quartus ENGINEERING

# Technical Brief: Introduction to Camera Models

Since the art of imaging was invented, humans have desired a way for recovering 3D information from 2D images. Many problems in the science and engineering fields find great benefit in the ability to reason size, determine depth, and determine one's location from image data. All of these processes are made possible by the use of calibrated camera models. Camera calibration, or more formally, geometric camera calibration, is the process of identifying the best fit camera model parameters for a physical camera. Before we can understand camera calibration, we first need to understand camera models and their usage. This brief document is one of a series of documents on camera calibration and the PixelTraq camera calibration process. Here, we introduce camera models and provide the relevant background to understand camera calibration.

## Camera Models

Camera models are mathematical representations of the way that light travels from the real world to the image plane of a camera. Inside a camera, there are typically many lenses that are optimized for realizing a specific projection function for a range of wavelengths of light. Due to manufacturing tolerances, the lenses of a real camera are all slightly shifted relative to their ideal positions and slightly deformed relative to their ideal shapes. Many advanced optical raytracing software packages exist that can model these imperfections, but doing so requires a great deal of computation and many parameters. Instead of modeling a camera as its physical components (i.e. lenses, sensor, etc.) engineers and scientists have determined that the geometric relationship between a camera's image and the world it sees can be modeled to high accuracy using a combination of relatively simple mathematical functions. All while using far fewer parameters than a full physical model would require. Various methods have been developed that characterize this behavior and decades of research have resulted in many well-established mathematical representations [1] [2] [3] [4].

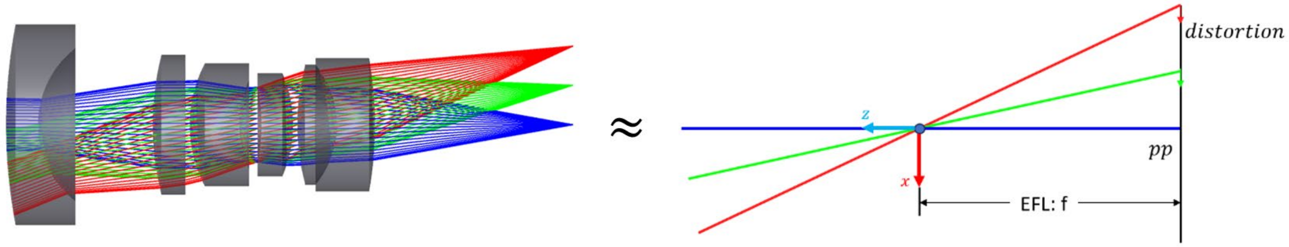

*Figure 1 Complex ray tracing lens model(left), simplified camera model(right)*

Fundamentally, camera systems achieve the function of *projecting* the 3D world onto a 2D image plane. We can say that this *projection* is the main characteristic of cameras. When we project something, we reduce how many dimensions it has, hence the projection from 3D to 2D. Camera modeling seeks to identify the underlying projection function from 3D to 2D. Once we have identified this, we can do many things such as invert the projection function and *back-project* rays associated with points in the image coordinates. These *projection* and *back-projection* operations are fundamental to many technologies such as image distortion correction, stereoscopic vision, visual SLAM, etc.

## Central and Non-Central Projection

The most common camera models can be split into two categories, central projection camera models and non-central projection camera models [5]. Central projection camera models have the common trait that they all model rays passing through a central point essentially replicating a classical pinhole camera. Although this may seem like a major assumption, most cameras can be very accurately modeled using central projection models. The variety of models in this class are generally differentiated by their representation of distortion.

Non-central projection camera models do not follow this pinhole projection characteristic or combine other elements with central projection concepts to form hybrid models. This discussion will focus primarily on central projection style models, but the PixelTraq calibration and camera calibration in general are still relevant to all types of camera models including non-central ones.

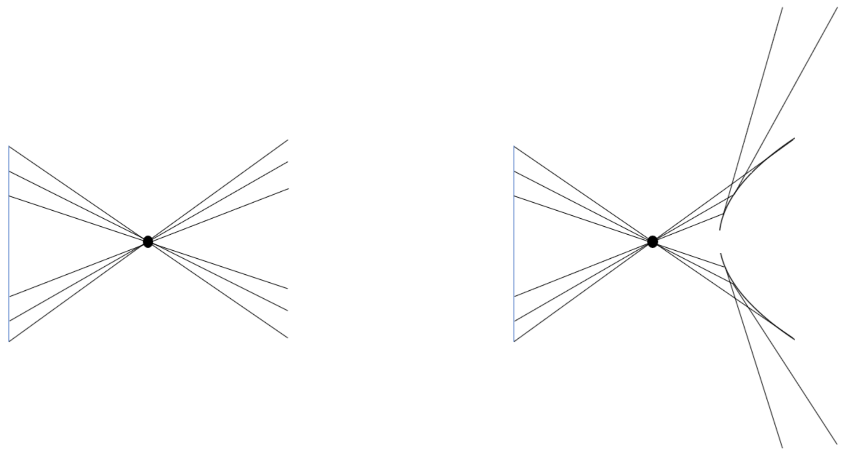

*Figure 2 Central Projection Camera (left) and Non-central Hybrid Catadioptric Projection (right)*

## Intrinsic and Extrinsic Parameters

Camera models are typically broken up into two sets of parameters known as the *intrinsic* and *extrinsic* parameters. As their names imply, the intrinsic parameters characterize the inherent optical properties of the camera and the extrinsic parameters characterize its external position and orientation.

Intrinsic parameters are the main thing that differentiates one camera model from another. The intrinsic parameters are used to implement the equations that determine what happens to light once it enters the camera. More detail will be provided on this in the next section, but some examples of typical intrinsic parameters are focal length, principal point and distortion parameters.

For all camera models, the extrinsic properties describe the three-dimensional pose of the camera in space relative to a fixed reference frame. There are many different equivalent parameterizations that can be used to describe the three-dimensional pose of a camera. One of the most common is a rotation matrix and a position vector or their combined form, a homogeneous transformation matrix [5]. Regardless of which parametrization is used, they are all equivalent and can be used interchangeably.

*Table 1 Examples of Extrinsic Parameters*

| Parametrization | Notation |
|---|---|
| **Rotation Matrix and Position Vector** | $R, p$ |
| **Homogeneous Transformation Matrix** | $T = \begin{bmatrix} R & p \\ 0 & 1 \end{bmatrix}$ |
| **Quaternion and Position Vector** | $q, p$ |

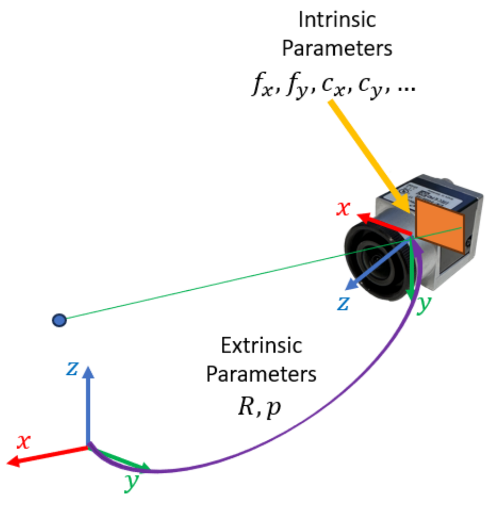

*Figure 3 Intrinsic and Extrinsic Parameters*

We call the frame with which the camera is described relative to, the world frame ($w$), and the frame of the camera itself, the camera frame ($c$). A typical notation has the frame with which we are describing another frame relative to as the first subscript and the frame we are describing as the second subscript. For example, $T_{wc}$ describes the pose of the camera frame with respect to the world frame. Conveniently, this subscript notation has some other advantages. Using homogenous coordinates $\vec{p} = [x, y, z, 1]^T$, we can transform points between two frames [5].

$$p_w = T_{wc} p_c$$

Or the using the inverse mapping

$$p_c = T_{wc}^{-1} p_w = T_{cw} p_w$$

Using this approach, we can transform points from the world frame into the camera frame or vice versa. The same operations can be performed using any of the previously discussed extrinsic parameter conventions, but the homogenous transformation matrices provide the most concise and convenient notation.

## The Pinhole Camera

The simplest camera model is known as the pinhole camera model. This model follows a perfect perspective projection relationship. In reality, no camera behaves exactly like a pinhole camera, but as it has the most straight-forward geometric interpretation, it represents the behavior of an ideal camera.

This type of projection is also sometimes called $ftan\theta$ due to the fact that a point in the image plane's radial distance from an on-axis projection is determined by this relationship. The diagram below shows the mapping of points in the world space to points on the image plane for a pinhole camera.

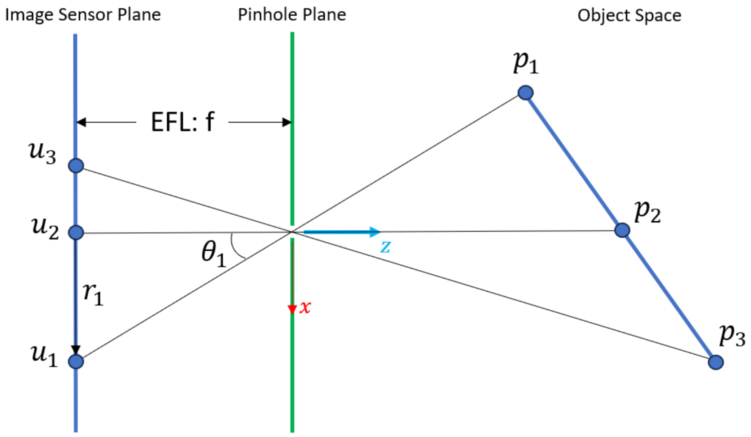

*Figure 4 Ray trace of projection for the pinhole camera*

Examining this diagram, we can define the first few commonly known intrinsic parameters. The point on the image sensor plane where the on-axis ray aligned to the pinhole plane lands (shown here as $u_2$) is known as the principal point. This can be defined in terms of X,Y coordinates in camera pixels and describes where the corner of the image plane lands relative to the center of projection. We will denote the principal point as $c = [c_x, c_y]$

The next intrinsic parameter of interest is the distance between the pinhole plane and the image sensor plane. This is known commonly as the effective focal length, EFL, or simply, focal length.

Looking at the relationship between a point $\overrightarrow{p_1} = [x_1, y_1, z_1]$ in object space and its projection $\overrightarrow{r_1} = [u_1, v_1]$ in the image space, we can determine the following geometric relationship:

$$r_{1x} = ftan\theta_1 = f\frac{x_1}{z_1}$$

Expanding this to two dimensions we find

$$r_{1x} = f\frac{x_1}{z_1}, r_{1y} = f\frac{y_1}{z_1}$$

Commonly this is represented in pixel coordinates by multiplying the focal length by the pixel pitch $f_x = s_x f$ and $f_y = s_y f$. It is common to add a skew term $\alpha$ to account for skew between the pixel axes. For convenience, we set $\hat{x} = x/z$ and $\hat{y} = y/z$ Finally, we can add in the principal point to get the full mapping from object space to image space:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x\hat{x} + \alpha\hat{y} + c_x \\ f_y\hat{y} + c_y \end{bmatrix} \ or \ \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & \alpha & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \frac{1}{z} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = K\frac{1}{z}p$$

This is known as the pinhole projection equation and the matrix $K$ is known as the intrinsic matrix. The resulting camera model has 5 parameters: two focal length values, two principal point values and 1 skew term.

One unique characteristic of the perspective projection of a pinhole camera is that it preserves linearity. Lines in the world space are also lines in the image. This makes it easy to determine if an image has been corrected to a perspective projection or not. Straight lines appearing as curves is a sign of distortion (relative to a perspective projection), which we will discuss next.

## Deviations from the Pinhole Camera

Now we will introduce models that include distortion. The simplest form of distortion is radial distortion. Radial distortion varies radially away from the principal point as is typically described by a polynomial function in $r$. This polynomial function contains only even powers of $r$ to ensure smoothness and symmetry about the optical axis.

$$R_n = 1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + \cdots$$

$$r = \sqrt{\hat{x}^2 + \hat{y}^2}$$

We can introduce this distortion into the previously defined pinhole model to get a radially distorted pinhole model as follows:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x \hat{x} R_n + \alpha \hat{y} + c_x \\ f_y \hat{y} R_n + c_y \end{bmatrix}$$

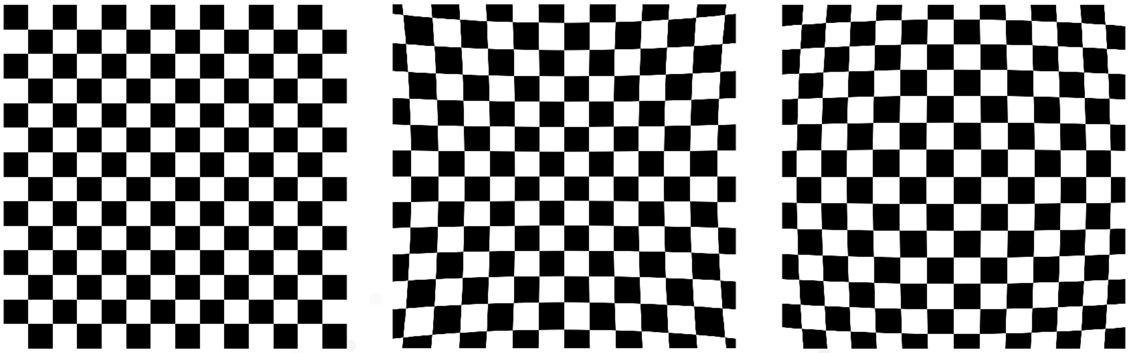

*Figure 5 No distortion (left), pincushion distortion (middle), and barrel distortion (right)*

By introducing the radial distortion model, we can model distortion that varies radially such as the pincushion or barrel distortion. Most implementations of radial distortion only use 3 terms, but it is possible to use more in some applications. Typically, this is not necessary and can result in overfitting of data when calibrating on datasets that are sparse as is true with most traditional calibration methods.

To model more general distortion, additional terms can be added including a denominator polynomial, $R_d$, and tangential distortion terms $T_x$ and $T_y$.

$$T_x = \left(2p_1\hat{x}\hat{y} + p_2(r^2 + 2\hat{x}^2)\right)(1 + p_3r^2 + p_4r^4 + \cdots)$$

$$T_y = \left(2p_2\hat{x}\hat{y} + p_1(r^2 + 2\hat{x}^2)\right)(1 + p_3r^2 + p_4r^4 + \cdots)$$

Some models such as the primary model featured in OpenCV simplifies and decouples the polynomial multiplier as follows:

$$T_{xCV} = \left(2p_1\hat{x}\hat{y} + p_2(r^2 + 2\hat{x}^2)\right) + s_1r^2 + s_2r^4$$

$$T_{yCV} = \left(2p_2\hat{x}\hat{y} + p_1(r^2 + 2\hat{x}^2)\right) + s_3r^2 + s_4r^4$$

As you can see, this is a slightly different implementation of the tangential distortion function.

This model can be made even more general by the addition of a denominator polynomial. This denominator polynomial is rarely used with the numerator. This is sometimes referred to as the division model [5].

$$R_d = 1 + d_1r^2 + d_2r^4 + d_3r^6 + \cdots$$

The most general form of the distorted $ftan\theta$ model is:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x & \alpha \\ 0 & f_y \end{bmatrix} \begin{bmatrix} \hat{x}\dfrac{R_n}{R_d} + T_x \\ \hat{y}\dfrac{R_n}{R_d} + T_y \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}$$

This form may be familiar to those who use the OpenCV camera model or are familiar with the work of Brown [2] [6]. With $R_d = 1$, this is known as the Brown-Conrady Model. With $R_n = 1$, this is known as the Division Model.

## Other Central Camera Models

The next common model type is targeted at addressing the distortion of certain types of cameras known as equidistant or $f\theta$ projection. Many wide field of view lens designs use this type of projection to fit very wide field of views onto a reasonably sized sensor.

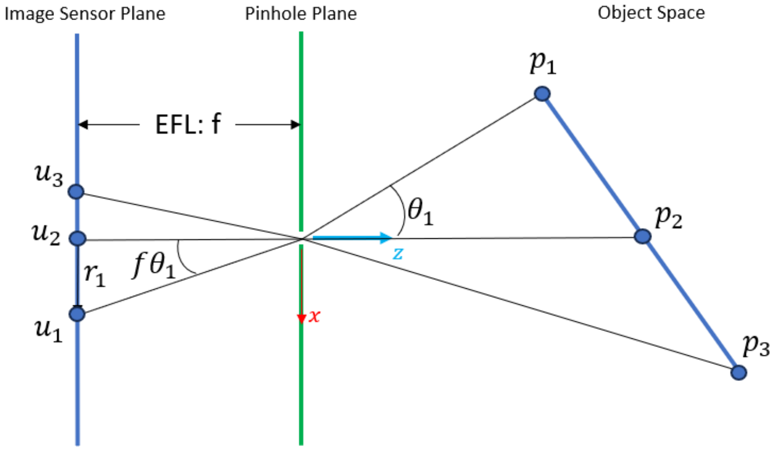

*Figure 6 Ray trace of projection for an f theta camera*

The $f\theta$ form of the projection model can be derived by dividing the original projection function by $tan\theta$ and multiplying by $\theta$. This is equivalent to multiplying by $\frac{\theta}{r}$.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x & \alpha \\ 0 & f_y \end{bmatrix} \frac{\theta}{r} \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}$$

With the addition of distortion from the nominal $f\theta$ projection, we arrive at another well-known model, the Kannala or Kannala Radial model. [3]

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x & \alpha \\ 0 & f_y \end{bmatrix} \frac{\theta}{r} \begin{bmatrix} \hat{x}R_\theta \\ \hat{y}R_\theta \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}$$

Note that for $f\theta$ projection models, it is common to use distortion terms in $\theta$ instead of in $r$.

$$R_\theta = 1 + k_1\theta^2 + k_2\theta^4 + k_3\theta^6 + \cdots$$

This model produces the well-known “fisheye” distortion shown in the image below.

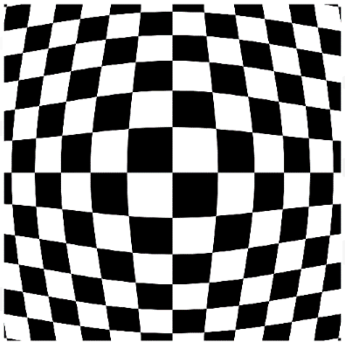

*Figure 7 Fisheye distortion*

In Kannala's paper, a more general version of this model is described which includes asymmetric radial, tangential, and Fourier terms. For brevity, we will not detail them out here, but the original paper discusses them in detail. This model is sometimes referred to as the Kannala Full, or just the Kannala model. A more general version of the equation is:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x & \alpha \\ 0 & f_y \end{bmatrix} \begin{bmatrix} R_\theta + R_{\theta,\phi} & -T_{\theta,\phi} \\ T_{\theta,\phi} & R_\theta + R_{\theta,\phi} \end{bmatrix} \frac{\theta}{r} \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} + \begin{bmatrix} c_x \\ c_y \end{bmatrix}$$

## Beyond the Central Projection Model

More complex camera systems exist that introduce mirrors and other optical elements into a traditional camera system. More specialized models exist for some of these systems where a central projection model may be too simplistic to capture the projection behavior of the system. Though these models are more complex, many of them can still be calibrated in the same manner that will be discussed in the following document in this series.

## Conclusion

We have discussed the fundamentals of camera models and described the most commonly used camera models in detail. These models were, the Pinhole model, Brown-Conrady Model, and the Kannala model. A variety of variations upon these models such as the division model or the Kannala Full model were also discussed. The next topic is naturally on how to calibrate a camera with these models and finally how to use the calibrated results. Please read our other Technical Briefs on camera calibration and applications.

# References

[1] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, no. 11, pp. 1330-1334, Nov 2000.

[2] D. C. Brown, "Close-Range Camera Calibration," *Photogrammetric Engineering,* vol. 37, pp. 855-866, 1971.

[3] J. Kannala and S. S. Brandt, "A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, no. 8, pp. 1335-1340, Aug 2006.

[4] J. Moravec, "A polynomial-division based correction model for camera calibration: a large comparative study," *Sādhanā ,* vol. 45, no. 92, 2020.

[5] R. Hartley and A. Zisserman, Multiple View Geometry, Cambridge University Press, 1999.

[6] "Camera Calibration and 3D Reconstruction," OpenCV, [Online]. Available: https://docs.opencv.org/3.4/d9/d0c/group__calib3d.html. [Accessed 3 12 2024].

## Glossary

back-projection – a mapping from an N-1 dimensional space to a N dimensional space. For a camera, this is the process of computing the object space ray associated with a pixel in a camera image. This can only be done up to a scale factor for single camera which is why this results in rays rather than 3D points.

camera calibration - the process of determining the optimal camera models of a mathematical camera model given a set of measurement data

camera model – a mathematical model that represents the forward and backward projection function of a camera system

extrinsic parameters – the set of parameters that allows of the unique determination of a camera 3d location and orientation in space

homogeneous coordinates – coordinates of a projective space appended with an extra parameter that is scale invariant. The extra parameter is often set to 1.

homogeneous transformation matrix – an affine transformation that transforms a homogenous point by rotating and translating

intrinsic parameters – the set of parameters that define the internal camera parameters that realize its projection function

projection – a mapping from an N dimensional space to an N-1 dimension. In the context of a camera, this is a mapping from the 3D world, to the 2D image coordinates.