Effective PnP algorithm

December 5, 2025

1 Problem statement

There is a spatial object/scene containing several spatial points. These points are given in their own world coordinate system. We take images of this object/scene, and the pixel locations are detected in the images. Therefore, 3D->2D correspondences are given between 3D space and image coordinates. They are denoted by $P_i = [X_i \quad Y_i \quad Z_i]^T$ and $p_i = [u_i \quad v_i]^T$ in the paper, respectively, $i \in \{1, ..., n\}$.

The task is to determine the rigid transformation, represented by orthonormal (rotation) matrix R and translation vector t, between the world and the camera coordinate systems. If n $3D \rightarrow 2D$ correspondences are given, the problem itself is called Perspective n Point (PnP) problem.

Theoretically, for the minimal problem, $n = 3$, but here we only considered the $n \geq 6$ cases due to their simplicity.

The goal is to find the rigid transformation, represented by a rotation (three degrees of freedom) and a translation (another three DoFs). From each projected 2D image location, we have a horizontal and vertical coordinate.

2 Algorithm overview

Here, we overview one of the many published solutions for the PnP problem. It is called the Effectice PnP (EPnp) algorithm, published as Vincent Lepetit; Francesc Moreno-Noguer; Pascal Fua. EPnP: An Accurate O(n) Solution to the PnP Problem International Journal Of Computer Vision. 2009. It is implemented in OpenCV.

We apply barycentric coordinates that can be used to represent the points that are projected to the images. We utilize the fact that the barycentric coordinates are not affected by the rigid transformation, therefore the same values can be used for barycentric coordinates both in world and camera coordinate systems.

2.1 Barycentric coordinates

The 3D and 2D point coordinates are given as input. In the world, four reference points are arbitrary selected, the the barycentric coordinates are calculated for each 3D point. There are four barycentric reference points are given, they are denoted by $c_i = [X_i \quad Y_i \quad Z_i]^T$, $i \in \{1, \ldots, 4\}$. Then a spatial point P_i can be obtained as

$$P_i = \sum_{j=1}^{4} \alpha_{ij} c_j,$$

subject to $\sum_{j=1}^{4} \alpha_{ij} = 1$.

2.2 Projection by a pin-hole camera

If it is assumed that the camera coordinate system are used, the projection can be calculated as follows:

$$\lambda_i x_i = K P_i = K \sum_{j=1}^{4} \alpha_{ij} c_j,$$

where

$$K = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}.$$

is the so-called calibration matrix with intrinsic camera parameters.
Then

$$\lambda_i p_i = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \sum_{j=1}^{4} \alpha_{ij} c_j,$$

It is equivalent to

$$\lambda_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \sum_{j=1}^{4} \alpha_{ij} \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_j \\ Y_j \\ Z_j \end{bmatrix},$$

if the coordinates are substituted.
Therefore,

$$\lambda_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \sum_{j=1}^{4} \alpha_{ij} \begin{bmatrix} fX_j + u_0 Z_j \\ fY_j + v_0 Z_j \\ Z_j \end{bmatrix}$$

Then three equations can be written for the coordinates:

$$\lambda_i u_i = f\sum_{j=1}^{4}\alpha_{ij}X_j + u_0\sum_{j=1}^{4}\alpha_{ij}Z_j,$$

$$\lambda_i v_i = f\sum_{j=1}^{4}\alpha_{ij}Y_j + v_0\sum_{j=1}^{4}\alpha_{ij}Z_j,$$

$$\lambda_i = \sum_{j=1}^{4}\alpha_{ij}Z_j.$$

2.3 Estimation of the reference points.

If the last equations is substituted to the first and second ones, the following two equations are obtained:

$$u_i\sum_{j=1}^{4}\alpha_{ij}Z_j = f\sum_{j=1}^{4}\alpha_{ij}X_j + u_0\sum_{j=1}^{4}\alpha_{ij}Z_j$$

$$v_i\sum_{j=1}^{4}\alpha_{ij}Z_j = f\sum_{j=1}^{4}\alpha_{ij}Y_j + v_0\sum_{j=1}^{4}\alpha_{ij}Z_j$$

Then the first equation can be written as

$$u_i\sum_{j=1}^{4}\alpha_{ij}Z_j - u_0\sum_{j=1}^{4}\alpha_{ij}Z_i = f\sum_{j=1}^{4}\alpha_{ij}X_j$$

$$f\sum_{j=1}^{4}\alpha_{ij}X_j + (u_0 - u_i)\sum_{j=1}^{4}\alpha_{ij}Z_j = 0$$

Similarly, the second equation is modified as

$$f\sum_{j=1}^{4}\alpha_{ij}Y_j + (v_0 - v_i)\sum_{j=1}^{4}\alpha_{ij}Z_j = 0$$

Now, let's stack the coordinates into a vector:

$$C = \begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \\ X_2 \\ \vdots \\ Z_4 \end{bmatrix}$$

Then, we have $2n$ equations in the form of

$$MC = 0,$$

where

$$f \sum_{j=1}^{4} \alpha_{ij} X_j + (u_0 - u_i) \sum_{j=1}^{4} \alpha_{ij} Z_j = 0,$$

$$f \sum_{j=1}^{4} \alpha_{ij} Y_j + (v_0 - v_i) \sum_{j=1}^{4} \alpha_{ij} Z_j = 0.$$

The i-th row pair of the coefficient matrix is as follows:

$$M_i = \left[\begin{array}{cccc} M_{i,1} & M_{i,2} & M_{i,3} & M_{i,4} \end{array} \right],$$

where

$$M_{i,j} = \left[\begin{array}{ccc} f\alpha_{ij} & 0 & (u_0 - u_i)\,\alpha_{ij} \\ 0 & f\alpha_{ij} & (v_0 - v_i)\,\alpha_{ij} \end{array} \right].$$

Then the full coefficient matrix can be written as

$$M = \left[\begin{array}{c} M_1 \\ M_2 \\ \vdots \\ M_n \end{array} \right].$$

If there are at least six points, thus $n \geq 6$, the solution for the reference points, stacked in C is given as the eigenvector corresponding to the smallest eigenvalue for $M^T M$.

2.4 Scale ambiguity

However, the scale of vector C cannot be retrieved. But as we know the distanced between the reference points, the scale ambiguity can be removed.

2.5 Final pose estimation

Now, we know the reference points in the camera coordinate system. They are also known in the world, the final task is to estimate the rigid transformation between those. These can be computed using the so-called point registration algorithm.

3 Overview of the EPnP algorithm.

The full algorithm can be summarized as follows:

1. Take four reference point for the barycentric system, arbitrary.
2. For the n 3D locations, define the barycentric coordinates. See the appendix how barycentric coordinates can be calculated.
3. Compose the matrix M from 2D pixel locations and barycentric coordinates.
4. Get the eigenvector, corresponding to the smallest eigenvalue of $M^T M$.It gives you βC.
5. Remove the scale ambiguity, represented by of parameter β,as the distances between the reference points are known.
6. Get the reference points c_j in the camera coordinate system from triplet rows of C.
7. Apply a point registration problem to the reference points in the world and camera systems. Then rigid parameters, aka. pose, R and t are obtained.

A Estimation of barycentric coordinates for given reference points

If there are four barycentric reference points, the original coordinates for a point P can be given as a weighted sum as follows:

$$P = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \sum_{j=1}^{4} \alpha_j C^j .$$

Here, $\sum_{j=1}^{4} \alpha_j = 1$. Therefore, we have three equations for four unknown parameters α_j , but there is a constraint.

Then, let

$$\alpha_4 = 1 - \alpha_1 - \alpha_2 - \alpha_3$$

Then

$$P = \alpha_1 C^1 + \alpha_2 C^2 + \alpha_3 C^3 + (1 - \alpha_1 - \alpha_2 - \alpha_3)\, C^4$$

Thus,

$$P = \alpha_1 \left(C^1 - C^4\right) + \alpha_2 \left(C^2 - C^4\right) + \alpha_3 \left(C^3 - C^4\right) + C^4$$

$$P - C_4 = \alpha_1 \left(C^1 - C^4 \right) + \alpha_2 \left(C^2 - C^4 \right) + \alpha_3 \left(C^3 - C^4 \right)$$

The solution is given by

$$\left[\begin{array}{ccc} C^1 - C^4 & C^2 - C^4 & C^3 - C^4 \end{array} \right] \left[\begin{array}{c} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{array} \right] = P - C_4.$$

Therefore,

$$\left[\begin{array}{c} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{array} \right] = \left[\begin{array}{ccc} C^1 - C^4 & C^2 - C^4 & C^3 - C^4 \end{array} \right]^{-1} \left(P - C_4 \right).$$