Stereo from Uncalibrated Cameras

Richard Hartley, Rajiv Gupta and Tom Chang
GE - Corporate Research and Development,
P.O. Box 8, Schenectady, NY, 12301.

Abstract

This paper considers the problem of computing placement of points in 3 dimensional space given two uncalibrated perspective views. The main theorem shows that the placement of the points is determined only up to an arbitrary projective transformation of 3-space. Given additional ground control points, however, the location of the points and the camera parameters may be determined. The method is linear and non-iterative whereas previously known methods for solving the camera calibration and placement to take proper account of both ground-control points and image correspondences are unsatisfactory in requiring either iterative methods or model restrictions.

As a result of the main theorem, it is possible to determine projective invariants of 3-D geometric configurations from two perspective views.

1 Introduction

A typical system for the construction of 3-D models from stereo imagery operates in three phases. In the first phase a set of *matched points* (i.e., pixels in the two views that are the images of the same point in the real world), are established between the two images. In the second phase, the identified matched points are used to derive the relative locations, orientations and other parameters of the cameras. This process usually requires iterative solution of a set of non-linear equations. In a third phase the locations of 3-D points are computed.

This paper describes a method of computing the 3-D point locations without explicit computation of the camera models. The method is related to work of Longuet-Higgins ([4]), but is different in that we do not assume a known calibration of the cameras.

1.1 Notation

The symbol $\mathbf{u}$ represents a column vector. We will use the letters u, v and w for homogeneous coordinates in image-space. In particular, the symbol $\mathbf{u}$ represents the column vector $(u, v, w)^\top$. Object space points will also be represented by homogeneous coordinates x, y, z and t, or more often x, y, z and 1. The symbol $\mathbf{x}$ will represent a point in three-dimensional projective space represented in homogeneous coordinates.

Since all vectors are represented in homogeneous coordinates, their values may be multiplied by any arbitrary non-zero factor. The notation $\approx$ is used to indicate equality of vectors or matrices up to multiplication by a scale factor.

Given a vector, $\mathbf{t} = (t_x, t_y, t_z)^\top$ it is convenient to introduce the skew-symmetric matrix

$$[\mathbf{t}]_\times = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} \tag{1}$$

For any non-zero vector $\mathbf{t}$, matrix $[\mathbf{t}]_\times$ has rank 2. Furthermore, the null-space of $[\mathbf{t}]_\times$ is generated by the vector $\mathbf{t}$. This means that $\mathbf{t}^\top[\mathbf{t}]_\times = [\mathbf{t}]_\times\mathbf{t} = 0$ and that any other vector annihilated by $[\mathbf{t}]_\times$ is a scalar multiple of $\mathbf{t}$.

1.2 Camera Model

The general model of a perspective camera that will be used here is that represented by an arbitrary 3×4 matrix, P, known as the *camera matrix*. The camera matrix transforms points in 3-dimensional projective space to points in 2-dimensional projective space according to the equation $\mathbf{u} = P\mathbf{x}$. The camera matrix P is defined up to a scale factor only, and hence has 11 independent entries. As shown by Strat ([6]), this model allows for the modeling of several parameters, in particular : the location and orientation of the camera; the principal point offsets in the image space; and unequal scale factors in two directions parallel to the axes in image space.

This accounts for 10 of the total 11 entries in the camera matrix. It may be seen that if unequal stretches in two directions **not** aligned with the image axes are allowed, then a further 11-th camera parameter may be defined. In practical cases, the focal length (magnification) of the camera may not be known, and neither may the principal point offsets. Strat [6] gives an example of an image where the camera parameters take on surprising values. Our purpose in treating general camera transforms is to avoid the necessity for arbitrary assumptions about the camera.

If the camera is not placed at infinity, then the left-hand 3×3 submatrix of P is non-singular. Then P can be written as $P = (M \mid -MT)$ where T is a vector representing the location of the camera. By the method of QR-factorization ([1]), M may be written as a product $M = KR$ where K is upper triangular and R is a rotation matrix. The matrix K represents the so-called internal parameters of the camera. If K is known *a priori*, then we say that the camera is *calibrated*. For calibrated cameras, a common simplification is to assume that the matrix K is the identity, so that M is a rotation matrix.

1.3 Overview

Corresponding to a pair of cameras, there exists a 3×3 matrix Q known as the essential matrix ([4, 3]), such that if $\mathbf{u}$ and $\mathbf{u'}$ are a pair of matched points expressed in homogeneous coordinates, then $\mathbf{u'}^\top Q \mathbf{u} = 0$. If a sufficient number of matched points are known, the matrix Q may be computed by the solution of a (possibly overdetermined) set of linear equations. If the internal calibration of the cameras is known, then it is possible to determine from Q the relative placement of the cameras and hence the relative locations of the 3-D points corresponding to the matched points. It is shown in [3] that this is also true even when the focal lengths of the two cameras are unknown. Unfortunately, for uncalibrated cameras, it is not possible to compute the camera parameters or the point locations unambiguously. However, we prove in Theorem 1 that the various solutions (i.e., the 3-D location of points and the camera transformations matrices) that are compatible with the given set of matched points are related with each other via a 3-dimensional projective transformation H. Then, we show how one can compute (non-uniquely) two camera transformations P_1 and P_2 from Q and use them to find a tentative set of 3-D points locations. Since both P_1 and P_2, and the set of points may be off by an unknown projective transformation H, ground control points are used to compute the true 3-D location of the points.

Thus, the 3-D point locations are found by considering both matched points and ground control points using linear methods. Other purely non-iterative methods (e.g. those by Sutherland [7] or Longuet-Higgins [4]) are not able to handle ground-control and matched points simultaneously. In our approach, we avoid the explicit computation of internal or external camera parameters though they may easily be obtained if needed.

2 Theory

We consider a general pair of camera matrices represented by $P_1 = (M_1 \mid -M_1T_1)$ and $P_2 = (M_2 \mid -M_2T_2)$. The form of the matrix Q may be given in terms of P_1 and P_2.

Lemma 1 *The essential matrix corresponding to the pair of camera matrices $(M_1 \mid -M_1T_1)$ and $(M_2 \mid -M_2T_2)$ is given by*

$$Q \approx M_2^* {M_1}^\top [M_1(T_2 - T_1)]_\times \ .$$

Here A^* represents the adjoint of a matrix A, that is, the matrix of cofactors. If A is an invertible matrix, then $A^* \approx (A^\top)^{-1}$. For a proof of Lemma 1 see [3].

As is indicated by the previous lemma, an essential matrix Q factors into a product $Q = RS$, where R is a non-singular matrix and S is skew-symmetric. The next lemma shows to what extent this factorization is unique.

Lemma 2 *Let the 3×3 matrix Q factor in two different ways as $Q \approx R_1S_1 \approx R_2S_2$ where each S_i is a non-zero skew-symmetric matrix and each R_i is non-singular. Then $S_2 \approx S_1$. Furthermore, if $S_i = [\mathbf{t}]_\times$ then $R_2 \approx R_1 + \mathbf{a}\mathbf{t}^\top$ for some vector $\mathbf{a}$.*

Proof: Since R_1 and R_2 are non-singular, it follows that $Q\mathbf{t} = 0$ if and only if $S_i\mathbf{t} = 0$. From this it follows that the null-spaces of the matrices S_1 and S_2 are equal, and so $S_1 \approx S_2$. Matrices R_1 and R_2 must both be solutions of the linear equation $Q \approx RS$. Consequently, they differ by the value $\mathbf{a}\mathbf{t}^\top$ as required. □

We now prove our main theorem which indicates when two pairs of camera matrices correspond to the same essential matrix.

Theorem 1 *Let $\{P_1, P_2\}$ and $\{P_1', P_2'\}$ be two pairs of camera transforms. Then $\{P_1, P_2\}$ and $\{P_1', P_2'\}$ correspond to the same essential matrix Q if and only if there exists a 4×4 non-singular matrix H such that $P_1H \approx P_1'$ and $P_2H \approx P_2'$.*

Proof : First we prove the **if** part of this theorem. To this purpose, let $\{\mathbf{x}_i\}$ be a set of at least 8 points in 3-dimensional space and let $\{\mathbf{u}_i\}$ and $\{\mathbf{u}_i'\}$ be the corresponding image-space points as imaged by the two camera P_1 and P_2. By the definition of the essential matrix, Q satisfies the condition $\mathbf{u}_i'^\top Q \mathbf{u}_i = 0$ for all i. We may assume that the points $\{\mathbf{x}_i\}$ have been chosen in such a way that the matrix Q is uniquely defined up to scale by the above equation. The point configurations that defeat this definition of the essential matrix are discussed in [4]. Suppose now that there exists a 4×4 matrix H taking P_1 to P_1' and P_2 to P_2' in the sense specified by the hypotheses of the theorem. For each i let $\mathbf{x}_i' = H^{-1}\mathbf{x}_i$. Then we see that

$$P_j'\mathbf{x}_i' = P_jHH^{-1}\mathbf{x}_i = P_j\mathbf{x}_i = u_i$$

for $j = 1, 2$. In other words, the image points $\{\mathbf{u}_i\}$ and $\{\mathbf{u}'_i\}$ are a matched point set with respect to the cameras P'_1 and P'_2. Thus the essential matrix for this pair of cameras is defined by the same relationship $\mathbf{u}'^{\top}_i Q \mathbf{u}_i = 0$ that defines the essential matrix of the pair P_1 and P_2. Consequently, the two camera pairs have the same essential matrix.

Now, we turn to the **only if** part of the theorem and assume that two pairs of cameras have the same essential matrix, Q. First, we consider the camera pair $\{(M_1 \mid -M_1 T_1), (M_2 \mid -M_2 T_2)\}$. It is easily seen that the 4×4 matrix

$$\begin{pmatrix} M_1^{-1} & T_1 \\ 0 & 1 \end{pmatrix}$$

transforms this pair to the camera pair

$$\{(I \mid 0), (M_2 M_1^{-1} \mid -M_2(T_2 - T_1))\}$$

where I and 0 are identity matrix and zero column vector respectively. Furthermore by the **if** part of this theorem (or as verified directly using Lemma 1), this new camera pair has the same essential matrix as the original.

Applying this transformation to each of the camera pairs

$$\{(M_1 \mid -M_1 T_1), (M_2 \mid -M_2 T_2)\}$$

and

$$\{(M'_1 \mid -M'_1 T'_1), (M'_2 \mid -M'_2 T'_2)\}$$

we see that there is 4×4 matrix transforming one pair to the other if and only if there is such a matrix transforming

$$\{(I \mid 0), (M_2 M_1^{-1} \mid -M_2(T_2 - T_1))\}$$

to

$$\{(I \mid 0), (M'_2 M'^{-1}_1 \mid -M'_2(T'_2 - T'_1))\}$$

Thus, we are reduced to proving the theorem for the case where the first cameras, P_1 and P'_1 of each pair are both equal to $(I \mid 0)$. Thus, let $\{(I \mid 0), (M \mid -MT)\}$ and $\{(I \mid 0), (M' \mid -M'T')\}$ be two pairs of cameras corresponding to the same essential matrix. According to Lemma 1, the Q-matrices corresponding to the two pairs are $M^*[T]_\times$ and $M'^*[T']_\times$ respectively, and these must be equal (up to scale). According to Lemma 2, $T \approx T'$ and $M'^* \approx M^* + \mathbf{a}T^\top$ for some vector $\mathbf{a}$. Taking the transpose of this last relation yields

$$M'^{-1} \approx M^{-1} + T\mathbf{a}^\top \qquad (2)$$

At this point we need to interpolate a lemma.

Lemma 3 *For any column vector* $\mathbf{t}$ *and row vector* $\mathbf{a}^\top$*, if* $I + \mathbf{t}\mathbf{a}^\top$ *is invertible then*

$$(I + \mathbf{t}\mathbf{a}^\top)^{-1} = I - k\mathbf{t}\mathbf{a}^\top$$

where $k = 1/(1 + \mathbf{a}^\top \mathbf{t})$.

Proof : The proof is done by simply multiplying out the two matrices and observing that the product is the identity. One might ask what happens if $\mathbf{a}^\top \mathbf{t} = 1$ in which case k is undefined. The answer is that in that case, $I - \mathbf{t}\mathbf{a}^\top$ is singular, contrary to hypothesis. Details are left to the reader. □

Now we may continue with the proof of the theorem. Referring back to (2), it follows that

$$\begin{aligned} M' &\approx (M^{-1} + T\mathbf{a}^\top)^{-1} \\ &\approx (M^{-1}(I + MT\mathbf{a}^\top))^{-1} \\ &\approx (I - kMT\mathbf{a}^\top)M \\ &\approx M - kMT(\mathbf{a}^\top M) \end{aligned}$$

and

$$\begin{aligned} M'T &\approx MT - kMT(\mathbf{a}^\top MT) \\ &\approx k'MT \quad \approx \quad MT \end{aligned} \qquad (3)$$

where $k' = 1 - k\mathbf{a}^\top MT$. Since $T' \approx T$ according to Lemma 2, $M'T' \approx MT$. From these results, it follows that

$$(M' \mid -M'T') \approx (M \mid -MT) \begin{pmatrix} I & 0 \\ k\mathbf{a}^\top M & k'' \end{pmatrix}$$

for some constant k''.

This completes the proof of the theorem. □

2.1 Choosing a Realization of Q.

Given a set of image correspondences $\mathbf{u}_i \leftrightarrow \mathbf{u}'_i$ defining an essential matrix Q, the previous theorem shows that one cannot unambiguously determine the position of the cameras, or the corresponding object-space points from Q. Since Q contains all the information that is available from the point correspondences, it follows that the position of the cameras and the object points can be determined only up to a 3-dimensional projective transform as specified by the matrix H. In order to determine the positions of the object-space points $\{x_i\}$ unambiguously, it is necessary for some ground-control points to be specified. Our strategy, therefore, is to select *any* pair of camera placements consistent with the essential matrix, Q. Later, a 3-dimensional projective transform will be carried out to transform to an absolute coordinate system.

The first task is to determine a pair of camera matrices corresponding to a given essential matrix, Q. To this purpose, suppose that the singular value decomposition [1] of Q is given by $Q = UDV^\top$, where D is the diagonal matrix $D = \mathrm{diag}(r, s, 0)$. In a practical case, the smallest singular value of Q will not be exactly equal to 0 because of numerical inaccuracies. However, setting the smallest singular value to 0 gives the matrix closest to Q in Euclidean norm that has the required rank 2. The following factorization of Q may now be verified by inspection.

$$Q = RS \;;\quad R = U\mathrm{diag}(r, s, \gamma)EV^\top \;;\quad S = VZV^\top$$

where

$$E = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \;;\; Z = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and γ is any non-zero number, but is best chosen to lie between r and s so that the condition number [1] of R is as good as possible. From Lemma 1 it follows that the pair of camera matrices

$$P_1 = (I \mid 0) \;;\; P_2 = (U\mathrm{diag}(r, s, \gamma)EV^\top \mid U(0, 0, \gamma)^\top)$$

correspond to the given essential matrix, Q. It is in no way intended that this should represent the true placement of the cameras, but it is related to the true camera placement by a 3-dimensional projective transformation.

2.2 Computation of 3-D Points.

The point in the object space that projects on to $\mathbf{u}_i = (u_i, v_i, 1)^\top$ and $\mathbf{u}'_i = (u'_i, v'_i, 1)^\top$ in the two images, under the transforms P_1 and P_2, can be computed as follows. The equations of the rays originating at the focal point of the two cameras and passing through the two matched points are given by

$$(w_i u_i,\ w_i v_i,\ w_i)^\top = P_1(x_i,\ y_i,\ z_i,\ 1)^\top$$

$$(w'_i u'_i,\ w'_i v'_i,\ w'_i)^\top = P_2(x_i,\ y_i,\ z_i,\ 1)^\top$$

The values of u_i, v_i, u'_i, v'_i, P_1 and P_2 are known, whereas x_i, y_i, z_i, w_i and w'_i are unknown. Thus we have 6 equations in 5 unknowns and the vector $\mathbf{x}_i = (x_i, y_i, z_i)$ that minimizes the error can be computed. This will correspond to the point of intersection of these two rays, if they do intersect in space, or the point midway between the points of their closest approach.

2.3 Absolute Point Placement

Since the relative 3-D points computed above may be off by a perspective transformation, ground control points are needed to transform the relative coordinates to absolute coordinates in some user-specified coordinate system. In order to determine absolute placements of the cameras, it is necessary to have at least 8 ground control points to resolve the ambiguity in camera placements derived from the matched point data. The method that is used here may be regarded in some ways as a generalization of the method of Sutherland [7] to more than one camera. Suppose that we have n cameras represented by matrices $P_1, P_2, \ldots, P_n$ and a set of ground control points $\{\mathbf{x}_i\}$, where ground control point $\mathbf{x}_i$ is visible in camera $P_{\sigma(i)}$, the corresponding image-coordinates being $\mathbf{u}_i = (u_i, v_i, 1)^\top$. It is assumed that there is a 4×4 non-singular matrix H that transforms each P_i to its true placement. This leads to a set of equations

$$\begin{pmatrix} w_i u_i \\ w_i v_i \\ w_i \end{pmatrix} = P_{\sigma(i)} H \begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix}$$

The only unknowns in this set of equations are the entries of the matrix H and the values w_i, the above equations may be written as a set of equations

$$\begin{aligned} w_i u_i &= A_i(h_{11}, h_{12}, \ldots, h_{44}) \\ w_i v_i &= B_i(h_{11}, h_{12}, \ldots, h_{44}) \\ w_i &= C_i(h_{11}, h_{12}, \ldots, h_{44}) \end{aligned}$$

where A, B and C are linear expressions in the entries h_{jk} of H. Since the w_i are unknown values, it is possible to eliminate them from the above equations by writing

$$\begin{aligned} C_i(h_{11}, \ldots, h_{44}) u_i &= A_i(h_{11}, \ldots, h_{44}) \\ C_i(h_{11}, \ldots, h_{44}) v_i &= B_i(h_{11}, \ldots, h_{44}) \end{aligned}$$

This gives a set of linear equations in the entries h_{jk} of H, which can be solved to find the matrix H. The solution will be determined only up to a scale factor, corresponding to the fact that H is itself only determined up to a scale factor. At least 15 equations are needed for a solution and each point gives two equations. If the image of a 3-D point is known in both images, then this gives rise to 4 equations, but it can be shown that only three of these are linearly independent. Once H is known, the true 3-D points may be computed by applying the inverse transformation, H^{-1} to the points $\mathbf{x}_i$ computed earlier.

3 Application to Invariants

It results from Theorem 1 that a configuration of 8 points or more (except for degenerate cases for which the essential matrix can not be determined ([4])) is determined up to a 3-dimensional projective transformation by two perspective views. In this case, any projective invariant ([5])of a set of 3-D points can be computed from the two views. For instance, six points in 3-dimensions determine 3 independent projective invariants[1]. The invariant can be computed as follows. Given a set of 8 matched points or more, the essential matrix Q can be computed. A realization of Q can be chosen as in section 2.1 and the point locations can be computed as in section 2.2. Now, from any subset of 6 points projective invariants may be computed which because of Theorem 1 will be invariants of the true locations of the points in space.

Many other 3 dimensional geometric configurations give rise to invariants which may also be computed using this method. Details are left to another paper.

4 Conclusions

The techniques presented in this paper have been implemented and tested by augmenting the STEREOSYS testbed ([2]). Our experiments reveal that these techniques result in fast processing and reliable point estimates. We thank Marsha Jo Hannah for the use of the STEREOSYS program and Pat Taylor for converting it to C++.

References

[1] K.E. Atkinson, "An Introduction to Numerical Analysis," John Wiley and Sons, 2nd Edition, 1989.

[2] M.J. Hannah, "A description of SRI's baseline stereo system," ARI International Artificial Intelligence Center Tech. Note 365, Oct. 1985. Workshop, College Park, MD, April 1980, pp. 201–208.

[3] R. Hartley, "Estimation of Relative Camera Positions for Uncalibrated Cameras,", Proc. of ECCV-92, G. Sandini Ed., LNCS-Series Vol. 588, Springer- Verlag, 1992.

[4] H.C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," Nature, Vol. 293, 10 Sept. 1981.

[5] J. L. Mundy and A. Zisserman, "Geometric Invariants in Computer Vision,", MIT Press, to appear, 1992.

[6] T.M. Strat, " Recovering the camera parameters from a transformation matrix," Proc. of DARPA Image Understanding Workshop, New Orleans, LA, pp. 264–271, 1984.

[7] I.E. Sutherland, "Three dimensional data input by tablet," Proceedings of IEEE, Vol. 62, No. 4, pp. 453–461, April 1974.

[1] Choose projective coordinates so that five of the points are $(1, 0, 0, 0)^\top$, $(0, 1, 0, 0)^\top$, $(0, 0, 1, 0)^\top$, $(0, 0, 0, 1)^\top$ and $(1, 1, 1, 1)^\top$. The coordinates of the final point are invariants.