EECS C106B / 206B Robotic Manipulation and Interaction **Spring 2021**

Lecture 13: (Uncalibrated Geometry and Stratification)

Scribes: Nico Carballal

13.1 Uncalibrated vs. Calibrated Camera Representation

13.1.1 Uncalibrated Camera

An uncalibrated camera distorts image plane coordinates **x** into pixel coordinates **x'**. This transformation from the center of the camera to the center of the image is characterized by the linear transform K.

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

so that

$$K\boldsymbol{x} = \boldsymbol{x'}$$

With **X** being defined by

$$\boldsymbol{X} = \begin{bmatrix} X & Y & Z & W \end{bmatrix}^T \varepsilon \mathbb{R}^4$$

the homogeneous coordinates of the calibrated camera have these characteristics

- Image plane coordinate
$$\boldsymbol{x} = \begin{bmatrix} x & y & 1 \end{bmatrix}^T$$
- Camera extrinsic parameters
$$g = (R, T)$$
- Perspective projection
$$\lambda \mathbf{x} = \begin{bmatrix} R & T \end{bmatrix} \mathbf{X}$$

Meanwhile, for the uncalibrated camera, distortion must be taken into account.

- Pixel coordinates
$$\boldsymbol{x'} = K\boldsymbol{x}$$
- Projection matrix
$$\lambda \boldsymbol{x'} = \Pi \boldsymbol{X} = \begin{bmatrix} KR & KT \end{bmatrix} \boldsymbol{X}$$

13.1.2 Taxonomy on Uncalibrated Reconstruction

If K is known, then reconstruction back to the calibrated case is simple

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

However, if K is unknown, some reconstruction of the actual image must be performed through determining the calibration of K. This can be done using a rig and using features from uncalibrated images (will be covered later). In addition, using partial knowledge to establish K, through parallel lines, vanishing points, and planar motion, can simplify the calibration. Some challenges may arise with ambiguities and stratification.

13.2 The Fundamental Matrix

13.2.1 Uncalibrated Epipolar Geometry

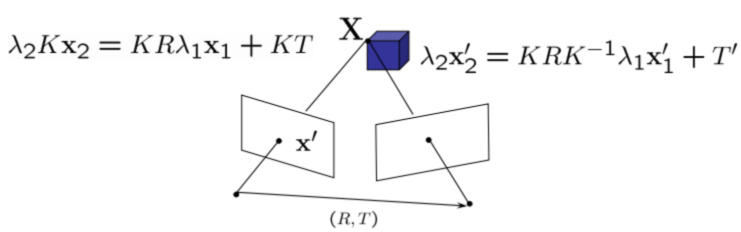


Figure 13.1: Epipolar constraints on an uncalibrated camera

Look at the above figure, we can establish that the three points $\boldsymbol{x_1}$, $\boldsymbol{x_2}$, and $\boldsymbol{X}$ are all co-planar. With this, we can establish the epipolar constraint using the distortion matrix K.

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

The fundamental matrix, $\boldsymbol{F}$, can be defined as

$$F = K^{-T}\hat{T}RK^{-1}$$

to allow the epipolar constraint to be displayed as

$$x_2'^T F x_1' = 0$$

Please note that F has two equivalent forms. Do **not** treat these as separate constraints:

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

13.2.2 Properties of the Fundamental Matrix

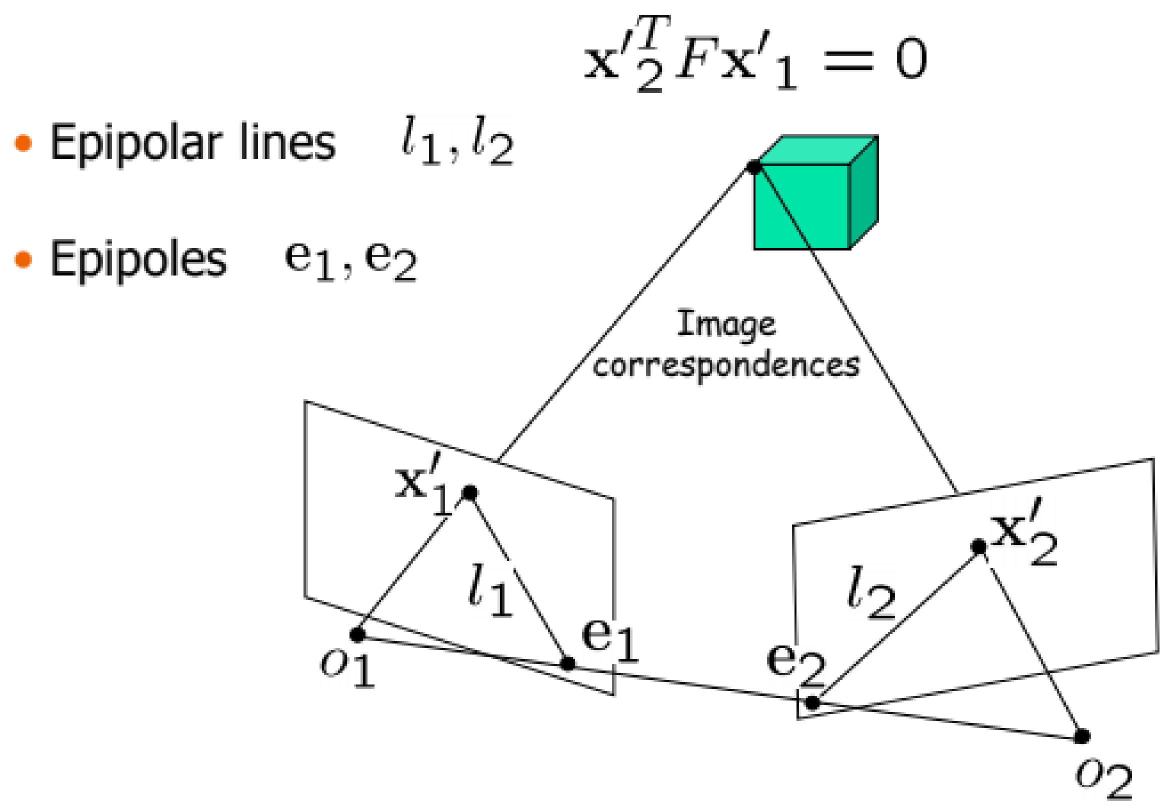


Figure 13.2: Epipoles and Epipolar lines with Image Correspondences

Due do the epipolar constraints placed upon two corresponding points, we can simplify the manner in which we search for image correspondence along separate camera views. As long as we can define the epipolar lines of one image (i.e. l_1), we can then search for corresponding features along the other epipolar line (l_2). Some additional properties of the Fundamental matrix are listed below.

- $l_1 \sim F^T \boldsymbol{x'_2}$ and $F\boldsymbol{e_1} = 0$
- $l_i^T \boldsymbol{x'}_i = 0$ and $l_i^T \boldsymbol{e}_i = 0$
- $l_2 \sim F\boldsymbol{x'_1}$ and $\boldsymbol{e}_2^T F = 0$

The matrix properties of the fundamental matrix can be described by its Singular Value decomposition $F = U\Sigma V^T$ with:

$$\Sigma = diag(\sigma_1, \sigma_2, 0)$$

as long as σ_1 and σ_2 are positive values. In addition, the matrix F can be highly variable, but its determinant must be equal to 0.

13.2.3 Estimating Fundamental Matrix

Since we know our epipolar error must theoretically be zero, we use our 8 point algorithm to solve the LLSE problem:

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

If

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

then the problem can be described by χF^s. The solution eigenvector is associated with the smallest eigenvalue of $\chi^T \chi$. After conducting SVD on F, project it onto the essential manifold to have the form with two non-zero singular values and one zero singular value. Once you get F, you cannot extract values for T, R, and K immediately. So, how do we get K?

13.3 Euclidean, Affine, and Projection Representations

13.3.1 Calibrated vs. Uncalibrated Space

If $S = K^{-T}K^{-1}$, then we can define the principle axes of the mapping of a unit sphere onto a unit ellipsoid by the eigenvalues of S. This concept can also be seen in the figure below as cubes can be mapped to the principle axes of the 2D parallelogram. If a shape is viewed as distorted, however, it is impossible to tell from just the distorted parallelogram if the camera is distorting the image or if the shape itself is distorted in the first place.

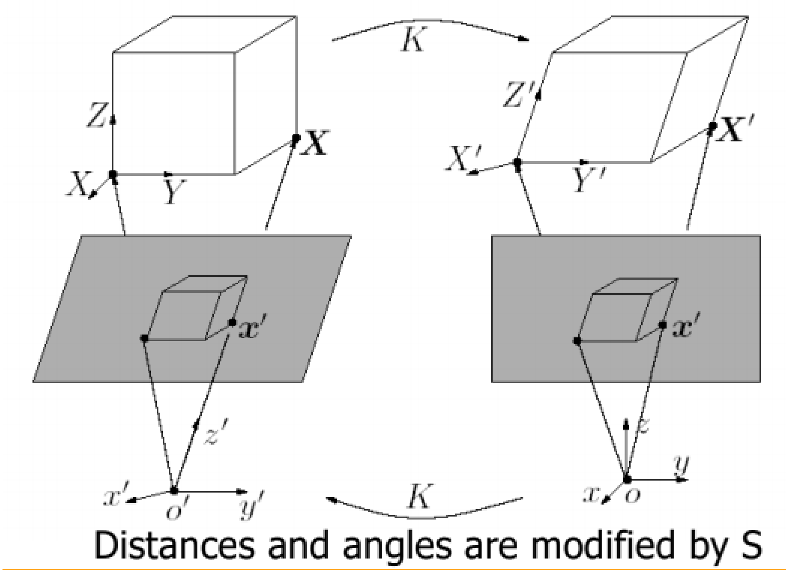


Figure 13.3: By just looking at the gray image, you can't tell if the camera is distorting the image or if the cube is just distorted.

13.3.2 Motion in the distorted space

Uncalibrated motion can be described by

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

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

Uncalibrated coordinates can be defined by G':

$$G' = \left\{ g' = \begin{bmatrix} KRK^{-1} & T' \\ 0 & 1 \end{bmatrix} \mid T' \in \Re^3, R \in SO(3) \right\}$$

13.3.3 What does F Tell Us

- F can be inferred from point matches
- Cannot extract motion, structure, and calibration from just one fundamental matrix
- F allows reconstruction up to a projective transformation
- F encodes all geometric information among two views when no additional information is available.

13.3.4 Decomposing the Fundamental Matrix

Using the definitions for F from before

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

we can decompose the matrix into a skew symmetric matrix and non singular matrix

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

The decomposition of F is shown to be not unique, and if unknown parameters (ambiguity) are defined by

$$v = [v_1, v_2, v_3]^T \epsilon \Re^3; v_4 \epsilon \Re$$

then the corresponding projection matrix Π is defined by

$$\Pi = [KRK^{-1} + T'v^T, v_4 T']$$

13.3.5 Projective Reconstruction

With the definition of Π above and the relationship $F = \hat{T}'R'$, we can compute our projection matrices Π_{ip} and structure $\boldsymbol{X}_p$ by choosing a solution for our first projection matrix Π_{1p} as $[I, 0]$. Since our second projection matrix is defined by a translation and rotation off the first, we can define that as so

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

$$\lambda_2 \boldsymbol{x_2'} = \Pi_{2p} \boldsymbol{X}_p = [((\hat{T'})^T F, T'] \boldsymbol{X}_p,$$

As a reminder, this is a just a choice of solution; it does not give you the Euclidean solution. This solution that we have convenient chosen, however, must be transformable to a Euclidean solution by some 4x4 transformation, H

$$\lambda_i \boldsymbol{x_i'} = \Pi_{ip} H^{-1} H \boldsymbol{X}_p$$

$$\lambda_i \boldsymbol{x_i'} = \tilde{\Pi}_{1p} \tilde{\boldsymbol{X}}_p$$

This may initially appear of little value; however, the proposed Canonical decomposition does give a solution that *can* be transformed and that is essential to comprehend.

Okay, so how do I recover the Euclidean coordinates? If you know the coordinates in the Euclidean spaces, you can solve using a linear least squares problem $M\boldsymbol{X}_p$.

$$(x_1 \pi_1^{3T}) \mathbf{X}_p = \pi_1^{1T} \mathbf{X}_p, \qquad (y_1 \pi_1^{3T}) \mathbf{X}_p = \pi_1^{2T} \mathbf{X}_p,$$
$$(x_2 \pi_2^{3T}) \mathbf{X}_p = \pi_2^{1T} \mathbf{X}_p, \qquad (y_2 \pi_2^{3T}) \mathbf{X}_p = \pi_2^{2T} \mathbf{X}_p,$$

This can give you the transformation from the projection solution $\boldsymbol{X}_p$ to the Euclidean solution $\boldsymbol{X}_e$ through

$$X_e = H X_p$$

This is not often not the case unfortunately, and we have to deal with ambiguities.

13.3.6 Ambiguities in the image formation

How do we undo all the ambiguities in image formation? We can interpret our projection matrix as follows.

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

This interpretation demonstrates how the image $\mathbf{X}$ can be distorted by some 4x4 matrix H, but it can then be undone by our projection matrix by left multiplying it by H^{-1}. In general, we know that H is of the following form

$$\begin{bmatrix} G & b \\ v^T & v_4 \end{bmatrix}$$

If we use the first frame as reference

$$\lambda_1 \boldsymbol{x'}_1 = K_1 \Pi_0 \boldsymbol{X}_e$$

$$\lambda_1 \boldsymbol{x'}_1 = K_1 \Pi_0 H^{-1} \boldsymbol{X}_e = \Pi_{1p} \boldsymbol{X}_p$$

then the general ambiguity is $H^{-1} = \begin{bmatrix} K_1^{-1} & 0 \\ v^t & v_4 \end{bmatrix}$, so

- H^{-1} can be further decomposed as

$$H^{-1} = \begin{bmatrix} K_1^{-1} & 0 \\ v^T & v_4 \end{bmatrix} = \begin{bmatrix} K_1^{-1} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} I & 0 \\ v^T & v_4 \end{bmatrix} \doteq H_a^{-1} H_p^{-1}$$

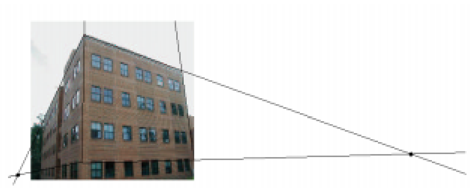
$$\mathbf{X}_p = H_p \overbrace{H_a \underbrace{g_e \mathbf{X}}_{\mathbf{X}_e}}^{\mathbf{X}_a}$$

13.3.7 Affine upgrade using vanishing points

Using the definition $H_p^{-1} = \begin{bmatrix} K_1^{-1} & 0 \\ v^t & v_4 \end{bmatrix}$ and

$$\boldsymbol{X}_a = H_p^{-1} \boldsymbol{X}_p$$

we want to find our unknown v. Since $\boldsymbol{X}_a$ is of the form $[X, Y, Z, 0]^T$ for the vanishing points of affine coordinates, we then know that mapping the vanishing points of parallel lines (where they meet) on the projected image to affine coordinates should result in the $\boldsymbol{X}_a = [X, Y, Z, 0]^T$. This will be a constraint for each vanishing point! After you get the projected transformation, you just have to look for the coordinates of where parallel lines intersect in the projected image to recover the projection ambiguities $[v_1, v_2, v_3, v_4]$



13.3.8 Euclidean upgrade

Using the previous strategy, we upgraded H_p. Now, we need to upgrade

$$H_a^{-1} = \begin{bmatrix} K_1^{-1} & 0 \\ 0 & 1 \end{bmatrix}$$

through using the intrinsic parameters of the camera, the upper triangular matrix K.

Some special cases exist to find K. For the case of special motions, there is no projective ambiguity, so you can estimate KK^T directly. For the multi-view case, you estimate projective and affine ambiguity together. If you have multiple views, there are absolute quadric constraints that can be used but are very difficult to solve, so realistically they are not used. Look at section 12.6 for more on calibration of K.

13.3.9 Geometric Stratification

The steps betweeen Euclidean (left), affine (middle), and projection (right) representations of the images are shown below. Picture H_a and H_p as aiding in transforming between these different representations.

	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.	$\Pi_{2p} = [KRK^{-1} + KTv^T, v_4 KT]$	$\boldsymbol{X}_p = H_p \boldsymbol{X}_a = \begin{bmatrix} I & 0 \\ -v^T v_4^{-1} & v_4^{-1} \end{bmatrix} \boldsymbol{X}_a$

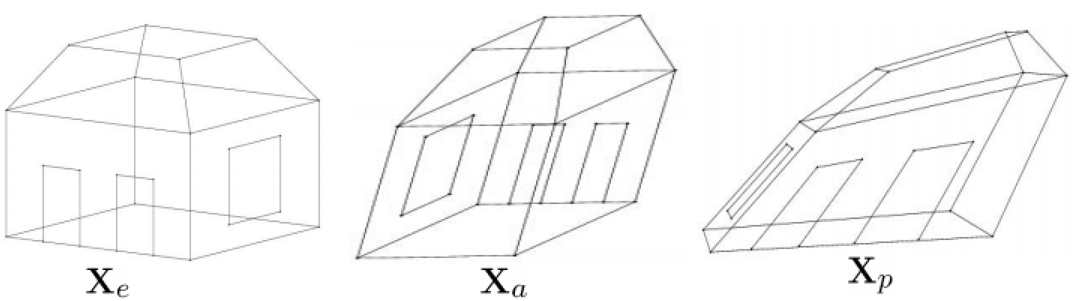


Figure 13.4: Table and Graphic of Different Camera Views

13.4 Summary

Review the below tables to review the last few lectures.

	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 13.5: Table of Distinctions between Calibrated and Uncalibrated Cases

	Euclidean	Affine	Projective
Structure	$\mathbf{X}_e = g_e \mathbf{X}$	$\mathbf{X}_a = H_a \mathbf{X}_e$	$\mathbf{X}_p = H_p \mathbf{X}_a$
Transformation	$g_e = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}$	$H_a = \begin{bmatrix} K & 0 \\ 0 & 1 \end{bmatrix}$	$H_p = \begin{bmatrix} I & 0 \\ -v^T v_4^{-1} & v_4^{-1} \end{bmatrix}$
Projection	$\Pi_e = [KR,\ KT]$	$\Pi_a = \Pi_e H_a^{-1}$	$\Pi_p = \Pi_a H_p^{-1}$
3-step upgrade	$\mathbf{X}_e \leftarrow \mathbf{X}_a$	$\mathbf{X}_a \leftarrow \mathbf{X}_p$	$\mathbf{X}_p \leftarrow \{\mathbf{x}_1', \mathbf{x}_2'\}$
Info. needed	Calibration K	Plane at infinity $\pi_\infty^T \doteq [v^T, v_4]$	Fundamental matrix F
Methods	Lyapunov eqn.	Vanishing points	Canonical decomposition
	Pure rotation	Pure translation	
	Kruppa's eqn.	Modulus constraint	
2-step upgrade	$\mathbf{X}_e \leftarrow \mathbf{X}_p$		$\mathbf{X}_p \leftarrow \{\mathbf{x}_i'\}_{i=1}^m$
Info. needed	Calibration K and $\pi_\infty^T = [v^T, v_4]$		Multiple-view matrix*
Methods	Absolute quadric constraint		Rank conditions*
1-step upgrade	$\{\mathbf{x}_i\}_{i=1}^m \leftarrow \{\mathbf{x}_i'\}_{i=1}^m$		
Info. needed	Calibration K		
Methods	Orthogonality & parallelism, symmetry or calibration rig		

Figure 13.6: Table Differentiating between Euclidean, Affine, and Projective Representations

13.5 Calibration Methods

13.5.1 Calibration with a Rig

Often, if you know the movement, it is much easier to solve for K. A rig allows you to solidify the location $[X, Y, Z]$ of an object, which allows you to solve for the projection matrix.

13.5.2 Calibration with a Planar Rig

This is much more popular than using a cube as a cube is difficult to perfectly manufacture. A special world frame is defined on the plane and is visually very obvious. Two linear constraints are thus placed on the calibration S per image, and you only need 2-3 images to calibrate the camera depending on the unknowns of K. This is the most used method.

13.5.3 Real-World Calibration (light-field/stage cameras)

Can be a much more difficult problem as you may need to calibrate relative poses of each cameras along with intrinsic features of each camera. This adds significantly to the amount of unknowns you are solving. You need very distinctive patterns often to calibrate these real-world cameras.

13.5.4 Calibration with Scene Structure

Although we already established the relation between the projection image and the affine image with the intersection of the vanishing point lines, we can also add additional constraints for an uncalibrated camera in some situations. For the building image on an earlier, we can determine that all the edge lines must be orthogonal to one another, giving us three more constraints. In a lot of cases, some of the parameters of the camera are known, so you only need a few equations about the image to define the calibration matrix.

13.5.5 Calibration with Motions: Pure Rotation

Very difficult to perform a pure rotation because you never know where the center of the camera is. Other special motions exist to figure out calibration, such as planar movement.