OpenLensIO Lens Model Version 0.9.0

Key authors: Martin Parsley, Robert North, Hendrik Fehlis, James Uren

23rd October 2024

Contents

Part I
The Model

1 Definitions and conventions

1.1 Definitions and measurement units

Pinhole camera: A simple camera model where all light passes through a single point.
Optical axis: The axis passing through the centre of projection and pinhole.

F	The focal length of lens, in mm, that projects the undistorted image.
w	Width of the image sensor, in mm.
R	Rotation part of the extrinsic matrix.
t	Translation part of the extrinsic matrix.
$[R\|t]_s$	Camera body component of the extrinsic matrix relative to the centre of the camera sensor (imaging plane).
$[\,]_W$	The subscript W denotes coordinates in the world frame. The units are metres and radians.
$[\,]_p$	The subscript p denotes coordinates in the pinhole-model camera frame. The units are metres and radians.
$[\,]_u, [\,]_d$	Subscripts u and d on vectors refer to undistorted and distorted coordinates respectively.
$[\,]_{uc}$	Subscript uc on vectors refer to undistorted coordinates centred at $\begin{bmatrix} \Delta\mathbb{P}_x & \Delta\mathbb{P}_y \end{bmatrix}^T$.
$[\,]_\Omega$	Subscript Ω on vectors denotes overscanned coordinates.
$\begin{bmatrix} \epsilon_x & \epsilon_y \end{bmatrix}_*^T$	In-image (screen) coordinates, where $*$ is a subscript referring to undistorted or distorted coordinates. Both the ϵ_x and ϵ_y are millimetres to correspond with w, and $\begin{bmatrix} 0 & 0 \end{bmatrix}_*^T$ is at the screen centre, as shown in Figure 1. ϵ_x is the coordinate in the horizontal direction +ve to the right, and ϵ_y in the vertical direction +ve downwards. The vector form is denoted $\epsilon_* = \begin{bmatrix} \epsilon_x & \epsilon_y \end{bmatrix}_*^T$.
r_*	The radius (norm) of in-image coordinates is $r_* = \sqrt{\epsilon_x^2 + \epsilon_y^2}$.

1.2 Coordinate frames

World frame

A right-handed world coordinate frame is used to define the pose of all other bodies. This is the world frame in which camera tracking system parameters are provided. The X_W, Y_W, Z_W vectors point in the world Right, Forward and Up directions respectively. Thus Z_W corresponds to height.

Camera frame

The camera frame is located at the effective pinhole location of the camera lens combination. This corresponds to entrance pupil and will vary based on the focus and zoom setting of the lens. The X_p, Y_p, Z_p vectors point Right, Down and Forward with respect to the lens body.

Image (screen) frame

In image space, the camera sensor's 2-dimensional array of photosite locations are defined in mm units relative to the image centre with ϵ_x pointing right and ϵ_y pointing down. Thus the image corresponds to the sensor area.

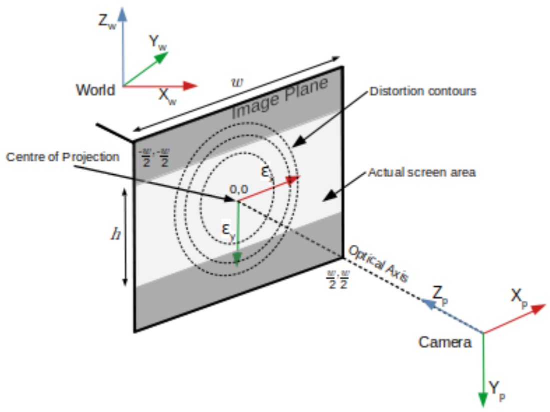

Figure 1: Coordinate frames including image coordinates. By defining the units in mm, means the actual screen area corresponds to the camera sensor area. The origin of the image coordinates is at the centre of the screen, with $-\frac{w}{2}, -\frac{w}{2}$ and $\frac{w}{2}, \frac{w}{2}$ at opposite diagonals for a square screen. As the axes are equal, the coordinates in the visible screen corners will extend from $\left(-\frac{w}{2}, -\frac{h}{2}\right)$ to $\left(\frac{w}{2}, \frac{h}{2}\right)$, while circular distortion appears circular.

1.3 Parameters of the OpenLensIO model

The OpenLensIO model defines a set of parameters to represent a compound lens with radial and decentering distortion. The static parameter w and 13 dynamic parameters F, $\Delta\mathbb{C}_x$, $\Delta\mathbb{C}_y$, $\Delta\mathbb{P}_x$, $\Delta\mathbb{P}_y$, $\mathtt{k}_{1\ldots6}$, z_{epd}, Φ and optionally Ω, $\mathit{\Omega}_{max}$, p_1, p_2. Dynamic parameters may vary based on the state of the lens (including zoom, focus and iris positions), whereas static parameters are independent of this. The OpenLensIO parameters are defined in the following sections.

1.4 Summary model equations

World coordinates are $\begin{bmatrix} x & y & z & 1 \end{bmatrix}_W^T$.

Conversion into camera frame is

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}_p = [R|t]_s \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}_W - \begin{bmatrix} 0 \\ 0 \\ z_{epd} \end{bmatrix}. \tag{1}$$

where $[R|t]_s$ is the extrinsic matrix component at the camera sensor, generally derived from the tracking system pose, and z_{epd} is the entrance pupil distance. We assume that transverse offsets of the entrance pupil (x_{epd} and y_{epd}) are negligible compared to the longitudinal movement (z_{epd}), and their effects are adequately modelled by the perspective shift terms $\Delta\mathbb{P}$, defined below.

In the pinhole camera model, the coordinates are projected on an image plane (normalised by depth) as follows;

$$\begin{bmatrix} x \\ y \end{bmatrix}_u = \frac{1}{z_p} \begin{bmatrix} x \\ y \end{bmatrix}_p, \tag{2}$$

whose corresponding undistorted screen coordinates in units of mm are

$$\epsilon_u = \begin{bmatrix} \epsilon_x \\ \epsilon_y \end{bmatrix}_u = F \begin{bmatrix} x \\ y \end{bmatrix}_u. \tag{3}$$

In computer graphics applications it is often convenient to convert between screen coordinates, for example when implementing distortion in a shader. The general form for the conversion of screen coordinates is given by

$$\epsilon_u = D\left(\epsilon_d - \Delta\mathbb{C}\right) + \Delta\mathbb{C} - \Delta\mathbb{P} \tag{4}$$

Where $\Delta\mathbb{C} = \left[\begin{array}{cc} \Delta\mathbb{C}_x & \Delta\mathbb{C}_y \end{array}\right]^T$ represents a shift of the distortion centre in units of mm, and $\Delta\mathbb{P} = \left[\begin{array}{cc} \Delta\mathbb{P}_x & \Delta\mathbb{P}_y \end{array}\right]^T$ represents a perspective shift from the centre of the screen, also in units of mm. In essence the distortion centre shift $\Delta\mathbb{C}$ allows for shifting the centre of distortion from the centre of the screen, while the perspective shift $\Delta\mathbb{P}$ allows for shifting the centre of perspective.

The perspective shift is applied on the distortion side of the model to enable use environments where perspective shift cannot be applied to the projection matrix, e.g. a virtual camera specified only in terms of pose and field of view.

$D\left(\epsilon\right)$ is the closed-form distortion function, defined below. This distorted to undistorted coordinates direction is convenient for implementing in a GPU shader. The inverse of this is

$$\epsilon_d = D^{-1}\left(\epsilon_u + \Delta\mathbb{P} - \Delta\mathbb{C}\right) + \Delta\mathbb{C}, \tag{5}$$

which can be solved using numerical iterative methods depending on the application.

1.5 Distortion model

The function $D\left(\epsilon\right)$ is the Brown-Conrady distortion model,

$$D\left(\boldsymbol{\epsilon}\right) = \mathrm{diag}\left[\begin{array}{c} R + 2p_1\epsilon_y + p_2\left(\frac{1}{\epsilon_x}r^2 + 2\epsilon_x\right) \\ R + 2p_2\epsilon_x + p_1\left(\frac{1}{\epsilon_y}r^2 + 2\epsilon_y\right) \end{array}\right]\boldsymbol{\epsilon} \tag{6}$$

where R is the radial distortion term

$$R = \frac{1 + \mathtt{k}_1 r^2 + \mathtt{k}_3 r^4 + \mathtt{k}_5 r^6}{1 + \mathtt{k}_2 r^2 + \mathtt{k}_4 r^4 + \mathtt{k}_6 r^6} \tag{7}$$

with radial distortion coefficients $\mathtt{k}_{1,3,5}$ in units of mm^{-n-1}, $\mathtt{k}_{2,4,6}$ in units of mm^{-n}, and p_n are the decentering coefficients in units of mm^{-2}, and where the radius $r = \sqrt{\epsilon_x^2 + \epsilon_y^2}$. Note that coefficient subscripts are alternating numerator and denominator in order that the distortion model is extensible to higher powers.

1.6 How overscan is applied.

A virtual camera may produce an image with unrendered areas when distortion is applied, for instance in the corners for a lens with barrel distortion. To remedy this, we apply overscan; where the virtual camera requires a larger image plane and thus a wider field of view than the pinhole camera. Post rendering, in order for the field of view of the final rendered image to match that of the real camera, it is necessary to zoom into the image by the inverse of the overscan factor.

The overscanned field of view of the virtual camera is given by

$$\theta_\Omega = 2\ tan^{-1}\left(\frac{w\Omega}{2F}\right), \tag{8}$$

where $w\Omega$ can be thought of as the overscanned sensor width.

Then post-rendering, the screen coordinates are scaled by the inverse of the overscan factor

$$\epsilon_\Omega = \frac{1}{\Omega}\left(D\left(\epsilon_d - \Delta\mathbb{C}\right) + \Delta\mathbb{C} - \Delta\mathbb{P}\right). \tag{9}$$

Overscan is typically either provided as a constant multiplier, or computed by approximation.

To help users that require a constant multiplier, OpenLensIO optionally provides Ω_{max} as a static parameter, the largest Ω across all zoom and focus settings. To preserve quality, it may be necessary to render the undistorted image with a higher screen percentage than the distorted image.

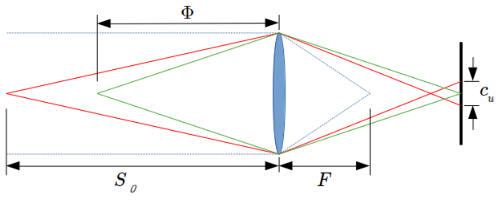

Figure 2: Diagram showing the circle of confusion, c_u for a simple paraxial optical system.

1.7 Other screen units

Equation (9) gives the screen coordinates in mm relative to the centre of the screen, however for implementation in shaders the screen coordinates need to be changed to be in shader units. For example, for a square texture of width w_{shader} with origin at the top left, the ϵ coordinates (in mm) can be converted into shader coordinates, ε_{shader}, by

$$\varepsilon_{shader} = w_{shader} \begin{bmatrix} \frac{\epsilon_x}{w} \\ \frac{\epsilon_y}{h} \end{bmatrix} + \frac{w_{shader}}{2}, \tag{10}$$

w and h are the sensor width and height in mm.

1.8 Aperture

The accuracy of the following accepted expression for circle of confusion is currently under investigation:

$$c_u = \frac{|S_o - \Phi|}{S_o} \frac{F^2}{N\,(\Phi - F)} \tag{11}$$

where c_u is the circle of confusion in mm, in undistorted screen space, N is the f-number of the lens (f-stop), F is the focal length, Φ is the focal distance and S_o is the object distance (which comes from the scene), as shown in Figure 2.

Note this requires the f-stop (aperture) of the lens. T-stop (transmission) cannot be used directly in this equation. Conversion from t-stop to f-stop is normally via a lookup table that can be provided by the lens manufacturer.

A Informative sections

A.1 Alternate model for distortion and pinhole camera.

Sometimes it is necessary to represent the virtual camera in terms of a perspective matrix. The following model is equivalent in terms of geometry, but gives different screen coordinates to render. This representation permits the centre of projection to be included in the perspective matrix.

In the context of the perspective matrix, θ_Ω cannot be derived and so cameras defined only by pose and angle of view should use the main model.

In this section we use the subscript a to signify the alternate representation.

$$\epsilon_{ua} = \left(F \begin{bmatrix} x \\ y \end{bmatrix}_u + \Delta\mathbb{P} \right). \tag{12}$$

$$\epsilon_{ua} = D\left(\epsilon_d - \Delta\mathbb{C}\right) + \Delta\mathbb{C} \tag{13}$$

$$\epsilon_d = D^{-1}\left(\epsilon_{ua} - \Delta\mathbb{C}\right) + \Delta\mathbb{C} \tag{14}$$

Here we have moved the perspective shift from the distortion equation.

With these equations, we can still define the angle of view

$$\frac{r_{ua}}{F} = tan(\frac{\alpha}{2}), \tag{15}$$

but now r_{ua} is the distance of a point ϵ_{ua} to the point $\Delta\mathbb{P}$.

Now overscan is applied differently. From this equation, the projection matrix with overscan can be derived:

$$\epsilon_{\Omega a} = \frac{1}{\Omega_a} \left(F \begin{bmatrix} x \\ y \end{bmatrix}_u + \Delta\mathbb{P} \right).$$

A similar undistort transform, with overscan is

$$\epsilon_{\Omega_a} = \frac{1}{\Omega_a}\left(D\left(\epsilon_d - \Delta\mathbb{C}\right)\right). \tag{16}$$

It can be proven that these equations can generate equivalent renders, when overscan is applied, **but** overscan computed on one form is not guaranteed to fill the screen of the other form.

A.2 Specifying ideal overscan

The minimum required overscan Ω and resulting field of view θ_Ω are defined as follows.

If we define the set of coordinates inside the distorted screen rectangle to be

$$\{\boldsymbol{\epsilon_d^i}\} = \{\epsilon_d : -\frac{w}{2} < \epsilon_d.x < \frac{w}{2} \wedge -\frac{h}{2} < \epsilon_d.y < \frac{h}{2}\}. \tag{17}$$

Then we define $\{\boldsymbol{\epsilon_u^i}\}$ to be the subset of corresponding undistorted coordinates

$$\{\boldsymbol{\epsilon_u^i}\} = \{\boldsymbol{\epsilon_u^i} : \boldsymbol{\epsilon_u^i} = \left(D\left(\boldsymbol{\epsilon_d^i} - \Delta\mathbb{C}\right) + \Delta\mathbb{P}\right)\}. \tag{18}$$

Then the smallest image width, w_Ω that ensures that the overscanned region contains all of $\{\boldsymbol{\epsilon_u^i}\}$ given by

$$w_\Omega = \quad 2\ max\left(\{max_i\left(\{\left|\epsilon_x^i\right|\}\right), \tfrac{w}{h} max_i\left(\{\left|\epsilon_y^i\right|\}\right)\}\right). \tag{19}$$

From this we can derive the overscan factor, Ω,

$$\Omega = \frac{w_\Omega}{w}. \tag{20}$$

This can then be used as defined in section 1.6.

B Future additions

B.1 Entrance pupil off-axis shift, and optical axis rotation.

As currently the entrance pupil off-axis shift and rotation has been ignored. The off-axis shift could be considered in order to better represent wide angle lenses when imaging something 1 metre or closer, and for rotation, may be necessary to express misalignment of lens, either internal to lens or lens to camera body.

B.2 Anamorphic lens distortion

To undistort an anamorphic lens, we could use a distortion model (just showing radial terms for readability, tangential terms can be added):

$$\begin{bmatrix} x \\ y \end{bmatrix}_u = \begin{bmatrix} \frac{x_d}{1+\left(\frac{k_1}{A_{sq}}\right)r^2+\left(\frac{k_2}{A_{sq}}\right)r^4+\left(\frac{A_x}{A_{sq}}\right)y_d^2} \\ \frac{y_d}{1+k_1r^2+k_2r^4+A_yx_d^2} \end{bmatrix} \tag{21}$$

where k_1, k_2 are the radial distortion terms, A_x, A_y are the Asymmetric distortion terms and A_{sq} is anamorphic squeeze. The rest of the pipeline remains unchanged.

B.3 Chromatic aberration

Axial (longitudinal)

Most noticeable at long focal lengths. A lens with specific refractive index cannot effectively “focus” all wavelengths of light at the same time. If blue is in focus, red and green will be out of focus and will appear in a different image location. Could be modelled by having three sets of focal distances.

Transverse (lateral)

Magnification and distortion of the lens is also dependent on wavelength. Could be modelled with three sets of distortion parameters, one for each primary colour.

B.4 Vignetting

Natural

Light that hits the sensor at an angle has lower intensity based on Cosine fourth law of optical falloff. Proposition coming in a future revision.

Mechanical

This occurs when physical lens components are in the way of the lens. To be considered (even if negligible) in a future revision.

Optical

This occurs when stacking multiple lenses which reduces the intensity of the light towards the outside of the image. This could be modelled as simple inverse squared or better a low order polynomial

$$v_n(x) = 1 - (\alpha_1 x^2 + \alpha_2 x^4 + \alpha_3 x^6) \tag{22}$$