IIST, Thiruvananthapuram	AVD624 - Computer Vision
Lecture Notes: Week - 2	
01/02/2022 - 04/02/2022	
Lecturer: Dr. Deepak Mishra	*Summarised by: Vatsalya Gupta*

This is a weekly summary of the lectures given by *Dr. Deepak Mishra*, Prof. and Head of Department of Avionics, for the course on *Computer Vision* conducted during the even semester 2021-22.

1 Camera Calibration

We want to determine the extrinsic and intrinsic parameters of a camera, given the coordinates of the control points. We also observe the coordinates of these 3D object points in the image. As seen in the previous lectures, Direct Linear Transform (DLT) maps object point *X* to the image point *x*, using the equation

$$x = KR[I_3 \mid -X_0] = PX$$

Here, *K* represents the intrinsics, i.e. camera-internal parameters, and X_0 and *R* represent the extrinsics, i.e. pose parameters of the camera. This equation has 11 unknowns, *c* (focal length), *s* (shear parameter), *m* (scale parameter), x_H, y_H (camera centre), 3 rotations and 3 translations. If we assume that the camera is already calibrated using a calibration matrix ${}^cK$, then we have only 6 unknown parameters left, and this can be solved using spatial resection. Otherwise, we can get our solution for 11 elements of P using DLT.

1.1 Elements of *P* using DLT

Each point gives us two observation equations, that means we need at least 6 points. So we define three column matrices *A, B and C* as the transpose of the respective 3 rows of the projection matrix.

$$x_i = \begin{bmatrix} A^T \\ B^T \\ C^T \end{bmatrix} X_i$$

Hence, for each point we have

$$a_{x_i}^T p = 0 \; and \; a_{y_i}^T p = 0$$

where, the parameter p is defined as the transpose of vector P. Thus, we can determine the projection matrix. In case of redundant observations, we can use Singular Value Decomposition (SVD) to find p such that it minimises $\Omega = w^T w$, given M (the matrix formed by stacking the point vectors). We get, $||P||_2 = \sum_{ij} p_{ij}^2 = ||p|| = 1$. Now, choosing p as the singular vector belonging to the smallest singular value minimises Ω and gives us an estimate of P.

Note - We will not get any solution if all points X_i are located on a plane or all points X_i and the projection center X_0 are located on a twisted cubic curve. So, more number of points may be needed to get the estimate.

1.2 Decomposition of *P*

We can obtain K, R and X_0 from the estimate obtained for the projection matrix P. We have,

$$P = [KR \mid -KRX_0] = [H \mid h]$$

The projection centre can be obtained using QR decomposition of H^{-1} which gives $R^T K^{-1}$, then the rotation and camera constant matrices can be obtained by choosing

$$R \leftarrow R(z,\pi)R\ ,\ K \leftarrow \frac{1}{K_{33}} K R(z,\pi)$$

1.3 Camera Calibration using Zhang's Method

Figure 1: Three perspective views of a checkerboard

We take example of a checkerboard. The main idea is to set the world coordinate system to the corner of the checkerboard. All points on the checkerboard lie in the XY plane, i.e., Z = 0. Deleting the third column of the extrinsic parameter matrix leads to the following simplification,

$$H = [h1, h2, h3] = \begin{bmatrix} c & cs & x_H \\ 0 & c(1+m) & y_H \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{bmatrix}$$

For multiple observed points in the same image of the checkerboard, we see that,

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = H_{3\times3} \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix}\ ,\ i = 1, ..., I$$

Using the same DLT method as discussed above, but this time we apply it for the 3x3 homography matrix. We can observe that *H* has 8 DoF. Since each point would give us 2 observation equations, we need at least 4 points to get an estimate of *H*.

Now to obtain the *K* matrix, we simplify the matrix multiplication and use the constraint that r_1, r_2, r_3 form an orthonormal basis.

$$r_1^T r_2 = 0 \implies h_1^T K^{-T} K^{-1} h_2 = 0$$

$$||r_1|| = ||r_2|| = 1 \implies h_1^T K^{-T} K^{-1} h_1 - h_2^T K^{-T} K^{-1} h_2 = 0$$

We can define a symmetric and positive definite matrix B as $K^{-T}K^{-1}$. Let b be a row vector defining the upper triangular elements of B. We can solve the following linear system to obtain b and hence K.

$$\begin{pmatrix} v_{12}^T \\ v_{11}^T - v_{22}^T \\ ... \\ v_{12}^T \\ v_{11}^T - v_{22}^T \end{pmatrix} b = Vb = 0$$

For non-trivial solution, we set $||b|| = 1$ and find the solution which minimises the squares error,

$$b^* = arg\ min_b \ ||\ Vb\ ||$$

Note - We need at least 4 points per plane to compute the matrix *H*. Each plane gives us two equations. Since B has 5/6 DoF, we need at least 3 different views of a plane.

2 Camera Pose Estimation

We know the intrinsic parameters of the camera from the methods discussed above. We need to determine the 6 DoF pose of a camera (position and orientation) with respect to the world frame. Given are 3D coordinates X_i of $I \geq 3$ object points and the corresponding image coordinates x_i. So we can formulate it as a Perspective-Three-Point (P3P) problem to get an estimate of the 6 extrinsic parameters.

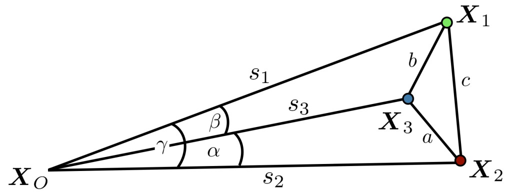

Figure 2: Camera centre and the correspondence points

We follow a two step process - estimate the length of projection rays, and then estimate the orientation. We can directly obtain the angle between rays since intrinsics are known.

$$\alpha = arccos({}^k x_2^s\ ,\ {}^k x_3^s)\ ,\ \beta = arccos({}^k x_3^s\ ,\ {}^k x_1^s)\ ,\ \gamma = arccos({}^k x_1^s\ ,\ {}^k x_2^s)$$

Now we need to get the length of the projection rays. Using law of cosines, we get

$$a^2 = s_2^2 + s_3^2 - 2s_2s_3cos\alpha$$
$$b^2 = s_3^2 + s_1^2 - 2s_3s_1cos\beta$$
$$c^2 = s_1^2 + s_2^2 - 2s_1s_2cos\gamma$$

Using the substitution $u = s_2/s_1$ and $v = s_3/s_1$, results in a fourth degree polynomial,

$$A_4v^4 + A_3v^3 + A_2v^2 + A_1v + A_0 = 0$$

Here A_i's are a function of a, b, c, α, β and γ. From here, we solve for v and get the values of s_1, s_2 and s_3. We can see that there will be four possible solutions, and these solutions will have different orientation of the $X_1X_2X_3$ triangle, since tilting it has no effect on our parameters.

■■■