Relative Positioning from Geometric Invariants

Roger Mohr, Luce Morin

Relative Positioning from Geometric Invariants

R. Mohr, L. Morin
LIFIA–IMAG
46 Avenue Félix Viallet
38031 Grenoble FRANCE

Abstract

In this paper, we give geometric constructive solutions to the 3D vision problems like positioning a point in the space from two views. Using reference points in the scene, no calibration is needed. The method involves only simple geometric computation. From our experiments we concluded that positioning 3D points relatively to references points is easy and provides more reliable results than absolute positioning as done usually.

1 Introduction

In a recent paper, Koendering and van Doorn [2] demonstrated that taking points in a sequence of images as reference points allowed to derive very easily qualitative 3D information of the scene. Their approach was however only restricted to affine geometry, a good approximation of the image formation process when the view angle is very small. Here we are going to extend their results using projective geometry. This will then allow us to extract quantitative 3D informations from two images.

One point of an image is associated with a corresponding line of sight in the 3D space. This correspondence between image point and line is computed through the calibration of the camera (see [6] [1] for some recent contributions). Using known reference points in the space, the parameter of the camera are estimated.

In the approach described here we are using reference points from the scene for locating other points in the space. So our approach can be compared with the standard one by the schema depicted in figure 1.

Projective geometry [5] is our central mathematical tool, since projective geometry deals with the properties that are invariant under homography and therefore under central projection. When the ca-
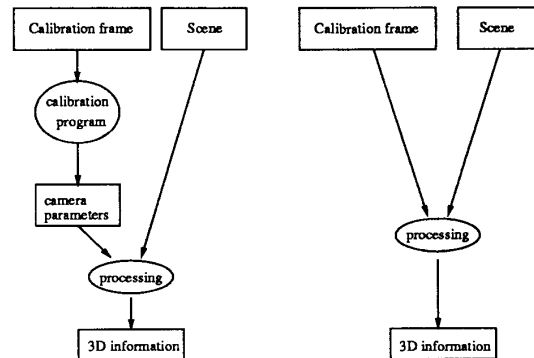

Figure 1: Traditional calibration versus geometrical approach

mera parameters are settled —which is not the case in our framework— the problem addressed here becomes much easier and the retrieved positions are referred to an absolute reference frame, whereas our method leads to relative positioning.

Next section introduces briefly the geometric tool needed to establish our main results: the cross-ratio. Besides, it defines "projective coordinates", a way to reference a point in a plane with projective invariant values.

Section 3 shows how to use these tools to solve the main problems we address here but considering only the case where at least four of the reference points are coplanar. For the more general case the reader is referred to [4].

Some experimental results are presented in section 4. They illustrate the robustness of the method. The last section explores the case of a camera motion which is not perfectly known, and we show that our method provides more robust position estimation than the standard positioning using the camera frame.

2 Projective geometry

2.1 Preliminary definitions

We provide here a short introduction to projective geometry definitions and vocabulary. The reader is referred to [5] for a gentle introduction or to [3] for vision oriented considerations on projective geometry. This subsection is helpful but not essential for the comprehension of the remainder.

We consider $I\!\!R^{n+1} - \{(0,...,0)\}$ with the equivalence relation:

$$(x_1,\ldots,x_{n+1}) \sim (x'_1,\ldots,x'_{n+1}) \Leftrightarrow \exists \lambda \neq 0 \text{ such that}$$

$$(x'_1,\ldots,x'_{n+1}) = \lambda(x_1,\ldots,x_{n+1}).$$

The quotient space obtained by this equivalence relation is the projective space $I\!\!P^n$. Thus the $n+1$-tuples of coordinates $(x_1,...,x_{n+1})$ and $(x'_1,\ldots,x'_{n+1})$ represent the same point in the projective space.

Any linear transformation in homogeneous coordinates is an homography. The matrix associated with a given homography is defined up to a non zero scale factor, that is $\lambda y = Ax$. Notice that a projection from the 3D space $I\!\!P^3$ onto the image plane $I\!\!P^2$ is therefore represented by a 4×3 matrix in homogeneous coordinates which leads us to a 11-dimensional space for these projections: exactly the same dimension as the space of all possible projections for uncalibrated camera.

The usual affine space $I\!\!R^n$ is mapped into $I\!\!P^n$ by the correspondence Ψ:

$$\Psi : (x_1,\ldots,x_n) \rightarrow (x_1,\ldots,x_n,1). \qquad (1)$$

Notice that only the points $(y_1,\ldots,y_n,0)$ are not reached by Ψ. These points are considered as points at infinity. They may be perceived as the limit of

$$(y_1,\ldots,y_n,\lambda) \sim (y_1/\lambda,\ldots,y_n/\lambda,1) \qquad (2)$$

while $\lambda \rightarrow 0$.

2.2 The basic invariant

The cross-ratio is the basic invariant in projective geometry: all other projective invariants can be derived from it.

- Cross-ratio of four points:

  Let A, B, C, D be four collinear points, we can define what we call their *cross-ratio* as:

$$[A,B,C,D] = \frac{\overline{CA}}{\overline{CB}} \times \frac{\overline{DB}}{\overline{DA}} \qquad (3)$$

where $\overline{AB}$ is the algebraic measure of AB.

Fundamental theorem: *Any homography preserves the cross-ratio.*

Thus central projection, linear scalings, skewings, rotations, translations...preserve the cross-ratio.

- Cross-ratio of four lines

  The cross-ratio of a pencil of four lines l_1, l_2, l_3, l_4 going through O is defined as the cross-ratio $[A,B,C,D]$ of the points of intersection of the four lines with any line l not passing through O.

2.3 Projective coordinates

In the affine space of size n, $n+1$ points provides a reference frame if no subset of $k+2$ points lies in a k-dimensional subspace. The first of these points is usually taken as the origin of a vectorial basis. In projective geometry it is well known that a reference frame for $I\!\!P^n$ has $n+2$ points, no $k+2$ of them lying in k-dimensional subspace. For the particular case of the projective plane, four points provide a reference frame if no three of them are coplanar.

We define how points can be designated through "coordinates" with respect to given reference points using cross-ratio. Obviously from the previous section these "coordinates" are invariant under any homography, and therefore under central projection and scaling along the image plane axes.

- Projective coordinates on a line

  Let A, B, C be three given points of a line l, and P be a point of l. P and A, B, C define a cross-ratio $\lambda = [A,B,C,P]$. Inversely, given a scalar λ there exists a unique point P such that the cross-ratio equals to λ. (x_1,x_2) such that $x_1/x_2 = \lambda$ are what we call projective coordinates of the point P in the coordinates system defined by (A,B,C).

- Projective coordinates on a plane

  In a projective plane $\mathcal{P}$, any four points A,B,C,D, no three collinear, define a projective coordinates system (see figure 2). Given a point P of $\mathcal{P}$, let (x_1,x_2,x_3) be a triple of real numbers defined up to a scaling factor and such that :

$$\frac{x_1}{x_2} = [CA,CB,CD,CP] \qquad (4)$$

$$\frac{x_2}{x_3} = [AB,AC,AD,AP] \qquad (5)$$

(x_1,x_2,x_3) are called the projective coordinates of P in the coordinates system (A,B,C,D). Naturally we also have $x_3/x_1 = [BC,BA,BD,BP]$.
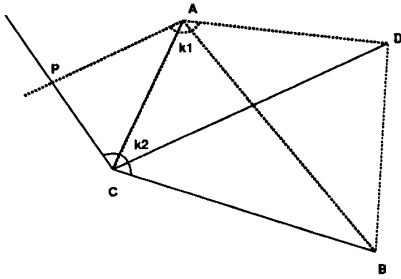
Figure 2: Projective coordinates in the plane

In a projective plane with four known points we can uniquely reference any point of the plane by their projective coordinates so defined. In fact only the two cross-ratios $k_1 = x_1/x_2$ and $k_2 = x_2/x_3$ are necessary to uniquely define a point.

3 Projective reconstruction

This section solves the problems mentioned in the introduction in the simplest case when at least four of the reference points are coplanar. First the back projection is solved and the technique introduced is used for solving the other ones.

Given two sets of 4 coplanar points $\{A, B, C, D\}$ and $\{E, F, G, H\}$ and their projections onto the image plane $\{a, b, c, d\}$ and $\{e, f, g, h\}$, we are going to show how to solve the problems mentioned in the first section.

Note that this configuration can be obtained with $A = E$ and $B = F$; this is the case each time we choose six vertices on a block as reference points, the six points lying on only two adjacent faces (see figure 3).
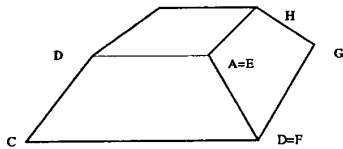

Figure 3: 6 points on adjacent faces provide 2 sets of 4 coplanar points

3.1 How to back project an image point?

Given an observed point m on the image plane, it is possible to determine the location of the viewing line

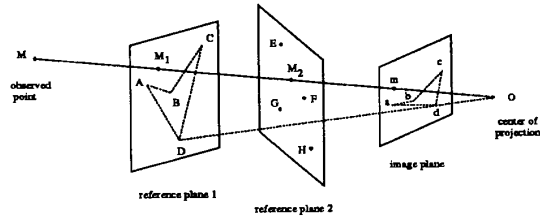*Om* through m with respect to the scene, without any calibration.

Figure 4: The back projection of the image point m

Proof: As shown in subsection 2.3, we can determine the projective coordinates (x_1, x_2, x_3) of m with respect to the coordinate system (a, b, c, d), from measures in the image plane. Let M_1 be the intersection in 3D space of the viewing line Om with the plane $ABCD$ (see figure 4). The projection of M_1 onto the image plane is m. Since the projective coordinates are invariant under projection (see section 2.3), the projective coordinates of M_1 with respect to (A, B, C, D) are (x_1, x_2, x_3). The position of M_1 in the reference plane $ABCD$ can then be determined from its projective coordinates as shown in section 2.3.

We can similarly determine M_2, the intersection of the viewing line Om with the plane $EFGH$. (M_1, M_2) determine the viewing line with respect to the scene, and this *without knowing the camera position in the scene*.

3.2 Where is the camera?

We just showed in 3.1 how to reconstruct the viewing line from a single shot. It is also possible to reconstruct the position of the optical center with respect to the scene, in a single shot, using the same method. If two points m and p are given in the image plane, it is possible to reconstruct the viewing line Om and Op. These two lines happen to intersect in space at O, which provides us with a method to compute the location of O.

Furthermore, if more than two points are available, it will be possible to derive a solution using the least squares estimate, which will be much more reliable against noise in the measures.

3.3 Where is the image point?

We suppose here that the position of camera center O is computed using the previous method. The image point m of a space point M is then easily constructed: First compute the intersection point of OM with a

known plane containing four reference points; then in this known plane take measures of the projective coordinates of this intersection as defined in section 2.3. Finally with these two known cross-ratios construct the point m in the image plane using the image projection of the four reference points.

3.4 How to locate a point in the scene?

The previous method can be applied to locate a point in a scene containing reference points. Here two snapshots of the same scene are needed. A given point M in the 3D space is projected respectively as m_1 and m_2 in each of these two images.

Let us suppose that the matching of m_1 and m_2 has been solved. For each image we are able to reconstruct the viewing line passing through M and therefore these two lines intersect on M. This method degenerates only when the two positions of the optical centers are aligned with the viewing line.

3.5 Where are the epipolar lines?

3.5.1 Finding one epipolar line

In stereo vision, the epipolar centers o and o' are the intersection of the line OO' with the image planes (see figure 5). These points are quite important as all epipolar lines are radiated from these centers. The epipolar constraint states that given a point M in 3D space with projections m in the first image plane and m' in the second one then the epipolar plane $OO'M$ intersects the image planes along the lines om and $o'm'$. Therefore given a point m in the first image plane, its image m' in the second image lies on the corresponding epipolar line. The line om (resp. $o'm'$) is what we call the epipolar line of the point m' (resp. m).
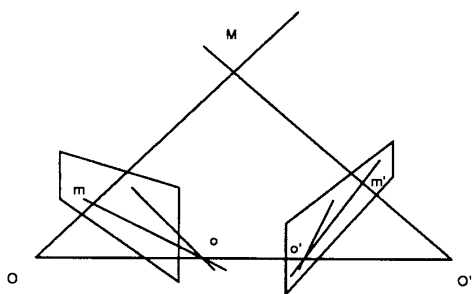

Figure 5: The epipolar geometry.

The same method defined in 3.1 enables to determine the epipolar lines.

The two sets of 4 coplanar points $\{A, B, C, D\}$ and $\{E, F, G, H\}$ respectively project as $\{a, b, c, d\}$ and $\{e, f, g, h\}$ onto the image plane. Given an observed point m on the image plane, it is possible to determine the location of the viewing line Om through m with respect to the scene.

Another snapshot from a new optical center O' provides a new image. We suppose here that this image contains also reference points (not necessary the same as for image 1, but within the same reference frame). It is then possible to determine the projection of the viewing line Om on the new image plane, which is the epipolar line of point m. Indeed, we know the projective coordinates of points $M1$ and $M2$ (as defined in 3.1), and we are therefore able to reconstruct points $m1$ and $m2$ on the new image plane. These two points determine the epipolar lines of point M on the new image plane.

3.5.2 The pencil of epipolar lines

Projective geometry allows us to prove easily the following result which constrains the position of the epipolar lines.

Theorem : two pencils of four corresponding epipolar lines have same cross-ratio.

Proof : from their definition, corresponding epipolar lines are the intersections of a 3D plane containing the base line OO' with the two image planes. Therefore the cross-ratio of the pencil of lines is the cross-ratio of the pencil of planes.

from there, if three corresponding epipolar lines are known, the epipolar geometry is completely defined: for any point m in image1, m defines the epipolar line going through m in image1. The cross-ratio of these known four lines can then be computed and we define the corresponding line in the second image as the one with the same cross-ratio.

4 Experimentation

In order to test the robustness of our method, we present here some few results. The scene is a simple polyhedral scene showing a cube in the foreground and with a background made up with two paper sheets containing black painted rectangles. Figure 6 shows two images of this scene after extracting edges, using polygonal approximation, and least square fitting.

Measures of reference points in the scene were performed with a regular ruler, which provides a 1mm accuracy. The two vertical planes were supposed to be orthogonal, in fact their angle is 92 degree.
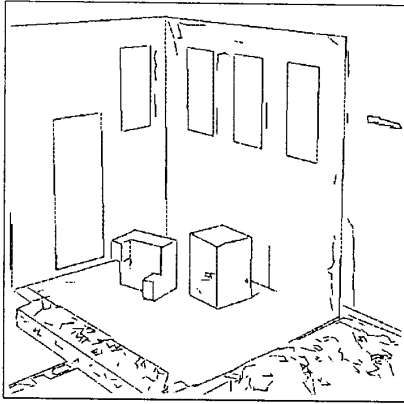
Figure 6: Contours image of the scene

A stereo reconstruction is performed using the two images displayed in figure 6. The second image is taken after rotating the scene about 40 degrees around a vertical axis. Matching between corresponding points was done by hand as matching was no the primary concern of this paper.

Table 1 describes the results of the retrieval of the cube vertices, using the method described in section 3.4. The exact size of the cube is 50mm. Hence the results are equal within 4% to the exact values. No comparison can be made with exact location of the cube since only its approximate position was known.

points	x	y	z	edges	length
0	78.9	140	48.5	0–1	50.5
1	79.1	141	-2	0–2	49.1
2	81.3	189	47.5	0–6	48.9
3	82.0	188	-1.5	1–3	47.1
4	33.2	195	48.5	2–3	49.5
5	34.4	194	-1.5	2–4	48.9
6	30.3	145	49.0	3–5	48.0
				4–5	49.8
				4–6	50.1

Table 1: Results for 3D reconstruction of the cube

So, if absolute errors are up to three millimeters, it has to be noticed that differences of such values (relative errors) are also bounded by three millimeters. The average value is lower, but this simple experiment had no intend to deliver reliable statistic on these results.

5 Case of a moving camera

In order to compare the error obtained with the geometrical approach and the traditional one, we have analyzed a standard simple case (see fig. 7). A point P lying in a plane is viewed from two linear cameras C_1 and C_2. The camera base–line is 8000 pixels long and the point is located 40000 pixels from the base–line. We assume the two cameras to be perfectly calibrated and their positions and orientations to be known with some uncertainty. The error on an image point position is set to 1 pixel. The camera positions are assumed to have a 100 pixels uncertainty. We assume that the field of view contains a reference line L passing through three known points.
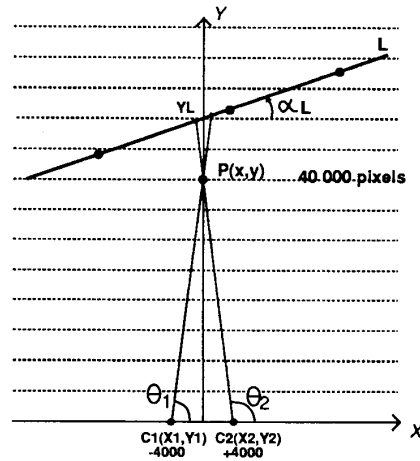

Figure 7: A particular case : planar stereo reconstruction.

Using the traditional approach, we can reconstruct the position of P from the two images using standard triangulation.

We can also compute the position of P using the cross–ratio method in a straightforward way.

From the Jacobian matrix associated to the reconstruction equations, we have analyzed for both methods the error on P as a function of the errors on the data.

Figure 8 shows for the traditional method the error on x and y coordinates of P as a function of the error on camera orientation ($\delta\theta = \delta\theta_1 = \delta\theta_2$).

Figure 9 refers to the geometrical method. It shows the error on x as a function of the orientation of the refence line, α_L. The different curves correspond to different values of Y_L, distance of the reference line from the cameras.

As shown in these two curves and as a conclusion of our other plots, the geometrical approach easily provides errors below 3.000 pixels, except in the proximity of degenerate cases (reference line oriented along the viewing line); in the traditional case (see fig. 8) this means an accuracy of 0.004 radians ($\sim$ 0.2 deg) measuring the angles θ_1 and θ_2. Moreover, as expected for relative positioning, the minimum errors occur when the point P is located near by the reference line.
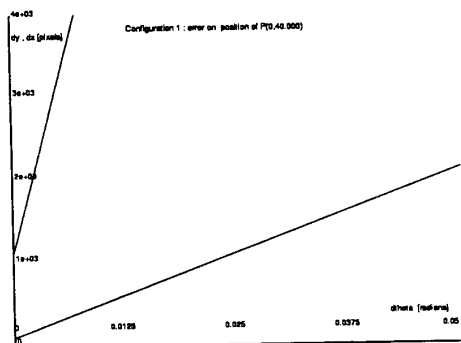

Figure 8: Error on x and y position of P as a function of the error on θ. The lower curve corresponds to the error on x.
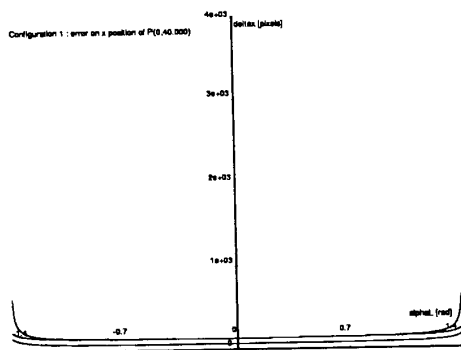

Figure 9: Error on x and y position of P as a function of α_L. Each curve corresponds to a different Y_L: from the bottom $Y_L = 40000 pixels$, $Y_L = 100000 pixels$, $Y_L = 25000 pixels$.

6 Conclusion

We demonstrated here how taking into account position points in the space allows relative positioning. Such an approach solves in an elegant way problems like back projection of a image point, 3D positioning from two frames, or how to determine the epipolar geometry. Furthermore such an approach has several benefits. Of course no calibration is needed (we assume however that the camera is a perfect pinhole camera) and only simple computations are required.

A study of error and experimentation shows that the precision obtained is better than the one obtained through the traditional approach. The geometrical approach provides explicit object based information, with uncertainty proportional to the distance so it is a really natural way to describe the environment in a wide set of robotic tasks.

Acknowledgements

References

[1] O.D. Faugeras and G. Toscani. Camera Calibration for 3D Computer Vision. In *Proceedings of International Workshop on Machine Vision and Machine Intelligence*, Tokyo, 1987.

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

[3] S.J. Maybank. The projective geometry of ambigious surfaces. Technical report, Long Range Laboratory, GEC, Wembley, Middlessex, UK, Sept. 1989.

[4] 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. 1991.

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

[6] R.Y. Tsai. A Versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf tv cameras and lenses. In *IEEE Journal of Robotics and automation*, volume 3, pages 323–344. IEEE, 1987.