CSE 252B: Computer Vision II

Lecturer: Serge Belongie
Scribe: Smriti Yamini

LECTURE 7
Uncalibrated Epipolar Geometry

7.1. Uncalibrated camera or distorted space?

So far we have been assuming $K = I$, which corresponds to the calibrated case. What happens when $K \neq I$? In 1992, Faugeras asked the question "What can be seen in 3-D with an uncalibrated stereo rig?" That is, what can we determine about the 3-D structure of the scene and the pose of the camera in the uncalibrated case? Hartley also posed the same question. The answer to the question is this: you can recover the structure of the scene in 3D (and the camera pose) up to a projective transformation.

There are in fact two equivalent ways to look at the problem setup:

- an uncalibrated camera moving in rectified space, or
- a calibrated camera moving in distorted space.

An uncalibrated camera with calibration matrix K, viewing points in a calibrated (Euclidean) world and moving with parameters $(R, \boldsymbol{T})$ is equivalent to a calibrated camera viewing points in distorted space moving with parameters $(KRK^{-1}, K\boldsymbol{T})$. This is illustrated in Figure 1. This distorted

[1]Department of Computer Science and Engineering, University of California, San Diego.

April 19, 2004

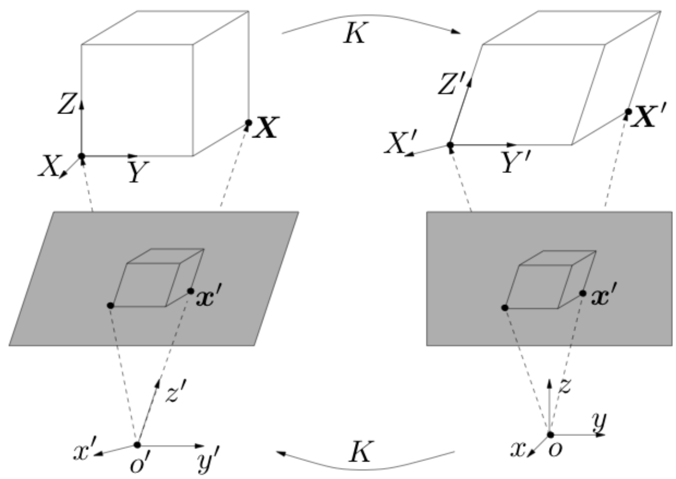

Figure 1. MaSKS Figure 6.4.

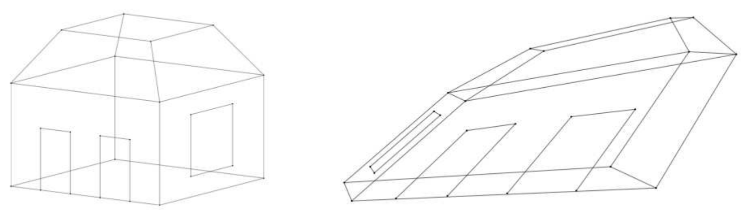

Figure 2. The Euclidean (left) and projective (right) structure of a house.

space is governed by the inner product

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle_S = \boldsymbol{u}^\top S \boldsymbol{v}, \quad \text{with} \quad S = K^{-T} K^{-1} = (KK^\top)^{-1}$$

We call S the *metric* of the space. In the Euclidean case, $S = I$ and $\langle \boldsymbol{u}, \boldsymbol{v} \rangle_S = \boldsymbol{u}^\top \boldsymbol{v}$. Figure 2 shows the difference between the Euclidean structure and the projective structure of a 3D object.

Recall the structure of the matrix K:

$$K = \begin{bmatrix} f s_x & s_\theta & o_x \\ 0 & f s_y & o_y \\ 0 & 0 & 1 \end{bmatrix}$$

Its purpose is to map metric coordinates (unit of metres) into image coordinates (unit of pixels). We use a prime to denote pixel coordinates:

$$\boldsymbol{x} = K^{-1} \boldsymbol{x}'$$

Applying the rotation matrix and the translation vector to some point $\boldsymbol{X}_0$ in Euclidean space, we get

$$\boldsymbol{X} = R\boldsymbol{X}_0 + \boldsymbol{T}$$

In the uncalibrated camera frame we have

$$K\boldsymbol{X} = KR\boldsymbol{X}_0 + K\boldsymbol{T} \quad \text{or} \quad \boldsymbol{X}' = KRK^{-1}\boldsymbol{X}'_0 + \boldsymbol{T}'$$

where $\boldsymbol{X}' = K\boldsymbol{X}$, $\boldsymbol{T}' = K\boldsymbol{T}$ and $\boldsymbol{X}'_0 = K\boldsymbol{X}_0$.

Applying the image formation model using homogeneous coordinates, we get

$$\begin{aligned}
\lambda\boldsymbol{x} &= K\Pi_0 g\boldsymbol{X}_0 \\
&= K\begin{bmatrix} I & \mathbf{0} \end{bmatrix}\begin{bmatrix} R & \boldsymbol{T} \\ \mathbf{0}^\top & 1 \end{bmatrix}\boldsymbol{X}_0 \\
&= KR\boldsymbol{X}_0 + K\boldsymbol{T} \\
&= KRK^{-1}\boldsymbol{X}'_0 + K\boldsymbol{T} \\
&= \Pi_0 g'\boldsymbol{X}'_0
\end{aligned}$$

where $g' = \begin{bmatrix} KRK^{-1} & \boldsymbol{T}' \\ \mathbf{0}^\top & 1 \end{bmatrix} \in \mathbb{R}^{4\times 4}$ is the distorted rigid transformation.

Summarizing, an uncalibrated camera moving in the calibrated space ($\lambda\boldsymbol{x}' = K\Pi_0 g\boldsymbol{X}_0$) is equivalent to a calibrated camera moving in a distorted space ($\lambda\boldsymbol{x}' = \Pi_0 g'\boldsymbol{X}'_0$).

7.2. Epipolar Constraint

Recall the form of the epipolar constraint in the calibrated case:

$$\boldsymbol{x}_2^\top E\boldsymbol{x}_1 = 0.$$

By direct substitution of the relation between metric and pixel coordinates $\boldsymbol{x} = K^{-1}\boldsymbol{x}'$, we get

$$\boldsymbol{x}_2'^\top K^{-\top}\widehat{T}RK^{-1}\boldsymbol{x}'_1 = 0$$

The matrix in the middle is known as the Fundamental matrix,

$$F = K^{-\top}\widehat{T}RK^{-1} = K^{-\top}EK^{-1}$$

Note that F reduces to the essential matrix when $K = I$.

7.2.1. Coplanarity constraint

We can also examine uncalibrated epipolar geometry in terms of the coplanarity constraint. The three vectors $\boldsymbol{x}'_2, \boldsymbol{T}' = K\boldsymbol{T}$ and $KR\boldsymbol{x}_1 = KRK^{-1}\boldsymbol{x}'_1$ in Figure 3 are coplanar. Hence their scalar triple product is 0:

$$\boldsymbol{x}_2'^{\top}\widehat{T}'KRK^{-1}\boldsymbol{x}'_1 = 0$$

The matrix in the middle looks slightly different than the previous expression for F, but we will see shortly that they are equivalent.

7.2.2. Algebraic derivation

Start with

$$\lambda_2\boldsymbol{x}_2 = R\lambda_1\boldsymbol{x}_1 + \boldsymbol{T}$$

with $\lambda\boldsymbol{x} = \boldsymbol{X}$. Multiply both sides by K,

$$\lambda_2K\boldsymbol{x}_2 = KR\lambda_1\boldsymbol{x}_1 + K\boldsymbol{T}$$

or

$$\lambda_2\boldsymbol{x}'_2 = KRK^{-1}\lambda_1\boldsymbol{x}'_1 + \boldsymbol{T}'$$

Taking the dot product with $\boldsymbol{T}' \times \boldsymbol{x}'_2 = \widehat{T}'\boldsymbol{x}'_2$ and dropping scalar factors, we get

$$\boldsymbol{x}_2'^{\top}\widehat{T}'KRK^{-1}\boldsymbol{x}'_1 = 0$$

since the vector $\widehat{T}'\boldsymbol{x}'_2$ is orthogonal to both $\boldsymbol{T}'$ and $\boldsymbol{x}'_2$.

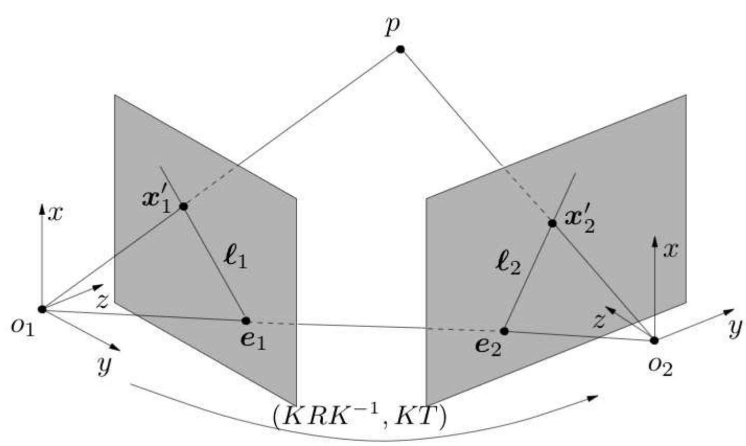

Figure 3. Uncalibrated epipolar geometry.

From MaSKS Lemma 5.4, we have the identity $K^{-T}\widehat{T}K^{-1} = \widehat{KT}$ when $\det(K) = +1$. We can therefore write

$$\begin{aligned} F &= K^{-T}\widehat{T}RK^{-1} \quad \text{(uncalibrated camera in calibrated space)} \\ &= K^{-T}\widehat{T}K^{-1}KRK^{-1} \\ &= \widehat{T'}KRK^{-1} \quad \text{(calibrated camera in uncalibrated space)} \end{aligned}$$

7.2.3. Epipolar Lines and Epipoles

The bilinear relationship

$$\boldsymbol{x}_2'^{\top}F\boldsymbol{x}_1' = 0$$

transfers a point in view 1 to a line in view 2. Equivalently, we may write

$$\boldsymbol{x}_2'^{\top}F\boldsymbol{x}_1' = \boldsymbol{x}_2'^{\top}\boldsymbol{l}_2 = 0$$

where $\boldsymbol{l}_2$ is the epipolar line in view 2. Similarly for $\boldsymbol{l}_1$, we have

$$\boldsymbol{l}_1^{\top}\boldsymbol{x}_1' = 0$$

Thus $\boldsymbol{l}_2 = F\boldsymbol{x}_1'$ and $\boldsymbol{l}_1 = F^{\top}\boldsymbol{x}_2'$.

As with E, we can get the epipoles from the left and right null space of F:

$$\boldsymbol{e}_2^{\top}F = \boldsymbol{0}, \quad F\boldsymbol{e}_1 = \boldsymbol{0}$$

from which it follows

$$\boldsymbol{e}_2 = K\boldsymbol{T} = \boldsymbol{T}', \quad \boldsymbol{e}_1 = KR^{\top}\boldsymbol{T}$$

7.3. Properties of F

Like the essential matrix, F has rank 2 since $\widehat{T'}$ is rank 2. The SVD of $F = U\Sigma V^{\top}$ has $\Sigma = \text{diag}\{\sigma_1, \sigma_2, 0\}$ with $\sigma_1 \geq \sigma_2$. (In the case of E, we had $\sigma_1 = \sigma_2$). This implies that *any* rank 2 matrix can be a fundamental matrix for some stereo rig.

We can apply the 8-point algorithm to estimate F using a design matrix $\chi \in \mathbb{R}^{n\times 9}$ with rows (carrier vectors) of the form

$$\boldsymbol{a} = \boldsymbol{x}_1' \otimes \boldsymbol{x}_2'$$

$$\boldsymbol{a} = [x_1'x_2', x_1'y_2', x_1', y_1'x_2', y_1'y_2', y_1', x_2', y_2', 1]^{\top} \in \mathbb{R}^9 \tag{7.1}$$

All variables are primed, as they are in pixel coordinates. If we assume the pixel coordinates are on the order of 10^2, then we encounter a practical problem: the entries in $\boldsymbol{a}$ range from 10^0 to 10^4, which makes χ ill-conditioned. Hartley proposed a simple means of overcoming this problem.

7.4. Hartley Normalization

In Hartley normalization we rescale the data using two matrices $H_i, i = 1, 2$, so as to produce coordinates that make the design matrix well-conditioned. We choose H_i such that the normalized coordinates $\widetilde{\boldsymbol{x}}_i = H_i \boldsymbol{x}'_i$ have zero mean and unit variance:

$$H_i = \begin{bmatrix} 1/\sigma_{x_i} & 0 & -\mu_{x_i}/\sigma_{x_i} \\ 0 & 1/\sigma_{y_i} & -\mu_{y_i}/\sigma_{y_i} \\ 0 & 0 & 1 \end{bmatrix}$$

In this expression, means and variances are given by:

$$\mu_{x_i} = \frac{1}{n}\sum_{j=1}^{n} x_i^j \qquad \sigma_{x_i}^2 = \frac{1}{n}\sum_{j=1}^{n}(x_i^j)^2 - \mu_{x_i}^2$$

$$\mu_{y_i} = \frac{1}{n}\sum_{j=1}^{n} y_i^j \qquad \sigma_{y_i}^2 = \frac{1}{n}\sum_{j=1}^{n}(y_i^j)^2 - \mu_{y_i}^2$$

Intuitively, H_i can be thought of as a guess at the calibration matrix, placing the centroid of the coordinates at the image center, assuming zero skew, and using the x and y variance to set the pixel aspect ratio.

After this transformation, we run the 8-point algorithm on $\widetilde{\boldsymbol{x}}_i$, $i = 1, 2$, to obtain the fundamental matrix $\widetilde{F}$ for the normalized data. Finally we obtain F by observing

$$\boldsymbol{x}_2'^{\top} F \boldsymbol{x}_1' = \widetilde{\boldsymbol{x}}_2^{\top} \underbrace{H_2^{-\top} F H_1^{-1}}_{\widetilde{F}} \widetilde{\boldsymbol{x}}_1 = 0$$

so $F = H_2^{\top} \widetilde{F} H_1$