Solving for 3D coordinates χ given n sets of 2D images coordinates

Here we describe the 3D reconstruction approach implemented in our toolbox in detail. For an example point in world coordinates $[X, Y, Z]$ recorded from 'n' cameras at a given time, our implementation follows the textbook (@Hartley2003) prescribed approach to 3D reconstruction. Given we have 'n' cameras we get 'n' sets of measurements of the point we are interested in reconstructing and is given by the below system of equations for each camera 'i'.

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = P_i \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

Here the projection matrix P_i, is of the form in equation (2). It is composed of intrinsic camera parameters namely focal length and principal point and extrinsic parameters namely rotation and translation vectors. These matrices are obtained from the camera calibration process.

$$P = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \tag{2}$$

We can combine the 'n' sets of linear equations into one of the form in (3) using cross product and solve for χ

$$A\chi = 0 \ ; \ \chi = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{3}$$

Since $[x_i, y_i, 1]$, is in the column space of P_i, cross product between vectors $[x_i, y_i, 1]$ and $[p_i^{1T}\chi, p_i^{2T}\chi, p_i^{3T}\chi]$ results in two independent equations shown below.

$$\begin{aligned} x_i \left(p_i^{3T}\chi\right) - \left(p_i^{1T}\chi\right) &= 0 \\ y_i \left(p_i^{3T}\chi\right) - \left(p_i^{2T}\chi\right) &= 0 \end{aligned} \tag{4}$$

Obtaining similar equations from all cameras and concatenating them along the rows of A results in a system of equations represented in equation (3) with

$$A = \begin{bmatrix} x_1 p_1^{3T} - p_1^{1T} \\ y_1 p_1^{3T} - p_1^{2T} \\ . \\ . \\ x_n p_n^{3T} - p_n^{1T} \\ y_n p_n^{3T} - p_n^{2T} \end{bmatrix} \tag{5}$$

Due to low image resolution and errors in camera calibration among others, there is no exact solution. Approximate solution is obtained by minimizing (3) in the least squared sense. To avoid trivial solution, constrained equation with Lagrange multiplier is constructed as in equation (6) and minimized (@Inkila2005).

$$L(x, \lambda) = |A\chi|^2 + \lambda \left(1 - |\chi|^2\right) \tag{6}$$

Taking partial derivative with respect to χ and setting it to 0 to minimize L, we get

$$A^T A\chi = \lambda\chi \tag{7}$$

Substituting (7) in equation (6), we get

$$L = |A\chi|^2 + \lambda \left(1 - |\chi|^2\right) = \chi^T A^T A\chi + \lambda \left(1 - \chi^T \chi\right) \tag{8}$$

$$L = \chi^T \lambda\chi + \lambda - \lambda\chi^T \chi = \lambda \tag{9}$$

From equation (9), we see that the χ that minimizes (3) in the least squared sense is the Eigen vector of $A^T A$ associated with its lowest Eigen value.

References

[1] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision.* Cambridge university press, 2003.

[2] Keijo Inkilä. Homogeneous least squares problem. *Photogrammetric Journal of Finland*, 19(2):34–42, 2005.