Triangulation

Richard I. Hartley[1] and Peter Sturm[2]

[1] GE-CRD, Rm K1-5C39, P.O. Box 8, Schenectady NY 12301, USA
Phone: (518)-387-7333, Fax: (518)-387-6845, Email: hartley@crd.ge.com
[2] Lifia-Inria, 46, Avenue Félix Viallet, 38031 Grenoble, France
Phone: (33) 76.57.43.28, Fax: (33) 76.57.46.02, Email: Peter.Sturm@imag.fr

Abstract. In this paper, we consider the problem of finding the position of a point in space given its position in two images taken with cameras with known calibration and pose. This process requires the intersection of two known rays in space, and is commonly known as triangulation. In the absence of noise, this problem is trivial. When noise is present, the two rays will not generally meet, in which case it is necessary to find the best point of intersection. This problem is especially critical in affine and projective reconstruction in which there is no meaningful metric information about the object space. It is desirable to find a triangulation method that is invariant to projective transformations of space. This paper solves that problem by assuming a Gaussian noise model for perturbation of the image coordinates. A non-iterative solution is given that finds a global minimum. Extensive comparisons of the new method with several other methods show that it consistently gives superior results.

1 The Triangulation Problem

We suppose that a point $\boldsymbol{x}$ in R^3 is visible in two images. The two camera matrices P and P' corresponding to the two images are supposed known. Let $\boldsymbol{u}$ and $\boldsymbol{u}'$ be projections of the point $\boldsymbol{x}$ in the two images. From this data, the two rays in space corresponding to the two image points may easily be computed. The triangulation problem is to find the intersection of the two lines in space. At first sight this is a trivial problem. Unfortunately, in the presence of noise these rays can not be guaranteed to cross, and we need to find the best solution under some assumed noise model.

A commonly suggested method [1] is to choose the mid-point of the common perpendicular to the two rays (the *mid-point method*). In the case of projective reconstruction, or affine reconstruction however, the camera matrices, will be known in a projective frame of reference, in which concepts such as common perpendicular, or mid-point have no sense. In this case, the mid-point method here will not work. This is confirmed by [1].

Beardsley et. al. suggest an alternative method [1, 2] based on "quasi-Euclidean" reconstruction. The disadvantage of this method is that an approximate calibration of the camera is needed. It is also clearly sub-optimal.

In this paper a new algorithm is described that gives an optimal global solution to the triangulation problem, equally valid in both the affine and projective

Hlaváč, Šára (Eds.): CAIP '95 Proceedings, LNCS 970

reconstruction cases. The solution relies on the concepts of epipolar correspondence and fundamental matrix [3]. The algorithm is non-iterative and simple in concept. In a series of experiments, it is extensively tested against many other methods of triangulation, and found to give consistent superior performance.

The triangulation problem is a small cog in the machinery of computer vision, but in many applications of scene reconstruction it is a critical one, on which ultimate accuracy depends [1].

2 Transformational Invariance

In the last few years, there has been considerable interest in the subject of affine or projective reconstruction [3, 6, 8, 10, 13, 11, 12]. In such reconstruction methods, a 3D scene is to be reconstructed up to an unknown transformation from the given class. Normally, in such a situation, instead of knowing the correct pair of camera matrices P and P', one has a pair PH^{-1} and $P'H^{-1}$ where H is an unknown transformation.

A desirable feature of a method of triangulation used is that it should be invariant under transformations of the appropriate class. Thus, denote by τ a triangulation method used to compute a 3D space point $\boldsymbol{x}$ from a point correspondence $\boldsymbol{u} \leftrightarrow \boldsymbol{u}'$ and a pair of camera matrices P and P'. We write

$$\boldsymbol{x} = \tau(\boldsymbol{u}, \boldsymbol{u}', P, P') \ .$$

The triangulation is said to be invariant under a transformation H if

$$\tau(\boldsymbol{u}, \boldsymbol{u}', P, P') = H^{-1}\tau(\boldsymbol{u}, \boldsymbol{u}', PH^{-1}, P'H^{-1}) \ .$$

This means that triangulation using the transformed cameras results in the transformed point. If the camera matrices are known only up to an affine (or projective) transformation, then it is clearly desirable to use an affine (resp. projective) invariant triangulation method to compute the 3D space points.

3 The Minimization Criterion

We assume that the camera matrices, and hence the fundamental matrix F [7], are known exactly, or at least with great accuracy compared with a pair of matching points in the two images. The two rays corresponding to a matching pair of points $\boldsymbol{u} \leftrightarrow \boldsymbol{u}'$ will meet in space if and only if the points satisfy the familiar relationship [9]

$$\boldsymbol{u}'F\boldsymbol{u} = 0 \ . \tag{1}$$

It is clear, particularly for projective reconstruction, that it is inappropriate to minimize errors in the 3D projective space, $\mathcal{P}^3$. Normally, errors occur not in placement of a feature in space, but in its location in the two images, due to digitization errors, or the exact identification of a feature in the image. It is common to assume that features in the images are subject to Gaussian noise which displaces the feature from its correct location in the image. We assume that noise model in this paper.

A typical observation consists of a noisy point correspondence $\boldsymbol{u} \leftrightarrow \boldsymbol{u}'$ which does not in general satisfy the epipolar constraint (1). We seek the points $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{u}}'$ that minimize the function

$$d(\boldsymbol{u}, \hat{\boldsymbol{u}})^2 + d(\boldsymbol{u}', \hat{\boldsymbol{u}}')^2 \ , \tag{2}$$

where $d(*, *)$ represents Euclidean distance, subject to the epipolar constraint

$$\hat{\boldsymbol{u}}'^{\top} F \hat{\boldsymbol{u}} = 0 \ . \tag{3}$$

Assuming a Gaussian error distribution, the points $\hat{\boldsymbol{u}}'$ and $\hat{\boldsymbol{u}}$ are the most likely values for true image point correspondences. Once $\hat{\boldsymbol{u}}'$ and $\hat{\boldsymbol{u}}$ are found, the point $\boldsymbol{x}$ may be found by any triangulation method, since the corresponding rays will meet precisely in space.

4 An Optimal Method of Triangulation

In this section, we describe a method of triangulation that finds the global minimum of the cost function (2) using a non-iterative algorithm. If the Gaussian noise model can be assumed to be correct, this triangulation method is then provably optimal. This new method will be referred to as the **Polynomial** method, since it requires the solution of a sixth order univariate polynomial.

4.1 Reformulation of the Minimization Problem

Given a measured correspondence $\boldsymbol{u} \leftrightarrow \boldsymbol{u}'$, we seek points $\hat{\boldsymbol{u}}'$ and $\hat{\boldsymbol{u}}$ that minimize the sum of squared distances (2) subject to the epipolar constraint (3). We may formulate the minimization problem differently as follows. We seek to minimize

$$d(\boldsymbol{u}, \boldsymbol{\lambda})^2 + d(\boldsymbol{u}', \boldsymbol{\lambda}')^2 \tag{4}$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$ range over all choices of corresponding epipolar lines. The point $\hat{\boldsymbol{u}}$ is then the closest point on the line $\boldsymbol{\lambda}$ to the point $\boldsymbol{u}$ and $\hat{\boldsymbol{u}}'$ is similarly defined.

Our strategy for minimizing (4) is as follows

1. Parametrize the pencil of epipolar lines in the first image by a parameter t. Thus an epipolar line in the first image may be written as $\boldsymbol{\lambda}(t)$.
2. Compute the corresponding epipolar line $\boldsymbol{\lambda}'(t)$ in the second image, using the fundamental matrix F,
3. Express $d(\boldsymbol{u}, \boldsymbol{\lambda}(t))^2 + d(\boldsymbol{u}', \boldsymbol{\lambda}'(t))^2$ explicitly as a function of t.
4. Find the value of t that minimizes this function.

In this way, the problem is reduced to that of finding the minimum of a function of a single variable, t. It will be seen that for a suitable parametrization of the pencil of epipolar lines the distance function is a rational polynomial function of t. Using techniques of elementary calculus, the minimization problem reduces to finding the real roots of a polynomial of degree 6.

4.2 Details of Minimization

If both of the image points correspond with the epipoles, then the point in space lies on the line joining the camera centres. In this case it is impossible to determine the position of the point in space. If only one of the corresponding point lies at an epipole, then we conclude that the point in space must coincide with the other camera centre. Consequently, we assume that neither of the two image points $\boldsymbol{u}$ and $\boldsymbol{u}'$ corresponds with an epipole.

In this case, we may simplify the analysis by applying a rigid transformation to each image in order to place both points $\boldsymbol{u}$ and $\boldsymbol{u}'$ at the origin, $(0,0,1)^\top$ in homogeneous coordinates. Furthermore, the epipoles may be placed on the x-axis at points $(1,0,f)^\top$ and $(1,0,f')^\top$ respectively. A value f equal to 0 means that the epipole is at infinity. Applying these two rigid transforms has no effect on the sum-of-squares distance function (2), and hence does not change the minimization problem. In this case, since $F(1,0,f)^\top = (1,0,f')F = 0$, the fundamental matrix has a special form

$$F = \begin{pmatrix} ff'd & -f'c & -f'd \\ -fb & a & b \\ -fd & c & d \end{pmatrix} .$$

Consider an epipolar line in the first image passing through the point $(0,t,1)^\top$ and the epipole $(1,0,f)^\top$. We denote this epipolar line by $\boldsymbol{\lambda}(t)$. The vector representing this line is given by the cross product $(0,t,1)^\top \times (1,0,f)^\top = (tf,1,-t)^\top$. The corresponding epipolar line in the other image is $\boldsymbol{\lambda}'(t) = F(0,t,1)^\top = (-f'(ct+d), at+b, ct+d)^\top$. The total squared distance is therefore given by

$$s(t) = d(\boldsymbol{u},\boldsymbol{\lambda}(t))^2 + d(\boldsymbol{u}',\boldsymbol{\lambda}'(t))^2 = \frac{t^2}{1+f^2t^2} + \frac{(ct+d)^2}{(at+b)^2+f'^2(ct+d)^2} . \qquad (5)$$

Our task is to find the minimum of this function. We therefore compute the derivative:

$$s'(t) = \frac{2t}{(1+f^2t^2)^2} - \frac{2(ad-bc)(at+b)(ct+d)}{((at+b)^2+f'^2(ct+d)^2)^2} . \qquad (6)$$

Maxima and minima of $s(t)$ will occur when $s'(t)=0$. Collecting the two terms in $s'(t)$ over a common denominator, and equating the numerator to 0 gives a condition

$$\begin{aligned} g(t) &= t((at+b)^2+f'^2(ct+d)^2)^2 - (ad-bc)(1+f^2t^2)^2(at+b)(ct+d) \\ &= 0 . \end{aligned} \qquad (7)$$

The minima and maxima of $s(t)$ will occur at the roots of this polynomial. This is a polynomial of degree 6, which may have up to 6 real roots, corresponding to 3 minima and 3 maxima of the function $s(t)$. The absolute minimum of the function $s(t)$ may be found by finding the roots of $g(t)$ and evaluating the function $s(t)$ given by (5) at each of the real roots. One should also check the asymptotic value of $s(t)$ as $t \to \infty$ to see if the minimum distance occurs when $t=\infty$, corresponding to an epipolar line $-fu = 1$ in the first image.

4.3 Local Minima

The fact that $g(t)$ in (7) has degree 6 means that $s(t)$ may have as many as three minima. This is indeed possible (see [4] for an example).

Even in the case of perfect matches local minima may occur [4]. Thus, an algorithm that attempts to minimize the cost function (2), or equivalently (4) by an iterative search beginning from an arbitrary initial point is in danger of finding a local minimum, even in the case of perfect point matches.

4.4 Minimizing the sum of the magnitudes of distances

Instead of minimizing the square sum of image errors, it is possible to adapt the polynomial method to minimize the sum of absolute values of the distances. This method will be called **Poly-Abs**.

The quantity to be minimized, expressed as a function of t, is

$$s_2(t) = d(\boldsymbol{u}, \boldsymbol{\lambda}) + d(\boldsymbol{u}', \boldsymbol{\lambda}') = \frac{|t|}{\sqrt{1 + f^2 t^2}} + \frac{|ct + d|}{\sqrt{(at + b)^2 + f'^2(ct + d)^2}}.$$

Development of this equation leads to a polynomial of degree 8 in t. We evaluate $s_2(t)$ at the roots of this polynomial to find the global minimum of $s_2(t)$.

5 Experimental Evaluation of Triangulation Methods

A large number of experiments were carried out to evaluate our triangulation methods. They were evaluated in comparison to various other methods which are not described here. Refer to [4] for a description. The tested methods were **Poly**, **Poly-Abs**, **Mid-point**, as well as two linear methods, **Linear-Eigen** and **Linear-LS**, and two non-linear methods, **Iterative-Eigen** and **Iterative-LS**.

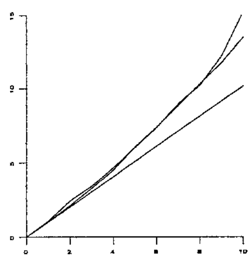
We concentrated on a configuration which simulates a situation similar to a robot moving down a corridor, looking straight ahead. The epipoles are close to the centre of the images. For points lying on the line joining the camera centres depth can not be determined, and for points close to this line, reconstruction becomes difficult. Simulated experiments were carried out for points at several distances in front of the front camera.

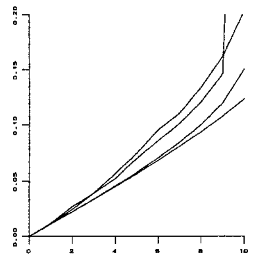
We also considered another typical configuration which simulates a pair of almost parallel cameras, as in an aerial imaging situation, and where the points were assumed to be approximately equidistant from both cameras. However, the most unstable situation is the first one, in which the epipoles are in the centre of the two images, and points lie close to the epipoles. Since this situation gave the most severe test to the algorithms, we will give results only for that configuration. Most of the methods worked relatively well for the second configuration.

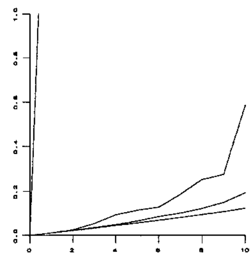
In each set of experiments, 50 points were chosen at random in the common field of view. For each of several noise levels varying from 1 to 10 pixels, each point was reconstructed 100 times, with different instances of noise chosen from a Gaussian random variable. For each reconstructed point both the 3D reconstruction error, and the 2D residual error (after reprojection) were measured. The errors shown are the average errors. The corresponding graphs for median errors (not shown) had the same general form and led to the same conclusions.

To measure the invariance to transformation, an affine or projective transformation was applied to each camera matrix. Then, triangulation was carried out, and finally the reconstructed points were retransformed into the original frame to compare with the correct values. We present here only the results of the most significant case, a projective transformation. Results for affine and Euclidean reconstruction are detailed in [4].

The measured error is denoted either as 2D error (meaning error of measured compared with reprojected points), or 3D error, meaning the error compared with the correct values of the points in space. Every data point in the graphs is the result of 5000 trials, and expresses the RMS or mean value over all the trials. The horizontal axis of each graph is the noise level, and the vertical axis measures the error, in pixels for 2D error, or in space units for 3D error.





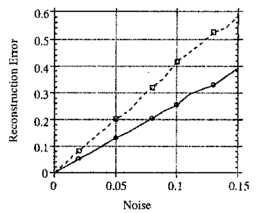


Graph 1 : **2D error for projective reconstruction.** *This graph shows the results for methods (from the bottom) Polynomial, Iterative-Eigen, and Iterative-LS. Poly-Abs (not shown) performed almost identically with Polynomial. This graph shows that Polynomial, or Poly-Abs, is the best method for projective reconstruction, whereas Iterative-Eigen and Iterative-LS (except for occasional non-convergence) perform almost as well.*
Graph 2 : **3D error for projective reconstruction.** *Poly-Abs performs marginally better than Polynomial. Then follow Iterative-Eigen (except that it fails for noise level of 10 pixels) and Iterative-LS.*
Graph 3 : **3D error for projective reconstruction, continued.** *The same as Graph 2 for the less well performing methods. From the bottom, are shown Poly-Abs (for reference), Linear-Eigen, Linear-LS and Mid-point (going off scale for noise of 1).*

6 Evaluation with real images

The algorithms were also carried out with real images, which were used for a set of experiments in [1]. The goal of these experiments was to determine how the triangulation method effects the accuracy of reconstruction.

Using a Euclidean model of the considered object as ground truth, we corrected the measured image point coordinates by an average of 0.02 pixels in order to obtain exactly corresponding model and matched image points. Then, a projective reconstruction of the points was computed [6, 5], and a projective transform H was determined that brought the projective reconstruction into agreement with the Euclidean model. Next, controlled zero-mean Gaussian noise was introduced into the point coordinates, triangulation was carried out in the projective frame, the transformation H was applied, and the error was measured in the Euclidean frame. Graph 4 shows the results of this experiment for two methods. The values are the average error over all points in 10 separate runs. It shows that the Polynomial method gives superior reconstruction results.



Graph 4 : **Reconstruction Error** *for the Mid-point (above) and Polynomial methods. The other methods gave results close to the Polynomial methods. Even for the best method the error is large for higher noise levels, because there is little movement between the images. However, for the actual coordinate error in the original matched points (about 0.02 pixels), the error is small.*

In the images used here, the two epipoles are distant from the image. For cases where the epipoles are close to the images, the results on synthetic images show that the advantage of the Polynomial methods will be more pronounced.

7 Discussion of Results and Conclusion

All the methods performed relatively for Euclidean reconstruction, as measured in terms of 3D error. In the case of 2D error, only Polynomial, Poly-Abs, Iterative-LS and Iterative-Eigen perform acceptably, and the last two have the disadvantage of occasional non-convergence. **Poly-Abs** seems to give slightly better 3D error performance than **Poly** but both of these seem to be excellent methods, not suceptible to serious failure and giving the best overall 3D and 2D error performance. The only distinct disadvantage is that they are not especially easily generalizable to more than two images. They are a bit slower than the other methods, but by a factor of 2 or 3 only, which is probably not significant.

We summarize the conclusions for the various methods.

Poly This is the method of choice when there are only two images and time is not an issue. It is clearly superior to all other methods, except perhaps **Poly-Abs**. In fact, it is optimum under the assumption of a Gaussian noise model. It is affine and projective-invariant.

Poly-Abs This finds the global minimum of sum of magnitude of image error, which may be a better model for image noise, placing less emphasis on larger errors. It seems to give slightly better 3D error results. Otherwise it does not behave much differently from **Poly**. It is affine and projective-invariant.

Mid-point This is not a method that one could recommend in any circumstances. Even for Euclidean reconstruction it is no better than other linear methods. It is neither affine nor projective-invariant.

Linear-Eigen, Linear-LS The main advantage is speed and simplicity. Linear-LS is affine-invariant, but neither of the two methods is projective-invariant.

Iterative-Eigen, Iterative-LS They give very good results, markedly better than the linear ones, but not quite as good as **Poly**. The big disadvantage is occasional non-convergence. They must be used with a back-up method. Both of the methods are affine-invariant. Iterative-LS should not be used for projective reconstruction, since it does not handle points at infinity well.

In summary, the **Polynomial** or **Poly-Abs** method is the method of choice for almost all applications. For Euclidean reconstruction, the linear methods are a possible alternative choice, as long as 2D error is not important. However, for affine or projective reconstruction, they may be orders of magnitude inferior.

Thanks to Paul Beardsley and Andrew Zisserman for making the calibration images and data available to us.

References

1. Beardsley, P.A., Zisserman, A., Murray, D.W.: Navigation using affine structure from motion. In ECCV '94, Vol. II, LNCS-Series Vol. 801, Springer, 85–96 (1994).
2. Beardsley, P.A., Zisserman, A., Murray, D.W.: Sequential updating of projective and affine structure from motion. Report, Oxford University (1994).
3. Faugeras, O.D.: What can be seen in three dimensions with an uncalibrated stereo rig ? In ECCV '92, Vol. II, LNCS-Series Vol. 588, Springer, 563–578 (1992).
4. Hartley, R., Sturm, P.: Triangulation. submitted to CVGIP-IU, 1995, available at ftp:ftp.imag.fr/pub/MOVI
5. Hartley, R.I.: Euclidean reconstruction from uncalibrated views. In Appl. of Invariance in Computer Vision, LNCS-Series Vol. 825, Springer, 237–256 (1994).
6. Hartley, R., Gupta, R., Chang, T.: Stereo from uncalibrated cameras. In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 761–764 (1992).
7. Hartley, R.I.: Estimation of relative camera positions for uncalibrated cameras. In ECCV '92, Vol. II, LNCS-Series Vol. 588, Springer, 579–587 (1992).
8. Koenderink, J.J., van Doorn, A.J.: Affine structure from motion. Journal of the Optical Society of America, A (1992).
9. Longuet-Higgins, H.C.: A computer algorithm for reconstructing a scene from two projections. Nature, 293, 133–135 (1981).
10. Mohr, R., Veillon, F., Quan, L.: Relative 3d reconstruction using multiple uncalibrated images. In Proc. IEEE Conf. on Comp. Vis. and Patt. Rec., 543–548 (1993).
11. Ponce, J., Marimont, D.H., Cass, T.A.: Analytical methods for uncalibrated stereo and motion recosntruction. ECCV '94, Vol. I, LNCS Vol. 800, 463–470 (1994).
12. Shapiro, L.S., Zisserman, A., Brady, M.: Motion from point matches using affine epipolar geometry. ECCV '94, Vol. II, LNCS-Series Vol. 801, Springer, 73–84 (1994).
13. Shashua A.: Projective depth: A geometric invariant for 3d reconstruction from two perspective/orthographic views and for visual recognition. ICCV, 583–590 (1993).