CSE-483: Mobile Robotics **Monsoon 2019**

Camera Modelling

Prepared by: Jayitha. C, Niharika. V

In this note we discuss different camera models and how their projection matrices are derived. We will also discuss intrinsic and extrinsic camera parameters based on which projection matrices are obtained. These notes are made under the assumption that the reader is familiar with the notation used in class, has a good understanding of how transformation matrices are derived and used and is familiar with various properties of lenses such as principal axis and focal length.

1 Recap

Frame of reference: All measurements are made with respect to a particular coordinate system called the frame of reference.

World Frame: A fixed coordinate system for representing objects(points, lines, surfaces, etc.) in the world

Translational Mapping In Fig 1, we have a position defined by the vector ${}^{B}P$. We wish to express this point in space in terms of frame $\{A\}$, when $\{A\}$ has the same orientation as $\{B\}$. In this case, $\{B\}$ differs from $\{A\}$ only by a translation, which is given by ${}^{A}P_{BORG}$, a vector that locates the origin of $\{B\}$ relative to $\{A\}$. Since the orientation of both the frames are the same, we can get the position of point ${}^{B}P$ with respect to frame $\{A\}$ using

$$ {}^{A}P = {}^{B}P + {}^{A}P_{BORG} $$

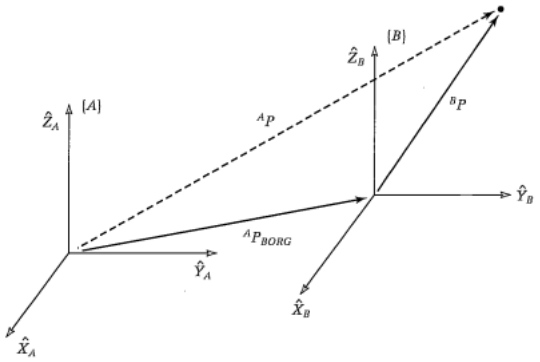

Figure 1: Translational Mapping

Rotational Mapping Similarly, to find the position of point ${}^{B}P$ with respect to $\{A\}$ i.e. ${}^{A}P$ when the origin of both the frames coincide but the axii don't, we use the following equation.

$$ {}^{A}P = {}^{A}_{B}R\,{}^{B}P $$

Where ${}^{A}_{B}R$ is the orientation of frame $\{B\}$ with respect to frame $\{A\}$. It is formed by horizontally stacking the projections of the axii in frame $\{B\}$ to $\{A\}$.

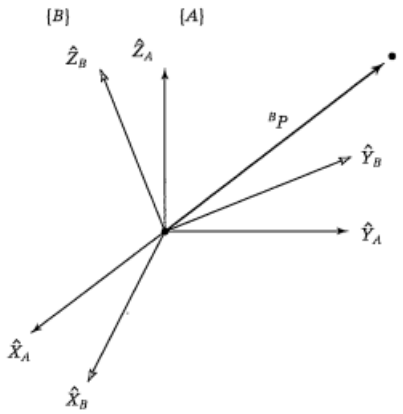

Figure 2: Rotational Mapping

General Frame Transformation If $\{B\}$ is both translated and rotated with respect to frame $\{A\}$ the the general transformation of a vector from frame $\{B\}$ to $\{A\}$ is given by

$$^AP = {}^A_BR\,{}^BP + {}^AP_{BORG}$$

If we homogenise position vectors, the we can clean up the above equation as follows

$$\begin{bmatrix} ^AP \\ 1 \end{bmatrix} = {}^A_BT \begin{bmatrix} ^BP \\ 1 \end{bmatrix}$$

where,

$${}^A_BT = \begin{bmatrix} {}^A_BR & {}^AP_{BORG} \\ 0\,0\,0 & 1 \end{bmatrix}$$

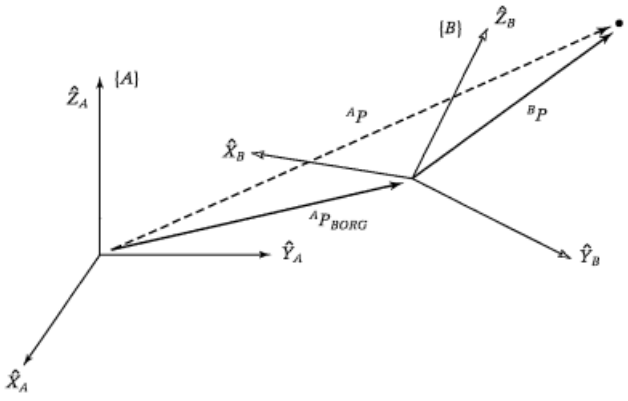

Figure 3: Transformation

With this we are now ready to move onto camera models. Before going into the details of the projection, we will first discuss camera parameters that will give us context to help us understand the following section.

2 Goal

The goal here is given points in the world frame $\{World\}$, we want to be able to map these points to pixels in an image of those points. In order to do this, we need to come up with a transformation. The series of transformation we'll go through to find this final transformation is as depicted in Fig 4 This

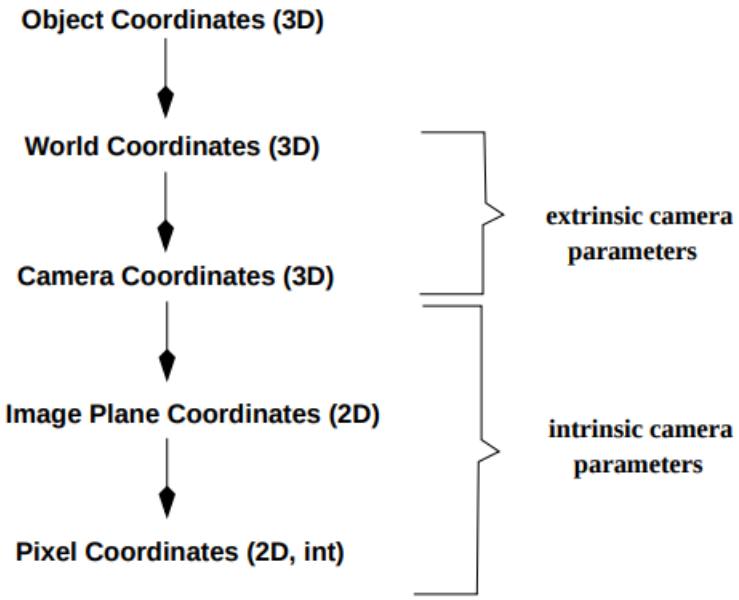

Figure 4: Series of Transformations

relation of 3D points and their 2D projections can be seen as a linear transformation from the projective space $(X_w, Y_w, Z_w, 1)^T$ to the projective plane $(u, v, w)^T$.

3 Camera Parameters

Camera models are structurally similar to the eye. Camera parameters can broadly be classified to be intrinsic or extrinsic.

3.1 Intrinsic Parameters

As seen in Fig 4, Intrinsic parameters are those parameters that are necessary to link the pixel coordinates of an image point with the corresponding coordinates in the camera reference frame. These are the parameters that characterize the optical, geometric, and digital characteristics of the camera:

- the focal length f of the camera lens
- the transformation between image plane coordinates and pixel coordinates i.e. scaling and the principal point
- the geometric distortion introduced by the optics

3.1.1 Distortion

To accurately represent a real camera, the camera model includes the radial and tangential lens distortion.

Radial Distortion Radial distortion occurs when light rays bend more near the edges of a lens than they do at its optical center. The smaller the lens, the greater the distortion. It causes the image magnification to decrease or increase as a function of the distance to the optical axis. We classify the radial distortion as **pincushion distortion** when the magnification increases and **barrel distortion** when the magnification decreases. Radial distortion is caused by the fact that different portions of the lens have differing focal lengths.

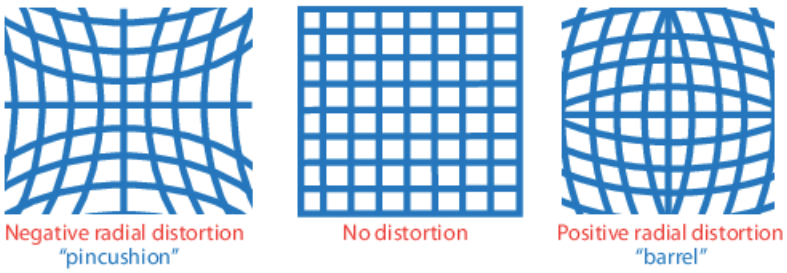

Figure 5: Radial Distortion

Tangential Distortion Tangential distortion occurs when the lens and the image plane are not parallel. The tangential distortion coefficients model this type of distortion.

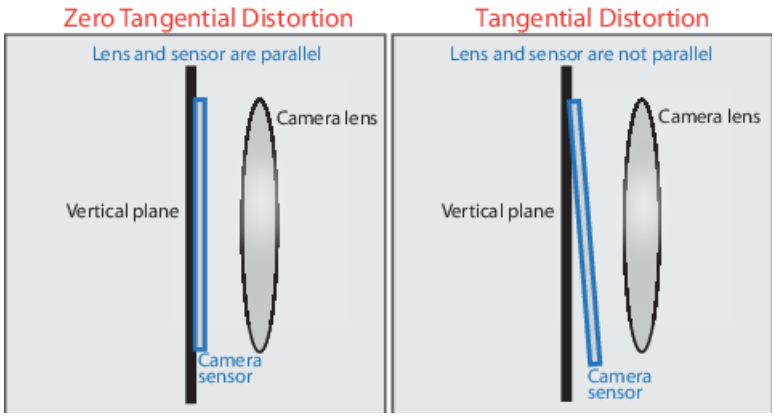

Figure 6: Tangential Distortion

3.2 Extrinsic Parameters

These are the parameters that identify uniquely the transformation between the unknown camera reference frame and the known world reference frame. Typically, determining these parameters means:

- finding the translation vector between the relative positions of the origins of the two reference frames (the world and camera frame) i.e. ${}^{C}P_{WORG}$ and
- finding the rotation matrix that brings the corresponding axes of the two frames into alignment (i.e., onto each other) ${}^{C}_{W}R$

Given that we know these two quantities, the transformation is then given by,

$$^{C}P = {}^{C}_{W}R\,{}^{W}P + {}^{C}P_{WORG}$$

With this broad understanding we are ready to start modelling out projection.

4 Camera Models

To begin modeling projections, we'll start with the Pinhole Camera Model.

4.1 Pinhole Camera Model

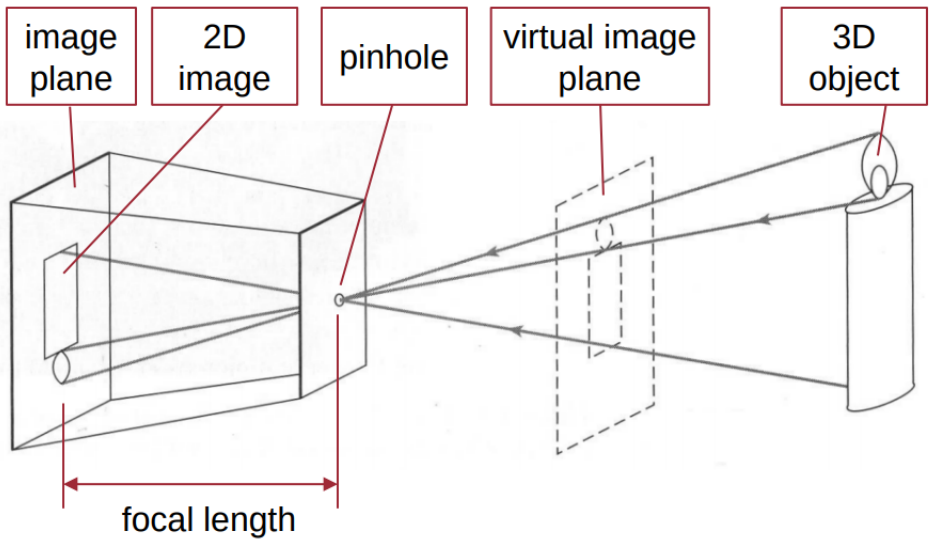

Figure 7: Pin Hole Camera Model

The pin hole camera model is the simplest camera system model. This model is made up of a system that can record an image of an object or scene in the 3D world. This camera system can be designed by placing a barrier with a small aperture between the 3D object and a photographic film or sensor. Without a barrier in place, every point on the film will be influenced by light rays emitted from every point on the 3D object. Due to the barrier, only a few of these rays of light passes through the aperture and hits the film. Therefore, we can establish a one-to-one mapping between points on the 3D object and the film. The result is that the film gets exposed by an image of the 3D object by means of this mapping.

The film is commonly called the **image or retinal plane**. The aperture is referred to as the **pinhole O or center of the camera**. The distance between the image plane and the pinhole O is the **focal length f**. Sometimes, the retinal plane is placed between O and the 3D object at a distance f from O. In this case, it is called the **virtual image or virtual retinal plane**. The optical axis intersects the image plane at a point **p** called the **principal point**

we can define a coordinate system centered at the pinhole O such that the Z axis is perpendicular to the image plane and points toward it. This coordinate system is often known as the **camera reference system or camera coordinate system**. Additionally, we can depict the position of any point on the image using 2 coordinates x and y. This coordinate system is commonly referred to as the **Image Coordinate System**. The line connecting any rea world point to the optical center is called a **projection line**. The point at which this projection line intersects the image plane is the point where we see

the image of the real world point. An idea pin hole model can be abstracted to Fig 8 although in the actual pin hole model the image would be upside down and on the other side of the camera center. It is

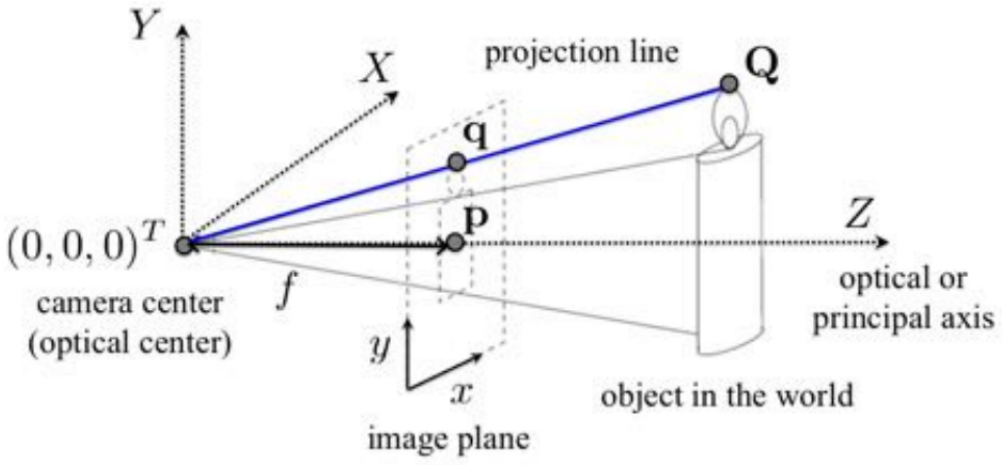

Figure 8: Ideal Pin Hole Camera Model

interesting to note that any point on a projection line will map to the same image coordinate.

For this model, let us derive the projection onto the image plane.

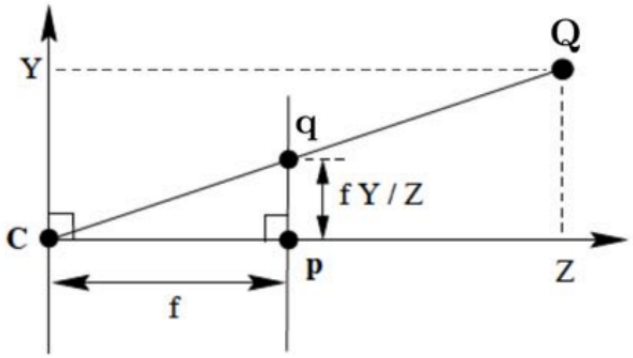

Figure 9: Ideal Pin Hole Camera Model Projection

Given the camera coordinate of a point $Q = (Xc, Yc, Zc)^T$, we want to find the image coordinates of projected point $q = (u, v)^T$. We will use the concept of similar triangles to find these coordinates. Triangle C-q-p is similar to triangle c-Q-Z. The length of C-p is f, Q-Z is Y, C-Z is Z. We want to find the length of p-q. From the property of similar triangles we know that

$$\frac{\bar{Cp}}{\bar{pq}} = \frac{\bar{CZ}}{\bar{QZ}}$$
$$\implies \frac{f}{\bar{pq}} = \frac{Z}{Y}$$
$$\implies v = \bar{pq} = \frac{fY}{Z}$$
$$\text{Similarly } u = \frac{fX}{Z}$$

But oh wait! The coordinates we've gotten are with respect to the principle point p and not the origin of

the image plane. To determine the coordinates of q with respect to the image plane we need to translate the coordinate we've got by the coordinates of the principal point i.e. (p_x, p_y).

$$Q = (X, Y, Z)^T \longrightarrow \left(\frac{fX}{Z} + p_x, \frac{fY}{Z} + p_y \right)^T$$

Before we proceed, we should remember that the point Q defined is with respect to the camera frame and not the world frame. Also we are mapping the point Q to the image frame and not pixels. To map to pixels all we'll have to do is scale these values with respect to the principal axis. We have to scale at all because pixels may be rectangular. If the scaling factors are s_x and s_y along the X and Y axis and we were to find the same using homogeneous coordinates we would obtain the following equation

$$q = (wu, wv, w)^T \longrightarrow \left(\frac{s_x f X}{Z} + p_x, \frac{s_y f Y}{Z} + p_y, 1 \right)^T \sim (fX + Zp_x, fY + Zp_y, Z)^T$$

This can be written as a matrix multiplication,

$$q = \begin{bmatrix} wu \\ wv \\ w \end{bmatrix} = \begin{bmatrix} s_x f & 0 & p_x \\ 0 & s_y f & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

Now, $q = (u, v)^T$ corresponds to the pixel coordinates of the point $Q = (X_c, Y_c, Z_c)$ We define $K = \begin{bmatrix} s_x f & 0 & p_x \\ 0 & s_y f & p_y \\ 0 & 0 & 1 \end{bmatrix}$ as the **Camera Calibration Matrix** or **Intrinsic Parameter Matrix**. This is because all the elements in K are intrinsic parameters of the camera and have nothing to do with the pose of the camera.

At this point, since we've staunchly stuck to the ideal pin hole model, we don't have to talk about radial distortion or tangential distortion. But ideal pin hole models aren't practically feasible. Why ? Because we cannot approximate a hole(aperture) to a point in 3D space. Also As the aperture size increases, the number of light rays that passes through the barrier consequently increases. With more light rays passing through, then each point on the film may be affected by light rays from multiple points in 3D space, blurring the image. Although we may be inclined to try to make the aperture as small as possible, a smaller aperture size causes less light rays to pass through, resulting in crisper but darker images. At this point we have successfully mapped point from the camera coordinate frame to their corresponding pixel value. This motivates us to introduce lenses, as they are capable of focusing light. If we replace the pinhole with a lens that is both properly placed and sized, then it satisfies the following property: all rays of light that are emitted by some point P are refracted by the lens such that they converge to a single point P'. Since all light rays parallel to the principal axis are refracted to pass through the focal point and those passing through the optical center are not deviated from their path, the similar triangles arguments is applicable here as well and we get the same equation.

Now that we've introduced lenses, we'll have to take care of tangential and radial distortions.

Briefly, to handle radially distorted points, the image coordinates would now be $(u(1+k_1r^2+k_2r^4), v(1+k_1r^2+k_2r^4))$ where $r^2 = u^2 + v^2$ if the image is symmetrically distorted. Here k_1 and k_2 are intrinsic parameters. To handle tangential distortion when the angle of skewness is θ, we can modify K as follows

$$K = \begin{bmatrix} s_x f & -s_x f \, cot\theta & p_x \\ 0 & \frac{s_y f}{sin\theta} & p_y \\ 0 & 0 & 1 \end{bmatrix}$$

But for the remainder of these notes we'll assume there are no distortions.

Till this point we've mapped points in camera coordinates to pixel space. What about points in the World coordinate space ? This is nothing but a transformation from one 3D space(World) to another(Camera). So if we have ${}^C_W T$ then we can map points from frame $\{World\}$ to frame $\{Camera\}$. The pose of the

camera constitute its extrinsic parameters. Putting everything we have so far together we get,

$$q = (wu, wv, w)^T = K {}^C_W T \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

As an exercise let us look at the dimensionality of each of the matrices involved in the above equation and also try to semantically attach some meaning. First, the result vector $q = (wu, wv, w)^T$ is a 3 x 1 vector. (u, v) give us the pixel coordinates of the world point. K is the calibration matrix which accounts for properties of the camera. K is a 3 x 3 matrix. ${}^C_W T$ takes into account the pose of the camera. And it is multiplied by a homogenised vector. It is a 3 x 4 matrix. Finally $(X_w, Y_w, Z_w, 1)$ is the homogenised coordinate of the point under consideration in the $\{World\}$ frame. It is a 4 x 1 vector. The reader is now requested to pause and go back to the goal and map the flow of this explanation to the series of transformations provided.

Now we'll look at other models to understand these parameters we've used better.

4.2 Orthographic Projection

In an orthographic projection we assume the the 3D scene we are trying to map is at infinity. Therefore all projection lines are parallel and directly map to image coordinates. Therefore

$$u = s_x X_c + p_x \text{ and } v = s_y Y_c + p_y$$

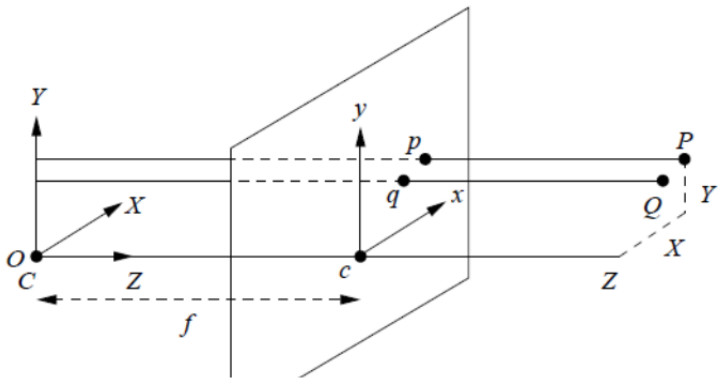

Figure 10: Orthographic Projection

This is particularly special because we see that the Z coordinate has simply been dropped and the focal length of the camera doesn't play a role.

4.3 Scaled Orthographic Projection

In scaled orthographic projection, we assume that the the image scene is a lot closer to the 3D scene than the camera is. If this is the case, then the image obtained will be scaled by a parameter $s = f/Z_0$ (Refer figure). Then

$$u = s_x X_c s + p_x \text{ and } v = s_y Y_c s + p_y$$

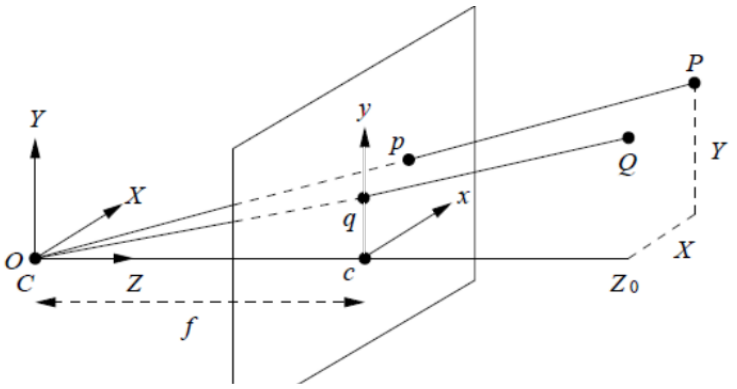

Figure 11: Scaled Orthographic Projection

4.4 Affine Camera Model

The affine camera model is exceedingly interesting, because unlike other models where we constraint the camera matrix, the affine model has no constraints on the model save one. if $T = K_W^C T$(soon to be called the homography matrix) and T is expanded as

$$T = \begin{bmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{21} & T_{23} & T_{24} \\ T_{31} & T_{32} & T_{33} & T_{34} \end{bmatrix}$$

Then the only constraint is that T_{31}, T_{32}, t_{33} are 0. The interested reader is encouraged to read up on the affine model. It is also interesting to see that all transformations we've seen so far have been affine transformation. In a sense affine transformations are generalizations of the projects we've seen.

Conclusion

In these notes we've discussed the pin hole camera model and how we can map a point from the world frame to an image frame and consequently to pixels given that we know the extrinsic and intrinsic parameters of the camera. We've also seen a few other models to further our understanding.

References

[1] Editable overleaf link to these notes Link

[2] *What Is Camera Calibration?* Link

[3] *Camera Models and Parameters*, University of Toronto Link

[4] *Geometric Camera Parameters* Link

[5] *Camera Matrix*, Carnegie Mellon University Link

[6] *Camera Models*, Ohio State University Link

[7] *Camera Models and Imaging*, National University of Singapore Link

[8] *Course Notes 1: Camera Models*, Stanford University Link