An Analytic Solution for the Perspective 4-Point Problem

Radu Horaud, Bernard Conio, Olivier Leboulleux, Bernard Lacolle

To cite this version:


HAL Id: inria-00589990
https://inria.hal.science/inria-00589990

Submitted on 16 Jan 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
An Analytic Solution for the Perspective 4-Point Problem

Radu Horaud, Bernard Conio, and Olivier Leboulleux, LIFIA
Bernard Lacolle, TIM3

LIFIA–IMAG, 46 Avenue Félix Viallet, 38031 Grenoble
TIM3–IMAG, BP 68, 38402 Saint-Martin d’Hères
FRANCE

Abstract

The perspective n-point (PnP) problem is the problem of finding the position and orientation of a camera with respect to a scene object from n correspondence points. In this paper we propose an analytic solution for the perspective 4-point problem. The solution is found by replacing the four points with a pencil of three lines and by exploring the geometric constraints available with the perspective camera model. We show how the P4P problem is cast into the problem of solving a biquadratic polynomial equation in one unknown. Although developed as part of an object recognition from a single view system [6], the solution might well be used for hand-eye coordination, landmark-guided navigation, and for fast determination of exterior camera parameters in general.

Key words: exterior camera calibration, 3D object recognition, location determination, visual navigation.

1 Introduction

One of the fundamental goals of Computer Vision is to discover properties that are intrinsic to a scene by analysing one or several images of this scene. Within this paradigm, an essential process is the determination of the position and orientation of the sensing device (the camera) with respect to objects in the scene. This problem is known as the exterior camera calibration problem and it has many interesting applications in Robotics and Cartography. Important applications in Robotics are: sensor calibration [13], object recognition and localisation from a single view [6], [9], [2], stereo sensor calibration [3], hand/eye coordination [14], and sensory based navigation [1].

In Cartography the problem is to determine the location in space from which an image or a photograph was obtained by recognizing a set of landmarks appearing in the image [4].

More formally, the problem may be stated as follows: Given a set of points with their coordinates in an object-centered frame and their corresponding projections onto an image plane and given the intrinsic camera parameters, find the transformation matrix (three rotations and three translations) between the object frame and the camera frame (see Figure 1).

This problem is referred to as the perspective n-point problem and is usually solved using least-squares techniques. An elegant least-squares solution has recently been proposed [9]. Least-squares techniques require the computation of numerical solutions. For these solutions to be stable a large set of data points are needed which inherently augments the complexity of the computation. This is not desirable especially when the computation resides in the inner loop of a recognition/localisation process [6], [9], [5], [12], [7], [8], [2].

For these reasons a certain number of researchers have tried to determine the minimum number of points necessary to find a solution, and associated with each set of points they have tried to find a closed form expression of the problem. A finite number of solutions is available only when the number of points is equal or greater than 3. The following is a brief review of the suggested solutions.

- Three points. Rives et al. [10] derive a set of three quadratic equations with three unknowns. These unknowns are the distances from the optical center of the camera to the three points. In theory there are 8 solutions. Fischler and Bolles [4] notice that for every real positive solution there is a real negative solution and hence a maximum of 4 solutions are in fact possible. They derive a closed form expression, namely a biquadratic polynomial.
polynomial can respond points. coplanar Finding in Figure lines provide fewer nar, narration PI

Four points. When the points are coplanar a set of solutions can be found by considering any three among the four points and verifying the solution with the fourth point: A unique solution is thus found in [4]. When the points are not coplanar a closed form expression does not appear to have been derived. Rives et al. [10] solve a set of six quadratic equations with four unknowns. Fischler and Bolles [4] attack the problem by finding solutions associated with subsets of three points and selecting the solutions that they have in common. They provide a geometric construction which shows that unlike the coplanar case, two solutions may be available.

Five points. For five points in general position the strategy mentioned above can as well be applied: Compare the solutions obtained with subsets of three or four points.

Six points or more will always produce a unique solution: For six points we obtain 12 linear equations which are enough to determine the 12 coefficients of the transformation matrix, 9 for the rotations and 3 for the translations.

Three lines. An alternative solution is to use lines instead of points. Horne [6] suggests a constructive method for the case of a pencil of three non coplanar lines. Dhome et al. [2] solve for the general case of three arbitrary lines. The solutions are given by the roots of a polynomial of order 8 in one unknown.

In this paper we derive an analytic solution for the case of four non coplanar points, namely a biquadratic polynomial in one unknown. Roots of such an equation can be found in closed form or by an iterative method. Finding a solution for four non coplanar points is equivalent to finding a solution for a pencil of three non coplanar lines: The three lines share one of the four points. Notice that these lines may or may not correspond to physical linear edges in the scene. Figure 2 (a to d) shows various line and point configurations which are amenable to solving the PnP problem. The shaded lines are not physical edges. The configurations shown in Figure 2 (e and f) lead to more complex solutions [2], [11].

Finding a closed form solution for four non coplanar points is important for several reasons. First, they provide fewer solutions than three points. Second, the solutions are more stable when the points are not coplanar, because they do not depend of the relative orientation of the image plane with respect to the scene plane containing the points. Third, the computation of such a solution is very fast and therefore it can be included in a runtime visual process.

2 The solution

In order to compute the transformation matrix \( A \) of Figure 1 we decompose it into two matrices, \( A_1 \) and \( A_2 \), and we define three frames: a camera centered frame, an image centered frame, and an object centered frame, e.g., Figure 3. \( A_1 \) is the transformation matrix from the image frame to the camera frame and \( A_2 \) is the transformation matrix from the object frame to the image frame. Therefore we have:

\[
A = A_1 A_2
\]

The four non coplanar points, \( M, M_1, M_2, M_3 \) are replaced by a pencil of three non coplanar line segments, as shown on Figure 5. We denote by \( L_1, L_2, \) and \( L_3 \) the unit vectors associated with the directions of these lines and let \( l_1, l_2, \) and \( l_3 \) be the unit vectors associated with their projections onto the image plane. The three frames used throughout the paper are defined as follows:

- The object coordinate system is defined as follows (see Figure 4). \( L_3 \) is the x-axis. Let \( P_3 \) be a unit vector in the plane perpendicular to \( L_3 \). The geometric meaning of \( P_3 \) will be soon made clear. Let
Figure 2: Various point and line configurations on a 3-D object. The solution proposed in this paper can be applied to configurations a, b, c, and d.

Let \( P_3 \) be the y-axis of the object frame; the z-axis is defined by the cross product \( L_3 \times P_3 \). It is easy to determine the object-frame coordinates of \( L_1 \) and \( L_2 \), e.g., Figure 4:

\[
L_1 = \sin \alpha_1 \ L_3 + \cos \alpha_1 \cos(\beta + \theta) \ P_3 + \\
\cos \alpha_1 \sin(\beta + \theta) \ L_3 \times P_3
\]

\[
L_2 = \sin \alpha_2 \ L_3 + \cos \alpha_2 \cos(\beta - \theta) \ P_3 + \\
\cos \alpha_2 \sin(\beta - \theta) \ L_3 \times P_3
\]

In these formulas \( \alpha_1 \) is the value of the angle between \( L_1 \) and \( L'_1 \) (the projection of \( L_1 \) onto the \( yz \)-plane). The value of the angle between \( L'_1 \) and \( L'_2 \) is \( 2\beta \). Notice that in these formulas \( \theta \) is an unknown which will be determined.

- The **camera coordinate system** has its origin at the focal point \( F \). The z-axis is the optical axis of the camera and the xy-plane is parallel to the image plane. The image is at distance \( f \) (the focal length) from the origin along the z-axis.

- We define now an **image coordinate system** which is rigidly attached to the projections of the object features, i.e., the image features. The projection of \( L_1 \) onto the image plane is \( l_1 \). The focal point \( F \) and \( l_1 \) define a plane called the interpretation plane. Notice that this plane is rigidly attached to the camera frame since the coordinates of \( l_1 \) are measured in this frame. All the spatial interpretations of \( l_1 \) and hence \( L_1 \), belong to this plane. Let \( P_1 \) be the normal unit vector associated with this plane. Hence \( P_3 \) mentioned above is the unit vector normal to the interpretation plane associated with \( l_3 \). Let \( J \) be the point of intersection of the image lines \( l_1, l_2, \) and \( l_3 \).

Figure 3: This figure shows the decomposition of the transformation matrix \( A \) into two matrices: \( A_1 \) which maps the image frame into the camera frame and \( A_2 \) which maps the object frame into the image frame.

Figure 4: An "object frame" is associated with the four points.
We are now ready to define the image coordinate system. The x-axis lies along the line from F to J. Let \( k' \) be the unit vector associated with this line. \( k' \) belongs to the three interpretation planes and hence \( k' \) is perpendicular to \( P_1, P_2, \) and \( P_3 \). Hence these vectors are coplanar. The image frame is defined by \( k', P_3, \) and their cross product, \( k' \times P_3 \).

We have:

\[
\begin{align*}
  k' &= \frac{\overrightarrow{FJ}}{||\overrightarrow{FJ}||} \\
  P_3 &= \frac{L_3 \times k'}{||L_3 \times k'||} \\
  k' \times P_3 &= \frac{k' \times (L_3 \times k')} {||L_3 \times k'||} = \frac{L_3 - (k' \cdot L_3)k'} {||L_3 \times k'||}
\end{align*}
\]

The translational parameters are given by the coordinates of J in the camera frame. Matrix \( A_1 \) is the following:

\[
A_1 = \begin{pmatrix}
  k'_x & (P_3)_x & (k' \times P_3)_x & (\overrightarrow{FJ})_x \\
  k'_y & (P_3)_y & (k' \times P_3)_y & (\overrightarrow{FJ})_y \\
  k'_z & (P_3)_z & (k' \times P_3)_z & (\overrightarrow{FJ})_z \\
  0 & 0 & 0 & 1
\end{pmatrix}
\]

2.2 The matrix \( A_2 \)

We recall that \( A_2 \) is the transformation matrix from the object frame (defined by \( L_3, P_3, \) and \( L_3 \times P_3 \)) to the image frame (defined by \( k', P_3, \) and \( k' \times P_3 \)). From Figure 6 which shows the interpretation plane associated with F and \( I_3 \) it is easy to derive an expression for \( L_3 \):

\[
L_3 = \cos \phi \ k' + \sin \phi \ k' \times P_3
\]

with the constraint \( 0 < \phi < \pi \)  

2.1 The matrix \( A_1 \)

In order to determine the coefficients of this matrix one has to express \( k', P_3 \) and \( k' \times P_3 \) in the camera frame.

Figure 5: The geometry of the 4-point perspective problem.

A transformation between any two frames is defined by three rotations and three translations. Such a transformation may be represented by a 4 by 4 matrix (standard homogeneous coordinates). There are nine coefficients that specify the three rotations and three coefficients that specify the three translations. Next we determine the coefficients of \( A_1 \) and \( A_2 \).

Figure 6: The interpretation plane associated with F and \( I_3 \)

The rotational coefficients of \( A_2 \) are the coordinates of \( L_3 \) in the image centered frame, i.e., eq. (6), the coordinates of \( P_3 \), i.e., 0, 1, and 0, and the coordinates of \( L_3 \times P_3 \). The translational coefficients are the coordinates of the vector \( JM \), e.g., Figure 5. Since the direction of this last vector is the direction of \( k' \), the matrix \( A_2 \) is:
Two equations (2)

Similarly, we have:

We express $L_1$, $L_2$, $P_1$, and $P_2$ in the image coordinate frame. By applying the transformation given by (7) to $L_1$ and $L_2$ (which are given by equations (2) and (3)) we obtain (the $k'$-components are not relevant for this computation):

We have already noticed that $P_1$ and $P_2$ are perpendicular to $k'$. We have:

Where $\gamma_1$ and $\gamma_2$ are given by (the unit vector normal to any interpretation plane, $P_1$ can be determined using equation (4)):

We inject the expressions of $L_1$, $L_2$, $P_1$, and $P_2$ in equations (8) and (9) and obtain:

We derive closed form expressions for these three unknowns.

### 2.3 Analytic expressions for $\theta$ and $\phi$

Two more geometric constraints are available: $L_1$ belongs to the interpretation plane associated with $l_1$. Hence $L_1$ is orthogonal to $P_1$:

Similarly we have:

We express $L_1$, $L_2$, $P_1$, and $P_2$ in the image coordinate frame. By applying the transformation given by (7) to $L_1$ and $L_2$ (which are given by equations (2) and (3)) we obtain (the $k'$-components are not relevant for this computation):

We have already noticed that $P_1$ and $P_2$ are perpendicular to $k'$. We have:

Where $\gamma_1$ and $\gamma_2$ are given by (the unit vector normal to any interpretation plane, $P_1$ can be determined using equation (4)):

We inject the expressions of $L_1$, $L_2$, $P_1$, and $P_2$ in equations (8) and (9) and obtain:

We determine $\theta$ and $\cos \theta$ as a function of $\phi$:

Finally using the constraint $\cos^2 \theta + \sin^2 \theta = 1$ we obtain:

With:

The roots of equation (20) can be found in closed form or by an iterative method. For its real roots verifying the obvious constraint $| \cos \phi | < 1$, we can compute $\theta$ through the formulas given by equations (18) and (19). With the values of $\phi$ and $\theta$ thus obtained we can compute $L_1$, $L_2$, and $L_3$ using equations (10), (11), and (6). The following constraint guarantees that $L_i$ lies in between $FM$ and $FM_i$ (Figure 6):

All these constraints allow us to eliminate roots which don't correspond to an admissible geometric configuration.
2.4 An expression for $d_x$

From Figure 6 it is easy to determine the length of $\overrightarrow{FM}$. We have:

$$||\overrightarrow{M_3}|| = ||\overrightarrow{FM}|| \cos \delta_2 + ||\overrightarrow{FM_3}|| \cos \delta_3$$

$$||\overrightarrow{FM}|| \sin \delta_2 = ||\overrightarrow{FM_3}|| \sin \delta_3$$

From these equations we obtain:

$$||\overrightarrow{FM}|| = ||\overrightarrow{M_3}|| - \sin \delta_2 \sin(\delta_2 + \delta_3)$$

We also have:

$$\sin \delta_3 = \frac{||\overrightarrow{F_3} \times \overrightarrow{M_3}||}{||\overrightarrow{F_3}|| ||\overrightarrow{M_3}||}$$

$$\sin(\delta_2 + \delta_3) = \sin \delta_1 = \frac{||\overrightarrow{F_3} \times \overrightarrow{F_3}||}{||\overrightarrow{F_3}|| ||\overrightarrow{F_3}||}$$

Finally we obtain:

$$||\overrightarrow{FM}|| = \frac{||\overrightarrow{F_3}|| ||\overrightarrow{F_3} \times \overrightarrow{M_3}||}{||\overrightarrow{F_3} \times \overrightarrow{F_3}||}$$

The translation vector $d_x$ is:

$$d_x = ||\overrightarrow{FM}|| = ||\overrightarrow{FM}|| - ||\overrightarrow{F_3}||$$

3 Special configurations

In the previous section we derived an analytical solution for the perspective 4-point problem in the general case. It is interesting to study some particular configurations associated with these four points.

3.1 Four coplanar points

In this section we show that the general formulation applies to four coplanar points. This situation corresponds to $2\beta = (\beta + \theta) + (\beta - \theta) = \pi$, e.g., Figure 4.

In this case we obtain:

$$K_1 = K_4 = K_5 = K_6 = 0$$

$$I_2 = I_4 = 0$$

$$I_1 = K_2^2$$

$$I_3 = K_2^2 + K_2^2$$

$$I_5 = -K_2^2$$

The solution is given by the following equation:

$$I_1 \cos^2 \phi + I_3 \cos^2 \phi + I_5 = 0$$

It is worth noticing that the discriminant of this equation is always positive:

$$\Delta = I_3^2 - 4I_1I_5 = (K_2^2 - K_2^2 + K_2^2)^2 + 4K_2^2K_2^2$$

3.2 Three colinear image points

Another particular situation is due to an accidental alignment: The image projections of three among the four points are colinear. Let's suppose for instance that the image points $J_1, J_2,$ and $J_3$ are colinear. In this case the interpretation planes $P_1$ and $P_3$ are colinear. Hence we have: $\cos \gamma_1 = 1$ and $\sin \gamma_1 = 0$. We obtain:

$$K_1 = K_3 = K_5 = 0$$

$$I_1 = I_2 = 0$$

$$I_3 = K_2^2 + K_2^2 - K_2^2$$

$$I_4 = -2K_2K_7$$

$$I_5 = K_2^2 - K_2^2 - K_2^2$$

The solution is given by:

$$I_3 \cos^2 \phi + I_4 \cos \phi + I_5 = 0$$

Real roots exist for this equation if and only if its discriminant is positive:

$$\Delta = \sin^2 \gamma_2 (\tan^2 \alpha_2 - \cos^2 2\beta) - \cos^2 \gamma_2 \geq 0$$

The accidental alignment described here corresponds to a 3-line spatial vertex being projected onto the image as a T-junction. The result of this section is that such a T-junction may have a three-dimensional interpretation.

3.3 A right vertex

Another particular case occurs when the four points form a right vertex, i.e., $L_1, L_2,$ and $L_3$ are mutually orthogonal:

$$L_i \cdot L_j = 0 \text{ for } i \neq j$$

In this case a simpler solution than the general case can be derived. Notice first that equation (6) can also be written for $L_1$ and $L_2$:

$$L_i = \cos \phi_i k' + \sin \phi_i k' \times P_3 \text{ for } i = 1, 2, 3$$

For notation homogeneity $\phi$ is replaced by $\phi_3$. We obtain:

$$L_i \cdot L_j = \cos \phi_i \cos \phi_j + \sin \phi_i \sin \phi_j (P_i \cdot P_j) \text{ with } i \neq j$$

We have already mentioned that $P_1, P_2$, and $P_3$ are coplanar (they are all orthogonal to $k'$). Hence, the three dot products $P_1 \cdot P_2$ cannot be simultaneously null. There are three possible situations:

1. $P_1 \cdot P_2 = 0$, $P_1 \cdot P_3 \neq 0$, and $P_2 \cdot P_3 \neq 0$. We obtain $\phi_1 = \phi_2 = 0$ which is impossible because it corresponds to two space points which project onto the image at the same location, e.g., Figure 6;
2. $P_1 \cdot P_2 = 0$, $P_1 \cdot P_3 = 0$, and $P_2 \cdot P_3 \neq 0$. $P_1$ is perpendicular to both $P_2$ and $P_3$, hence they are colinear, $|P_2 \cdot P_3| = 1$. We obtain:

$$\cos \phi_1 \cos \phi_2 = 0 \quad (34)$$

$$\cos \phi_1 \cos \phi_3 = 0 \quad (35)$$

$$\sin \phi_2 \sin \phi_3 \pm \cos \phi_2 \cos \phi_3 = 0 \quad (36)$$

Which gives the solutions: $\phi_1 = \pi/2 \phi_2 + \phi_3 = \pi/2$, and

3. $P_1 \cdot P_2 \neq 0$, $P_1 \cdot P_3 \neq 0$, and $P_2 \cdot P_3 \neq 0$. We obtain:

$$\tan \phi_2 = \tan \phi_3 \frac{P_1 \cdot P_3}{P_1 \cdot P_2} \quad (37)$$

$$\tan \phi_1 = \tan \phi_3 \frac{P_2 \cdot P_3}{P_1 \cdot P_2} \quad (38)$$

$$\tan^2 \phi_3 = -\frac{P_1 \cdot P_2}{(P_1 \cdot P_3)(P_2 \cdot P_3)} \quad (39)$$

In this case solutions exist under the constraint: $(P_1 \cdot P_2)(P_1 \cdot P_3)(P_2 \cdot P_3) < 0$.

This last constraint can be used to select image junctions that have a right vertex interpretation. Figure 7 shows the image of four objects. Figure 8 shows the lines extracted from the above image (top) and the 3-line junctions having a right vertex interpretation.

![Image of a few objects](image_url)

Figure 7: An image of a few objects

4 Discussion

In this paper we derived an analytic solution for computing the exterior camera parameters from four correspondence points in general positions. This solution is of the same complexity as for three points and is particularly simple for such configurations as four coplanar points or four points forming a right vertex. Such an analytic formulation allows fast numerical computation which is desirable in many applications such as on line calibration (hand/eye coordination, navigation) and/or object recognition and positioning from a single view.

References


Figure 8: The lines extracted from the image and the set of 3-line junctions that have a right vertex interpretation.