EECS 106B: Uncalibrated Geometry and Stratification

Chris Mitchell and Beau Kuhn

7 April 2020

1 Introduction

1.1 Overview

- Calibration with a rig
- Uncalibrated Epipolar Geometry
- Ambiguities in Image Formation
- Stratified Reconstruction
- Autocalibration with partial scene knowledge

2 Uncalibrated Camera

If a camera is fully calibrated, we can find the transform between real world coordinates $\mathbf{X} = [X, Y, Z, 1]^T$ and the image plane coordinates $\mathbf{x} = [x, y, 1]^T$ using only the extrinsic camera parameters R and T, the rotation matrix and translation vector, respectively. These align the camera origin with the projection of the image on the image plane, such that the following holds:

$$\lambda \mathbf{x} = [R, T]\mathbf{X}$$

Picking λ such that the last element of the vector $\mathbf{x} = 1$ creates this **perspective projection** relationship.

An uncalibrated camera requires an additional transformation $\mathbf{x'} = K\mathbf{x}, K \in R^{3x3}$. Now the projection matrix from the real world to the image pixels is:

$$\lambda \mathbf{x'} = \Pi \mathbf{X} = [KR, KT]\mathbf{X}$$

If K is known, we essentially have the calibrated case. However, if it is unknown, we need a way to recover or estimate it for image analysis based on what you know:

- Calibration with complete scene knowledge (a rig) – estimate
- Uncalibrated reconstruction despite the lack of knowledge of K
- Autocalibration (recover from uncalibrated images)
- Use partial knowledge, e.g. parallel lines, vanishing points, planar motion, constant intrinsic

2.1 Uncalibrated Epipolar Geometry

Epipolar geometry is the geometry of stereo vision, when we have images of the same point, but taken from different locations, whether from the same or different camera. For our purposes, let's assume that both images have the same calibration matrix K. Our extrinsic values R and T represent the transformation between the two camera locations and angles. Combining this with the above equations, we generate the following relationship between a point in frame 1 $\mathbf{x}_1$ and its matching point in frame 2, $\mathbf{x}_2$:

$$\lambda_2 K\mathbf{x}_2 = [KR, KT]\lambda_1\mathbf{x}_1 + KT$$
$$\lambda_2 \mathbf{x'}_2 = [KRK^{-1}, T']\lambda_1\mathbf{x'}_1$$

Combining these relationships and the image geometry, we get the **epipolar constraint**:

$$x_2'^T K^{-T}\hat{T}RK^{-1}x_1' = 0$$

The exact proof is left as an exercise to the reader. Note here that $\hat{T}$ is the vector T with the cross product hat operation applied. We call the matrix $K^{-T}\hat{T}RK^{-1}$ the fundamental matrix F. Note that defining $T' = KT$, then we also have $F = \hat{T}'KRK^{-1}$

2.2 Properties of the Fundamental Matrix

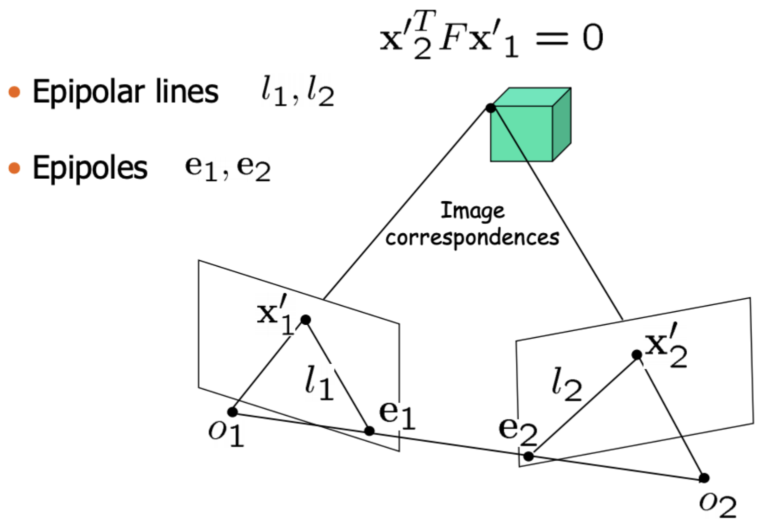

Figure 1: Epipolar Relationship

A nonzero matrix $F \in \mathrm{R}^{3x3}$ is a fundamental matrix if F has a singular value decomposition (SVD); i.e., $F = U\Sigma V^T$ with $\Sigma = diag(\sigma_1, \sigma_2, 0)$ for some $\sigma_1, \sigma_2 \in R_+$. There is little structure in the matrix F except that: $det(F) = 0$

2.3 Estimating Fundamental Matrix

We can estimate the fundamental matrix with the following optimization using n known correspondence points:

$$\min_F \sum_{j=1}^{n} \mathbf{x'}_2^{jT} F \mathbf{x}_1^j$$

We can unroll F into a vector F^S and create a vector a_j for each point such that $a_j^T * F^S$

$$a_j = \mathbf{x}_1^j \mathbf{x}_2^{jT}$$

$$F^S = [f_1, f_4, f_7, f_2, f_5, f_8, f_3, f_6, f_9]^T$$

Now the optimization function is:

$$\min_F^S ||A * F^S||^2$$

where A is a matrix with rows a_j's.

As F has 9 values, but any scaled F satisfies the epipolar constraint, we only need 8 values to find a valid F, which we just scale to match properly. For this reason we call the algorithm for recovering F the **8 Point Algorithm**:

- Solve the least squares optimization to minimize $||AF^S||$. This would be the right most singular vector of A, or the eigenvector corresponding to the smallest eigenvalue of A^TA, both stemming from SVD.
- Reformat F^S into F and compute SVD
- Project onto essential manifold: F must be rank 2, so we must enforce this constraint. Using our decomposed SVD, take the third singular value and replace it with zero.
- Reconstruct F using the original U and V matrices but with the new singular value diagonal matrix. This will produce us an optimal estimate for F.

2.4 Physical Intuition of K

$$\boldsymbol{x}' = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = K\boldsymbol{x} = \underbrace{\begin{bmatrix} fs_x & fs_\theta & o_x \\ 0 & fs_y & o_y \\ 0 & 0 & 1 \end{bmatrix}}_{K} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Linear transformation

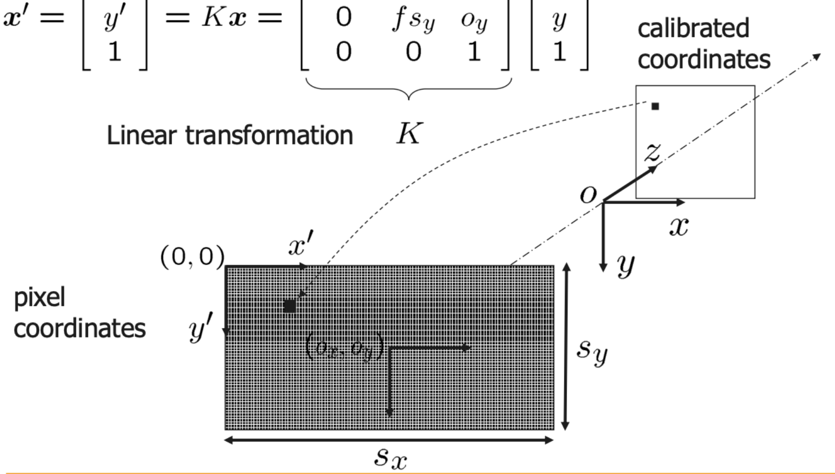

Figure 2: The physical connection to the K values

Notice that while we solved for K using 8 points, we can actually sometimes make stronger assumptions to solve for K analytically. K can be represented as well with the matrix above, where:

- f corresponds to the image focal length
- s_x and s_y correspond to pixel scaling (which may or may not be the same)
- s_θ corresponds to any skewing between the image plane and pixel plane

- o_x and o_y correspond to a translational alignment shift

We can effectively view this as a basis transformation, with calibrated space being the standard basis and uncalibrated being the real, warped image. Notice the following:

$$\mathbf{X}(t) = R(t)\mathbf{X}(t_0) + T(t)$$

$$K\mathbf{X}(t) = KR(t)\mathbf{X}(t_0) + KT(t)$$

$$\mathbf{X'}(t) = KR(t)K^{-1}\mathbf{X'}(t_0) + KT(t)$$

Therefore, we can find a relationship between the uncalibrated coordinates that actually is a conjugate of the Euclidean group:

$$G' = \{g' = \begin{pmatrix} KRK^{-1} & T' \\ \\ 0 & 1 \end{pmatrix} |T' \in R^3, R \in SO(3)\}$$

We can use this conjugate of the Euclidean group to map motion from the real world to the camera frame as well, using the same methods we used for Euclidean motion but with a transformed g matrix.

2.5 What does F tell us?

F is the epipolar geometry represented in mathematical form and can be inferred from using point matches via eight-point algorithm. Using F, we can reconstruct a projective transformation since F encodes information among two perspective views. Take note that we cannot extract motion, structure, or calibration from a single fundamental matrix F.

2.6 Decomposing the Fundamental Matrix

$$F = K^{-T}\hat{T}RK^{-1} = \hat{T'}KRK^{-1}$$

Decomposing F into a skew-symmetric matrix and nonsingular matrix:

$$F \mapsto \Pi = [R', T'] \implies F = \hat{T'}R'$$

Note: Decomposition of F is not unique:

$$x_2'\hat{T'}(\hat{T'}v^T + KRK^{-1})x_1' = 0$$
$$T' = KT$$

Unknown parameters - ambiguity

$$v = [v_1, v_2, v_3]^T \in R^3, v_4 \in R$$

Corresponding projection matrix

$$\Pi = [KRK^{-1} + T'v^T, v_4T']$$

3 Projective Reconstruction

Projective Reconstruction is the process of using cameras with differing perspectives in order to create a mathematical model that represents a particular 3D space, or scene. (Hartley R. (2014) Projective Reconstruction. In: Ikeuchi K. (eds) Computer Vision. Springer, Boston, MA).

Both Π_{ip} and $\tilde{\Pi}_{ip}$ are consistent with the epipolar geometry - give the same fundamental matrix - Given projection matrices recover projective structure - This is a linear problem and can be solved using linear least-squares: $MX_p = 0$ - Projective reconstruction - projective camera matrices and projective structure

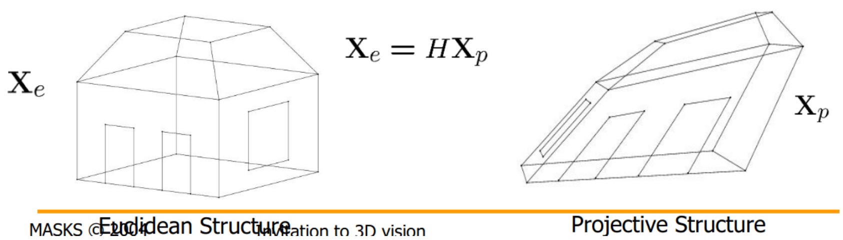

Figure 3: Summary of Epipolar Relationships

3.1 Euclidean (Homogenous) vs Projective Reconstruction

- Euclidean Reconstruction: true metric properties of objects lengths (distances), angles, parallelism are preserved, i.e. not changed under rigid body transform
- Euclidean Geometry: Properties of rigid bodies under rigid body transformations, similarity transformation.
- Projective Reconstruction: lengths, angles, parallism are NOT preserved, i.e. ARE changed under rigid body transform

  We get distorted images of objects - their distorted 3D counterparts - 3D projective reconstruction - Projective Geometry

3.2 Homogenous and Projective Coordinates

- 3D coordinates are related by: $X_c = RX_w + T$
- Homogeneous coordinates are related by the matrix transformation $[R, T; 0, 1]$, which is also a rigid body transform, with $R \in R^{3x3}$ or R^{2x2} and $T \in R^3$ or $T \in R^2$.

- Homogeneous Coordinates are vectors $(x, y, z) \in R^3$ that have an additional 1 appended to them to distinguish between vectors and points, i.e. $(x, y, z, 1) \in R^4$ and $(x, y, z, 0) \in R^4$, respectively. When in homogenous coordinates, vectors with ending with a 0 are called **ideal points** and are rays parallel in the projective space, or image plane, while **points at infinity** never intersect the image.

- Projective coordinates are the images of vectors in either R^3 or R^4 when viewed with respect to a particular projective transformation matrix:

$$(x, y, 1) \mapsto (Wx, Wy, W) \text{ and } (x, y, z, 1) \mapsto (W'x, W'y, W'z, W')$$

3.3 Vanishing Points - Points at Infinity

You can represent a 3D point as $\mathbf{X} = [X, Y, Z, 1]^T = [X_0, Y_0, Z_0, 1]^T + \lambda * [V_1, V_2, V_3, 0]^T$. If you normalize $\mathbf{X}$ by λ and take the limit as λ approaches infinity, you get $X = [V_1, V_2, V_3, 0]^T$. The vanishing point is similar but in the image plane. This is where parallel lines in the real world may converge in the image plane due to the projective nature of the transformation.

3.4 Ambiguities in the Image Formation

Note that the projective transformation is not invertible. Many different transform/point pairs exist that can produce some **x**':

$$\mathbf{x}' = \Pi\mathbf{X} = (\Pi * H) * (H^{-1}\mathbf{X}) = \tilde{\Pi}\tilde{\mathbf{X}}$$

Using the identity as projective reference frame can decompose H some into:

$$H^{-1} = \begin{pmatrix} K_1^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I & 0 \\ v^T & v_4 \end{pmatrix}$$

We can then exploit partial scene knowledge to help remove some of these unknowns. This includes vanishing points, lack of skew, and a known principal point. This also includes knowing something about types of motion, such as pure rotation, translation, planar, or rectilinear motion.

3.5 Geometric Stratification

	Camera projection	3-D structure
Euclid.	$\Pi_{1e} = [K, 0],\ \Pi_{2e} = [KR,\ KT]$	$\boldsymbol{X}_e = g_e \boldsymbol{X} = \begin{bmatrix} R_e & T_e \\ 0 & 1 \end{bmatrix} \boldsymbol{X}$
Affine	$\Pi_{2a} = [KRK^{-1},\ KT]$	$\boldsymbol{X}_a = H_a \boldsymbol{X}_e = \begin{bmatrix} K & 0 \\ 0 & 1 \end{bmatrix} \boldsymbol{X}_e$
Project.	$\Pi_{2p} = [KRK^{-1} + KTv^T, v_4 KT]$	$\boldsymbol{X}_p = H_p \boldsymbol{X}_a = \begin{bmatrix} I & 0 \\ -v^T v_4^{-1} & v_4^{-1} \end{bmatrix} \boldsymbol{X}_a$

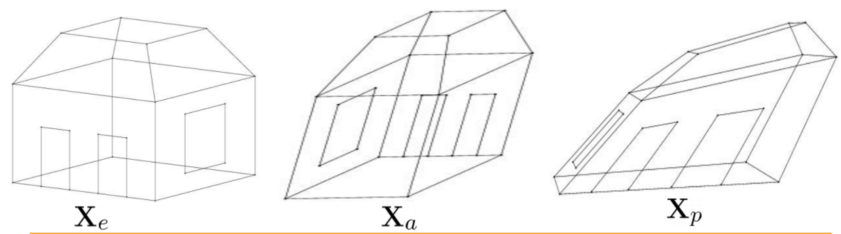

Figure 4: Stratification based on different transformations

4 Other methods of calibration

Sometimes due to the exact environment (or rig), we can find other constraints on the calibration matrix K and limit the degrees of freedom.

4.1 Calibration with a Planar Rig

For this case, we can effectively say $Z = 0$. We can define a 3 x 3 matrix H such that $\lambda \mathbf{x'} = H\mathbf{X}$, where for simplicity while maintaining homogenous coordinates, $\mathbf{X} = [X, Y, 1]^T$. This produces two linear constraints:

$$H \doteq K[r_1, r_2, T] \quad \in \mathbb{R}^{3\times 3} \quad K^{-1}[h_1, h_2] \ \sim \ [r_1, r_2]$$

$$h_1^T K^{-T} K^{-1} h_2 = 0, \quad h_1^T K^{-T} K^{-1} h_1 = h_2^T K^{-T} K^{-1} h_2.$$

4.2 Calibration with Scene Structure: Vanishing Points

Recall that the vanishing points are the intersection of real world parallel lines in the projective image plane. We can then determine vanishing points of three orthogonal directions such that $v_i = KRe_i$. Then for $i \neq j$, $S = K^{-T} * K^{-1}$, we have $v_i^T S v_j = 0$. S has 5 degrees of freedom, and the vanishing points here can give us 3 linear constraints. We do need additional assumptions to recover K, such as zero skew and an aspect ratio of 1.

4.3 Calibration with Motions - Pure Rotation

Consider the case where $T = 0$. Then we have $\hat{x}_2 R x_1 = 0$. From here, we can derive:

$$\hat{x'}_2 K R K^{-1} x'_1 = 0$$

Let $C = KRK^{-1}$. From this geometry, we can produce three linear constraints:

$$S^{-1} - CS^{-1}C^T = 0, S^{-1} = KK^T$$

Given two rotations around linearly independent axes, we can estimate S and K using linear techniques.

4.4 Calibration with Motions: General Motions

The fundamental matrix satisfies **Kruppa's Equation**:

$$FKK^T F^T = \hat{T}' KK^T \hat{T'}^T$$

If we know the fundamental matrix up to a scale, we can use this to produce two nonlinear constraints on S. Note that solutions to this equation can be sensitive to noises.

4.4.1 Calibration with Motions: Special Motions

For some special motions, we can determine the scale λ of the fundamental matrix:

$$FKK^T F^T = \lambda^2 hatT' KK^T \hat{T'}^T$$

By determining the scale, each Kruppa equation gives two linear constraints on S

5 Summary

	Calibrated case	Uncalibrated case
Image point	$\boldsymbol{x}$	$\boldsymbol{x}' = K\boldsymbol{x}$
Camera (motion)	$g = (R, T)$	$g' = (KRK^{-1}, KT)$
Epipolar constraint	$\boldsymbol{x}_2^T E \boldsymbol{x}_1 = 0$	$(\boldsymbol{x}_2')^T F \boldsymbol{x}_1' = 0$
Fundamental matrix	$E = \widehat{T} R$	$F = \widehat{T'} KRK^{-1}, T' = KT$
Epipoles	$E\boldsymbol{e}_1 = 0,\ \boldsymbol{e}_2^T E = 0$	$F\boldsymbol{e}_1 = 0,\ \boldsymbol{e}_2^T F = 0$
Epipolar lines	$\boldsymbol{\ell}_1 = E^T \boldsymbol{x}_2,\ \boldsymbol{\ell}_2 = E\boldsymbol{x}_1$	$\boldsymbol{\ell}_1 = F^T \boldsymbol{x}_2',\ \boldsymbol{\ell}_2 = F\boldsymbol{x}_1'$
Decomposition	$E \mapsto [R,\ T]$	$F \mapsto [(\widehat{T'})^T F,\ T']$
Reconstruction	Euclidean: $\boldsymbol{X}_e$	Projective: $\boldsymbol{X}_p = H\boldsymbol{X}_e$

Figure 5: Summary of Epipolar Relationships