Relative 3D Reconstruction Using Multiple Uncalibrated Images

R. Mohr F. Veillon L. Quan

LIFIA/INRIA
46, Avenue Felix Viallet, 38031 Grenoble, FRANCE

Abstract

In this paper, we show how relative 3D reconstruction for point correspondences of multiple images from uncalibrated cameras can be achieved through reference points. The original contributions with respect to other related works in the field are mainly a direct global method for relative 3D reconstruction, a geometrical method to select a correct set of reference points among all image points. Experimental results from both simulated and real image sequences are presented.

1 Relative positioning

From a single image, no depth can be computed without *a priori* information. Even more, no invariant can be computed from a general set of points [3]. This problem becomes feasible using multiple images. The process is composed of two major steps. First image features are matched in the different images. Then, from such a correspondence, depth is easily computed using standard triangulation. This approach suffers from several drawbacks: firstly the calibration process is an error sensitive process; secondly it cannot always be performed on line, particularly when the imaging system is obtained by a dynamic system with zooming, focusing and moving. Similarly stereo vision with a moving camera is impossible as the standard tool for locating the position of a camera with translation and rotation does not reach the required precision for calibrating such a multistereo system. Introducing in each image beacons with exact known position may overcome these drawbacks: calibration and reconstruction are then solved in the same process [2, 1]. But for many problems it is impossible to provide such carefully positioned reference points.

The alternative approach is to use points in the scene as reference frame without knowing their coordinates nor the camera parameters. This has been investigated by several researchers these past few years, for instance in [10, 11, 9, 17, 8, 14, 16].

This year, three independent teams approached the same problem of 3D reconstruction from uncalibrated cameras, and all three with the same projective basis. Faugeras [4] published an insightful algebraic method to do 3D projective reconstruction. He demonstrated that once the epipolar geometry is somehow determined, 3D projective structure can be reconstructed up to a collineation by assigning 5 reference points to the standard projective basis. One month later, Hartley published his paper [7] in which he described the similar approach in a slightly different way. The same month appeared our technical report [12] performed differently with our first experimental results.

The original contributions of this paper are mainly twofold. First, we describe a direct 3D relative reconstruction method, which differs from that of Faugeras and Hartley in that our method is formulated globally as a least squares estimation method which does not need to first estimate the epipolar geometry, and also the method makes full use of redundancy of multiple images. Secondly, we provide a geometrical way to choose among the set of points those which can be selected as reference points. The selected reference points should not be degenerated, i.e. no four of them coplanar. This result allows to derive a computational way to choose the correct reference points.

We assume that the reader is familiar with elementary projective geometry, as it can be found in the first chapters of [15] (see also [5]).

2 Using scene reference points

This section provides the basic equations of 3D reconstruction problem, together with the self calibration problem. This derivation was developed independently from these recently published by Faugeras in [4]. The basic starting point is similar to this work, however the way to solve it was influenced by the way photogrammetrists simultaneously calibrate their camera and reconstruct the scene, by using carefully located beacons (cf. [1]).

1063-6919/93 $03.00 © 1993 IEEE

We consider m views of a scene ($m \geq 2$); it is assumed that n points have been matched in all the images, thus providing $n \times m$ image points. The assumption that the scene points appear in all the images is not essential but only simplifies the explanation here. $\{M_i, i = 1, \ldots, n\}$ is the (unknown) set of 3D points projected in each image, represented by a column vector of its four yet unknown homogeneous coordinates.

2.1 The basic equations

For each image j, the point M_i, represented by a column vector of its homogeneous coordinates $(x_i, y_i, z_i, t_i)^T$ or its usual non homogeneous coordinates $(X_i, Y_i, Z_i)^T = (\frac{x_i}{t_i}, \frac{y_i}{t_i}, \frac{z_i}{t_i})^T$, is projected as the point m_{ij}, represented by a column vector of its three homogeneous coordinates $(u_{ij}w_{ij}, v_{ij}w_{ij}, w_{ij})^T$ or its usual non homogeneous coordinates $(u_{ij}, v_{ij})^T$. Let P_j be the 3×4 projection matrix of the jth camera.

We have for homogeneous coordinates

$$\rho_{ij} m_{ij} = P_j M_i, i = 1, \ldots, n, j = 1, \ldots, m \qquad (1)$$

where ρ_{ij} is an unknown scaling factor which is different for each image point.

Equation 1 is usually written in the following way, hiding the scaling factor, using the non homogeneous coordinates of the image points:

$$u_{ij} = \frac{p_{11}^{(j)} x_i + p_{12}^{(j)} y_i + p_{13}^{(j)} z_i + p_{14}^{(j)} t_i}{p_{31}^{(j)} x_i + p_{32}^{(j)} y_i + p_{33}^{(j)} z_i + p_{34}^{(j)} t_i} \qquad (2)$$

$$v_{ij} = \frac{p_{21}^{(j)} x_i + p_{22}^{(j)} y_i + p_{23}^{(j)} z_i + p_{24}^{(j)} t_i}{p_{31}^{(j)} x_i + p_{32}^{(j)} y_i + p_{33}^{(j)} z_i + p_{34}^{(j)} t_i} \qquad (3)$$

As we have n points and m images, this leads us to $2 \times n \times m$ equations. The unknowns are $11 \times m$ for the P_j which are defined up to a scaling factor, plus $3 \times n$ for the M_i. So if m and n are large enough we have a redundant set of equations.

It is easy to understand that the solution for the equation 1 is not unique. Let A be a spatial collineation represented by its 4×4 invertible matrix. If $P_j, j = 1, \ldots, m$ and $M_i, i = 1, \ldots, n$ are a solution to 1, so are obviously $P_j A^{-1}$ and AM_i, as $\rho_{ij} m_{ij} = (P_j A^{-1})(AM_i), i = 1, \ldots, n, j = 1, \ldots, m$

Therefore is established the first result: The solution of the system 1 can only be defined up to a collineation.

As a consequence of this result, a basis for any 3D collineation can be arbitraryly chosen in the 3D space. For a projective space $\mathbb{P}^3$, 5 algebraically free points form a basis, i.e. a set of 5 points, no four of them coplanar. We will come back to how to choose for such a basis later in 3.1. For convenience, we assume here that the first five points M_i can be chosen to form such a basis; their coordinates can be assigned to the canonical ones: $(1, 0, 0, 0)^T$, $(0, 1, 0, 0)^T$, $(0, 0, 1, 0)^T$, $(0, 0, 0, 1)$ and $(1, 1, 1, 1)^T$.

The remaining part of this section is devoted to the problem of building from these now fixed reference points an explicit solution.

2.2 Direct nonlinear reconstruction

From the above section, the most direct way is to try to solve this system of nonlinear equations. As the projective coordinates of the spatial points are defined up to a constant, so for each point, the constraint $x_i^2 + y_i^2 + z_i^2 + t_i^2 = 1$ can be added. Since the system is an overdetermined one, we can hope to solve it by standard least squares technique. The problem can be formulated as minimizing, over

$$(x_i, y_i, z_i, t_i, p_{11}^{(j)}, \ldots p_{34}^{(j)}) \text{ for } i = 1, \ldots, m, j = 1, \ldots, n;$$

$$F = \sum_{k=1}^{2 \times m \times n + n} \left(\frac{f_k(u_{ij}, v_{ij}; x_i, y_i, z_i, t_i, p_{11}^{(j)}, \ldots p_{34}^{(j)})}{\sigma_k}\right)^2$$

where $f_k(\cdot)$ is either

$$u_{ij} - \frac{p_{11}^{(j)} x_i + p_{12}^{(j)} y_i + p_{13}^{(j)} z_i + p_{14}^{(j)} t_i}{p_{31}^{(j)} x_i + p_{32}^{(j)} y_i + p_{33}^{(j)} z_i + p_{34}^{(j)} t_i}$$

$$\text{or } v_{ij} - \frac{p_{21}^{(j)} x_i + p_{22}^{(j)} y_i + p_{23}^{(j)} z_i + p_{24}^{(j)} t_i}{p_{31}^{(j)} x_i + p_{32}^{(j)} y_i + p_{33}^{(j)} z_i + p_{34}^{(j)} t_i}$$

subject to $x_i^2 + y_i^2 + z_i^2 + t_i^2 - 1 = 0$ for $i = 1, \ldots, m$.

σ_k is the standard deviation of each image measure, u_{ij} or v_{ij}, suppposed normally distributed and uncorrelated. On the other hand, it can also be considered as the weight for each function. So the problem is a general weighted least squares estimation, thus the constraints $x_i^2 + y_i^2 + z_i^2 + t_i^2 - 1 = 0$ can be easily transformed into corresponding penalty functions in order that the whole problem is an unconstrained least squares problem. As for the multiplicative scalar of each projection matrix, we can for example impose $p_{34}^{(j)} = 1$ for $j = 1, \ldots, n$ with no loss of generality. Therefore this system leads to $m + 2 \times n \times m$ equations in $11 \times m + 3 \times n$ unknowns,

This can be solved by the standard nonlinear least squares routine due to Levenberg-Marquardt [13].

Statistically, it is equivalent to the maximum likelihood estimator. The alternative of minimizing $F(\cdot)$ as above is to minimize over $(x_i, y_i, z_i, t_i, p_{11}^{(j)}, \dots p_{34}^{(j)})$,

$$G = \sum_{k=1}^{2\times m\times n+n} (\frac{g_k(u_{ij}, v_{ij}; x_i, y_i, z_i, t_i, p_{11}^{(j)}, \dots p_{34}^{(j)})}{\sigma_k})^2$$

where $g_k(\cdot)$ is either

$$u_{ij}(p_{31}^{(j)}x_i + p_{32}^{(j)}y_i + p_{33}^{(j)}z_i + p_{34}^{(j)}t_i) - (p_{11}^{(j)}x_i + p_{12}^{(j)}y_i + p_{13}^{(j)}z_i + p_{14}^{(j)}t_i)$$

or

$$v_{ij}(p_{31}^{(j)}x_i + p_{32}^{(j)}y_i + p_{33}^{(j)}z_i + p_{34}^{(j)}t_i) - (p_{21}^{(j)}x_i + p_{22}^{(j)}y_i + p_{23}^{(j)}z_i + p_{24}^{(j)}t_i)$$

$g_k(\cdot)$ is a simple algebraic transformation of $f_k(\cdot)$, this transforms the real Euclidean *distance* error into an *algebraic distance* which degrades the error function. However, in doing so, the degree of nonlinearity of equations is greatly reduced, especially the Jacobian matrix of $g_k(\cdot)$ is nicely reduced. This may lead to faster convergence but leaves the solution a little bit degraded, since the distance error is only algebraic, not Euclidean. This point will be discussed later and get confirmed in our experimentation in Section 4.

Since the standard projective basis are assigned to the reference points, the solution provides at the same time the projective shape and each camera's projection matrix. A projective shape is defined up to a collineation, at this stage, no metric information is present, only projective properties are preserved. For example, aligned points remain aligned, coplanar points remain coplanar and conics are transformed into conics, a circle may be represented by an hyperpole ...

Next, a pure projective shape can be transformed into its affine or Euclidean representation. However to do this, supplementary affine and Euclidean information shoud be incorporated. That is, we should determine a collineation A, a matrix of 4×4, which brings the canonical basis $e_i, i = 1, \dots, 5$ to any five points $a_i = (a_{i1}, a_{i2}, a_{i3}, a_{i4})^T = Ae_i$

If these five points are only affinely known, that is, 4 of them can be assigned the standard affine coordinates, the fifth point should have its affine coordinates with respect to these 4 points, that is the 5 points can have the following coordinates $(1, 0, 0, 1)^T$, $(0, 1, 0, 1)^T$, $(0, 0, 1, 1)^T$, $(1, 1, 1, 1)$ and $(\alpha, \beta, \gamma, 1)^T$.

That is, to get the affine representation, affine knowledge (α, β, γ) has to be available. Then by solving the linear equations system above, we obtain the collineation which transforms a pure projective shape into an affine shape.

To have the usual Euclidean shape representation, the Euclidean coordinates should be known for the 5 points, that should be like $(x_i, y_i, z_i, 1)^T$, $i = 1, \dots 5$ then, solve for the corresponding collineation which transforms a pure projective shape into a usual Euclidean shape.

However we can also at the beginning assign the reference points to their Euclidean coordinates, in this case, the 3D reconstruction thus obtained is directly its Euclidean shape.

3 Geometrical reconstruction

In this section, we will show some very interesting geometric properties once the epipolar geometry has been established. In particular, we can determine if any fourth point is coplanar with the plane defined by any three other points. That leads to an automatic selection of general reference points from image planes and point reconstruction in a geometric way.

3.1 The coplanarity test

As we assume here that the epipolar constraint is known, we know the essential matrix E which contains all this information [4, 7]. E is a 3×3 matrix such that from the point $m = (x, y, t)^T$ in image 1, the corresponding epipolar line l' in image 2 has its coefficients satisfying $l' = (a', b', c')^T = Em$.

Now, consider Figure 1. It displays two images of four 3D points A, B, C, D, projected in the two images. The dashed lines correspond to some of the epipolar lines going through each of the vertices of the quadrangles. The epipolar constraint specifies that the epipolar line corresponding to c passes through c', and conversely.

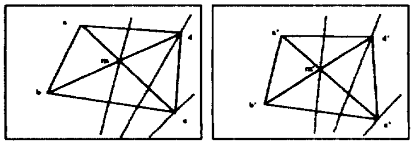

Figure 1: Match of diagonal intersections with epipolar constraint

If A, B, C, D are coplanar, then the diagonals intersect in this 3D space plane in a point M which is projected respectively as m and m'. Therefore m and m' have to satisfy the epipolar constraint too, as it is displayed in Fig. 1. Conversely consider the case where A, B, C, D are not coplanar. The diagonals are no more in the same plane and therefore do not intersect in the space. So m is the image of two 3D points, M_1 lying on (AC), and N_1 lying on (BD). Similarly m' is the image of M_2 and N_2. If the central point O' of the second image is not in the plane defined by (ACO), nor in the plane (BDO), then the 2 view lines (Om) and $(O'm')$ do not intersect, and therefore the points m and m' are not in epipolar correspondence.

The condition that O' does not lie in the plane (OAC) is equivalent to the condition that the epipole in the first image does not lay on (ac), which is therefore checked easily. Notice that in such a case, we can choose as diagonals (AB) and (CD) instead of (AC) and (BD). Therefore the only condition we reach for applying this method is to have none of the projections a, b, c, d being the epipole. So we proved that

Theorem: *If neither a, b, c, nor d are the epipole point of image 2 with respect of image 1, then it exists at least one diagonal intersection m such that m and its corresponding intersection m' satisfy the epipolar constraint if and only if A, B, C, D are coplanar.*

In fact this theorem leads to a useful and straightforward construction technique. Observing three points and a line in an image, it is possible to reconstruct the intersection of this line with the plane defined by the three points, provided the essential matrix.

Such a technical result is particularly useful for computing construction directly in the image without going through the 3D reconstruction. It allows for instance to compute several invariants for stereo images (*cf.* [6]).

3.2 Search for a 5 point basis

The above result can be directly applied to automatically select the necessary reference points from image points for projective reconstruction without any *a priori* spatial knowledge. Basically, we will be able to get rid of the coplanar reference points. In this case, one possible version of the algorithm can be 1. choose for M_1 and M_2 the farthest points pair in one of the image; 2. choose for M_3 the farthest point from M_1M_2; 3. sort the other points according to the *distances* to the plane determined by the triangle $M_1M_2M_3$, choose for M_4 the one which has the maximum distance. The *distance* is not the orthogonal distance from the point to the plane as we expect (not possible at this step), it is the projection on the second image of the distance from the point to the plane along the first viewing line of that point; 4. Sort the remaining points according to the maximum distance to any face of the tetrahedron M_1, M_2, M_3, M_4, choose for M_5 the point which has the maximum distance.

This will provide us with a reasonably scattered points set.

4 Experimental results

4.1 Qualitative results

All our experiences are conducted with a Pulnix 765 camera, a lens of 18mm kinoptics and FG150 Imaging technology grab board. The camera is assumed to be a perfect pin-hole one, distorsion is not compensated. One of the experiments is performed on a wooden house. The camera is set in about 2m from the object. We tracked over about one hundred images covering roughly two sides of the wooden house. The points tracked are the curvature extrama of smoothed contour chains by B-splines. In this experience, we also wanted to validate the reconstruction with points only present in part of the sequence. In total, 73 points were tracked, but almost half of them are present between only two successive views. Final reconstruction is done with five views of them. In Figure 2, three images of the sequence are displayed.

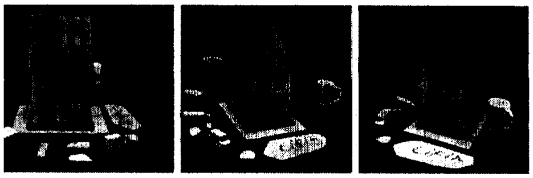

Figure 2: The wooden house image sequence

The reconstruction, illustrated in Figure 3, has an excellent qualitative aspect.

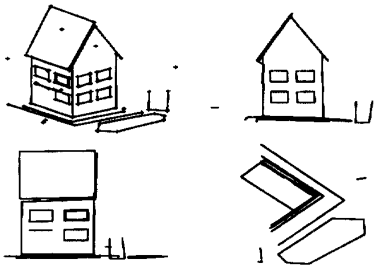

Figure 3: The reconstructed wooden house

As we have mentionned, we have choice between minimizing $F(\cdot)$ and $G(\cdot)$. Experiences confirm that

estimated coordinates	measured coordinates	absolute error
(11.678 -0.013 8.256)	(11.8 0 8.1)	(0.122 0.013 0.156)
(7.697 -0.035 10.663)	(7.85 0 10.5)	(0.153 0.035 0.163)
(6.646 -0.300 24.051)	(6.85 -0.4 23.8)	(0.204 0.1 0.251)
(11.666 0.007 10.766)	(11.8 0 10.5)	(0.134 0.007 0.266)
(7.773 -0.166 8.241)	(7.85 0 8)	(0.077 0.166 0.241)
(-0.065 1.300 7.922)	(0 1.35 7.8)	(0.065 0.05 0.122)
(-0.139 7.261 7.860)	(0 7.3 7.75)	(0.139 0.039 0.11)
(0.082 1.372 10.407)	(0 1.4 10.35)	(0.082 0.028 0.057)
(0.007 7.250 10.325)	(0 7.35 10.3)	(0.007 0.10 0.025)
(-1.488 -0.298 13.299)	(-1.7 -0.5 12.8)	(0.212 0.202 0.501)
(-1.086 18.143 12.934)	(-1.7 18.2 12.8)	(0.614 0.057 0.134)

while minimizing $G(\cdot)$, with very few, about 5 iterations instead of about 10 or even more, we can obtain a quite satisfactory solution. But since the distance error is only algebraic, not Euclidean, the solution is always slightly degraded. In our experiments, we put them in a sequential order, beginning with minimizing $G(\cdot)$, and ending with minimizing $F(\cdot)$.

All experimental results are performed by Levenberg-Marquardt's algorithm. Practical experimentation shows that the algorithm works very well. The convergence does not depend too much on the initial starting points, it converges with almost ***any initialization*** although we have no mean to formally prove its convergence. For our experiments, each projection matrix is systematically initialized as unity identity matrix and each point is started from $(0.5, 0.5, 0.5, 0.5)$. Some other experiments, for instance on a paper house and the calibration patterns, are also performed and have been reported in the technical report [12].

4.2 Quantitative results

The accuracy of the tracked points is generally within two pixels, but some of them may have more than that. To have an idea of the precision of the reconstructed points, we measured some points' coordinates of the wooden house by a ruler. The following numerical table 4.2 shows the absolute errors of the reconstruction of some selected points. While taking into account of rough measures' performance by the ruler, the absolute error is within one millimeter, it is a very acceptable result.

As the least squares estimator can be considered as an maximum likelyhood one if we admit that the images points are normally distributed, that is what we assumed at the beginning. The confidence limits of the reconstructed points can be estimated from the corresponding covariance matrix provided by Levenberg-Marquardt's algorithm. In Figure 4, the confidence region ellipsoid of each point corresponding to 68.3 percent confidence region is displayed. For simplicity, each associated ellipsoid is displayed by its corresponding bounding parallelepiped.

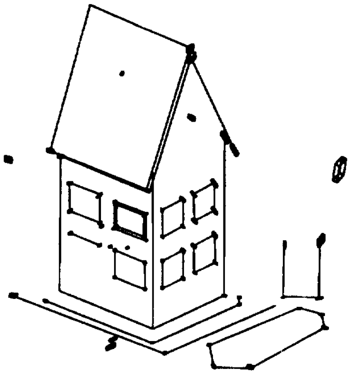

Figure 4: Confidence region

It is very important to note that in this figure the point with the largest confidence region is just the point which lies on a cup, therefore is not a real 3D "corner" in the original image. The general correct rigid "corner" points have very small confidence regions.

5 Discussion

The qualitative results are excellent. If we assume that the exact location of the reference points are known, quantitative results can be obtained; they are better than those provided by stereovision, but still not excellent enough for some industrial applications. One first way to improve the location accuracy is to have better location measures in the image. Another source of inaccuracy is the lens distorsion which is not yet taken into account.

For the same problem, Faugeras [5] and independently Hartley [7] provide an elegant linear projective reconstruction which heavily relies on the computation of the epipolar geometry and the associated fundamental matrix. The results we obtained with their method was much less accurate then the one we got with our approach; but as we were unable to reproduce their accuracy in the computation of the fundamental matrix, no comparison can be made right now. A common testbed will be set in the near future in order to be able to compare with the linear methods proposed independently by Faugeras [5] and Hartley [7].

Acknowledgement

We are pleased to acknowledge Horst Beyer for pointing out to us that photogrammetrists use Levenberg-Marquardt for closely related problems, Olivier D. Faugeras and his associates for fruitful discussions on the topic and our colleague B. Boufama who provided the tracking algorithm.

References

[1] H.A. Beyer. Accurate calibration of CCD cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 96–101, 1992.

[2] D.C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37(8):855–866, 1971.

[3] J.B. Burns, R. Weiss, and E.M. Riseman. View variation of point set and line segment features. In *Proceedings of DARPA Image Understanding Workshop, Pittsburgh, Pennsylvania, USA*, pages 650–659, 1990.

[4] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 563–578. Springer-Verlag, May 1992.

[5] O.D. Faugeras. *3D Computer Vision*. M.I.T. Press, 1992.

[6] P. Gros and L. Quan. Projective Invariants for Vision. Technical Report RT 90 IMAG - 15 LIFIA, IRIMAG–LIFIA, Grenoble, France, December 1992.

[7] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764, 1992.

[8] J.J. Koenderink and A. J. van Doorn. Affine structure from motion. Technical report, Utrecht University, Utrecht, The Netherlands, October 1989.

[9] C.H. Lee and T. Huang. Finding point correspondences and determining motion of a rigid object from two weak perspective views. *Computer Vision, Graphics and Image Processing*, 52:309–327, 1990.

[10] R. Mohr and E. Arbogast. It Can Be Done without Camera Calibration. Technical Report RR 805-I- IMAG 106 LIFIA, IMAG - LIFIA, February 1990.

[11] R. Mohr, L. Morin, C. Inglebert, and L. Quan. Geometric solutions to some 3D vision problems. In R. Storer J.L. Crowley, E. Granum, editor, *Integration and Control in Real Time Active Vision*, ESPRIT BRA Series. Springer-Verlag, 1991.

[12] R. Mohr, L. Quan, F. Veillon, and B. Boufama. Relative 3D reconstruction using multiples uncalibrated images. Technical Report RT 84-I-IMAG LIFIA 12, IRIMAG–LIFIA, 1992.

[13] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling W.T. *Numerical Recipes in C*. Cambridge University Press, 1988.

[14] L. Quan and R. Mohr. Affine shape representation from motion through reference points. *Journal of Mathematical Imaging and Vision*, 1:145–151, 1992.

[15] J.G. Semple and G.T. Kneebone. *Algebraic Projective Geometry*. Oxford Science Publication, 1952.

[16] G. Sparr. Projective invariants for affine shapes of point configurations. In *Proceeding of the* DARPA–ESPRIT *workshop on Applications of Invariants in Computer Vision, Reykjavik, Iceland*, pages 151–170, March 1991.

[17] C. Tomasi and T. Kanade. Factoring image sequences into shape and motion. In *Proceedings of IEEE Workshop on Visual Motion, Princeton, New Jersey*, pages 21–28, Los Alamitos, California, USA, October 1991. IEEE Computer Society Press.