CSE 252B: Computer Vision II

Lecturer: Serge Belongie

Scribes: Paul Hammon and Evan Schumacher

LECTURE 8
Projective Reconstruction

8.1. The Fundamental Matrix vs. the Essential Matrix

The fundamental matrix, F, is an extension of the the essential matrix, E, to the case of an uncalibrated camera. The epipolar constraint that gives rise to F is also an extension of the epipolar constraint that that leads to E. The difference is that in the uncalibrated case we must take in to account the intrinsic camera parameters K which result in the "distorted space." As seen in the last lecture, F is given by:

$$(8.1) \qquad F = K^{-T}\widehat{T}RK^{-1} = \widehat{T'}KRK^{-1}$$

where $\boldsymbol{T'} = K\boldsymbol{T}$. Note that if $K = I$ then $F = E$.

We can use the 8-point algorithm to get F as we did with E, but unfortunately we cannot simply extract $R, K, \boldsymbol{T}$ from F as we do with E. What we would like is to be able to get a projection matrix, Π, from F such that:

$$F = \widehat{T'}KRK^{-1} \mapsto \Pi = [KRK^{-1}, K\boldsymbol{T}]$$

The reason this is not possible can be seen by comparing the degrees of freedom (DOF) in F with the degrees of freedom required to represent $R, K, \boldsymbol{T}$.

---

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

To express $R, K, \boldsymbol{T}$ we need 10 DOF, 2 for translation, 5 for intrinsic parameters, and 3 for rotation. F has nine values specified up to a scale factor, so it has at most only 8 DOF (not taking into account its special structure), 2 DOF short of being able to provide $R, K, \boldsymbol{T}$.

8.1.1. Ambiguities and constraints in image formation

When given an equation involving matrix products it is possible for ambiguities to crop up because the product can be interspersed with a multiplication of a matrix and its inverse. For example,

$$M = BC = (BH^{-1})(HC) = B'C'$$

In this case (B', C') and (B, C) cannot be distinguished using the measurement M. This is an issue for us since the equation that relates image and real world coordinates

$$(8.2) \qquad\qquad \lambda \boldsymbol{x}' = K \Pi_0 g \boldsymbol{X}$$

could hide three such ambiguities, as evidenced by the following:

$$(8.3) \qquad\qquad \lambda \boldsymbol{x}' = K \Pi_0 g \boldsymbol{X} = K R_0^{-1} R_0 \Pi_0 H^{-1} H g g_w^{-1} g_w \boldsymbol{X}$$

In this case, R_0 and g_w can be fixed by choosing Euclidean coordinate frames. However, H causes ambiguity to the projection matrix Π, leading to a distortion of the world coordinate frame that $\boldsymbol{X}$ lives in, which distorts our reconstruction. Fixing this distortion is known as "rectification," and is equivalent to identifying the metric of the space, or "calibrating" the space. Being able to find H corresponds to discovering the metric of the space as in:

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle_S = \boldsymbol{u}^\top S \boldsymbol{v}.$$

This is analogous to calibrating the camera to provide a rectified image.

8.2. Stratified Reconstruction

There are different levels of reconstruction of a scene possible as shown in figure 1. Without knowing K we can only get the projective reconstruction, but this can be upgraded to affine (parallelism preserved) and Euclidean (parallelism and orthogonality preserved) reconstructions. Today we are answering the question of how do we get the projective reconstruction given a set of corresponding points in two views and the camera projection matrices Π_{ip}. Note: The i in the subscript represents which camera view we are referencing, and the p stands for projective. Without loss of generality, and to simplify the math, we use the following convention:
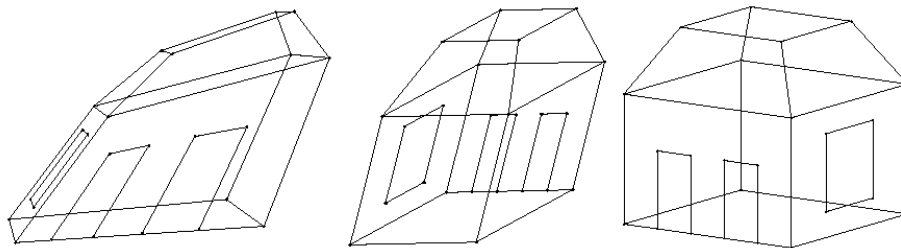
$$\Pi_{1p} = [I, \boldsymbol{0}]$$

Figure 1. Three cases of reconstruction (from left to right): Projective, Affine, and Euclidean, from *MaSKS fig 6.10*

8.2.1. Projective Reconstruction

Given a set of point correspondences between two views $\{(\boldsymbol{x}_1', \boldsymbol{x}_2')\}$ we can get the Projective Structure $\boldsymbol{X}_p$ by using the following technique: first get F from the point correspondences, then use F to get Π_{2p} (remember $\Pi_{1p} = [I, 0]$), and finally triangulate to get $\boldsymbol{X}_p$. One issue that must be dealt with is that, as stated earlier, we cannot uniquely get $K, R, \boldsymbol{T}$ from F. However, the following theorem guarantees that all instances of Π_{2p}'s that come from a given F are related to one another by a projective transform of the form $H_p \in GL(4)$.

Theorem 8.4. *(MaSKS Theorem 6.3):* (Π_{1p}, Π_{2p}) *and* $(\Pi_{1p}, \tilde{\Pi}_{2p})$ *are two pairs of projection matrices that yield the same fundamental matrix F iff there exists a nonsingular transformation matrix H_p such that $\tilde{\Pi}_{2p} \sim \Pi_{2p}H_p^{-1}$ or equivalently, $\Pi_{2p} \sim \tilde{\Pi}_{2p}H_p$.*

Proof. Let $\Pi_{2p} = [C, \boldsymbol{c}]$, $\tilde{\Pi}_{2p} = [B, \boldsymbol{b}]$, where $C, B \in \mathbb{R}^{3 \times 3}$ and $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{R}^3$. If both Π_{2p} and $\tilde{\Pi}_{2p}$ give rise to the same F, then

$$(8.5) \qquad\qquad \widehat{c}C \sim \widehat{b}B$$

Because $\boldsymbol{c}^\top \widehat{c} = \boldsymbol{0}^\top$ and $\boldsymbol{b}^\top \widehat{b} = \boldsymbol{0}^\top$, $\boldsymbol{c}$ and $\boldsymbol{b}$ span the left nullspaces of the lefthand and righthand sides, respectively, this implies that $\boldsymbol{c} \sim \boldsymbol{b}$. Thus,

$$(8.6) \qquad\qquad \widehat{c}C = k\widehat{c}B \quad \text{or} \quad \widehat{b}C = k'\widehat{b}B$$

If $\widehat{c}$ were rank 3 and $k = 1$, then this would imply that $C = B$. It has been shown previously, however, that $\widehat{c}$ is rank 2, and for a rank-deficient matrices $\widehat{c}C = \widehat{c}B \not\Rightarrow C = B$.

Since $\widehat{c}$ has $\boldsymbol{c}$ in its nullspace, and since $\boldsymbol{b} \sim \boldsymbol{c}$, one can show that

$$(8.7) \qquad\qquad C \sim B + \boldsymbol{b}\boldsymbol{v}^\top \quad \text{where} \quad \boldsymbol{v} \in \mathbb{R}^3$$

This can be re-written as

(8.8)
$$\underbrace{[C, \boldsymbol{c}]}_{\Pi_{2p}} \sim \underbrace{[B, \boldsymbol{b}]}_{\tilde{\Pi}_{2p}} \underbrace{\begin{bmatrix} I & \boldsymbol{0} \\ \boldsymbol{v}^\top & v_4 \end{bmatrix}}_{H_p}$$

Thus, we have shown that $\Pi_{2p} \sim \tilde{\Pi}_{2p} H_p$ for H_p of the form in the above equation. $\square$

8.2.2. Canonical Choices of the Projection Matrices

We want to fix the choice of H_p such that F only maps to one pair (Π_{1p}, Π_{2p}) with no parametrization in terms of $\boldsymbol{v}$ or v_4. The canonical choice for these two projection matrices is stated below and then expanded upon later.

(8.9)
$$F \mapsto \Pi_{1p} = [I, \boldsymbol{0}], \quad \Pi_{2p} = \left[\widehat{T'}^\top F, \boldsymbol{T'}\right]$$

where $\boldsymbol{T'} = K\boldsymbol{T}$ and $\|\boldsymbol{T'}\| = 1$. This choice of projection matrices depends only on F because $\boldsymbol{T'}^\top$ is related to F by $\boldsymbol{T'}^\top F = \boldsymbol{0}$. (Recall $\boldsymbol{T'} \sim \boldsymbol{e}_2$, the epipole in the 2nd view.)

Now let's examine whether this projection will give back F. We know that F is constructed from hatting the rightmost column of Π_{2p} and premultiplying it by the leading 3×3 block as follows:

$$\left[\widehat{T'}^\top F, \boldsymbol{T'}\right] \to \widehat{T'}\widehat{T'}^\top F$$

Note that $\widehat{T'}\widehat{T'}^\top = I - T'T'^\top$ when $\|\boldsymbol{T'}\| = 1$ (see *MaSKS* exercise 5.3). This implies that:

$$\widehat{T'}\widehat{T'}^\top F = \left(I - T'T'^\top\right) F = F - T'T'^\top F = F - T'\boldsymbol{0} = F$$

8.2.3. Determination of Canonical Projection Matrices

Now we will examine how to do the projective reconstruction, which is the process from which the 3D coordinates, $\boldsymbol{X}_p$, are recovered from pairs of corresponding uncalibrated image points, $\boldsymbol{x}_1'$ and $\boldsymbol{x}_2'$.

Using Π_{1p} and Π_{2p} as described previously, we have:

(8.10)
$$\lambda_1 \boldsymbol{x}_1' = \Pi_{1p} \boldsymbol{X}_p = [I, \boldsymbol{0}] \boldsymbol{X}_p$$

(8.11)
$$\lambda_2 \boldsymbol{x}_2' = \Pi_{2p} \boldsymbol{X}_p = \left[\widehat{T'}^\top F, \boldsymbol{T'}\right] \boldsymbol{X}_p$$

Premultiplying by $\widehat{\boldsymbol{x}_1'}$ and $\widehat{\boldsymbol{x}_2'}$ in the previous equations yields $\widehat{\boldsymbol{x}_i'}\Pi_{ip}\boldsymbol{X}_p = \boldsymbol{0}$ for $i = 1, 2$. For two corresponding points $\boldsymbol{x}_1' = [x_1, y_1, 1]^\top$ and $\boldsymbol{x}_2' =$
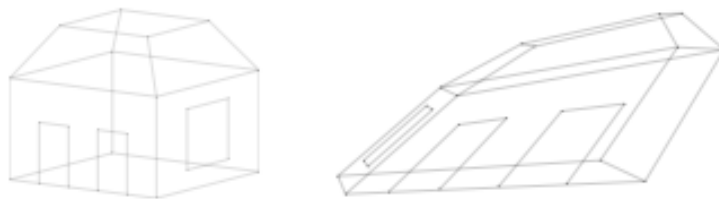
Figure 2. An illustration of how the true Euclidean structure might be skewed in the process of recovering its projective structure in the uncalibrated case. (Figure from *MaSKS* p.191.)

$[x_2, y_2, 1]^\top$, where Π_{ip} is written as a set of three row vectors $[\boldsymbol{\pi}_i^1, \boldsymbol{\pi}_i^2, \boldsymbol{\pi}_i^3]^\top$, these constraints result in four equations:

$$(8.12) \qquad \left(x_1 \boldsymbol{\pi}_1^{3\top}\right) \boldsymbol{X}_p = \boldsymbol{\pi}_1^{1\top} \boldsymbol{X}_p$$

$$(8.13) \qquad \left(y_1 \boldsymbol{\pi}_1^{3\top}\right) \boldsymbol{X}_p = \boldsymbol{\pi}_1^{2\top} \boldsymbol{X}_p$$

$$(8.14) \qquad \left(x_2 \boldsymbol{\pi}_2^{3\top}\right) \boldsymbol{X}_p = \boldsymbol{\pi}_2^{1\top} \boldsymbol{X}_p$$

$$(8.15) \qquad \left(y_2 \boldsymbol{\pi}_2^{3\top}\right) \boldsymbol{X}_p = \boldsymbol{\pi}_2^{2\top} \boldsymbol{X}_p$$

Which can be re-written as a homogeneous linear system $M\boldsymbol{X}_p = \boldsymbol{0}$, $M \in \mathbb{R}^{4 \times 4}$:

$$(8.16) \qquad \begin{bmatrix} x_1 \boldsymbol{\pi}_1^{3\top} - \boldsymbol{\pi}_1^{1\top} \\ y_1 \boldsymbol{\pi}_1^{3\top} - \boldsymbol{\pi}_1^{2\top} \\ x_2 \boldsymbol{\pi}_2^{3\top} - \boldsymbol{\pi}_2^{1\top} \\ y_2 \boldsymbol{\pi}_2^{3\top} - \boldsymbol{\pi}_2^{2\top} \end{bmatrix} \boldsymbol{X}_p = \boldsymbol{0}$$

We can solve for $\boldsymbol{X}_p$ using the SVD as described in previous lectures.[1] An example of what the projective structure of a scene might look like can be seen in Figure 2. Note, however, that this only works exactly for the case where there is no noise. In particular, the smallest singular value of M will only be zero in the exact case. In the case of noisy data, this initialization can serve as a starting point for a nonlinear optimization method. See MaSKS Example 6.4 on p. 190 for a numerical example of this process.

---

[1]Recall that we made a big deal about the twisted pair ambiguity in the case of calibrated reconstruction, but in the uncalibrated case we don't. Why is that?