective reconstruction to be less accurate, whereas the perspective reconstruction should give more accurate results. This expectation was confirmed, although the weak perspective reconstruction is only slightly less accurate than the perspective one. Table 2 gives only 6 point coordinates with their errors, but the mean errors are computed with all the scene points.

Case 3: large field of view. The ratio size/distance is about 1. The errors here are larger in both cases (Table 3), but even the weak perspective reconstruction resembles the actual scene (see Fig 1d).

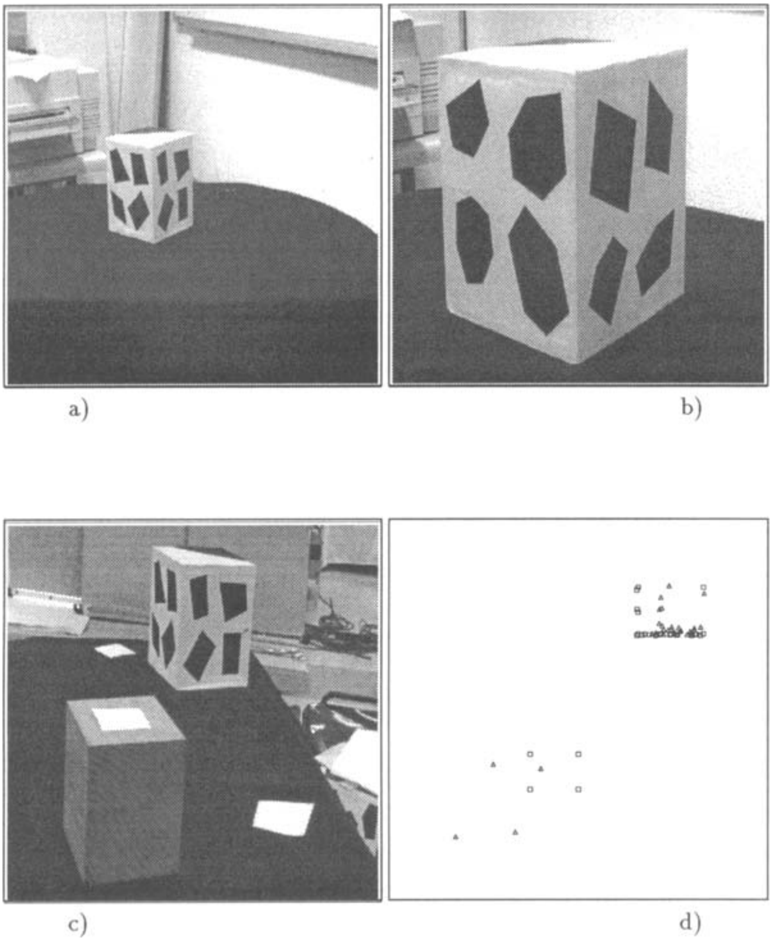

Fig. 1. a) A view of the scene for case 1; b) a view of the scene for case 2; c) a view of the scene for case 3; d) The reconstructed scene for case 3: squares mark the location of the points in the perspective reconstruction, and the remaining points are the location in the weak perspective reconstruction (top view).

measured coordinates			*weak perspective errors(cm)*			*perspective errors(cm)*		
X	Y	Z	ΔX	ΔY	ΔZ	ΔX	ΔY	ΔZ
0.00	12.00	12.50	0.43	0.10	0.18	0.02	1.77	0.27
6.10	0.00	11.00	0.02	0.16	0.03	0.47	0.15	0.21
11.00	0.00	8.10	0.05	0.24	0.04	0.51	0.24	0.56
0.00	2.45	11.40	0.28	0.04	0.15	0.01	0.33	0.01
0.00	3.00	6.60	0.24	0.03	0.08	0.01	0.18	0.29
0.00	4.40	1.00	0.09	0.10	0.18	0.15	0.22	0.03
mean errors			0.14	0.16	0.13	0.21	0.34	0.24

Table 1. Errors in the reconstructed $3D$ coordinates for case 1.

measured coordinates			*weak perspective errors(cm)*			*perspective errors(cm)*		
X	Y	Z	ΔX	ΔY	ΔZ	ΔX	ΔY	ΔZ
0.00	3.05	1.00	0.57	0.51	0.44	0.02	0.05	0.11
8.60	0.00	9.00	0.09	0.49	0.05	0.07	0.02	0.04
14.90	0.00	0.00	0.43	1.80	1.11	0.12	0.08	0.08
0.00	11.20	7.50	0.54	0.12	0.12	0.12	0.01	0.07
4.40	0.00	5.95	0.32	0.33	0.22	0.06	0.08	0.02
12.20	0.00	8.00	0.17	1.31	0.08	0.01	0.04	0.04
mean errors			0.55	0.57	0.2	0.06	0.06	0.05

Table 2. Errors in the reconstructed $3D$ coordinates for case 2.

measured coordinates			*weak perspective errors(cm)*			*perspective errors(cm)*		
X	Y	Z	ΔX	ΔY	ΔZ	ΔX	ΔY	ΔZ
-20.00	-31.00	9.00	7.22	2.54	0.92	2.79	2.85	0.18
12.00	0.00	12.50	0.87	1.74	0.24	1.10	0.31	0.13
4.40	0.00	7.60	1.41	1.22	1.32	0.16	0.10	0.30
-11.00	-31.00	9.00	7.38	3.65	1.44	2.11	2.86	0.16
11.00	0.00	8.10	0.94	1.09	1.12	0.59	0.27	0.11
2.20	0.00	7.10	2.20	1.36	1.41	0.04	0.07	0.32
mean errors			2.99	1.92	1.68	0.66	0.69	0.14

Table 3. Errors in the reconstructed $3D$ coordinates for case 3.

4 A cooperation between the two algorithms

The results of section 3 show that as the size/distance ratio increases, the quality of the reconstruction obtained by the weak perspective algorithm degrades (as is expected). On the other hand, the non-linear nature of the perspective reconstruction makes it impossible to guarantee convergence. To overcome this problem, we combined the two approaches by using the weak perspective reconstruction as an initialization for the perspective iterative algorithm.

Our reconstruction is therefore performed in two steps:

1. The $3D$ structure of the scene is computed using weak perspective algorithm. This reconstruction is not accurate, but it is not far from the correct one.
2. The reconstruction is improved by using the iterative perspective algorithm.

After computing the $3D$ structure of the scene using the weak perspective algorithm, we consider two cases: either no additional knowledge about the scene is available, or the Euclidean coordinates of five points are known.

In the first case, we don't use any a priori knowledge, only the output of the weak perspective algorithm provides the initial guess for the iterative perspective algorithm. Without fixing any parameter, the iterative algorithm has too many degrees of freedom and usually does not converge. By choosing five points (no four of them coplanar) as a relative basis, we get a projective reconstruction up to this basis [4]. Using the approximation of the Euclidean $3D$ coordinates of the five points computed in the first step, we obtain an approximately Euclidean reconstruction. Unlike the reconstruction computed in the first step, this solution is guaranteed to be the correct projective one. Furthermore, this kind of reconstruction is of special interest for some tasks in robotics, when only Euclidean approximation of the scene is needed.

In the second case, the output of the weak perspective algorithm is also used to initialize the iterative perspective algorithm. Now we are given the Euclidean $3D$ coordinates of five basis points. In this case, the weak perspective algorithm ensures the convergence of the iterative perspective algorithm.

5 Conclusions

We studied the scope and limitation of the assumption of weak perspective projection and its use in three dimensional reconstruction from a sequences of images. The conclusions of our comparative study can be summarized as follows:

- Given a small field of view, the reconstruction assuming weak perspective seems more robust, this gets confirmed by our experiments.
- Given a medium field of view, the reconstruction assuming weak perspective projection was slightly less accurate (but much easier to obtain) than the one assuming perspective projection (Section 3).
- Given a large field of view, the results of weak perspective algorithm significantly degrades.

Robustness and accuracy are improved by using a joint algorithm described in the paper. It computed the weak perspective reconstruction in its first step, then improved this reconstruction in the second step.

References

1. R. Deriche and T. Blaszka. Recovering and characterizing image features using an efficient model based approach. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 530–535, June 1993.
2. Q.T. Luong. *Matrice Fondamentale et Autocalibration en Vision par Ordinateur.* Thèse de doctorat, Université de Paris-Sud, Orsay, France, December 1992.
3. R. Mohr, L. Quan, F. Veillon, and B. Boufama. Relative 3D reconstruction using multiples uncalibrated images. Technical Report RT 84-I-IMAG LIFIA 12, LIFIA–IRIMAG, 1992.
4. R. Mohr, F. Veillon, and L. Quan. Relative 3D reconstruction using multiple uncalibrated images. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, New York, USA*, June 1993.
5. D. Weinshall. Model-Based Invariants for 3-D Vision. *International Journal of Computer Vision*, 10(1):27–42, 1993.