3.2 The single-view geometry

Remarks

- The half point redundancy of the six points with respect to the strict minimum of 5.5 points in the calibration finds the ramification in the six-point algorithm in Section 3.5.2.
- A more elaborate calibration procedure usually includes a nonlinear optimization after the DLT solutions with additional nonlinear distortion parameters. The methods of using a planar calibration pattern introduced in [254, 215] are also practical.

3.2.4 The three-point pose algorithm

Given a set of point correspondences \( x_i \leftrightarrow u_i \) between the 3D reference points \( x_i \) and 2D image points \( u_i \), and also given the intrinsic parameters of the camera \( K \). The pose estimation consists of determining the position and orientation of the calibrated camera with respect to the known reference points. The camera pose is called space resection in photogrammetry. The difference between the camera calibration and the camera pose is that the intrinsic parameters of the camera needs to be estimated for the calibration and is known for the pose.

The distance constraint

An uncalibrated image points in pixels \( u_i \) and its calibrated counterpart \( \bar{x}_i \) is related by the known calibration matrix \( K \) such that \( u_i = K \bar{x}_i \). The calibrated point \( \bar{x}_i = K^{-1} u_i \) is a three vector representing a 3D direction in the camera-center coordinate frame. For convenience, we assume the direction vector is normalized to a unit vector such that \( \bar{x} \equiv \bar{x} / \| \bar{x} \| \). A 3D point corresponding to the back-projection of an image point/direction \( \bar{x} \) is determined by a depth \( \lambda \) as \( \lambda \bar{x} \). The depth \( \lambda \) is the camera-point distance.

In summary, \( u \) is an image point in pixels; \( \bar{x} \) is the direction vector of an image point for a calibrated camera; \( x' = \lambda \bar{x} \) is a space point corresponding to the image point \( u \) in the camera-centered coordinate frame; and \( x \) is the space point corresponding to the image point \( u \) in the world coordinate frame.

The distance between two 3D points represented by the vectors \( p \) and \( q \) is given by the cosine rule:

\[
\| p - q \|^2 = \| p \|^2 + \| q \|^2 - 2 p^T q.
\]

Applying this to the normalized direction vectors representing the 3D points in the camera frame, and using the fact that \( \| \bar{x}_p \| = 1 \), gives:

\[
d_{pq}^2 = \lambda_p^2 + \lambda_q^2 - c_{pq} \lambda_p \lambda_q,
\]
where $c_{pq} = 2\bar{x}_i^T\bar{x}_j = 2\cos(\theta_{pq})$ is a known constant from the image points, and $d_{pq}$ is the known distance between the space points.

The three points

If we are given three points, we have three pairs of points and three quadratic equations

$$f_{12}(\lambda_1, \lambda_2) = 0,$$
$$f_{13}(\lambda_1, \lambda_3) = 0,$$
$$f_{23}(\lambda_2, \lambda_3) = 0$$

of the form

$$f_{ij}(\lambda_i, \lambda_j) \equiv \lambda_i^2 + \lambda_j^2 - 2\lambda_i\lambda_j\cos\theta_{ij} - d_{ij}^2 = 0$$

for the three unknown distances $\lambda_1, \lambda_2, \lambda_3$.

The polynomial system has a Bézout bound of $2 \times 2 \times 2 = 8$ solutions. But the quadratic equations do not have linear terms, so $\lambda_i \mapsto -\lambda_i$ preserves the form and the eight solutions should appear in four pairs. There are many \emph{ad hoc} elimination techniques in the literature [206, 54] to effectively obtain a polynomial of degree four. Our favorite one is the classical Sylvester resultant, which first eliminates $\lambda_3$ from $f_{13}(\lambda_1, \lambda_3)$ and $f_{23}(\lambda_2, \lambda_3)$ to obtain a new polynomial $h(\lambda_1, \lambda_2)$ in $\lambda_1$ and $\lambda_2$. Then, we can further eliminate $\lambda_2$ from $f_{12}(\lambda_1, \lambda_2)$ and $h(\lambda_1, \lambda_2)$, and obtain a polynomial in $\lambda_1$ of degree eight with only even terms. We let $x = \lambda_1^2$, the polynomial of degree four is of the following form:

$$a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0,$$
where the coefficients are given in [182] for instances.

The equation has at most four solutions for $x$ and can be solved in closed-form. Since $\lambda_i$ is positive, $\lambda_1 = \sqrt{x}$. Then $\lambda_2$ and $\lambda_3$ are uniquely determined from $\lambda_1$.

**The absolute orientation**

The camera-point distances $\lambda_i$ are then converted into the camera-centered 3D coordinates, $x'_i = \lambda_i K^{-1} u_i$, of the reference points in space. We are now given a set of corresponding 3D points, $x'_i \leftrightarrow x_i$. We would like to compute a rigid transformation up to a scale, or a similarity transformation $R$, $t$, and $s$ such that the set of points $x_i$ maps to the set of points $x'_i$,

$$x'_i = sRx_i + t.$$

This is also called the *absolute orientation* in the photogrammetry literature.

As there are totally seven degrees of freedom for the similarity transformation, we need $2 \frac{1}{2}$ points or at least three points if we do not have a third of a point. It is elementary Euclidian geometry to find a closed-form solution from three points: the rotation maps the normal of the plane determined by the given three points, the scale is the ratio of the vector lengths relative to the centroid, and the translation is the de-rotated centroid. If more than three points are available, the best least-square rotation is obtained in closed-form using quaternions [45, 87]. The determination of the translation and the scale follow immediately from the rotation.

**Algorithm 2 (The three-point pose)** Given the calibration matrix $K$ of the camera and three 3D-2D point correspondences $x_i \leftrightarrow u_i$ for $i = 1, \ldots, 3$, compute the rotation $R$ and translation $t$ of the camera with respect to the points $x_i$.

1. Convert 3D points $x_i$ in coordinates into pair-wise distances $d_{ij}$. Convert 2D image points $u_i$ into the pair-wise angular measures $\cos \theta_{ij}$ with the calibration matrix $K$.
2. Compute the coefficients of the fourth degree polynomial in $x$ from the quadratic equations.
3. Solve the equation in closed-form. For each solution of $x$, get all the camera-point distances $\lambda_i$.
4. Convert back the distances $\lambda_i$ into the 3D points $x'_i$.
5. Estimate the similarity transformation, the scale, the rotation and the translation, between the two sets of 3D points $x_i$ and $x'_i$.

There are at most four solutions to $R$ and $t$.

**The critical configurations**

There are usually multiple solutions to the pose from the minimum of three points. All critical configurations for which multiple distinct or coinciding (unstable) solutions occur are known in [225, 251].