CS6733 Class Notes, Fall 2007

1 Camera Calibration

Camera calibration is used to find the mapping from 3D to image space coordinates. There are 2 approaches:

- Method I: Find both extrinsic and intrinsic parameters of the camera system. However, this can be difficult to do. The instinsic parameters of the camera may be unknown (i.e. focal length, pixel dimension) and the 6-DOF transform also may be difficult to calculate directly.
- Method 2: An easier method is the "Lumped" transform. Rather than finding individual parameters, we find a composite matrix that relates 3D to 2D. Given the equation below:

$$^{image}P \;=\; ^{image}T_{persp}\,^{persp}T_{camera}\,^{camera}T_{world}\,^{world}P$$

  we can lump the 3 T matrices into a 3x4 calibration matrix C:

$$^{image}P \;=\; C\,^{world}P$$

$$C \;=\; ^{image}T_{persp}\,^{persp}T_{camera}\,^{camera}T_{world}$$

- C is a <u>single</u> 3×4 transform that we can calculate empirically.

$$\overbrace{\begin{bmatrix} C \end{bmatrix}}^{3\times4} \underbrace{\overbrace{\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}}^{4\times1}}_{\text{3-D homo. vec}} = \underbrace{\overbrace{\begin{bmatrix} u \\ v \\ w \end{bmatrix}}^{3\times1}}_{\text{2-D homo. vec}} \triangleq \underbrace{\begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix}}_{\text{Pixels}} \quad \begin{array}{l} where \\ u' = \frac{u}{w} \\ v' = \frac{v}{w} \end{array}$$

- Multiplying out the equations, we get:

$$c_{11}x + c_{12}y + c_{13}z + c_{14} = u$$

$$c_{21}x + c_{22}y + c_{23}z + c_{24} = v$$

$$c_{31}x + c_{32}y + c_{33}z + c_{34} = w$$

- Substituting $u = u'w$ and $v = v'w$, we get:
    1. $c_{11}x + c_{12}y + c_{13}z + c_{14} = u'(c_{31}x + c_{32}y + c_{33}z + c_{34})$
    2. $c_{21}x + c_{22}y + c_{23}z + c_{24} = v'(c_{31}x + c_{32}y + c_{33}z + c_{34})$
- How to interpret <u>1</u> and <u>2</u>:
    1. If we know all the c_{ij} and x, y, z, we can find u', v'. This means that if we know calibration matrix C and a 3-D point, we can predict its image space coordinates.
    2. If we know $x,\ y,\ z,\ u',\ v'$, we can find c_{ij}. Each 5-tuple gives 2 equations in c_{ij}. This is the basis for empirically finding the calibration matrix C (more on this later).
    3. If we know $c_{ij},\ u',\ v'$, we have 2 equations in x, y, z. They are the equations of 2 planes in 3-D. 2 planes form an intersecton which is a line. These are the equations of the line emanating from the center of projection of the camera, through the image pixel location u', v' and which contains point x, y, z.
- We can set up a linear system to solve for c_{ij}: $AC = B$

$$\begin{bmatrix} x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 & -u'_1x & -u'_1y & -u'_1z \\ 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -v'_1x & -v'_1y & -v'_1z \\ x_2 & y_2 & z_2 & 1 & 0 & 0 & 0 & 0 & -u'_2x & -u'_2y & -u'_2z \\ 0 & 0 & 0 & 0 & x_2 & y_2 & z_2 & 1 & -v'_2x & -v'_2y & -v'_2z \\ . \\ . \\ . \\ . \\ . \\ . \\ . \end{bmatrix} \underbrace{\begin{bmatrix} c_{11} \\ c_{12} \\ c_{13} \\ c_{14} \\ c_{21} \\ c_{22} \\ c_{23} \\ c_{24} \\ c_{31} \\ c_{32} \\ c_{33} \end{bmatrix}}_{\text{We can assume } c_{34}=1} = \begin{bmatrix} u'_1 \\ v'_1 \\ u'_2 \\ v'_2 \\ u'_3 \\ v'_3 \\ . \\ . \\ . \\ u'_N \\ v'_N \end{bmatrix}$$

- Each set of points x, y, z, u', v' yields 2 equations in <u>11</u> unknowns (the c_{ij}'s).
- To solve for C, A needs to be invertible (square). We can <u>overdetermine</u> A and find a Least-Squares fit for C by using a pseudo-inverse solution.

  If A is $N \times 11$, where $N > 11$,

$$AC = B$$
$$A^TAC = A^TB$$
$$C = \underbrace{(A^TA)^{-1}}_{\text{pseudo inv.}}\ A^TB$$