Camera Calibration

D.A. Forsyth, UIUC
Camera calibration

- Issues:
  - what are intrinsic parameters of the camera?
  - what is the camera matrix? (intrinsic+extrinsic)

- General strategy:
  - view calibration object; identify image points in image
  - obtain camera matrix by minimizing error
  - obtain intrinsic parameters from camera matrix
Optimization problem: Notation

The optimization problem is relatively straightforward to formulate. Notation is the main issue. We have \( N \) reference points \( s_i = [s_{x,i}, s_{y,i}, s_{z,i}] \) with known position in some reference coordinate system in 3D. The measured location in the image for the \( i \)’th such point is \( \hat{t}_i = [\hat{t}_{x,i}, \hat{t}_{y,i}] \). There may be measurement errors, so the \( \hat{t}_i = t_i + \xi_i \), where \( \xi_i \) is an error vector and \( t \) is the unknown true position. We will assume the magnitude of error does not depend on direction in the image plane (it is isotropic), so it is natural to minimize the squared magnitude of the error

\[
\sum_{i} \xi_i^T \xi_i.
\]

The main issue here is writing out expressions for \( \xi_i \) in the appropriate coordinates. Write \( T_i \) for the intrinsic matrix whose \( u, v \)’th component will be \( i_{uv} \); \( T_e \) for the extrinsic transformation, whose \( u, v \)’th component will be \( e_{uv} \). Recalling that \( T_i \) is lower triangular, and engaging in some manipulation, we obtain
\[
\sum_i \xi_i^T \xi_i = \sum_i (t_{x,i} - p_{x,i})^2 + (t_{y,i} - p_{y,i})^2
\]

where

\[
p_{x,i} = \frac{i_{11}g_{x,i} + i_{12}g_{y,i} + i_{13}g_{i,3}}{g_{i,3}}
\]

\[
p_{y,i} = \frac{i_{22}g_{x,i} + i_{23}g_{i,3}}{g_{i,3}}
\]

and

\[
g_{x,i} = e_{11}s_{x,i} + e_{12}s_{y,i} + e_{13}s_{z,i} + e_{14}
\]

\[
g_{y,i} = e_{21}s_{x,i} + e_{22}s_{y,i} + e_{23}s_{z,i} + e_{24}
\]

\[
g_{z,i} = e_{31}s_{x,i} + e_{32}s_{y,i} + e_{33}s_{z,i} + e_{34}
\]

(which you should check as an exercise). This is a constrained optimization problem because \( T_e \) is a Euclidean transformation. The constraints here are

\[
1 - \sum_v e_{j,1v}^2 = 0 \quad \text{and} \quad 1 - \sum_v e_{j,2v}^2 = 0 \quad \text{and} \quad 1 - \sum_v e_{j,3v}^2 = 0
\]

\[
\sum_v e_{j,1v}e_{j,2v} = 0 \quad \text{and} \quad \sum_v e_{j,1v}e_{j,3v} = 0 \quad \text{and} \quad 1 - \sum_v e_{j,2v}e_{j,3v} = 0
\]
Optimizing

- Chuck in constrained optimizer and run
  - doesn’t work well - need a start point
Write $C_j^T$ for the $j$’th row of the camera matrix, and $S_i = [s_{x,i}, s_{y,i}, s_{z,i}, 1]^T$ for homogeneous coordinates representing the $i$’th point in 3D. Then, assuming no errors in measurement, we have

$$\hat{t}_{x,i} = \frac{C_1^T S_i}{C_3^T S_i} \text{ and } \hat{t}_{y,i} = \frac{C_2^T S_i}{C_3^T S_i},$$

which we can rewrite as

$$C_3^T S_i \hat{t}_{x,i} - C_1^T S_i = 0 \text{ and } C_3^T S_i \hat{t}_{y,i} - C_2^T S_i = 0.$$ 

2 Homogeneous equations per point - so with enough, can solve for $C$
RECALL: The camera matrix - II

- Turn previous expression into HC’s
  - HC’s for 3D point are \((X,Y,Z,T)\)
  - HC’s for point in image are \((U,V,W)\)

\[
\begin{pmatrix}
U \\
V \\
W
\end{pmatrix} = C
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
X \\
Y \\
Z \\
T
\end{pmatrix}
\]

Transforms points from object coordinates into world coordinates most likely a rotation and translation

Transforms camera coordinates (f is hidden in here)
Camera matrix and focal point

• Trick:
  • camera matrix maps focal point to $(0, 0, 0)$
  • which means given a camera matrix, it is easy to extract focal point
  • which means that we can extract translation from camera matrix

Remember this: Given a $3 \times 4$ camera matrix $\mathbf{P}$, the homogeneous coordinates of the focal point of that camera are given by $\mathbf{X}$, where $\mathbf{PX} = [0, 0, 0]^T$
Camera matrix and focal point

- Recall that extrinsic transformation is a rotation+translation

Remember this: Assume camera matrix $\mathcal{P}$ has null space $\lambda \mathbf{u} = \lambda \begin{bmatrix} f^T, 1 \end{bmatrix}^T$. Then we must have $\mathcal{T}_e \mathbf{u} = [0, 0, 0, 1]^T$, so we must have

$$
\mathcal{T}_e = \begin{bmatrix}
\mathcal{R} & -\mathcal{R}f \\
0^T & 1
\end{bmatrix}
$$
This means that, if we know $\mathcal{R}$, we can recover the translation from the focal point. We must now recover the intrinsic transformation and $\mathcal{R}$ from what we know.

$$\lambda \mathcal{P} = \mathcal{T}_i \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathcal{R} \\ -\mathcal{R}f \\ 0^T \\ 1 \end{bmatrix} = \begin{bmatrix} \mathcal{T}_i \mathcal{R} \\ -\mathcal{T}_i \mathcal{R}f \end{bmatrix}$$

We do not know $\lambda$, but we do know $\mathcal{P}$. Now write $\mathcal{P}_l$ for the left $3 \times 3$ block of $\mathcal{P}$, and recall that $\mathcal{T}_i$ is upper triangular and $\mathcal{R}$ orthonormal. The first question is the sign of $\lambda$. We expect $\text{Det}(\mathcal{R}) = 1$, and $\text{Det}(\mathcal{T}_i) > 0$, so $\text{Det}(\mathcal{P}_l)$ should be positive. This yields the sign of $\lambda$—choose a sign $s \in \{-1, 1\}$ so that $\text{Det}(s\mathcal{P}_l)$ is positive.

We can now factor $s\mathcal{P}_l$ into an upper triangular matrix $\mathcal{T}$ and an orthonormal matrix $\mathcal{Q}$. This is an RQ factorization (Section 25.2). Recall we could not distinguish between scaling caused by the focal length and scaling caused by pixel scale, so that

$$\mathcal{T}_i = \begin{bmatrix} as & k & c_x \\ 0 & s & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

In turn, we have $\lambda = s(1/t_{33})$, $c_y = (t_{23}/t_{33})$, $s = (t_{22}/t_{33})$, $c_x = (t_{13}/t_{33})$, $k = (t_{12}/t_{33})$, and $a = (t_{11}/t_{22})$. 