Computer Vision CITS4240

School of Computer Science & Software Engineering
The University of Western Australia

Camera calibration

We will now look at image formation and camera geometry in detail to determine how one *calibrates* a camera to determine the relationship between what appears on the image (or retinal) plane and where it is located in the 3D world.

Imagine we have a three dimensional coordinate system whose origin is at the centre of projection (also called the *optical centre*) and whose Z axis is along the optical axis, as shown in Figure 1. This coordinate system is called the *standard coordinate system* of the camera. A point M on an object with coordinates (X, Y, Z) will be imaged at some point $m = (x, y)$ in the image plane. These coordinates are with respect to a coordinate system whose origin is at the intersection of the optical axis and the image plane, and whose x and y axes are parallel to the X and Y axes. The relationship between the two coordinate systems (c, x, y) and (C, X, Y, Z) is given by

$$x = \frac{Xf}{Z} \qquad \text{and} \qquad y = \frac{Yf}{Z}, \tag{1}$$

where f is the *effective* focal length[1] of the camera and is often measured in pixel units.

This can be written linearly in homogeneous coordinates as

$$\begin{bmatrix} sx \\ sy \\ s \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix},$$

where $s \neq 0$ is a scale factor.

Now, the actual pixel coordinates (u, v) are defined with respect to an origin in the top left hand corner of the image plane, and will satisfy

$$\begin{aligned} u &= u_0 + x \quad \text{and} \\ v &= v_0 + y \end{aligned} \tag{2}$$

---

[1]This is not the same as the focal length (e.g., a 9mm lens) marked on the lens of the camera.
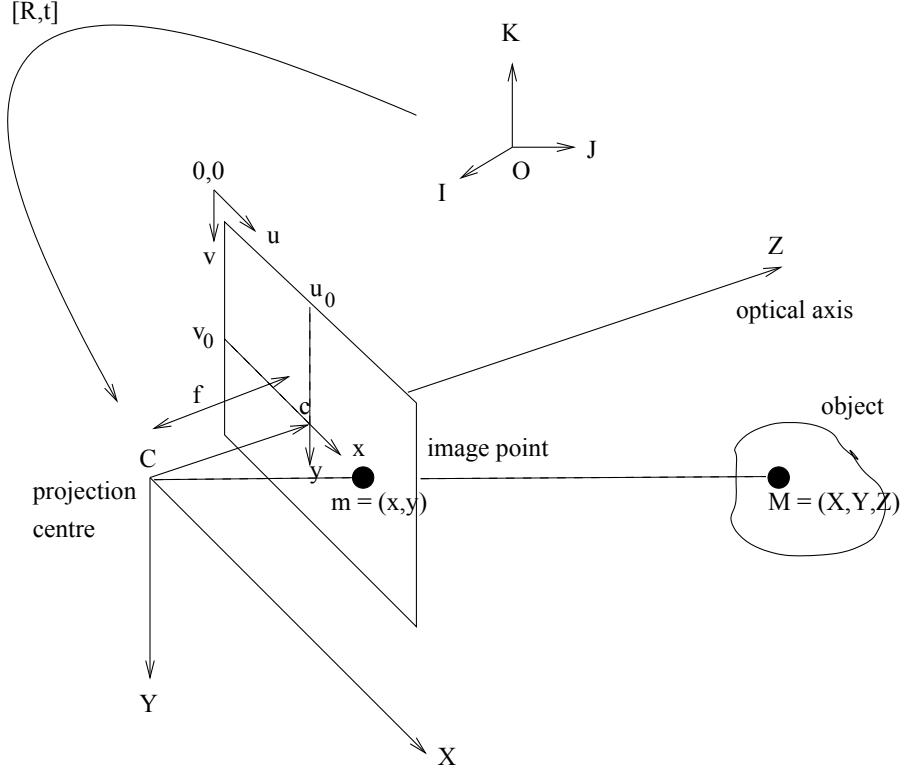
Figure 1: The coordinate systems involved in camera calibration.

We can express the transformation from three dimensional world coordinates to image pixel coordinates using a 3×4 matrix. This is done by substituting equation (1) into equation (2) and multiplying through by Z to obtain

$$
\begin{aligned}
Zu &= Zu_0 + Xf \\
Zv &= Zv_0 + Yf.
\end{aligned} \tag{3}
$$

In other words,

$$
\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} f & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix},
$$

where the scaling factor s has value Z. In short hand notation, we write this as

$$
\tilde{\mathbf{u}} = K\tilde{X},
$$

where $\tilde{\mathbf{u}}$ represents the homogeneous vector of image pixel coordinates, K is the perspective projection matrix, and $\tilde{\mathbf{X}}$ is the homogeneous vector of world coordinates. Thus, a camera can be considered as a system that performs a linear projective transformation from the projective space $\mathcal{P}^3$ into the projective plane $\mathcal{P}^2$.

There are three camera parameters, namely the focal length f, the parameter u_0 which is the u pixel coordinate at the *principal point*[2], and the parameter v_0 which is the v pixel

---

[2]The principal point is the image of the camera's optical centre.

coordinate at the principal point. Some old-fashioned CCD cameras have non-square pixels. These types of cameras are said to have aspect ratio other than 1 and would give rise to different scalings in the u and v-axes (e.g., a sphere would appear as an ellipse in the image). For these types of cameras, two terms, f_u and f_v, would be required to describe the effective focal length. The term f_u is the effective focal length in the u pixel units and f_v is the effective focal length in the v pixel units. As all modern cameras have unit aspect ratio, we can safely assume $f_u = f_v = f$. The parameters f, u_0 and v_0 do not depend on the position and orientation of the camera in space, and are thus called the *intrinsic* parameters.

In general, the three dimensional world coordinates of a point will not be specified in a frame whose origin is at the centre of projection and whose Z axis lies along the optical axis. Some other, more convenient frame, will more likely be specified, and then we have to include a change of coordinates from this other frame to the standard coordinate system. Thus we have

$$\tilde{\mathbf{u}} = K T \tilde{\mathbf{X}},$$

where T is a 4×4 homogeneous transformation matrix:

$$T = \left[\begin{array}{cc} R & \mathbf{t} \\ \mathbf{0}_3^\top & 1 \end{array} \right].$$

The top 3×3 corner is a rotation matrix R and encodes the camera orientation with respect to a given world frame; the final column is a homogeneous vector $\mathbf{t}$ capturing the camera displacement from the world frame origin. The matrix T has six degrees of freedom, three for the orientation, and three for the translation of the camera. These parameters are known as the *extrinsic* camera parameters.

The 3×4 camera matrix K and the 4×4 homogeneous transform T combine to form a single 3×4 matrix $\mathbf{C}$, called the *camera calibration matrix*. We can write the general form of $\mathbf{C}$ as a function of the intrinsic and extrinsic parameters:

$$\mathbf{C} = \left[\begin{array}{cc} f\mathbf{r}_1^\top + u_0\mathbf{r}_3^\top & ft_x + u_0t_z \\ f\mathbf{r}_2^\top + v_0\mathbf{r}_3^\top & ft_y + v_0t_z \\ \mathbf{r}_3^\top & t_z \end{array} \right], \tag{4}$$

where the vectors $\mathbf{r}_1, \mathbf{r}_2$, and $\mathbf{r}_3$ are the row vectors of the matrix R, and $\mathbf{t} = (t_x, t_y, t_z)^\top$. The matrix $\mathbf{C}$, like the matrix K, has rank three.

Orthographic projection

Consider a translation of f along the Z axis of the standard coordinate frame, so that the origin and the centre of the image plane are coincident, and the focal point is now positioned at $Z = -f$. Since there is no rotation involved in this transformation, it is easy to see that the camera calibration matrix is just

$$\mathbf{C} = \left[\begin{array}{cccc} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & f \end{array} \right],$$

3

where we are assuming that the pixel width and height are both 1. Now since $\mathbf{C}$ is defined up to a scale factor, this is the same as

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{f} & 1 \end{bmatrix}.$$

Now, if we let f go to infinity, the matrix becomes

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This defines the transformation $u = X$ and $v = Y$ and is known as an *orthographic projection* parallel to the Z axis. It appears as the limit of the general perspective projection as the focal length f becomes large with respect to the distance Z of the camera from the object.

Solving for the calibration matrix

Calibration is the process of estimating the intrinsic and extrinsic parameters of the camera. It can be thought of as a two stage process:

- estimating the matrix $\mathbf{C}$, and

- estimating the intrinsic and extrinsic parameters from $\mathbf{C}$.

In many cases, particularly for stereo, the second stage is not necessary.

We assume that we are given the 3D coordinate vectors $\mathbf{X}_i = (X_i, Y_i, Z_i)^\top$ of N reference points as well as the 2D retinal coordinates (u_i, v_i) of their images. In general, we have at least 6 points, preferably more, and they are arranged in a special pattern, such as that shown in Figure 2.

There are several methods for obtaining the coefficients of the matrix $\mathbf{C}$. We will outline both linear and non-linear methods.

Linear methods for estimating C

Linear method 1

Recall that in homogeneous coordinates we have a linear relationship between the image points with coordinates $(u_i, v_i)^\top$ and the 3D reference point coordinates $(X_i, Y_i, Z_i)^\top$ given by

$$\begin{bmatrix} s_i u_i \\ s_i v_i \\ s_i \end{bmatrix} = \begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}, \tag{5}$$
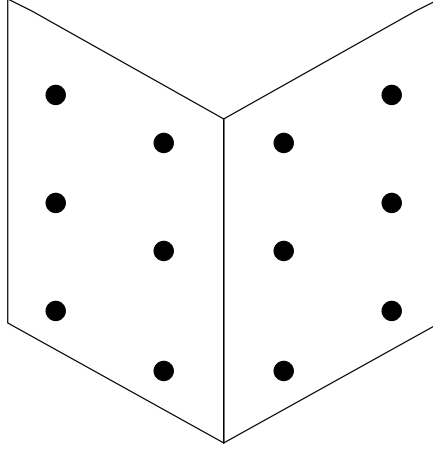
Figure 2: The pattern of points on a calibration frame.

or, in more compact form:

$$s_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1^\top & q_{14} \\ \mathbf{q}_2^\top & q_{24} \\ \mathbf{q}_3^\top & q_{34} \end{bmatrix} \tilde{\mathbf{X}}, \tag{6}$$

where $\mathbf{q}_j^\top = (q_{j1}, q_{j2}, q_{j3})$, for $j = 1, \cdots, 3$. Because of the arbitrary scale factor involved, we may set $q_{34} = 1$. From this equation we can write

$$u_i = \frac{X_i q_{11} + Y_i q_{12} + Z_i q_{13} + q_{14}}{X_i q_{31} + Y_i q_{32} + Z_i q_{33} + 1}$$

and

$$v_i = \frac{X_i q_{21} + Y_i q_{22} + Z_i q_{23} + q_{24}}{X_i q_{31} + Y_i q_{32} + Z_i q_{33} + 1}.$$

This implies that

$$X_i q_{11} + Y_i q_{12} + Z_i q_{13} + q_{14} - u X_i q_{31} - u Y_i q_{32} - u Z_i q_{33} = u_i$$

and

$$X_i q_{21} + Y_i q_{22} + Z_i q_{23} + q_{24} - v X_i q_{31} - v Y_i q_{32} - v Z_i q_{33} = v_i.$$

So given a set of N 3D world points and their image coordinates, we can build up the following matrix equation:

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1 X_1 & -u_1 Y_1 & -u_1 Z_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1 X_1 & -v_1 Y_1 & -v_1 Z_1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -u_N X_N & -u_N Y_N & -u_N Z_N \\ 0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -v_N X_N & -v_N Y_N & -v_N Z_N \end{bmatrix} \begin{bmatrix} q_{11} \\ q_{12} \\ q_{13} \\ q_{14} \\ q_{21} \\ q_{22} \\ q_{23} \\ q_{24} \\ q_{31} \\ q_{32} \\ q_{33} \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ u_N \\ v_N \end{bmatrix}, \tag{7}$$

where the large matrix on the left consisting of known image and world coordinates is $2N \times 11$.

With 11 unknowns and each point providing 2 constraint equations, we need at least six points to solve the equation.

The best least squares estimate of the q_{ij} is obtained using the pseudo-inverse as derived below.

We can write (7) in matrix form as

$$A\mathbf{q} = \mathbf{b}, \tag{8}$$

where A is the $2N \times 11$ matrix, $\mathbf{q} = (q_{11}, \cdots, q_{33})^\top$ is a 11-vector, and $\mathbf{b}$ is the $2N$-vector in (7).

To estimate the unknown elements embodied in the vector $\mathbf{q}$, we formulate the problem as one of minimizing $\|A\mathbf{q} - \mathbf{b}\|^2$, where $\|.\|$ denotes the 2-norm[3] of the vector.

Let

$$L = \|A\mathbf{q} - \mathbf{b}\|^2 = (A\mathbf{q} - \mathbf{b})^\top (A\mathbf{q} - \mathbf{b}).$$

Then our objective is

$$\min_{\mathbf{q}} L = \min_{\mathbf{q}} (A\mathbf{q} - \mathbf{b})^\top (A\mathbf{q} - \mathbf{b})$$

Differentiating L with respect to $\mathbf{q}$ and setting to 0 gives

$$A^\top (A\mathbf{q} - \mathbf{b}) = 0$$
$$\implies \quad A^\top A\mathbf{q} = A^\top \mathbf{b}$$
$$\implies \quad \mathbf{q} = (A^\top A)^{-1} A^\top \mathbf{b} \tag{9}$$

In general, the term $(A^\top A)^{-1} A^\top$ is commonly known as the *pseudo-inverse* of the matrix A and is often denoted by A^+. i.e.,

$$A^+ = (A^\top A)^{-1} A^\top.$$

It is clear from (9) that $\mathbf{q}$ can only be estimated if $A^\top A$ is invertible. With A being a $2N \times 11$ matrix, this means that $A^\top A$ must be of full rank (i.e. 11). That is, we need $N \geq 6$ and the rows of A must not be linearly dependent so that the rank of A would not collapse to less than 11. Thus, we must ensure that the reference points $\{\mathbf{X}_i \,|\, 1 \leq i \leq N, \, N \geq 6\}$ are in *general position*. This means that the chosen reference points must not lie in a certain configuration, which can be defined mathematically but is beyond the scope of this lecture. If six or more points are chosen at random, and do not lie on a plane, then we can be confident that this situation will not occur.

---

[3]For example, $\|(a, b, c)\| = \sqrt{a^2 + b^2 + c^2}$. So $\|(a, b, c)\|^2 = a^2 + b^2 + c^2 = \begin{pmatrix} a & b & c \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$.

Linear method 2

A slightly different way of solving for the unknowns q_{ij} is to relax the condition $q_{34} = 1$ in (5), giving

$$
\begin{bmatrix}
X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 & -u_1 \\
0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 & -v_1 \\
. & . & . & . & . & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . & . & . & . \\
X_N & Y_N & Z_N & 1 & 0 & 0 & 0 & 0 & -u_NX_N & -u_NY_N & -u_NZ_N & -u_N \\
0 & 0 & 0 & 0 & X_N & Y_N & Z_N & 1 & -v_NX_N & -v_NY_N & -v_NZ_N & -v_N
\end{bmatrix}
\begin{bmatrix}
q_{11} \\ q_{12} \\ q_{13} \\ q_{14} \\ q_{21} \\ q_{22} \\ q_{23} \\ q_{24} \\ q_{31} \\ q_{32} \\ q_{33} \\ q_{34}
\end{bmatrix}
=
\begin{bmatrix}
0 \\ 0 \\ . \\ . \\ . \\ . \\ . \\ . \\ . \\ 0 \\ 0
\end{bmatrix}.
\tag{10}
$$

In matrix form, this is

$$A\mathbf{q} = \mathbf{0}, \tag{11}$$

where, in this case, A is a $2N \times 12$ data matrix, $\mathbf{q}$ is a 12-vector containing the unknowns q_{ij}, and $\mathbf{0}$ is a $2N$-vector of zeros. A trivial solution to (11) is $\mathbf{q} = \mathbf{0}$, which is not physically significant. To obtain the non-trivial solution, we can use a technique known as constrained optimization. That is, we formulate the problem as one of minimizing

$$\|A\mathbf{q}\|^2 \tag{12}$$

subject to the constraint $\|\mathbf{q}\|^2 - 1 = 0$ (This constraint will prevent $\mathbf{q}$ from becoming a zero vector.).

Let $\lambda > 0$ be the Lagrange multiplier. Then the Lagrangian to be minimized is

$$L(\mathbf{q}, \lambda) = (A\mathbf{q})^\top(A\mathbf{q}) - \lambda(\mathbf{q}^\top\mathbf{q} - 1).$$

Differentiating L with respect to $\mathbf{q}$ and setting to 0 gives

$$\frac{\partial L}{\partial \mathbf{q}} = A^\top A\mathbf{q} - \lambda\mathbf{q} = 0$$
$$\implies \quad A^\top A\mathbf{q} = \lambda\mathbf{q}. \tag{13}$$

Differentiating L with respect to λ and setting to 0 gives

$$\frac{\partial L}{\partial \lambda} = \mathbf{q}^\top\mathbf{q} - 1 = 0$$
$$\implies \quad \mathbf{q}^\top\mathbf{q} = 1. \tag{14}$$

Pre-multiplying both sides of (13) by $\mathbf{q}^\top$ gives

$$\mathbf{q}^\top A^\top A\mathbf{q} = \lambda\mathbf{q}^\top\mathbf{q}$$
$$\implies \quad (A\mathbf{q})^\top(A\mathbf{q}) = \lambda \cdot 1$$
$$\implies \quad \|A\mathbf{q}\|^2 = \lambda \tag{15}$$

We see that the left hand side of (15) is the same expression that we try to minimize in (12). This says that to minimize $\|A\mathbf{q}\|^2$, we should minimize λ.

Now equation (13) tells us that $\mathbf{q}$ should be an eigenvector of the matrix $A^\top A$ with λ being the corresponding eigenvalue. Equation (15), on the other hand, says that we should minimize λ as much as possible (ideally λ should be 0). These two pieces of information together simply say that $\mathbf{q}$ should be the eigenvector that corresponds to the *smallest* eigenvalue of $A^\top A$.

Thus, we can solve for $\mathbf{q}$ via the eigen-decomposition of $A^\top A$.

Suffice it to say, it leads to a closed form solution. What is of more interest though, is the question of the rank of the matrix A, since this reinforces our understanding of how the reference points should be chosen. Let us reconsider equation (11). We can think of what we are after in that equation is the non-trivial null vector $\mathbf{q}$ of the matrix A.

We know from standard linear algebra that if we have an $n \times m$ matrix A then

$$\text{rank}(A) + \text{null}(A) = m,$$

where $\text{null}(A)$ represents the dimension of the null space of A. In our case $n \geq m = 12$ and there are three cases to consider:

- $\text{rank}(A) = 12$. Then the null space has dimension 0, and there is only one solution to the system, namely $\mathbf{q} = 0$, which is not very meaningful. See also the discussion later about this case.

- $\text{rank}(A) = 11$. Then the null space has dimension 1 and there is a unique solution (up to a scale factor).

- $\text{rank}(A) < 11$. Then the null space has dimension 2 or more. The null vector $\mathbf{q}$ we are seeking can be any vector in this 2-dimensional space. This means that there is an infinite number of solutions to equation (11). One way in which this can happen is if all the world reference points are in a plane.

So, we come to the same conclusion as our previous analysis of the pseudo-inverse and non-singularity of $A^\top A$: we need to ensure that we use at least 6 world reference points and that they are in *general position*.

Note that the rank of A is often 12 rather than 11 in calibration with real data — noise inflates the rank of the matrix!!! When noise is present in our data points, the end result is that the smallest eigenvalue of $A^\top A$ is not zero but a small positive number[4]. We can often pinpoint how much noise there is in our data (i.e., the world and image coordinates of the reference points) by inspecting the ratio between the smallest and the largest eigenvalues of the data matrix $A^\top A$.

Non-linear methods for estimating C

It is possible to re-cast the problem of solving equation (5) as a non-linear minimization problem, where we attempt to minimize the distance in the image plane between the point

---

[4]Note that $A^\top A$ is positive definite and symmetric. See the earlier lecture note on the positive definite property of a matrix.

coordinates (u_i, v_i) and the re-projected points of $(X_i, Y_i, Z_i, 1)$ using the estimate of $\mathbf{C}$. We can do this by defining the quantity

$$e = \sum_{i=1}^{N} \left\| \frac{\mathbf{q}_1^\top \mathbf{X}_i + q_{14}}{\mathbf{q}_3^\top \mathbf{X}_i + q_{34}} - u_i \right\|^2 + \left\| \frac{\mathbf{q}_2^\top \mathbf{X}_i + q_{24}}{\mathbf{q}_3^\top \mathbf{X}_i + q_{34}} - v_i \right\|^2. \qquad (16)$$

This quantity, e, is known as the *reprojection error*. Unlike the linear methods which yield closed-form solutions, a non-linear method only refines an initial estimate of the solution in an iterative fashion, leading to a better estimate of the solution. In that regards, we can treat a non-linear method as a postprocess that refines the initial estimate of the solution obtained from a linear method.

Using the initial estimate of $\mathbf{q} = (q_{11}, \cdots, q_{34})^\top$ that we obtain from any of the linear methods mentioned above, we can compute an initial value of the reprojection error, e, using (16). Applying any standard minimization technique, such as gradient descent or Newton's method, we can compute a new and improved estimate of $\mathbf{q}$ that gives a smaller reprojection error e in each iteration.

It is important that the initial estimate of the solution $\mathbf{q}$ is very close to the true solution. Otherwise, the non-linear method that we adopt may lead to a local minimum rather than the global minimum. One will also need to set a suitable stopping criterion for terminating the iteration (e.g., terminate iteration when $e < 10^{-6}$).

In general, non-linear methods lead to much more robust solutions.

Decomposing C to obtain intrinsic parameters

Clearly, not every 3×4 matrix can be written in the form of equation (4) above. Indeed, this matrix depends upon nine parameters, $f, u_0, v_0, t_x, t_y, t_z$, and the three independent degrees of freedom associated with the rotation matrix $R = [\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3]^\top$. A general 3×4 projective matrix has *eleven* degrees of freedom: it has 12 entries, but an arbitrary scale factor is involved, so one of the entries can be set to 1 without loss of generality.

With the elements of $\mathbf{C}$ written in the form as (6), there exist 4 sets of intrinsic and extrinsic parameters such that $\mathbf{C}$ can be written in the form of equation (4) if and only if the following two conditions are satisfied:

1. $\|\mathbf{q}_3\| = 1$, and

2. $(\mathbf{q}_1 \wedge \mathbf{q}_3) \cdot (\mathbf{q}_2 \wedge \mathbf{q}_3) = 0$.

To see why this is true, we will consider the following proof, and, in the process, compute explicitly the nine unknown parameters:

If $\mathbf{C}$ is in the form of equation (4), then $\mathbf{q}_3 = \mathbf{r}_3$, and since $\mathbf{r}_3$ is a row of a rotation matrix, its norm is 1. Moreover,

$$\begin{aligned} (\mathbf{q}_1 \wedge \mathbf{q}_3) \cdot (\mathbf{q}_2 \wedge \mathbf{q}_3) &= (f\mathbf{r}_1 + u_0\mathbf{r}_3) \wedge \mathbf{r}_3 \cdot (f\mathbf{r}_2 + v_0\mathbf{r}_3) \wedge \mathbf{r}_3 \\ &= (f\mathbf{r}_1 \wedge \mathbf{r}_3) \cdot (f\mathbf{r}_2 \wedge \mathbf{r}_3) \\ &= 0. \end{aligned} \qquad (17)$$

So in the case when $\mathbf{C}$ is a matrix in the form of equation (4), both of these conditions clearly hold.

To prove the converse, we note that if $\|\mathbf{q}_3\|^2 = 1$ then the third row of $\mathbf{C}$ must be a unit vector. The second condition

$$(\mathbf{q}_1 \wedge \mathbf{q}_3) \cdot (\mathbf{q}_2 \wedge \mathbf{q}_3) = 0,$$

on the other hand, implies that $\mathbf{q}_1$, $\mathbf{q}_2$, and $\mathbf{q}_3$ can take on the more strict form that they are mutually orthogonal to each other. However, they can also take the more general form as given below:

$$\mathbf{q}_1 = \alpha_1 \mathbf{r}_1 + \beta_1 \mathbf{r}_3$$
$$\mathbf{q}_2 = \alpha_2 \mathbf{r}_2 + \beta_2 \mathbf{r}_3$$
$$\mathbf{q}_3 = \mathbf{r}_3,$$

where $\alpha_1, \alpha_2, \beta_1$, and β_2 are arbitrary constants. Clearly, for modern cameras, we would expect $\alpha_1 \approx \alpha_2$. The two conditions in (17) impose no constraints on the 4th column of the $\mathbf{C}$ matrix, leaving t_x and t_y to take on values that fit with other parameters.

In practice, $\mathbf{C}$ is defined up to an unknown scale factor so $\|\mathbf{q}_3\| \neq 1$. To decompose $\mathbf{C}$ to obtain the intrinsic and extrinsic parameters, we must firstly estimate the unknown scale factor, s, that 'undoes' the scaling. We incorporate s into the $\mathbf{C}$ matrix and rewrite (4) as follows:

$$s \begin{bmatrix} \mathbf{q}_1^\top & q_{14} \\ \mathbf{q}_2^\top & q_{24} \\ \mathbf{q}_3^\top & q_{34} \end{bmatrix} = \begin{bmatrix} f\mathbf{r}_1^\top + u_0\mathbf{r}_3^\top & ft_x + u_0t_z \\ f\mathbf{r}_2^\top + v_0\mathbf{r}_3^\top & ft_y + v_0t_z \\ \mathbf{r}_3^\top & t_z \end{bmatrix}. \qquad (18)$$

Then $\|s\,\mathbf{q}_3\| = \|\mathbf{r}_3\|$ implies that $\|s\,\mathbf{q}_3\| = 1$. Thus,

$$s = \pm\frac{1}{\|\mathbf{q}_3\|}. \qquad \text{(2 solutions)}$$

Next,

$$t_z = s\,q_{34} = \pm\frac{q_{34}}{\|\mathbf{q}_3\|}. \qquad \text{(2 solutions)}$$

and

$$\mathbf{r}_3 = s\,\mathbf{q}_3. \qquad \text{(2 solutions)}$$

As $\mathbf{r}_1$ is orthogonal to $\mathbf{r}_3$, the dot product of $f\mathbf{r}_1 + u_0\mathbf{r}_3$ and $\mathbf{r}_3$ will annihilate the f term, leaving only the u_0 term. So

$$u_0 = s^2(\mathbf{q}_1 \cdot \mathbf{q}_3). \qquad \text{(1 solution)}$$

Similarly,

$$v_0 = s^2(\mathbf{q}_2 \cdot \mathbf{q}_3). \qquad \text{(1 solution)}$$

Again, the mutual orthogonality of $\mathbf{r}_1$, $\mathbf{r}_2$, and $\mathbf{r}_3$ can be exploited in the cross product of $f\mathbf{r}_1 + u_0\mathbf{r}_3$ and $\mathbf{r}_3$ for estimating f:

$$f = \pm s^2\|\mathbf{q}_1 \wedge \mathbf{q}_3\|. \qquad \text{(2 solutions for each value of } s) \quad (*)$$

The minus sign in the formula $(*)$ above takes care of the case when f is negative (this depends on which direction the optical axis is pointing). Alternatively,

$$f = \pm s^2\|\mathbf{q}_2 \wedge \mathbf{q}_3\|. \qquad \text{(2 solutions for each value of } s) \quad (*')$$

Ideally, the values of f obtained from $(*)$ and $(*')$ should be identical. If these two values are similar then it is reasonable to take the average from the two formulae as the final value for f; otherwise, the accuracy of the data and your calibration/computation procedure must be re-examined. You may also need to reconsider your assumption about the pixels being square (rather than rectangular).

Finally, having computed f, t_z, u_0 and v_0, we can now compute $\mathbf{r}_1, \mathbf{r}_2, t_x$ and t_y:

$$
\begin{aligned}
\mathbf{r}_1 &= s\,(\mathbf{q}_1 - u_0\mathbf{q}_3)/f &&\text{(2 solutions)} \\
\mathbf{r}_2 &= s\,(\mathbf{q}_2 - v_0\mathbf{q}_3)/f &&\text{(2 solutions)} \\
t_x &= (s\,q_{14} - u_0 t_z)/f &&\text{(2 solutions)} \\
t_y &= (s\,q_{24} - v_0 t_z)/f &&\text{(2 solutions)}
\end{aligned}
$$

Because of the choices of ± 1 in front of the solution for s and the combination of a $\pm$ sign for f, we see that there are four possible sets of solutions for the intrinsic and extrinsic parameters. These four sets of solutions correspond to whether the origin of the coordinates is in front of the camera ($t_z > 0$) or behind it ($t_z < 0$), and to the choice of the direction of the optical axis. Note that you may rule out two sets of the solutions if you enforce a right-hand coordinate system, which imposes that $\det(R)$ must be equal to 1 (rather than -1).

References

[1] Olivier Faugeras. "Three Dimensional Computer Vision". MIT Press, 1993.

[2] S. Ganapathy. "Decomposition Of Transformation Matrices For Robot Vision". *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 130-139, 1984.