CS-565 Computer Vision

Nazar Khan

Department of Computer Science
University of the Punjab

18. Camera Calibration

Camera Matrix

- A 3D point in homogeneous world coordinates $(X_w, Y_w, Z_w, 1)^T$ is mapped to a 2D image point with homogeneous pixel coordinates $(u, v, w)^T$ as

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \underbrace{\begin{bmatrix} f_x & -f_x \cot\theta & u_0 \\ 0 & f_y / \sin\theta & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{intrinsic}} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\text{projection}} \underbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\text{extrinsic}} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}}_{\text{full projection matrix}} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

- 12 parameters in total but with 1 free scaling parameter. So 11 degrees of freedom: 6 extrinsic plus 5 intrinsic.

Camera Matrix

- The projection equation

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \underbrace{\begin{bmatrix} f_x & -f_x \cot\theta & u_0 \\ 0 & f_y / \sin\theta & v_0 \\ 0 & 0 & 1 \end{bmatrix}}_{K} \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\text{projection}} \underbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\begin{bmatrix} R & \mathbf{t} \end{bmatrix}} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

can be simply written as

$$\begin{aligned} \tilde{\mathbf{m}} &= K \begin{bmatrix} R & \mathbf{t} \end{bmatrix} \tilde{\mathbf{M}} \\ &= K \left(R\mathbf{M} + \mathbf{t} \right) \end{aligned}$$

- In other words

$$P = K \begin{bmatrix} R & \mathbf{t} \end{bmatrix}$$

Factorization of the Camera Matrix

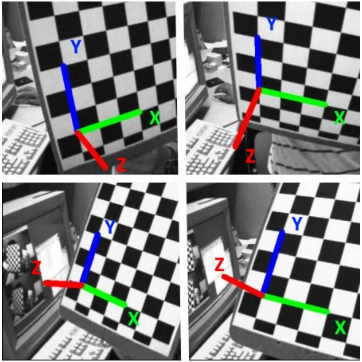
- Since the camera center $\mathbf{C}$ is the null-vector of P,

$$
\begin{aligned}
& P\tilde{\mathbf{C}} = \mathbf{0} \\
\implies & K \begin{bmatrix} R & \mathbf{t} \end{bmatrix} \mathbf{C} = \mathbf{0} \\
\implies & KR\mathbf{C} + K\mathbf{t} = \mathbf{0} \\
\implies & \mathbf{C} = -R^T\mathbf{t}
\end{aligned}
$$

$R^{-1} = R^T$ since rotation matrices are orthogonal.

Camera Calibration Using 2D Checkerboard

- Camera calibration is the process of finding the 12 values of P.
- We can use a checkerboard of known size and structure.
- Fix one corner as origin of world coordinate system.
- Then all points on the checkerboard lie in the XY-plane ($Z = 0$).
- Z-axis is orthogonal to the checkerboard.



Checkerboard-to-Image Plane Homography

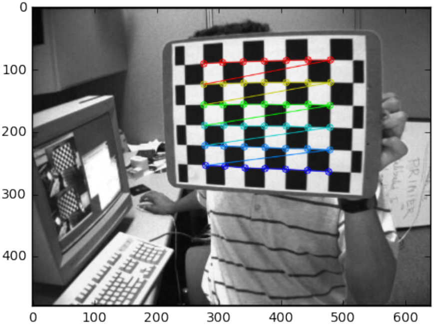
- For any point $\tilde{\mathbf{M}}$ on the checkerboard, its image $\tilde{\mathbf{m}}$ is obtained as

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} K & \mathbf{0}_{3\times 1} \end{bmatrix} \begin{bmatrix} R & \mathbf{t} \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z = 0 \\ 1 \end{bmatrix} = \underbrace{K \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix}}_{\text{Homography } H} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

- That is $\tilde{\mathbf{m}}_i = H\tilde{\mathbf{M}}_i$.
- Homography for checkerboard image j can be computed via DLT using at least 4 correspondences $\tilde{\mathbf{m}}_i \longleftrightarrow \tilde{\mathbf{M}}_i$.

Checkerboard-to-Image Plane Homography

- The $\mathbf{m}_i$ can be found using corner detection.



- The corresponding $\mathbf{M}_i$ can be computed since the size and structure of the checkerboard are known (squares of measured dimensions).

Checkerboard-to-Image Plane Homography

- For the estimated homography $H = \begin{bmatrix}\mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3\end{bmatrix}$, we can write

$$\begin{bmatrix}\mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3\end{bmatrix} = K\begin{bmatrix}\mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t}\end{bmatrix}$$

- So $\mathbf{r}_i \equiv K^{-1}\mathbf{h}_i$.
- Since $\mathbf{r}_1$ and $\mathbf{r}_2$ are columns of a rotation matrix

$$\begin{aligned}
\mathbf{r}_1^T\mathbf{r}_2 = 0 &\implies \mathbf{h}_1^T K^{-T}K^{-1}\mathbf{h}_2 = 0\\
\mathbf{r}_1^T\mathbf{r}_1 = 1 &\implies \mathbf{h}_1^T K^{-T}K^{-1}\mathbf{h}_1 = 1\\
\mathbf{r}_2^T\mathbf{r}_2 = 1 &\implies \mathbf{h}_2^T K^{-T}K^{-1}\mathbf{h}_2 = 1
\end{aligned}$$

- The term $K^{-T}K^{-1}$ is a symmetric and positive definite matrix that we denote by B.

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13}\\ b_{12} & b_{22} & b_{23}\\ b_{13} & b_{23} & b_{33}\end{bmatrix}$$

Intrinsic Matrix

- So for each homography (*i.e.*, checkerboard image), we have two constraints

$$\mathbf{h}_1^T B \mathbf{h}_2 = 0$$
$$\mathbf{h}_1^T B \mathbf{h}_1 - \mathbf{h}_2^T B \mathbf{h}_2 = 0$$

- For all images, these constraints can be written as a linear system $V\mathbf{b} = 0$ from which $\mathbf{b} = \begin{bmatrix} b_{11} & b_{12} & b_{22} & b_{13} & b_{23} & b_{33} \end{bmatrix}^T$ can be found via SVD.
- Matrix K can be recovered through the Cholsky decomposition of B.

$$\text{chol}(B) = LL^T$$

where L is a lower-triangular matrix.
- This means that $K = L^{-T}$.

Cholsky Decomposition

- Cholsky decomposition of a 3×3 matrix B can be computed *iteratively* as

$$L = \begin{bmatrix} \sqrt{B_{11}} & 0 & 0 \\ B_{21}/L_{11} & \sqrt{B_{22} - L_{21}^2} & 0 \\ B_{31}/L_{11} & \left(B_{32} - L_{31}L_{21}\right)/L_{22} & \sqrt{B_{33} - L_{31}^2 - L_{32}^2} \end{bmatrix}$$

- In terms of formulae for each term

$$L_{j,j} = (\pm)\sqrt{B_{j,j} - \sum_{k=1}^{j-1} L_{j,k}^2},$$

$$L_{i,j} = \frac{1}{L_{j,j}}\left(B_{i,j} - \sum_{k=1}^{j-1} L_{i,k}L_{j,k}\right) \quad \text{for } i > j.$$

Extrinsic Matrix

- Once K is found, the extrinsic parameters *for a given image of the calibration grid* can be found easily.

$$\mathbf{r}_1 \equiv K^{-1}\mathbf{h}_1$$
$$\mathbf{r}_2 \equiv K^{-1}\mathbf{h}_2$$
$$\mathbf{t} \equiv K^{-1}\mathbf{h}_3$$

  and after normalising $\mathbf{r}_1$ and $\mathbf{r}_2$, we can find the third column of R as $\mathbf{r}_3 \equiv \mathbf{r}_1 \times \mathbf{r}_2$.
- Notice that K is computed from B which is computed using the constraint $\mathbf{h}_1^T B\mathbf{h}_1 = \mathbf{h}_2^T B\mathbf{h}_2$.
- This means that $\mathbf{r}_1$ and $\mathbf{r}_2$ will have the same magnitude m which represents the homogeneous scale of matrix K.

Extrinsic Matrix

- So the actual rotation and translation vectors are

$$
\begin{aligned}
\mathbf{r}_1 &\leftarrow \frac{1}{m}\mathbf{r}_1 \\
\mathbf{r}_2 &\leftarrow \frac{1}{m}\mathbf{r}_2 \\
\mathbf{r}_3 &\leftarrow \mathbf{r}_1 \times \mathbf{r}_2 \\
\mathbf{t} &\leftarrow \frac{1}{m}\mathbf{t}
\end{aligned}
$$

Augmented Reality

- A calibrated camera provides an interface between the $2D$ and $3D$ worlds.
- We can exploit it to *augment 2D images* with a *false 3D reality*.
- This is known as *augmented reality*.

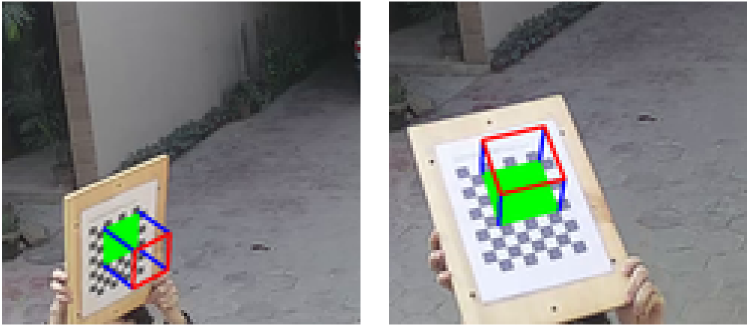


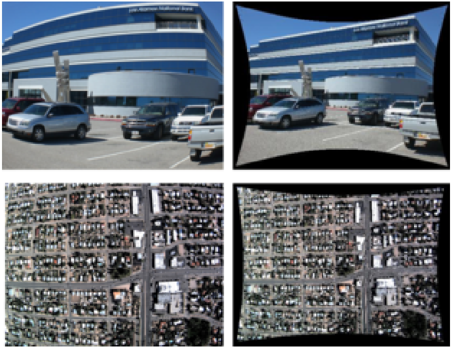
Figure: A $3D$ cube augmented onto the plane of the calibration checkerboard in two different orientations. Author: N. Khan (2018)

Summary

- We have seen how to calibrate a camera when represented as a 3×4 linear transformation.
- Multiple images of a checkerboard of known structure can be used to find $2D - 3D$ correspondences.
- Linear algebra modules can be used to recover camera parameters.
  - SVD
  - Cholsky decomposition
  - Cross-product

Summary

- Due to the presence of a lens, a real camera transformation has *additional non-linear transformations*.
  - Radial distortion



- Non-linear camera calibration incorporates such non-linear effects as well.