ECE 661
Purdue Univ
Avi Kak
Lecture 18

Modeling The Camera

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

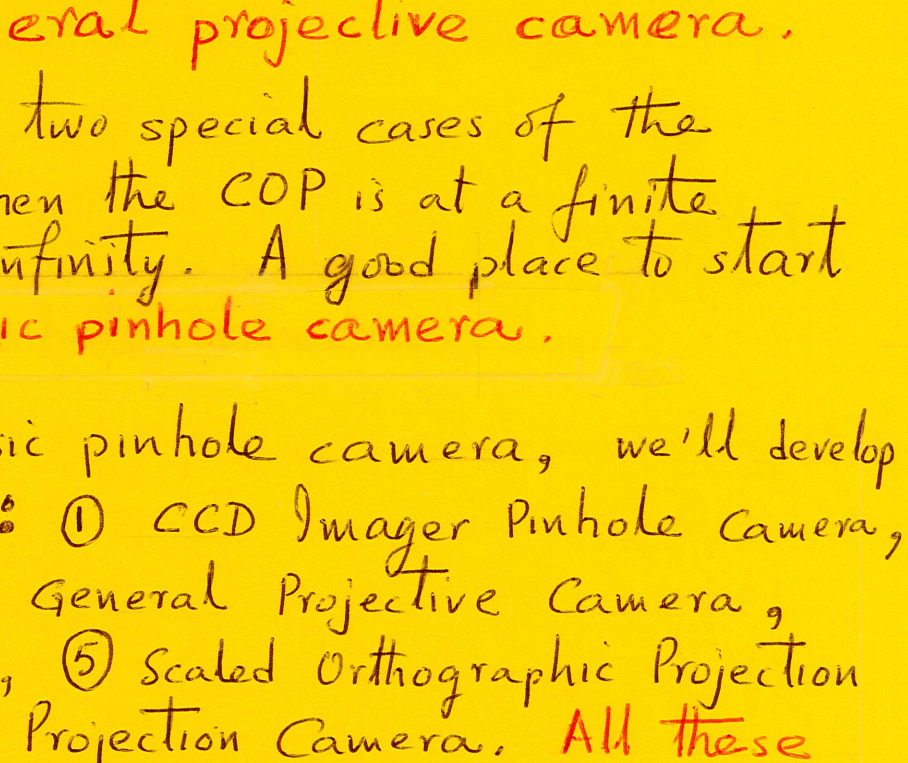
- A camera is a mapping from the 3D world to a 2D image.
- Our primary focus will be on cameras that carry out central projection. Central projection means that all rays that connect a world point with its corresponding image point pass through a common point known as the Center of Projection.
- As you will see later, cameras based on central projection are special cases of the general projective camera.
- At this time we are interested in two special cases of the central projection cameras: ① When the COP is at a finite location; and ② When COP is at infinity. A good place to start for understanding both is the basic pinhole camera.
- Starting with a model for the basic pinhole camera, we'll develop models for the following cameras: ① CCD Imager Pinhole Camera, ② Finite Projective Camera, ③ General Projective Camera, ④ Orthographic Projection Camera, ⑤ Scaled Orthographic Projection camera, and ⑥ Weak Perspective Projection Camera. All these camera models are obtained by placing different kinds of constraints on the structure and the rank of a 3×4 matrix known as the Camera Projection Matrix.

The Basic Pinhole Camera

- Comparing similar triangles, the relationship between the world point (X,Y,Z) and the corresponding image point (x,y) is given by:

$$x = \frac{fX}{Z} \qquad y = \frac{fY}{Z}$$

nonlinear



The above relationship is based on measuring the pixel coordinates with respect to the Principal Point. More commonly, though, the pixel coordinates are measured with respect to a corner of the image frame. Using the corner P as the image origin, we can write

$$x = \frac{fX}{Z} + p_x \qquad y = \frac{fY}{Z} + p_y$$

nonlinear

where (p_x, p_y) are the coordinates of the Principal Point in the image plane

- Using the homogeneous coordinate representation $(X, Y, Z, 1)^T$ for the world point, we can write the following **linear** relationship between the pixel coords on the left and the corresponding world point coords on the right:

$$\begin{pmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{pmatrix} = \begin{bmatrix} f & 0 & p_x & 0 \\ 0 & f & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad \rightarrow \quad \vec{x} = K\,[I_{3\times3} | \vec{0}]\,\vec{X} \quad \text{with } K = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix}$$

$\vec{x}$: pixel coords

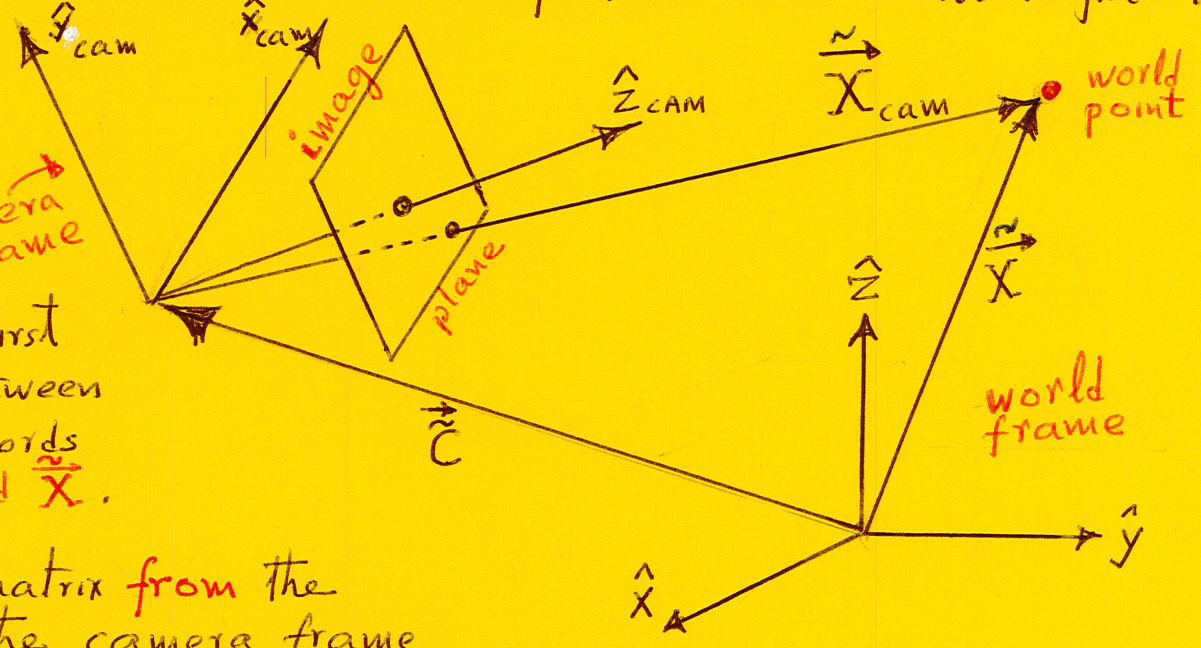
Camera Calibration Matrix for the parameters intrinsic to the camera.

- The world frame that is shown at the bottom of the previous page is in reality **the Camera Coordinate Frame** — because it is anchored at the camera. Therefore, what we have derived should be shown as:

$$\vec{x} = K\,[I_{3\times3} | \vec{0}]\,\vec{X}_{cam}$$

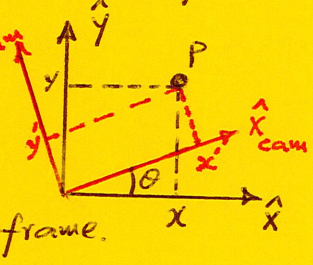
where $\vec{X}_{cam}$ are the Camera Frame Coords of World $\vec{X}$.

- Let's now separate the camera coordinate frame and the world frame:



- In order to find the relationship between $\vec{X}_{cam}$ and $\vec{X}$, in camera modeling work, it is often best to first establish the relationship between the corresponding physical coords that we denote by $\tilde{\vec{X}}_{cam}$ and $\tilde{\vec{X}}$.

- Let R denote the rotation matrix **from** the physical world coords **to** the camera frame coords. To understand what we mean by 'from' and 'to' clauses, assume that the two origins are at exactly the same point. Our definition of R would imply that $\tilde{\vec{X}}_{cam} = R\,\tilde{\vec{X}}$. To see this more clearly, consider the 2D case at right: $\tilde{\vec{X}}_{cam} \equiv \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \equiv R\,\tilde{\vec{X}}$



So when R multiplies the coordinates of a point P in the world frame, we get the coordinates of the same point in the camera frame.

- Let $\tilde{\vec{C}}$ denote the translational vector **from** the world frame origin **to** the camera frame origin. If we ignore the rotational misalignment between the two frames, we have $\tilde{\vec{X}}_{cam} = \tilde{\vec{X}} - \tilde{\vec{C}}$.

- When we have both a translation $\tilde{\vec{C}}$ and a rotation R **from** the world frame **to** the camera frame, we have $\tilde{\vec{X}}_{cam} = R(\tilde{\vec{X}} - \tilde{\vec{C}})$. For the world point $\tilde{\vec{X}}$, the subtraction $\tilde{\vec{X}} - \tilde{\vec{C}}$ gives us the coords of the same world point in the camera frame at a vector distance $\tilde{\vec{C}}$ but with the camera frame still aligned with the world frame. That multiplying this by R should yield $\tilde{\vec{X}}_{cam}$ follows from our previous explanation of how R works.

- The equation $\tilde{\vec{X}}_{cam} = R(\tilde{\vec{X}} - \tilde{\vec{C}})$ can be adapted to homogeneous coordinates by:

$$\vec{X}_{cam} = \begin{pmatrix} \tilde{\vec{X}}_{cam} \\ 1 \end{pmatrix} = \begin{bmatrix} R & -R\tilde{\vec{C}} \\ \vec{0}^T & 1 \end{bmatrix} \begin{pmatrix} \tilde{\vec{X}} \\ 1 \end{pmatrix} = \begin{bmatrix} R & -R\tilde{\vec{C}} \\ \vec{0}^T & 1 \end{bmatrix} \vec{X}$$

$\rightarrow \tilde{\vec{X}}_{cam} = R(\tilde{\vec{X}} - \tilde{\vec{C}})$

- Substituting the relationship between the homogeneous vectors $\vec{x}_{cam}$ and $\vec{X}$ in the pinhole camera model shown in the second bullet of the previous page, we can write for the pixel coords:

$$\vec{x} = K\left[I_{3\times3} \middle| \vec{0}\right]\begin{bmatrix} R & -R\tilde{\vec{C}} \\ \vec{0}^T & 1 \end{bmatrix}\vec{X}$$

$$\left[I_{3\times3} \middle| \vec{0}\right]\begin{bmatrix} R & -R\tilde{\vec{C}} \\ \vec{0}^T & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} r_{11} & r_{12} & r_{13} & (-R\tilde{\vec{C}})_x \\ r_{21} & r_{22} & r_{23} & (-R\tilde{\vec{C}})_y \\ r_{31} & r_{32} & r_{33} & (-R\tilde{\vec{C}})_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} r_{11} & r_{12} & r_{13} & (-R\tilde{\vec{C}})_x \\ r_{21} & r_{22} & r_{23} & (-R\tilde{\vec{C}})_y \\ r_{31} & r_{32} & r_{33} & (-R\tilde{\vec{C}})_3 \end{bmatrix} = R\left[I \middle| -\tilde{\vec{C}}\right]$$

This relationship is expressed more compactly as

(A) $$\vec{x} = KR\left[I_{3\times3} \middle| -\vec{C}\right]\vec{X}$$

- $\vec{x}$: pixel coords (homogeneous)
- K: camera parameters (intrinsic)
- R: world-to-camera rotation (physical)
- $\vec{C}$: world-to-camera translation (physical)
- $\vec{X}$: world point (homogeneous)

- The equation shown above will serve as our fundamental equation for modeling a pinhole camera. However, note that there is another form of this equation that sometimes works better for camera calibration. To derive the other form, let's revisit the relationship we derived previously between the world point $\tilde{\vec{X}}$ and the physical camera frame coords, $\tilde{\vec{X}}_{cam}$, of the same world point. The relationship derived on the previous page was based on **first** finding $\tilde{\vec{X}}_{cam}$ by only taking into account the translation $\tilde{\vec{C}}$ and **then** accounting for the world-to-camera rotation. **Let's now reverse these two steps.** In order to find the camera frame coords of the same physical point that is at $\tilde{\vec{X}}$ in the world frame, we write

$$\tilde{\vec{X}}_{cam} = R\tilde{\vec{X}} + \vec{t}$$

- R has the same meaning as before. It's the world-to-camera rotation
- $\vec{t}$ is the translational displacement that must be applied to the coordinate values
- Both R and $\vec{t}$ exist in physical coordinates

Using homogeneous coords, we can write this as:

$$\vec{X}_{cam} = \begin{pmatrix} \tilde{\vec{X}}_{cam} \\ 1 \end{pmatrix} = \begin{bmatrix} R & \vec{t} \\ \vec{0}^T & 1 \end{bmatrix}\begin{pmatrix} \tilde{\vec{X}} \\ 1 \end{pmatrix} = \begin{bmatrix} R & \vec{t} \\ \vec{0}^T & 1 \end{bmatrix}\vec{X}$$

Note that this form becomes the same as the one shown at the bottom of the previous page if we set $\boxed{\vec{t} = -R\tilde{\vec{C}}}$.

- Substituting the new relationship between $\vec{X}_{cam}$ and $\vec{X}$ in the pinhole camera model in the second bullet on the previous page:

(B) $$\vec{x} = K\left[I_{3\times3} \middle| \vec{0}\right]\begin{bmatrix} R & \vec{t} \\ \vec{0}^T & 1 \end{bmatrix}\vec{X} = K\left[R \middle| \vec{t}\right]\vec{X} = KR\left[I_{3\times3} \middle| R^T\vec{t}\right]\vec{X}$$

Taking Into Account the Sampling of the Image Plane

- For images recorded with digital cameras, we must also factor in the sampling rates in the image plane. Let's start by assuming that the 2D array of photo sensors in the image plane is perfectly rectangular, with m_x cells per unit length along x and m_y cells per unit length along y.
- We may now associate the "discretized" location (I, J) with the pixel at (x, y), with $\boxed{I = m_x x}$ and $\boxed{J = m_y y}$. Obviously, the products $m_x x$ and $m_y y$ do not by themselves yield integer values. However, to get integer indices associated with the cells, all we have to do is to ignore the fractional parts of $m_x x$ and $m_y y$.

- Using our two foundational equations for x and y at the bottom of page 16-1, we can write

$$I = m_x \frac{fX}{Z} + m_x p_x \qquad J = m_y \frac{fy}{Z} + m_y p_y$$

If we use $x_o = m_x p_x$ and $y_o = m_y p_y$ as the "integer" coords of the Principal Point in the image plane, we can write:

$$I = \frac{m_x f X + x_o Z}{Z} \qquad J = \frac{m_y f y + y_o Z}{Z}$$

These two equations can be turned into the following composite relationship between the homogeneous representation of the "discrete" pixel coords and the world point in the camera coordinate frame:

$$\begin{pmatrix} IZ \\ JZ \\ Z \end{pmatrix} = \begin{pmatrix} m_x f X + x_o Z \\ m_y f y + y_o Z \\ Z \end{pmatrix} = \begin{bmatrix} m_x f & 0 & x_o & 0 \\ 0 & m_y f & y_o & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ y \\ Z \\ 1 \end{pmatrix} = \begin{bmatrix} \alpha_x & 0 & x_o & 0 \\ 0 & \alpha_y & y_o & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ y \\ Z \\ 1 \end{pmatrix}$$

with $\boxed{\alpha_x = m_x f}$ and $\boxed{\alpha_y = m_y f}$. You can think of α_x as the focal length in terms of the number of pixels using the x-sampling rate, and of α_y as the focal length in terms of the number of pixels using y-sampling rate.

- $\begin{pmatrix} IZ \\ JZ \\ Z \end{pmatrix}$ is obviously the homogeneous representation of the "discrete" pixel coordinates (I, J)! Going forward, we will represent this homogeneous vector by $\vec{x}$. So we again have

$$\boxed{\vec{x} = K \left[I_{3\times 3} \middle| \vec{0} \right] \vec{X}_{cam}} \quad \text{with} \quad \boxed{K = \begin{bmatrix} \alpha_x & 0 & x_o \\ 0 & \alpha_y & y_o \\ 0 & 0 & 1 \end{bmatrix}}$$

This relationship between the homogeneous vectors $\vec{x}$ and $\vec{X}_{cam}$ is the same as what was shown in the first bullet on page 18-2. The only difference here is the form of the intrinsic-parameter camera calibration matrix K.

- Therefore, our overall relationship between a world point $\vec{X}$ and the corresponding pixel $\vec{x}$ remains the same as shown earlier at (A) and (B) on page 18-3 — except for the intrinsic calibration matrix K being as shown above. We reproduce below one of those two relationships:

$$\vec{x} = \underbrace{K R \left[I_{3\times 3} \middle| -\vec{C} \right]}_{\text{Projection Matrix } P \text{ of size } 3\times 4} \vec{X}$$

Note that P is homogeneous

- The 3x4 projection matrix P has 10 DoF (Degrees of Freedom). Of these K accounts for 4, R for 3, and $\vec{C}$ for the remaining 3. Now consider ~~the fact~~ that since the 3x4 P has a maximum of 11 DoF, as to what might be the practical significance of the remaining DoF in P. Consider the following form for K:

$$K = \begin{bmatrix} \alpha_x & s & x_o \\ 0 & \alpha_y & y_o \\ 0 & 0 & 1 \end{bmatrix} \leftarrow \text{5 DoF}$$

$s \equiv$ the skew parameter

With this K, the projection matrix P will have the maximum allowable DoF of 11. As it turns out, the additional parameter s in K is required if the image sensor array is not perfectly rectangular. In digital cameras there is always a bit of "skew" in the xy-layout of the sensor cells.

Finite Projective Cameras

- In general — and especially if one allows for virtual cameras — the relationship between a world point at $\vec{X}$ and its corresponding pixel at $\vec{x}$ is expressed as

$$\underset{3\times1}{\vec{x}} = \underset{3\times4}{P}\ \underset{4\times1}{\vec{X}}$$

P: the camera projection matrix

In general, we will express P as

$$P = [\vec{p}_1\ \ \vec{p}_2\ \ \vec{p}_3\ \ \vec{p}_4]$$

when we need to identify it four columns separately. And, sometimes, we will show P as

$$P = [M \mid \vec{p}_4]$$

where M is the 3x3 submatrix formed by the first three columns of P.

- We now claim that any 3x4 real matrix whose first three columns constitute a nonsingular submatrix is a valid camera. We refer to the set of all such cameras as Finite Projective Cameras.

- To prove the claim made above, we first note that for all such P:

$$P = [M \mid \vec{p}_4] = M\,[I_{3\times3} \mid M^{-1}\vec{p}_4]$$

Next, we note that any square matrix can be subject to RQ decomposition. That is, any square matrix can be expressed as a product of an upper triangular matrix and an orthogonal matrix. A variant of the same decomposition expresses a square matrix as a product of an upper triangular matrix and an orthonormal matrix. We can treat the upper triangular matrix as the intrinsic calibration matrix K of some camera and the orthonormal matrix as the world-to-camera rotation matrix R.

- You are probably more familiar with the QR decomposition of a square matrix that is achieved by the Gram-Schmidt algorithm. This algorithm consists of scanning the rows of a matrix, and, for the current row, finding a new basis vector that is perpendicular to all previously constructed basis vectors. The algo expresses the current row as a linear sum of all the previously constructed basis vectors and the new basis vector. At the end, these coefficients of expansion form an upper triangular matrix R, and the basis vectors the orthogonal matrix Q.
- The RQ decomposition, on the other hand, is best achieved for 3x3 matrices by using Givens rotations.

$$Q_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \quad Q_y = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \quad Q_z = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

When a 3x3 matrix A is multiplied on the right by one of these three Q matrices, it leaves one of the columns of A unchanged and creates a linear combination of the other two columns. In the linear combinations thus produced, we can set the value of θ so as to drive one of the elements of A to zero to achieve the desired triangulation.

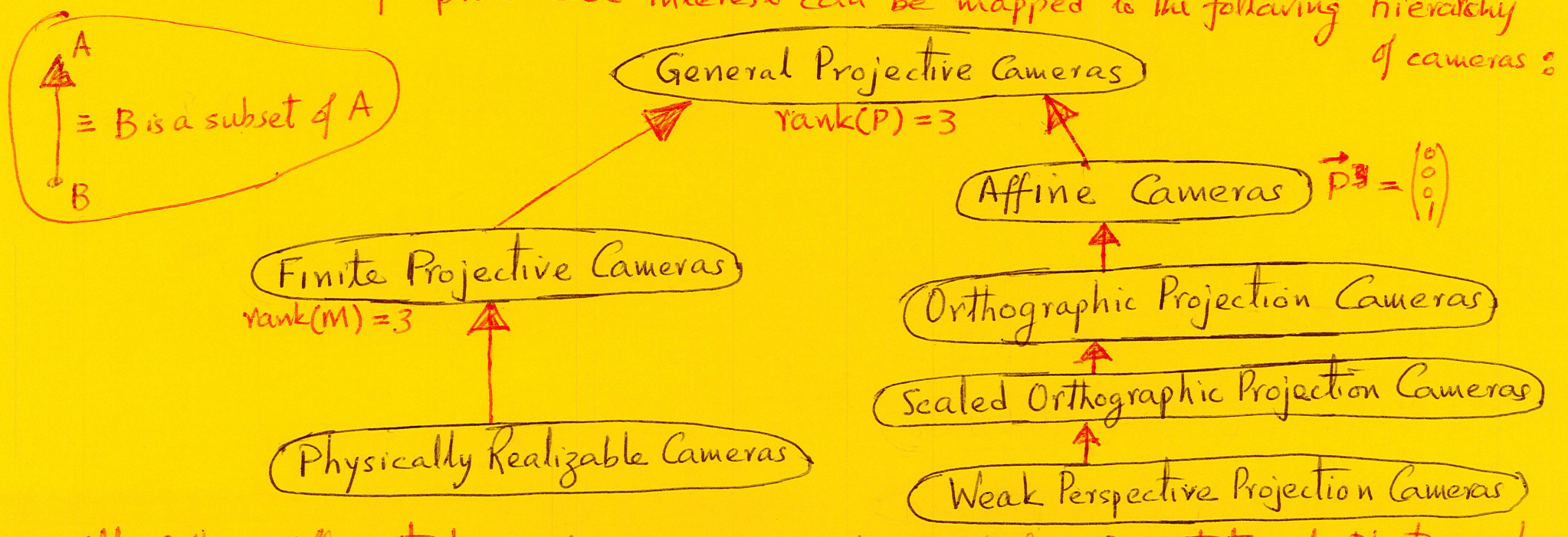
- Obviously, all digital cameras are examples of Finite Projective Cameras. However, the Finite Projective set includes cameras that cannot be realized physically but that may nonetheless be useful as virtual cameras. When a a finite projective camera cannot be realized physically it may be on account of where the COP is in relation to the object or on account of where

the image is formed on the image plane, or on account of the values given to the intrinsic parameters to the camera.

General Projective Camera

- The Finite Projective Camera is obviously a more general category than the pinhole camera (with and without the skew parameter).
- An even more general category of cameras is the General Projective Camera. The only constraint we now place on the projection matrix P is that its rank be exactly 3.
- A 3x4 projection matrix P whose rank is less than 3 will not able to produce an image. For example, if a given 3x4 matrix P has a rank of only 2, the output pixel homogeneous vector will have only 2 DOF. Since the output coordinates are homogeneous, this would correspond to the output being a straight line.
- As you'll see in the next lecture, the center of projection of a general projective camera (that is NOT a finite projective camera) is at infinity.

NOTE: The remaining three camera models mentioned on page 16-1 — Orthographic Projection Camera, Scaled Orthographic Projection Camera, and Weak Perspective Projection Camera — will be taken up in the next lecture. You'll see in the next lecture that all Central Projection cameras of practical interest can be mapped to the following hierarchy of cameras:



- All of these different types of cameras can be used for Computational Photography