ECE 661
Purdue U.
Avi Kak
Lecture 23

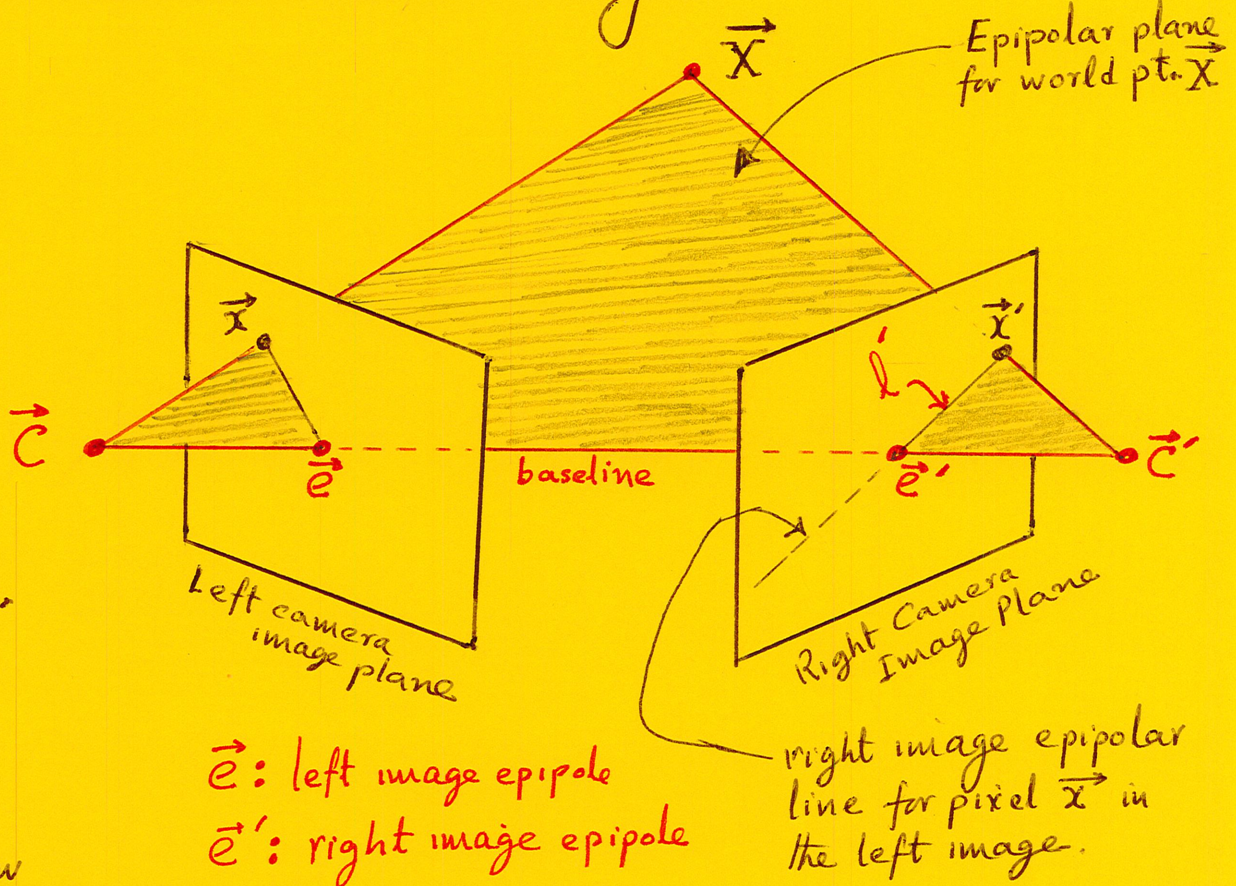
Epipolar Geometry and The Fundamental Matrix

Reference: Multiple View Geometry in Computer Vision by Hartley & Zisserman

- We will now concern ourselves with the imaging of scenes with a stereo pair of cameras. Epipolar Geometry is the projective geometry that is intrinsic to such a pair of cameras and the Fundamental Matrix is the algebraic representation of the Epipolar Geometry.
- Let P, K, C denote the camera matrix, the intrinsic calibration matrix, and the camera center for the left camera. And let P′, K′, C′ represent the same properties of the right camera.
- Given two cameras looking at the same scene, there exists a 3X3 matrix F of rank 2 that captures the most fundamental relationship between the pixels $\vec{x}$ and $\vec{x}'$ in the two cameras for the same scene point $\vec{X}$: $\boxed{\vec{x}'^T F \vec{x} = 0}$. We call F the Fundamental Matrix. To derive F we must first understand the epipolar geometry.

Epipolar Geometry

- For any given world point $\vec{X}$, its epipolar plane (defined as the plane that passes through $\vec{X}$ and the camera centers $\vec{C}$ and $\vec{C}'$) is fixed.
- The baseline is the line that joins the two camera centers. Where the baseline pierces the two image planes, we have the two epipoles. Location of the epipoles depends on how the cameras are oriented vis-à-vis each other.



- For any given pixel $\vec{x}$ in the left image, its corresponding pixel $\vec{x}'$ in the right image must lie on a line $\vec{l}'$ that is called the epipolar line. Obviously, the line $\vec{l}'$ depends on $\vec{x}$. All epipolar lines in the right image for all left-image pixels must pass through the right-image epipole $\vec{e}'$.

- By construction, $\vec{e}' = P'\vec{C}$ and $\vec{e} = P\vec{C}'$

- You know from Lecture 19 (page 19-5) that if we backproject the pixel $\vec{x}$, the world point $P^+\vec{x}$ is guaranteed to exist on the ray from $\vec{C}$ through $\vec{x}$. The image of this point in the right camera is at $P'P^+\vec{x}$. This pixel must lie on the epipolar line $\vec{l}'$ in the right camera image. $\left(P^+ = P^T(PP^T)^{-1}\right)$

- That gives us two points on $\vec{l}'$: the epipole $\vec{e}'$ and the pixel at $P'P^+\vec{x}$. Therefore, $\boxed{\vec{l}' = \vec{e}' \times P'P^+\vec{x}}$ where '$\times$' is the vector cross-product operator.

- We note that with $\vec{A} = \begin{pmatrix} a_x \\ a_y \\ a_z \end{pmatrix}$ and $\vec{B} = \begin{pmatrix} b_x \\ b_y \\ b_z \end{pmatrix}$, the vector cross-product $\vec{A} \times \vec{B}$ can be expressed as the matrix-vector product $[\vec{A}]_\times \vec{B}$ where $[\vec{A}]_\times = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix}$

- Therefore, $\vec{l}' = [\vec{e}']_\times P'P^+\vec{x} \equiv F\vec{x}$ where $\boxed{F = [\vec{e}']_\times P'P^+}$.

- From Lecture 2, the right-image pixel $\vec{x}'$ that corresponds to the left-image pixel $\vec{x}$ must satisfy $\boxed{\vec{x}'^T \vec{l}' = 0}$ we have $\boxed{\vec{x}'^T F \vec{x} = 0}$. We refer to F as the Fundamental Matrix for a given pair of cameras.

Fundamental Matrix — A Special Case

- As you now know, the fundamental matrix F for a pair of cameras P and P' is given by $F = [\vec{e}']_\times P'P^+$. Let's now see what F simplifies to for the special case when the world frame coincides with left camera coord. frame.

- In this case we have $P = K[I|\vec{0}]$ and $P' = K'[R|\vec{t}]$ where $(R, \vec{t})$ is the location/orientation of the right camera with respect to the left camera.

- Since we need the pseudoinverse P^+ in F, let's first see what $P^+ = P^T(PP^T)^{-1}$ simplifies to. We note $PP^T = K[I|\vec{0}]\begin{bmatrix} I \\ \vec{0}^T \end{bmatrix}K^T = KK^T$. Therefore $(PP^T)^{-1} = K^{-T}K^{-1}$ [This result is based on the following two properties of matrix inverses: 1 For any nonsingular matrix A, the inverse of the transpose equals the transpose of the inverse. That is $A^{-T} = (A^T)^{-1} = (A^{-1})^T$; and 2 When two nonsingular matrices A & B are of equal rank $(AB)^{-1} = B^{-1}A^{-1}$.] So, $P^+ = P^T(PP^T)^{-1} = \begin{bmatrix} K^T \\ \vec{0}^T \end{bmatrix} K^{-T}K^{-1} = \begin{bmatrix} K^T K^{-T} K^{-1} \\ \vec{0}^T \end{bmatrix} = \begin{bmatrix} K^{-1} \\ \vec{0}^T \end{bmatrix}$

- Let's now see what the $P'P^+$ part of F simplifies to. We can write: $P'P^+ = K'[R|\vec{t}]\begin{bmatrix} K^{-1} \\ \vec{0}^T \end{bmatrix} = K'RK^{-1}$.

- The rest of F is $[\vec{e}']_\times$ which is the same as $[P'\vec{C}]_\times$. We note $\vec{C} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$ since the left-camera center is at the world origin. Therefore, $[P'\vec{C}]_\times = \left[K'[R|\vec{t}]\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}\right]_\times = [K'\vec{t}]_\times$.

- Bringing together the two parts of F, we get $\boxed{F = [K'\vec{t}]_\times K'RK^{-1}}$

Properties of the Fundamental Matrix

NOTE: So far we have used the phrases "left camera" and "right camera" to refer to the two cameras of a stereo pair. (The "left" and the "right" are for an observer standing between the cameras and looking out in the same general direction as the cameras.) A more correct usage is to talk about an ordered pair of cameras and refer to the individual cameras as the first camera and ~~the~~ the second camera. With this new way of referring to the cameras, here are some genuinely useful properties of the Fundamental Matrix:

1. F can be of great help in solving the stereo correspondence problem. This is the problem of finding the $\vec{x}'$ in the second image that corresponds to a given $\vec{x}$ in the first image. If we know F, we can confine our search to the line $\boxed{\vec{l}' = F\vec{x}}$ in the second image. For the opposite case, given any pixel $\vec{x}'$ in the second image, the corresponding pixel in the first image is on the epipolar line $\boxed{\vec{l} = F^T\vec{x}'}$.

2. The determinant of F is always zero: $\boxed{\det(F) = 0}$. This follows from the fact that for all $n \times n$ matrices, the determinant obeys the multiplicative property: $\det(A \cdot B) = \det(A) \cdot \det(B)$. In $F = [\vec{e}']_\times P' P^+$, the matrix $[\vec{e}']_\times$ is skew symmetric and therefore $\boxed{\det([\vec{e}']_\times) = 0}$. Hence $\det(F) = 0$.

3. F has 7 DoF. Since F is homogeneous, that removes 1 DoF. And the fact that $\det(F) = 0$ removes another DoF. With only 9 elements, that leaves F with just 7 DoF.

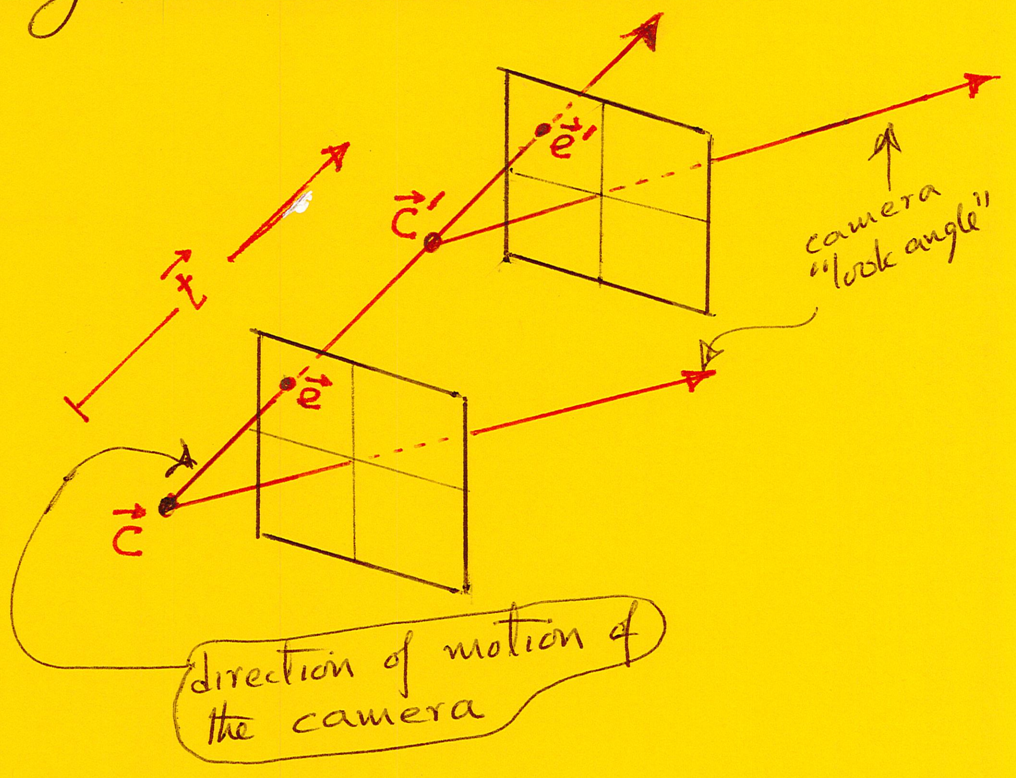
4. Each known correspondence $(\vec{x}, \vec{x}')$ gives us one equation of the form $\vec{x}'^T F \vec{x} = 0$ in which we know the values for $\vec{x}$ and $\vec{x}'$ and treat the 9 elements of F as the unknowns. Since F has 7 DoF, it takes at least 7 correspondences to estimate F directly from the two images.

5. The second-image epipole $\vec{e}'$ is the left null-vector of F and the first-image epipole $\vec{e}$ is its right null-vector: $\boxed{\vec{e}'^T F = 0}$ and $\boxed{F\vec{e} = 0}$. To prove $\vec{e}'$ is the left null-vector, we note that $\vec{x}'$ for a given $\vec{x}$ is on the right-image line $\vec{l}' = F\vec{x}$. Since $\vec{e}'$ is also on this line, we have $\vec{e}'^T \vec{l}' = \vec{e}'^T F \vec{x} = 0$. Since $\vec{e}'^T F \vec{x} = 0$ must be true for every pixel $\vec{x}$ in the first image, it must be the case that $\vec{e}'^T F = 0$. To prove that $\vec{e}$ is the right null-vector of F, we note that $\vec{x}$ for a given $\vec{x}'$ must lie on the line $\vec{l} = F^T \vec{x}'$ in the first image. Since $\vec{e}$ is also on this line, $\vec{l}^T \vec{e} = \vec{x}'^T F \vec{e} = 0$. Since $\vec{x}'^T F \vec{e} = 0$ must be true for every $\vec{x}'$ in the second image, we have $F\vec{e} = 0$.

6. If F is the fundamental matrix for a given ordered pair of cameras, the fundamental matrix becomes F^T if you reverse the order of the cameras.

7. $\text{rank}(F) = 2$. Since $\det(F) = 0$, $\text{rank}(F) < 3$. Also, since $\vec{l}' = F\vec{x}$ must have 2 DoF as an epipolar line, $\text{rank}(F) > 1$.

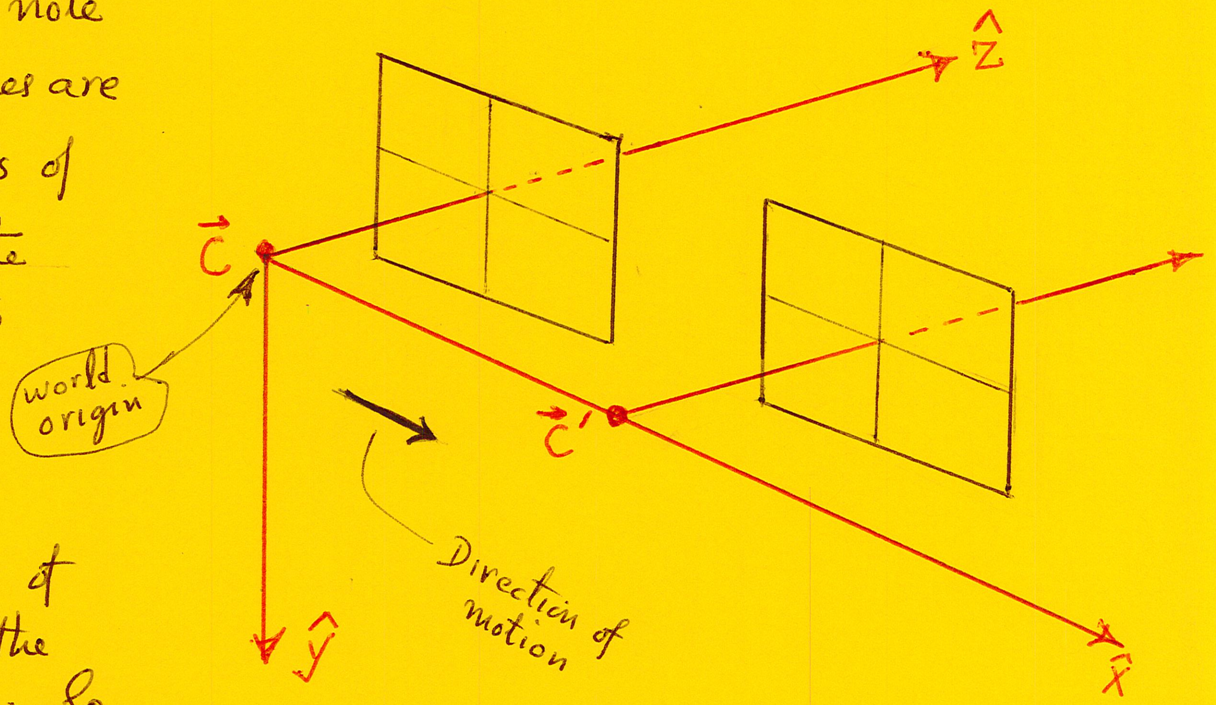
Fundamental Matrix for the Case of Translational Motions by the Same Camera



- Shown at right is a camera that has executed a translational motion $\vec{t}$ without changing its orientation. We record two images, one before the motion, and one after.

- Based on the derivation on page 23-2, we have $F = [K\vec{t}]_X K I K^{-1}$ which reduces to $\boxed{F = [K\vec{t}]_X}$. Using the arguments in the next to the last bullet on p. 23-2, we can also express this fundamental ~~express~~ matrix as $F = [\vec{e}']_X$. Don't forget that these results apply only when the world frame coincides with with the camera frame at the first position.

- Let's now specialize the result obtained above to the case when the camera moves in a direction parallel to the image plane as shown below. The most interesting thing to note here is that both the epipoles are at infinity along the X-axis of the world frame. That is, the projection of the point $\vec{C}$ into the the image plane of the second camera through the center $\vec{C}'$ is at $\begin{pmatrix}1\\0\\0\end{pmatrix}$. The same goes for the projection of $\vec{C}'$ into the image plane for the for the first camera position. So we can write $\boxed{\vec{e} = \vec{e}' = \begin{pmatrix}1\\0\\0\end{pmatrix}}$.



- Substituting $\vec{e}' = \begin{pmatrix}1\\0\\0\end{pmatrix}$ in the result in the second bullet above, we get

$$F = [\vec{e}']_X = \begin{bmatrix} 0 & -e_z & e_y \\ e_z & 0 & -e_x \\ -e_y & e_x & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

- Let's now see what sort of a constraint this fundamental matrix places on the search for corresponding pixels in the two images. Let $\vec{x} = \begin{pmatrix}u\\v\\w\end{pmatrix}$ and $\vec{x}' = \begin{pmatrix}u'\\v'\\w'\end{pmatrix}$ be a pair of corresponding pixels. We can therefore write:

$$\vec{x}'^T F \vec{x} = 0 \Rightarrow (u' v' w') \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{pmatrix}u\\v\\w\end{pmatrix} \Rightarrow (u' v' w') \begin{pmatrix}0\\-w\\v\end{pmatrix} = 0 \Rightarrow v'w = vw' \Rightarrow \boxed{\frac{v'}{w'} = \frac{v}{w}}$$

All that the fundamental matrix is telling us is that, given a pixel $\vec{x}$ in the first image, the search for its corresponding pixel in the second image should be confined to a line that has the same y-coordinate as the pixel in the first image. We could obviously have reached the same conclusion just by looking at the geometrical construction shown on the previous page. Nonetheless, the message here is that the fundamental matrix helps us limit the search to just a line in the second image. And that is true in general.

Essential Matrix

- A fundamental matrix is called the Essential Matrix when the two images are expressed in normalized coordinates. The normalized coordinates of a pixel at $\vec{x}$ are given by $\boxed{\hat{\vec{x}} = K^{-1}\vec{x}}$
- Since $\vec{x} = P\vec{X} = K[R|\vec{t}]\vec{X}$, representing the pixels with the coordinates $K^{-1}\vec{x}$ amounts to recording the image with a camera for which $K = I$. Since, ordinarily, $K = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}$, when $K = I$, that means the image origin is at the Principal Point, there is no skew, and that $\alpha_x = m_x f = 1$ and $\alpha_y = m_y f = 1$.
- Consider two images, both in normalized coordinates, as recorded by the two cameras $P = [I|\vec{0}]$ and $P' = [R|\vec{t}]$. We are obviously assuming that the world origin coincides with the first camera frame of reference. The fundamental matrix for these cameras is their Essential Matrix. It is given by the same formula as at the bottom of page 23-2. So $\boxed{E = [\vec{t}]_\times R}$ E is Essential Matrix
- To see the relationship between the Essential Matrix E and the Fundamental Matrix F for a given pair of cameras, let $\hat{\vec{x}}$ and $\hat{\vec{x}}'$ represent the normalized coordinates in the two images. $\hat{\vec{x}} = K^{-1}\vec{x}$ and $\hat{\vec{x}}' = K'^{-1}\vec{x}'$. As corresponding pixels, $\hat{\vec{x}}$ and $\hat{\vec{x}}'$ must satisfy $\hat{\vec{x}}'^T E \hat{\vec{x}} = 0$. So $\boxed{\vec{x}'^T K'^{-T} E K^{-1}\vec{x} = 0}$. This implies $\boxed{F = K'^{-T} E K^{-1}}$. In other words, $\boxed{E = K'^T F K}$

An Important Invariance of the Fundamental Matrix

- You'll recall that the camera projection matrix $P = K[R|\vec{t}]$ was a simplified form of $P = K[I|\vec{0}]\begin{bmatrix} R & \vec{t} \\ \vec{0}^T & 1 \end{bmatrix}$ where the Euclidean transformation $\begin{bmatrix} R & \vec{t} \\ \vec{0}^T & 1 \end{bmatrix}$ takes a point from the world frame into the camera coordinate frame where it is imaged by $K[I|\vec{0}]$. We are at liberty to interpose an additional 4x4 projective transformation H between the world points and where P takes over. If we incorporate H within the camera projection framework, the new camera matrix would become PH. (H is a 4x4 homography in general)
- A fundamental matrix F is invariant to right-multiplication of camera matrices P and P′ by a projective transformation H. To prove this, let's start with the original camera pair (P, P′). Substituting $\vec{x} = P\vec{X}$ and $\vec{x}' = P'\vec{X}$

in $\vec{x}'^T F \vec{x} = 0$, we get $\vec{X}^T P'^T F P \vec{X} = 0$. Let's denote the projectively modified cameras by P_{new} and P'_{new}: $P_{new} = PH$ and $P'_{new} = P'H$. Substituting $P = P_{new} H^{-1}$ and $P' = P'_{new} H^{-1}$ in $\vec{X}^T P'^T F P \vec{X} = 0$, we get $\boxed{X^T H^{-T} P'^T_{new} F P_{new} H^{-1} X = 0}$. For convenience, let's write $\vec{X}_{new} = H^{-1} X$. Our equality can now be expressed as $\boxed{\vec{X}^T_{new} P'^T_{new} F P_{new} \vec{X}_{new} = 0}$. There exists a corresponding pair of pixels given by $\vec{x}_{new} = P_{new} \vec{X}_{new}$ and $\vec{x}'_{new} = P'_{new} \vec{X}_{new}$. In terms of these pixels we have $\boxed{\vec{x}'^T_{new} F \vec{x}_{new} = 0}$ implying that F continues to serve as a fundamental matrix.

- This invariance of F implies that if we try to infer the values of P and P′ from a given F, we can only do so up to a missing right multiplier H for the two camera matrices, (where H is a 4x4 projective transformation of 3D space). (just from the standpoint of how the camera sees the 3D space)
- This leads us to the notion of canonical configuration for an ordered pair of cameras. Two cameras are in canonical configuration if the projection matrix of the first camera is given by $P = [I|\vec{0}]$.
That leads to the question: Can we find ~~the~~ two cameras in the canonical configuration for a given F? The answer is 'Yes'. For a given F, the two cameras are $P = [I|\vec{0}]$ and $P' = [sF|\vec{e}']$ where s is any 3x3 skew-symmetric matrix that results in P′ to be a rank 3 matrix and where $\vec{e}'$ is the left null-vector of F ($\vec{e}'^T F = 0$). As long as P′ is of rank 3, any choice for s works.

Estimating the Fundamental Matrix from Pixel Correspondences

- As you saw in the 4th property on page 23-3, each $(\vec{x}, \vec{x}')$ correspondence gives us one equation $\vec{x}'^T F \vec{x} = 0$ for estimating the elements of F. With $\vec{x} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$, $\vec{x}' = \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix}$ and $F = \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}$, the equation becomes $\boxed{x'x f_{11} + x'y f_{12} + x' f_{13} + y'x f_{21} + y'y f_{22} + y' f_{23} + x f_{31} + y f_{32} + f_{33} = 0}$
We can express this equation as $A\vec{f} = 0$ where A is the 1x9 matrix given by $A = [x'x \;\; x'y \;\; x' \;\; y'x \;\; y'y \;\; y' \;\; x \;\; y \;\; 1]$ and $\vec{f} = (f_{11}\; f_{12}\; f_{13}\; f_{21}\; f_{22}\; f_{23}\; f_{31}\; f_{32}\; f_{33})^T$.
- Given N correspondences $(\vec{x}_i, \vec{x}'_i)$, $i = 1, 2, \ldots, N$, let $A_i \vec{f} = 0$ be the equation for the i^{th} correspondence. Stacking these equations together, we get $\mathbb{A}\vec{f} = \vec{0}$ where $\mathbb{A}$ is N×9. Since $\vec{f}$, being homogeneous, only needs to be calculated up to a scale factor, we need at least 8 equations. Therefore, you need at least 8 correspondences for estimating F. Note however the very important rule of thumb: The number of equations you need is roughly five times the number of unknowns when using any least-squares algorithm for solving a system of homogeneous or inhomogeneous equations. So, practically speaking, you are going to need about 40 correspondences. (When manually specified, you may be able to get by with far fewer than 40 correspondences.)
- Next, as described in Lecture 10, construct a linear least-squares solution for $\vec{f}$ from $\mathbb{A}\vec{f} = 0$. This solution must be conditioned to enforce the requirement rank(F) = 2. (Without conditioning, you'll end up with multiple apparent epipoles in the second image.) You condition the solution F by doing its SVD to obtain UDV^T, by zeroing out the smallest singular value in D, and forming the product $UD'V^T$ where D′ is the modified form of D.
- If the world points for all your correspondences happen to be coplanar, your F will be unreliable