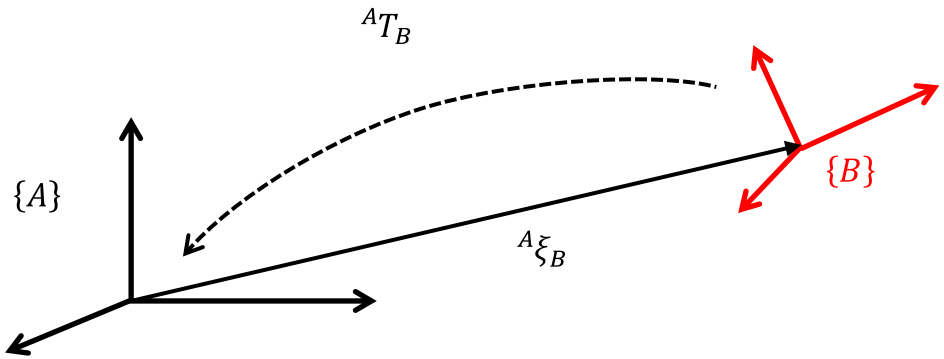
Lecture 1.4
The perspective camera model

Thomas Opsahl

Recap

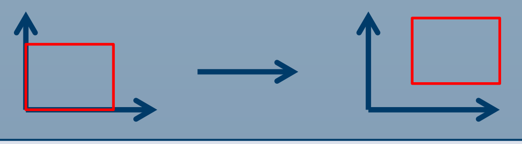
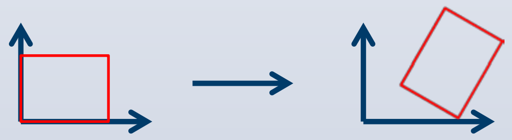
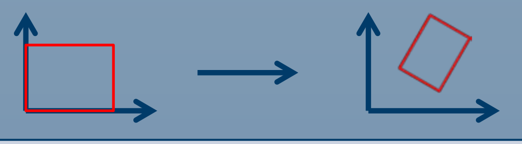
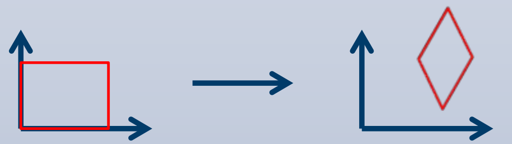
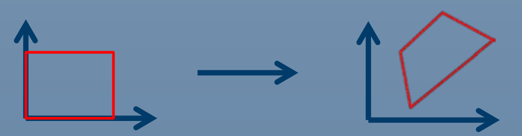


- The pose of a coordinate frame $\{B\}$ relative to a coordinate frame $\{A\}$, denoted ${}^{A}\xi_{B}$, can be represented as a homogeneous transformation ${}^{A}T_{B}$

$$ {}^{A}\xi_{B} \quad \mapsto \quad {}^{A}T_{B} = \begin{bmatrix} {}^{A}R_{B} & {}^{A}\boldsymbol{t}_{B} \\ \boldsymbol{0} & 1 \end{bmatrix} $$

$$ {}^{A}\xi_{B} \cdot {}^{B}\boldsymbol{p} = {}^{A}\boldsymbol{p} \quad \mapsto \quad {}^{A}\tilde{\boldsymbol{p}} = {}^{A}T_{B}\, {}^{B}\tilde{\boldsymbol{p}} = \begin{bmatrix} {}^{A}R_{B} & {}^{A}\boldsymbol{t}_{B} \\ \boldsymbol{0} & 1 \end{bmatrix} \begin{bmatrix} {}^{B}\boldsymbol{p} \\ 1 \end{bmatrix} $$

Recap

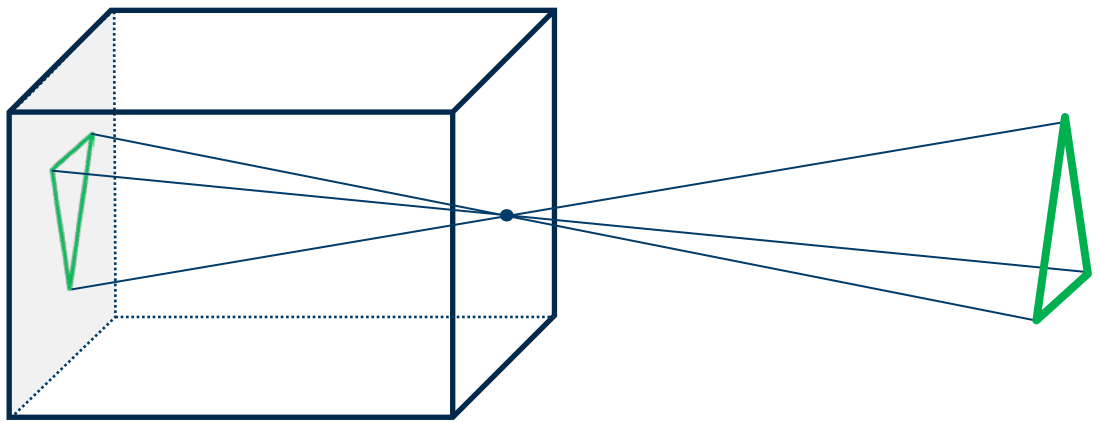
Transformation of $\mathbb{P}^2$	Matrix	#DoF	Preserves	Visualization
Translation	$\begin{bmatrix} I & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$	2	Orientation + all below	
Euclidean	$\begin{bmatrix} R & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$	3	Lengths + all below	
Similarity	$\begin{bmatrix} sR & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$	4	Angles + all below	
Affine	$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix}$	6	Parallelism + all below	
Homography /projective	$\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$	8	Straight lines	

Recap

Transformation of $\mathbb{P}^3$	Matrix	#DoF	Preserves
Translation	$\begin{bmatrix} I & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$	3	Orientation + all below
Euclidean	$\begin{bmatrix} R & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$	6	Lengths + all below
Similarity	$\begin{bmatrix} sR & \boldsymbol{t} \\ \mathbf{0}^T & 1 \end{bmatrix}$	7	Angles + all below
Affine	$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}$	12	Parallelism + all below
Homography /projective	$\begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \\ h_{41} & h_{42} & h_{43} & h_{44} \end{bmatrix}$	15	Straight lines

The perspective camera

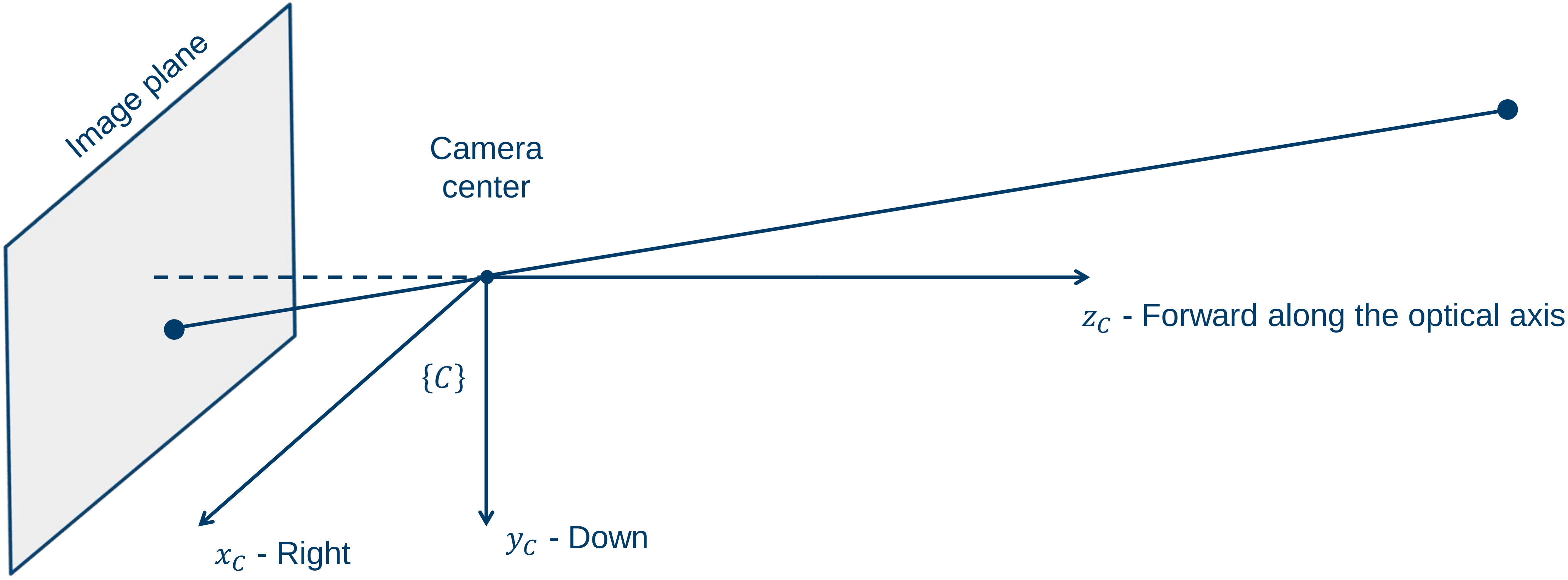
- The perspective camera – or pinhole camera – is a simple imaging device



- The perspective camera model is a mathematical model describing the correspondence between observed points in the world and pixels in the captured image

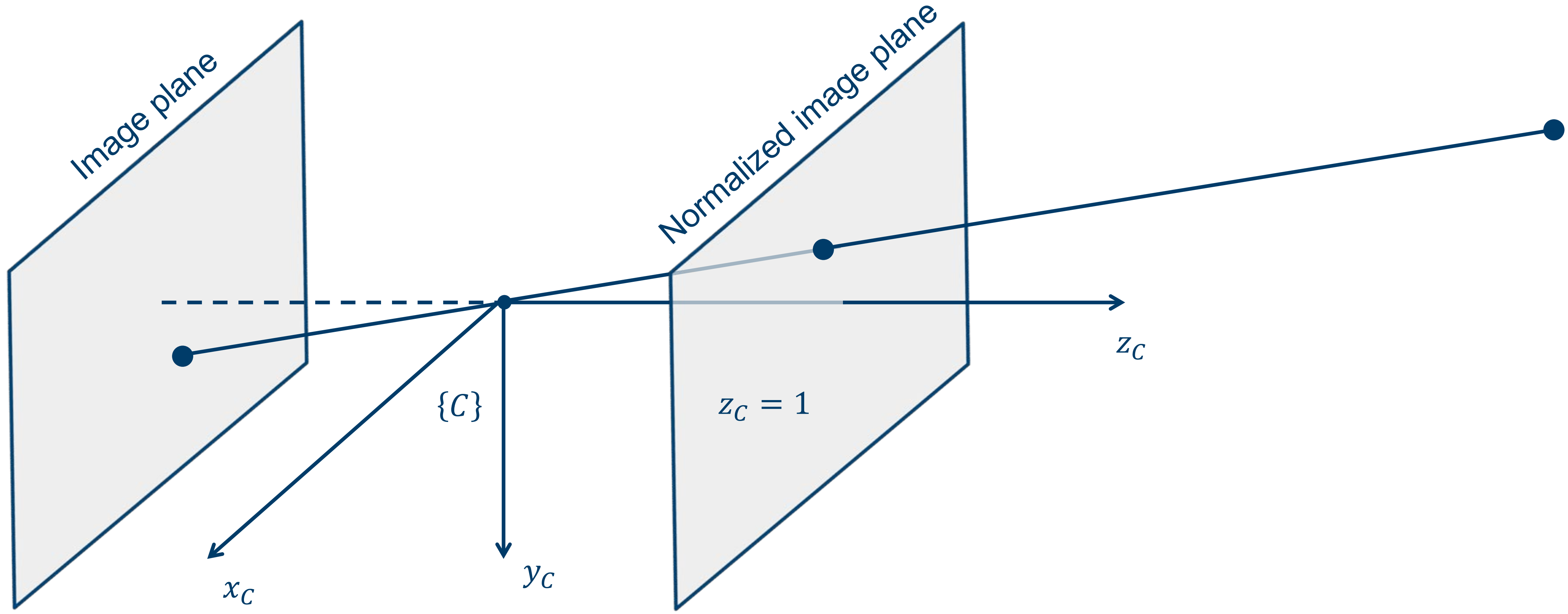
- To describe the transformation from 3D points in the world to 2D points in an image, we need to represent the camera by a coordinate frame

The perspective camera model

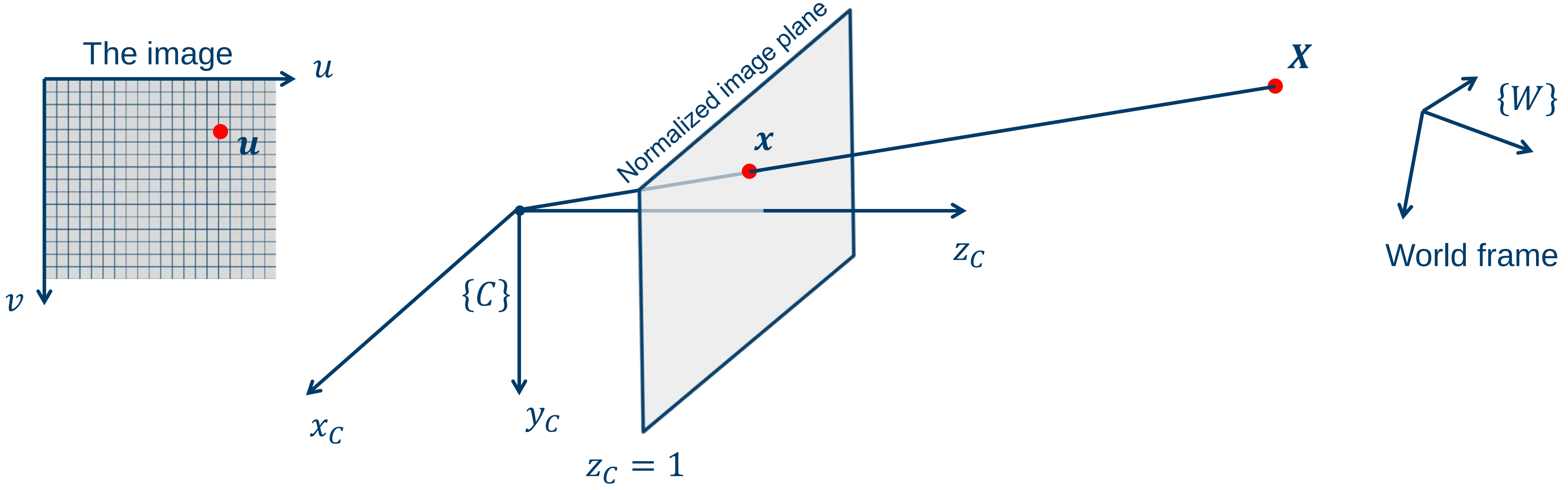


Camera coordinate frame

The perspective camera model



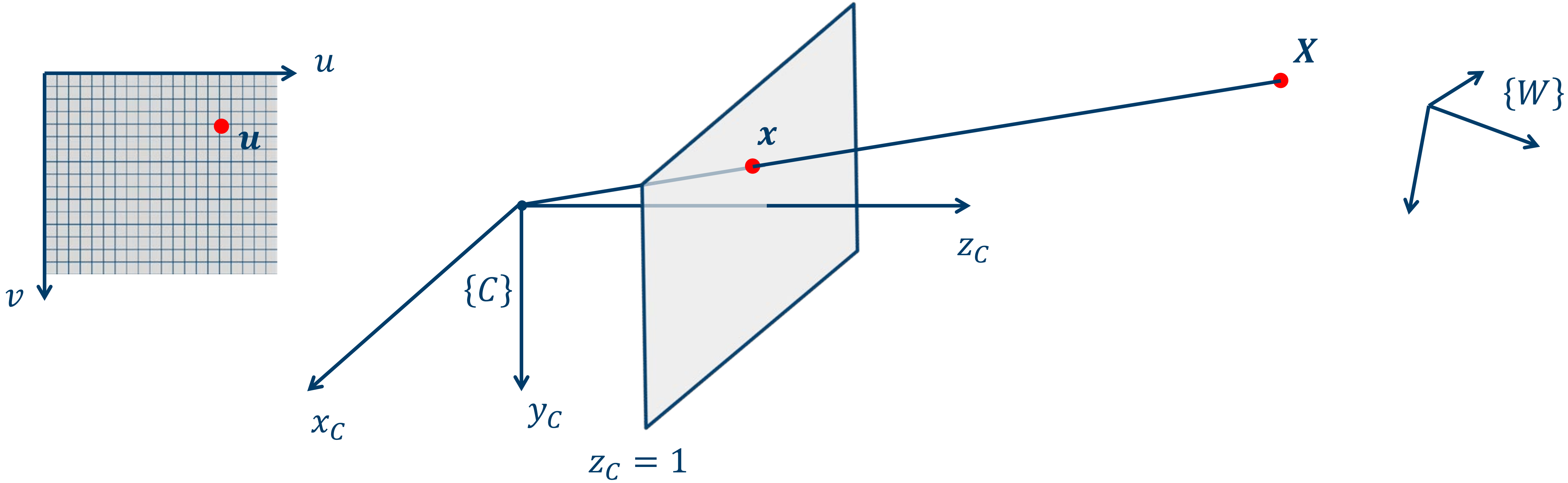
The perspective camera model



- It is natural to divide the perspective camera model into two parts
  - Extrinsic: ${}^{W}\boldsymbol{X} \mapsto {}^{C}\boldsymbol{x}$ 3D→2D
  - Intrinsic: ${}^{C}\boldsymbol{x} \mapsto \boldsymbol{u}$ 2D→2D

- Both parts are commonly represented by a homogeneous matrix

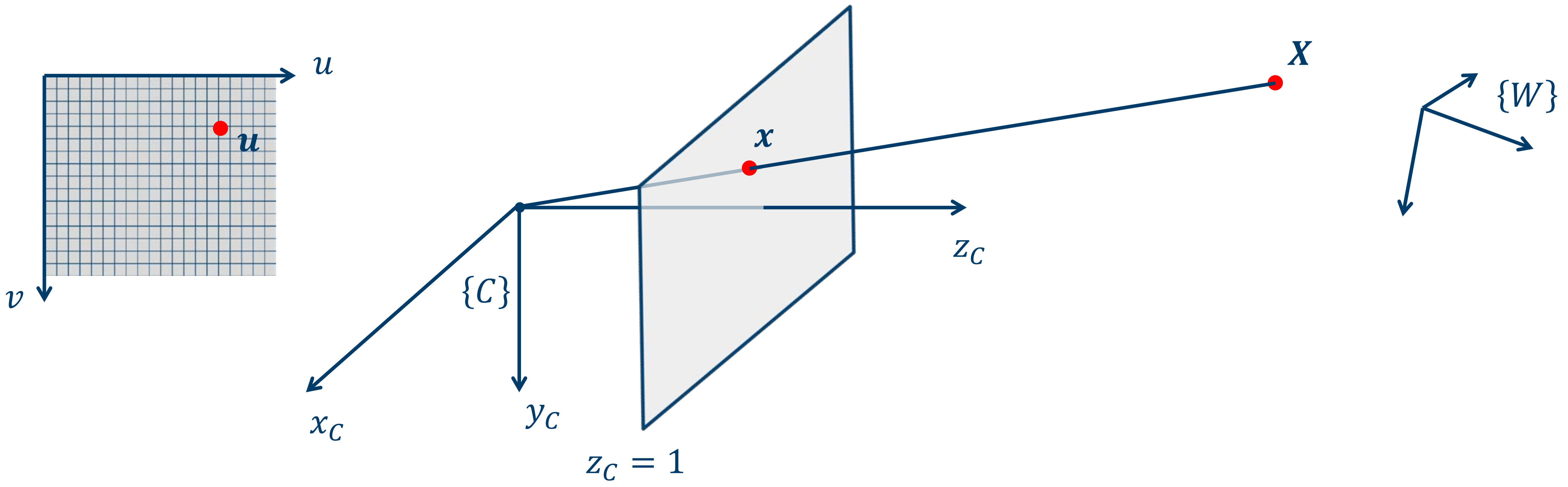
The perspective camera model



- The perspective camera model is typically presented like this

$$\tilde{\boldsymbol{u}} = K[R \quad \boldsymbol{t}]^W\tilde{\boldsymbol{X}}$$

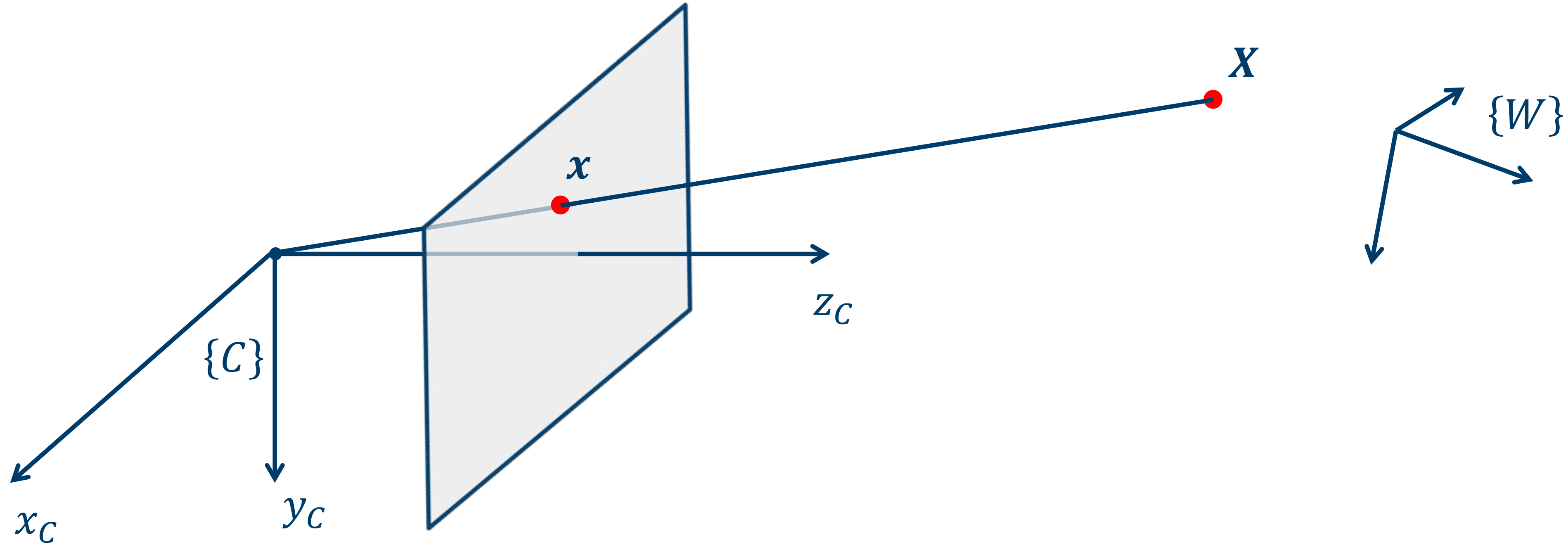
The perspective camera model



- A more detailed version reveals the typical parameters used to characterize perspective cameras

$$\tilde{\boldsymbol{u}} = \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_{3\times3} & \boldsymbol{t}_{3\times1} \\ \boldsymbol{0}_{1\times3} & 1 \end{bmatrix} {}^{W}\tilde{\boldsymbol{X}}$$

Understanding the extrinsic part of the model



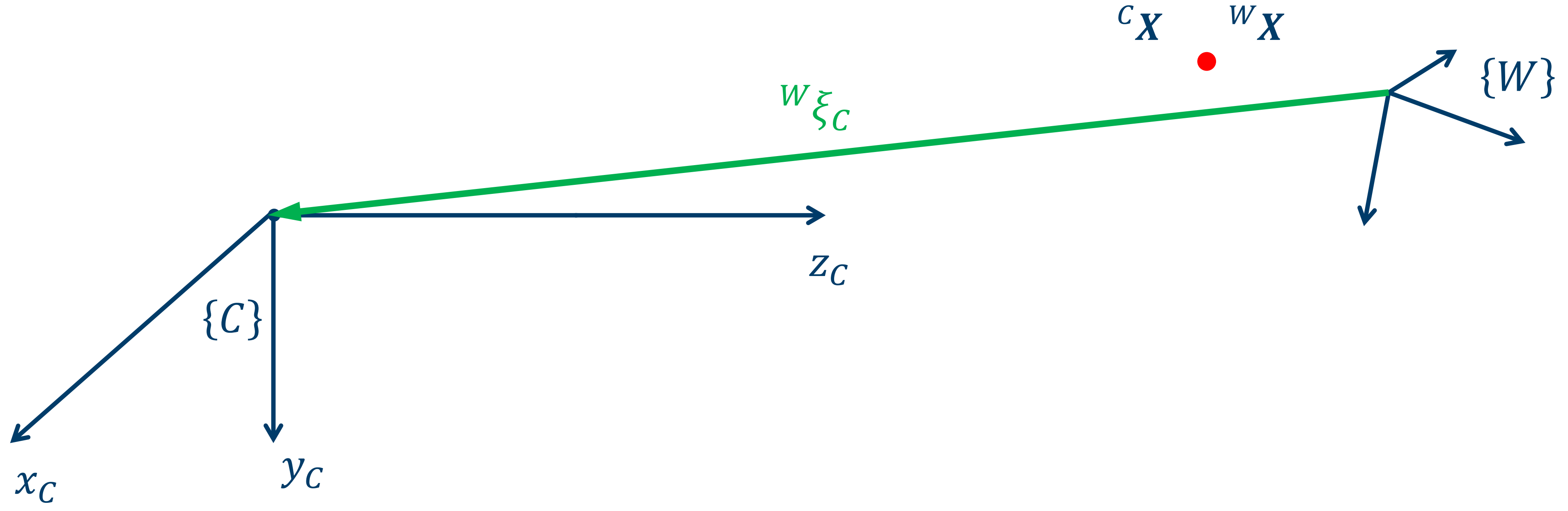
- The extrinsic part of the perspective camera model is composed by

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R_{3\times 3} & \boldsymbol{t}_{3\times 1} \\ \boldsymbol{0}_{1\times 3} & 1 \end{bmatrix}$$

The perspective projection from 3D to 2D

The Euclidean transformation of points from $\{W\}$ to $\{C\}$

Understanding the extrinsic part of the model

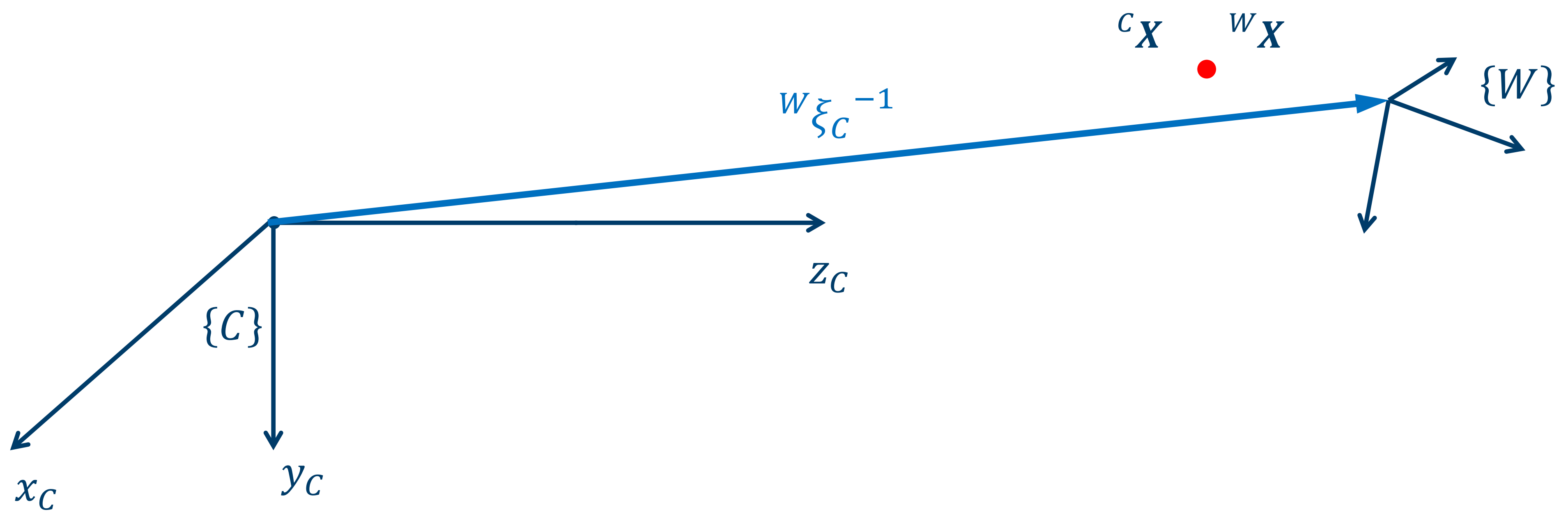


- Recall that ${}^{W}\xi_{C}$ – the pose of the camera relative to the world frame – can be represented by a homogeneous transformation of points from $\{C\}$ to $\{W\}$

$$
{}^{W}\xi_{C} = \begin{bmatrix} {}^{W}R_{C} & {}^{W}\boldsymbol{t}_{C} \\ \boldsymbol{0}_{1\times 3} & 1 \end{bmatrix}
$$

$$
{}^{W}\tilde{\boldsymbol{X}} = {}^{W}\xi_{C}\,{}^{C}\tilde{\boldsymbol{X}}
$$

Understanding the extrinsic part of the model

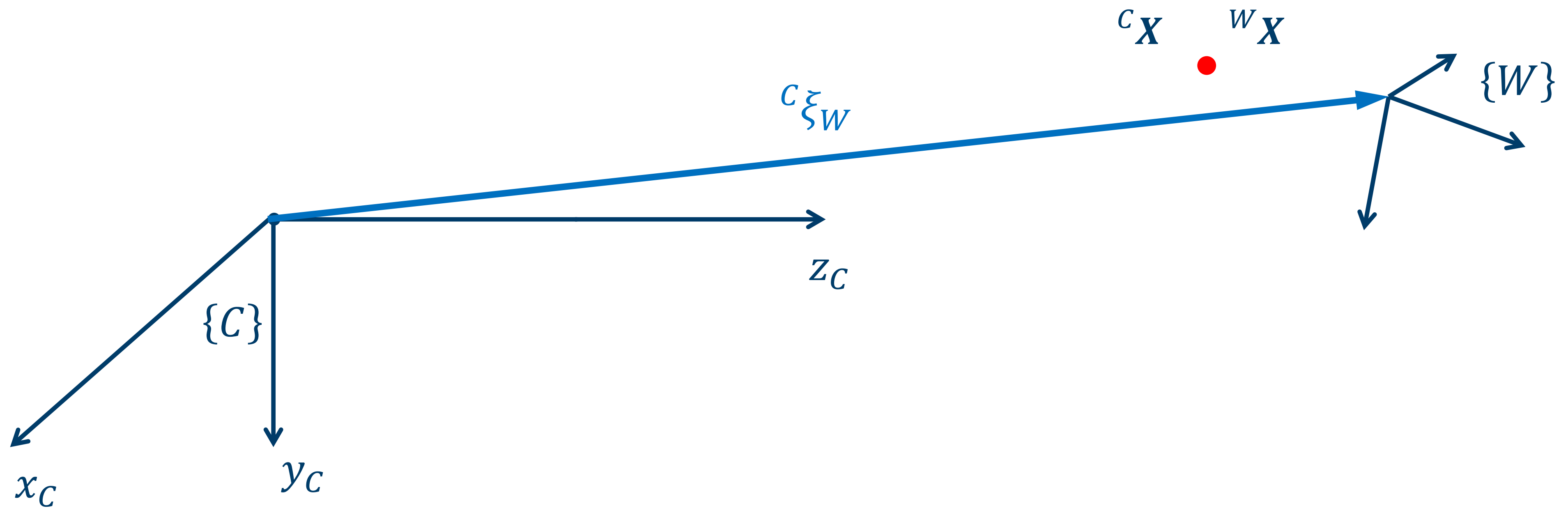


- Hence we can express the Euclidean transformation from $\{W\}$ to $\{C\}$ in terms of the cameras pose relative to the world frame

$$\begin{bmatrix} R_{3\times3} & \boldsymbol{t}_{3\times1} \\ \boldsymbol{0}_{1\times3} & 1 \end{bmatrix} = {}^{W}\xi_{C}{}^{-1} = \begin{bmatrix} {}^{W}R_{C} & {}^{W}\boldsymbol{t}_{C} \\ \boldsymbol{0}_{1\times3} & 1 \end{bmatrix}^{-1} = \begin{bmatrix} {}^{W}R_{C}{}^{T} & -{}^{W}R_{C}{}^{T}\,{}^{W}\boldsymbol{t}_{C} \\ \boldsymbol{0}_{1\times3} & 1 \end{bmatrix}$$

$${}^{C}\tilde{\boldsymbol{X}} = {}^{W}\xi_{C}{}^{-1}\,{}^{W}\tilde{\boldsymbol{X}}$$

Understanding the extrinsic part of the model

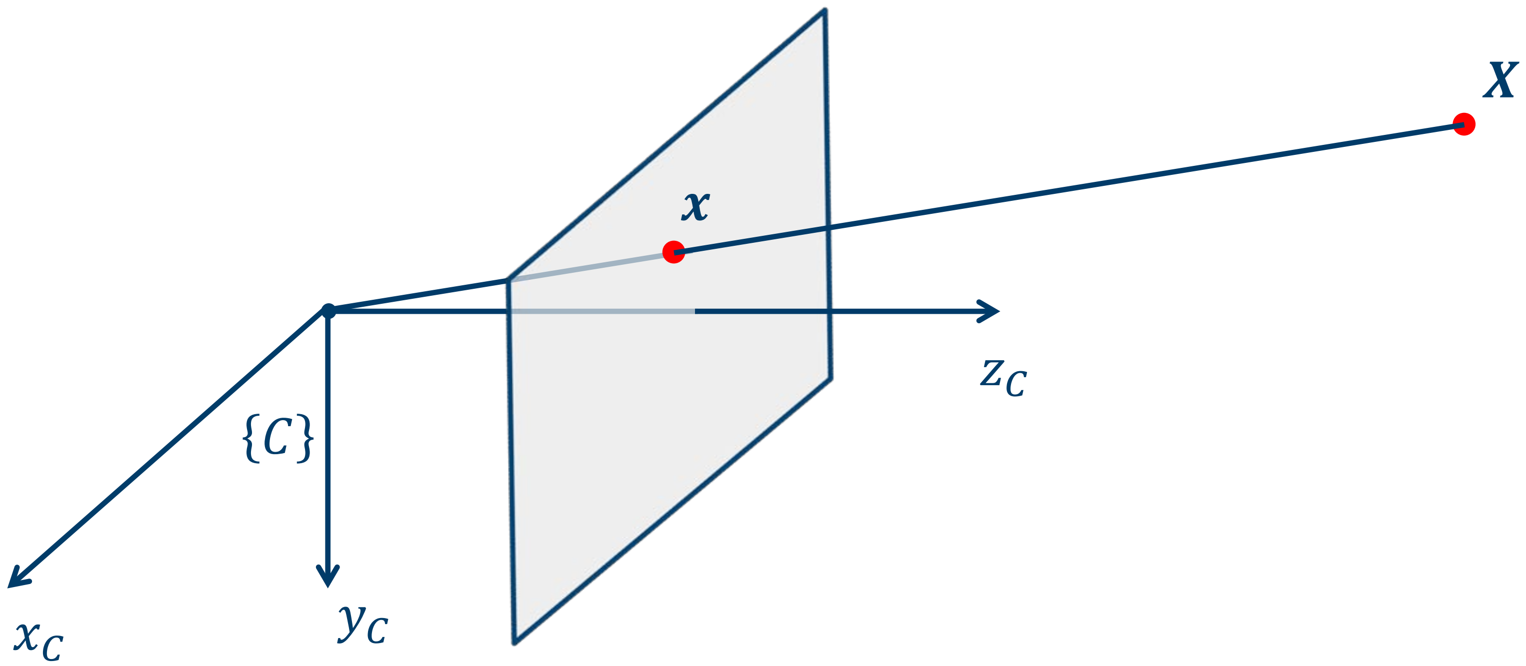


- But it directly represents the pose of the world frame relative to the camera frame

$$\begin{bmatrix} R_{3\times3} & \boldsymbol{t}_{3\times1} \\ \boldsymbol{0}_{1\times3} & 1 \end{bmatrix} = {}^{C}\xi_{W} = \begin{bmatrix} {}^{C}R_{W} & {}^{C}\boldsymbol{t}_{W} \\ \boldsymbol{0}_{1\times3} & 1 \end{bmatrix}$$

$${}^{C}\tilde{\boldsymbol{X}} = {}^{C}\xi_{W}\,{}^{W}\tilde{\boldsymbol{X}}$$

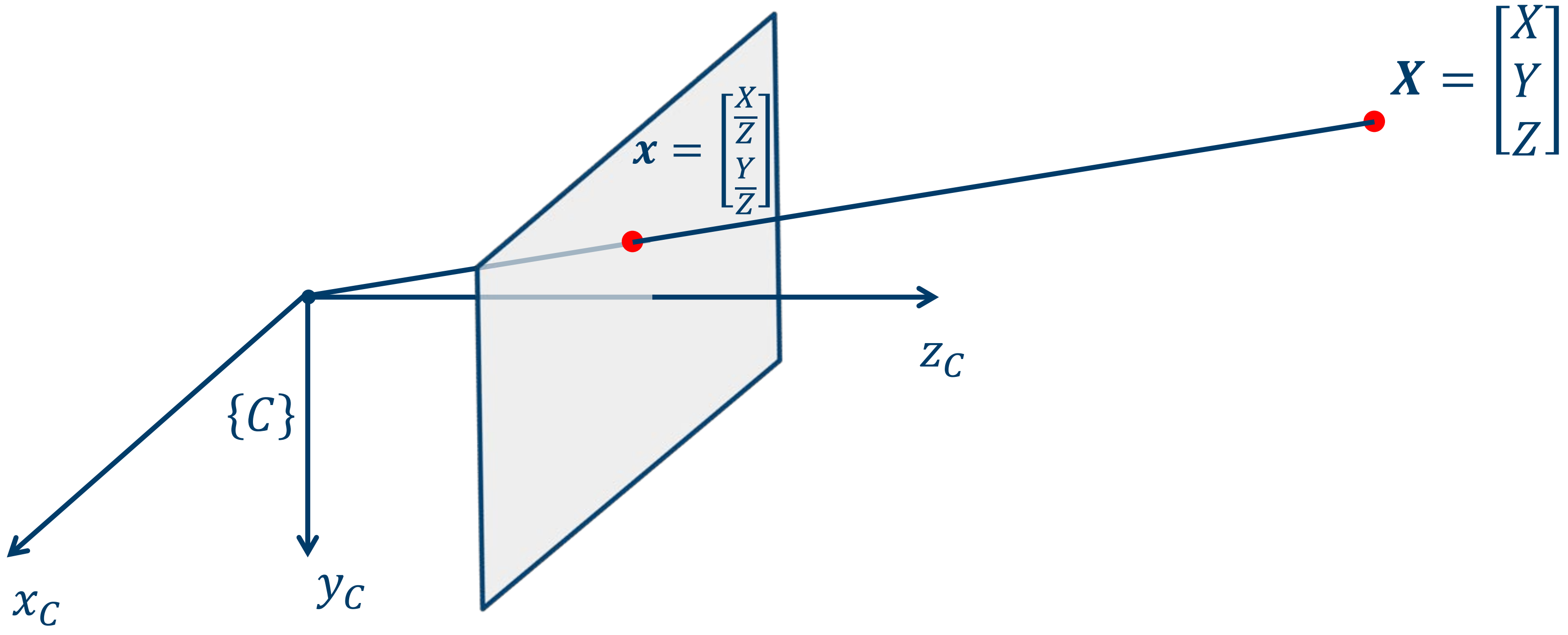
Understanding the extrinsic part of the model



- The perspective projection from 3D to 2D can be represented by the following homogeneous matrix

$$\underbrace{\tilde{\boldsymbol{x}}}_{\in \mathbb{P}^2} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \underbrace{\tilde{\boldsymbol{X}}}_{\in \mathbb{P}^3}$$

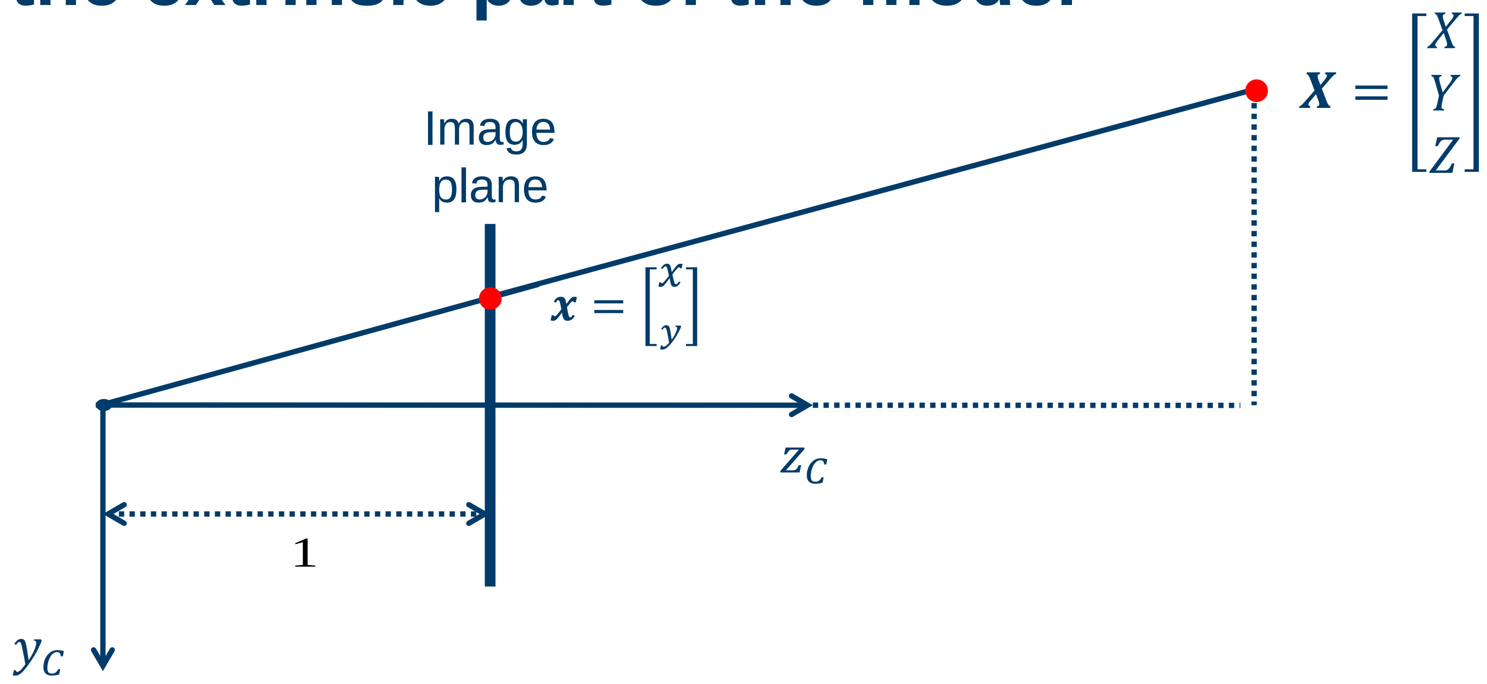
Understanding the extrinsic part of the model



- In coordinates

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \frac{X}{Z} \\ \frac{Y}{Z} \\ 1 \end{bmatrix}$$

Understanding the extrinsic part of the model

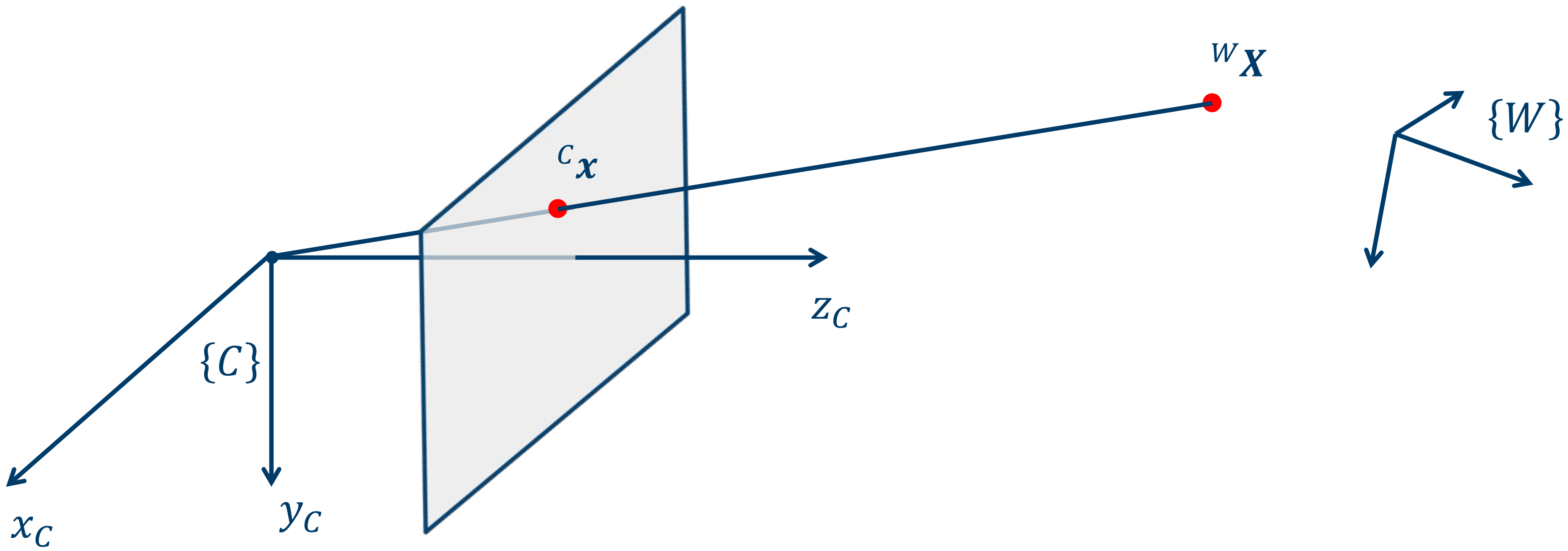


- To see that this is exactly what we want the perspective projection to do, we can take an isolated look at the y- and z- coordinates

- From the two similar triangles in the illustration we see that

$$\frac{y}{Y} = \frac{1}{Z} \Leftrightarrow y = \frac{Y}{Z}$$

Understanding the extrinsic part of the model

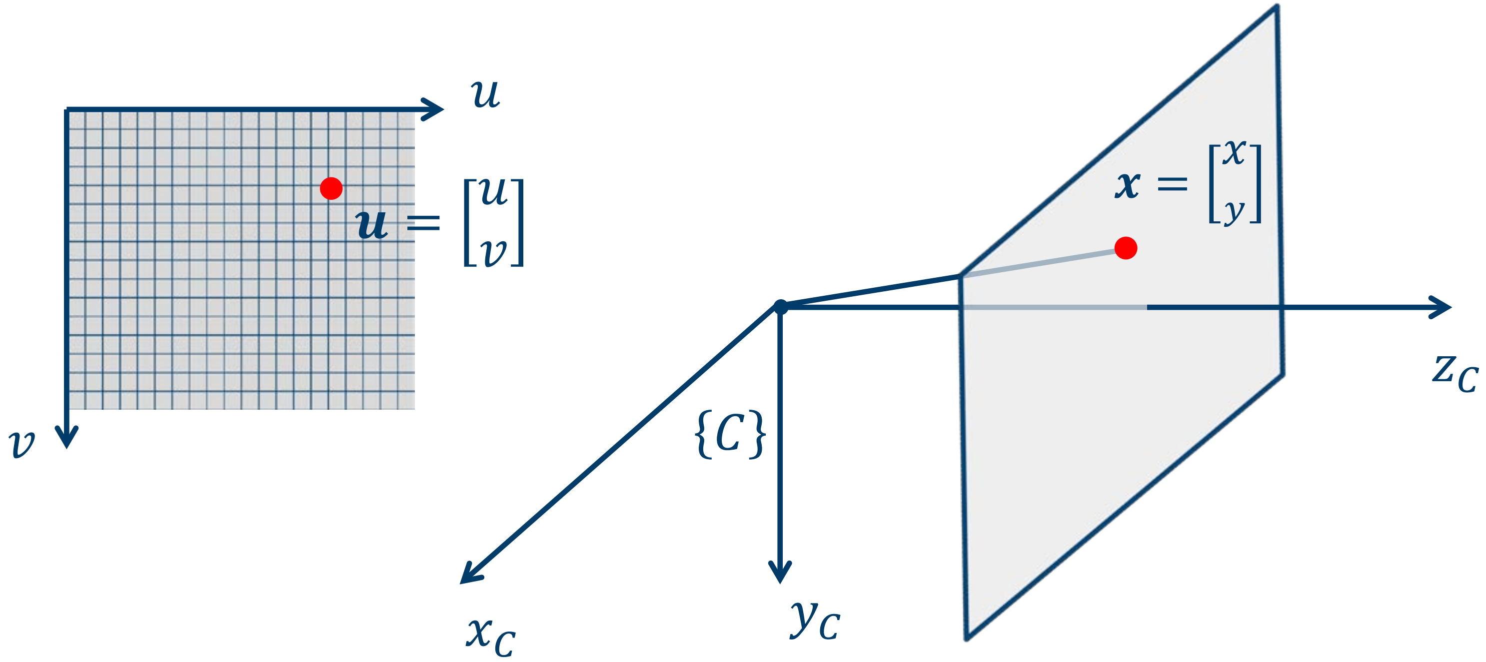


- Combining the perspective projection and the Euclidean coordinate transformation we arrive at the compact representation of the extrinsic part of the perspective camera model

$$[R \quad \boldsymbol{t}] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix}$$

$${}^{C}\tilde{\boldsymbol{x}} = [R \quad \boldsymbol{t}]\,{}^{W}\tilde{\boldsymbol{X}}$$

Understanding the intrinsic part of the model

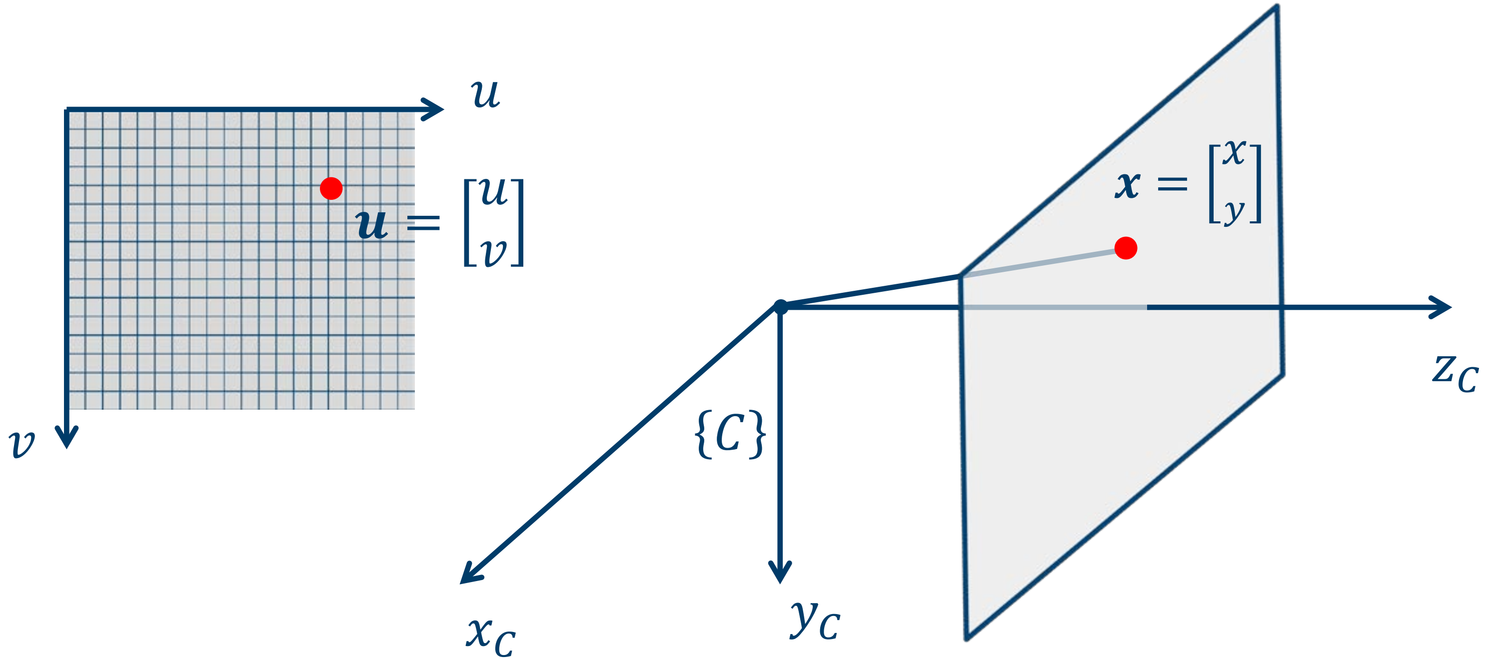


- The intrinsic part of the perspective camera model describes the transformation from normalized image coordinates to image coordinates (often pixels, but not always)

- This transformation is represented by a homogeneous matrix commonly referred to as *the camera calibration matrix* K

$$\tilde{\boldsymbol{u}} = K\tilde{\boldsymbol{x}}$$

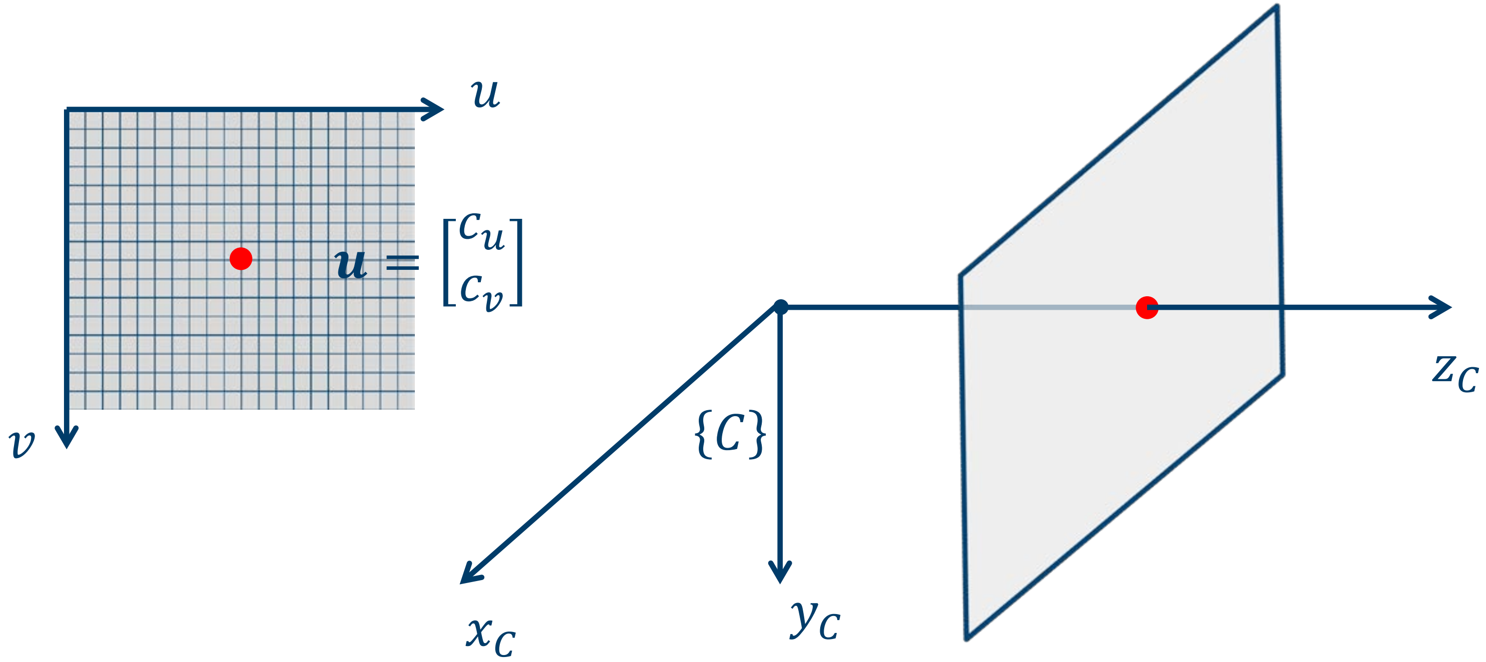
Understanding the intrinsic part of the model



- The camera calibration matrix has 5 parameters describing different physical aspects of the relationship between the image projected onto the normalized image plane and the sensor array that produces the image

$$K = \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix}$$

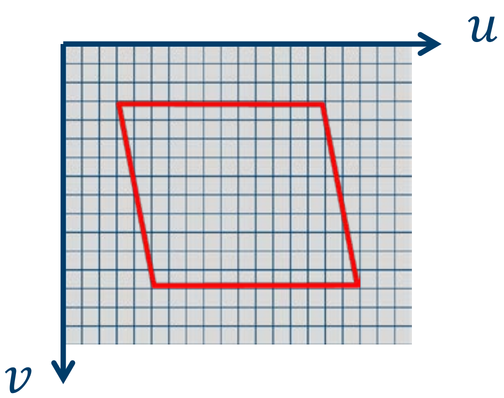
Understanding the intrinsic part of the model

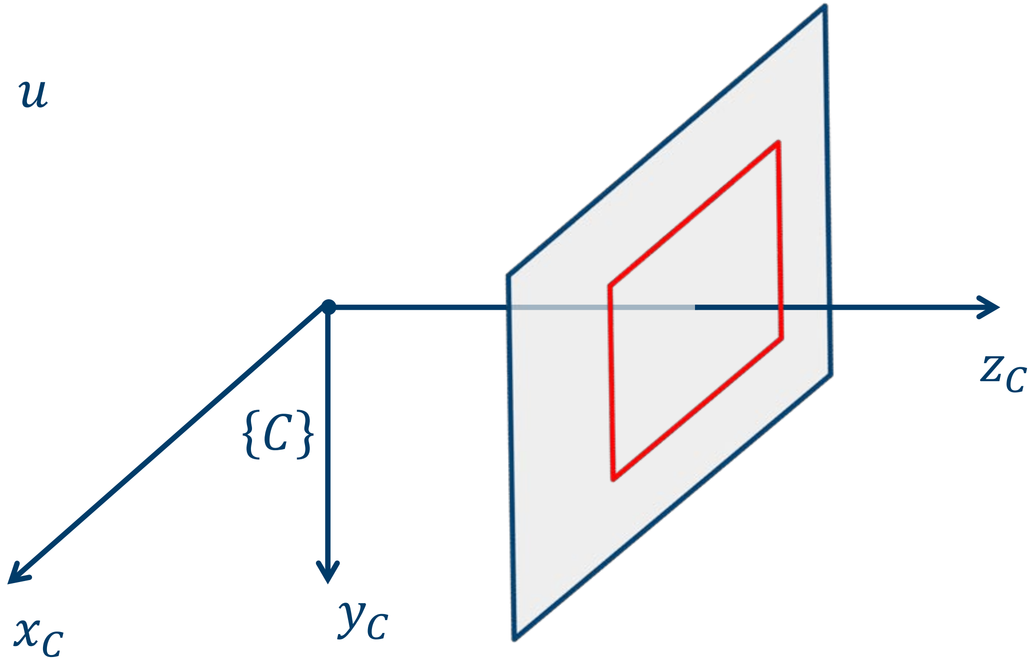


$$K = \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix}$$

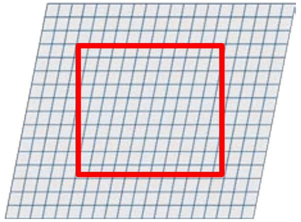
- The optical center (c_u, c_v) is where the optical axis intersects the image plane
- Often approximated by the center of the image, but the true value depends on how the sensor array lines up with the optical axis

Understanding the intrinsic part of the model





$$K = \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix}$$

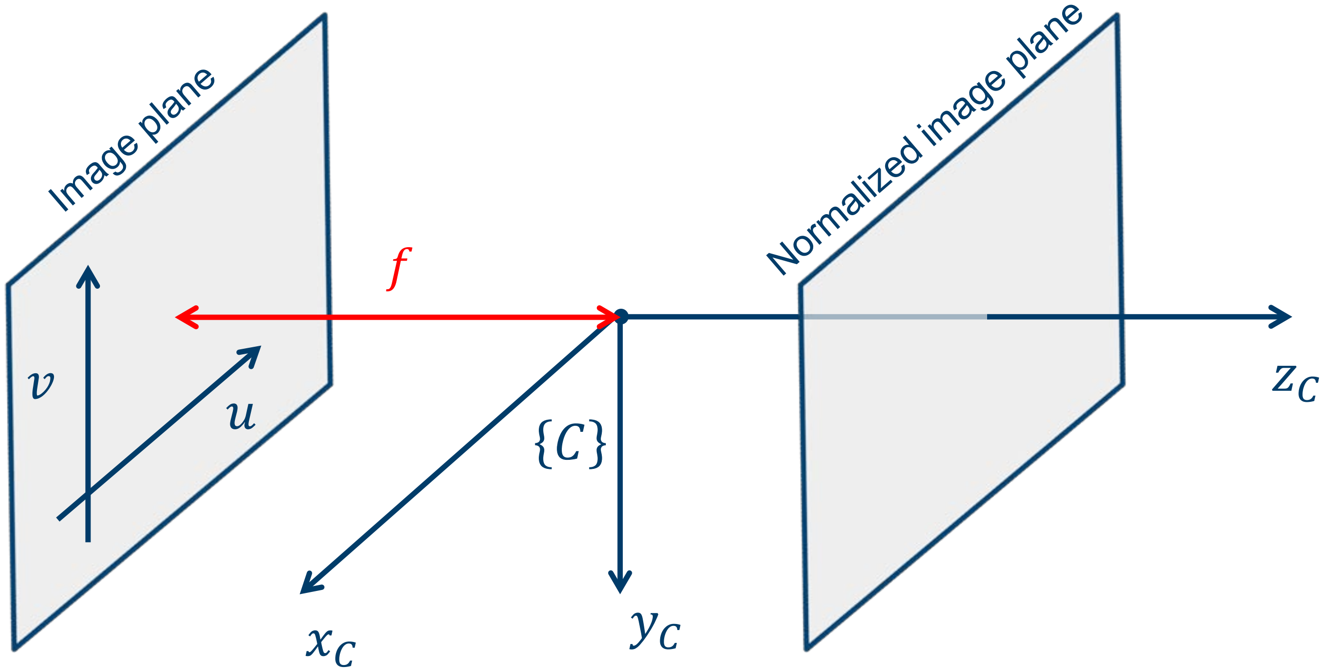


Detector array

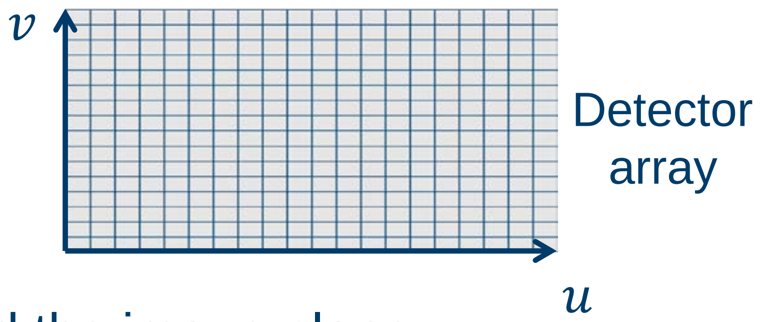
- The skew parameter s is required to describe cases when detector array has a non-orthogonal structure or when the array is not orthogonal to the optical axis
  - The illustration above shows how a rectangle projected onto a non-orthogonal detector array in the image plane can look like a rhombus in the image coordinates

- For most modern cameras this effect can be ignored, so we set $s = 0$

Understanding the intrinsic part of the model



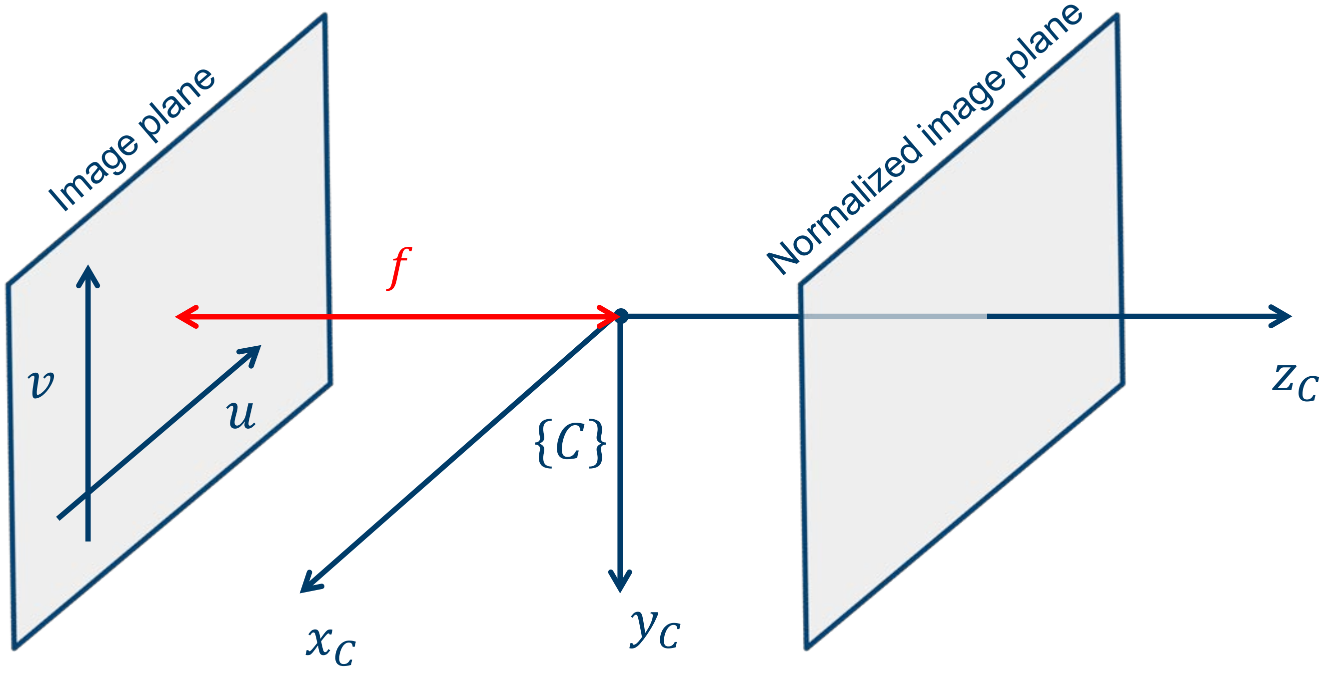
$$K = \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix}$$



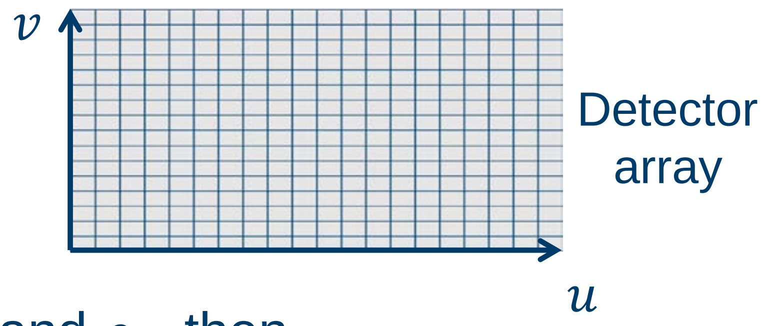
- The focal length f is the distance between the camera center and the image plane

- The parameters f_u and f_v are scaled versions of f reflecting that the density of detector elements can be different in the u- and v- direction of the image plane

Understanding the intrinsic part of the model



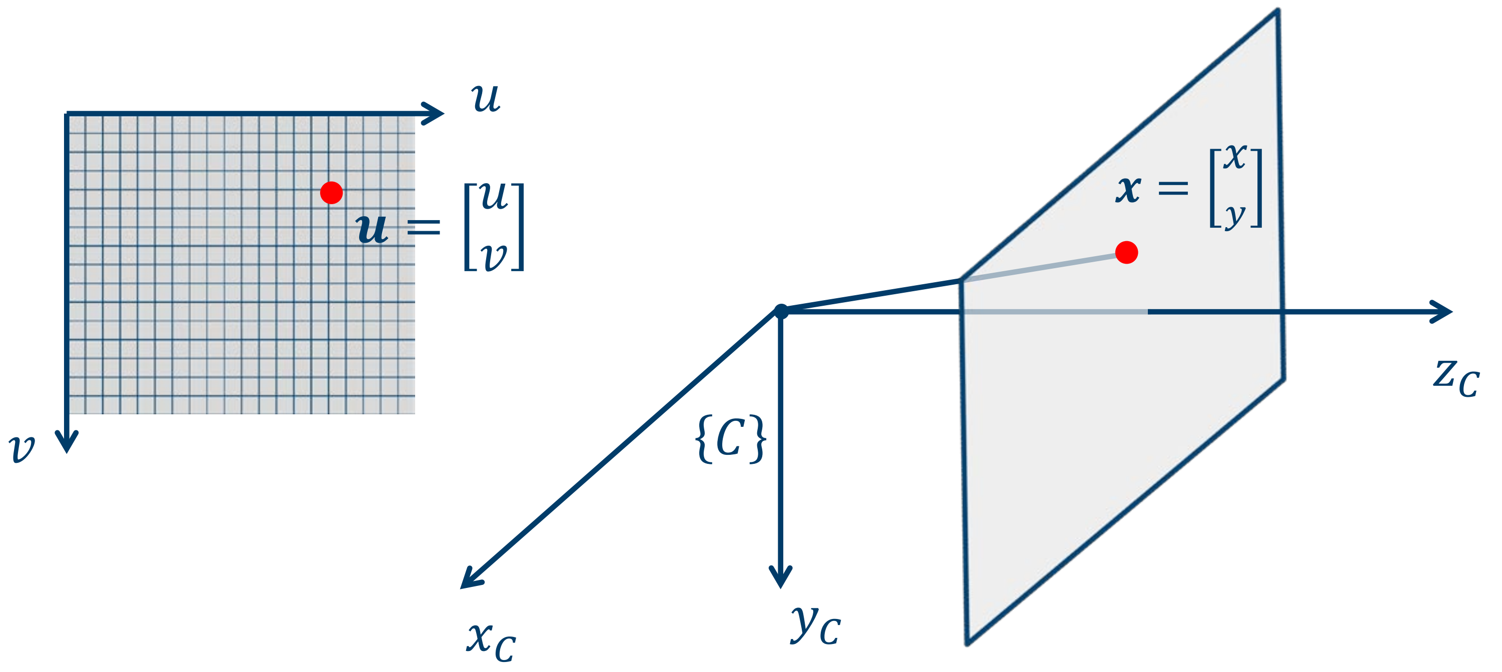
$$K = \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix}$$



- If we denote the detector density in the u- and v- direction by ρ_u and ρ_v, then

$$\begin{aligned} f_u &= \rho_u \cdot f \\ f_v &= \rho_v \cdot f \end{aligned} \quad \Rightarrow \quad \frac{f_u}{\rho_u} = \frac{f_v}{\rho_v} \quad \Leftrightarrow \quad f_v = \frac{\rho_v}{\rho_u} f_u$$

Understanding the intrinsic part of the model



$$\tilde{\boldsymbol{u}} = K\tilde{\boldsymbol{x}}$$

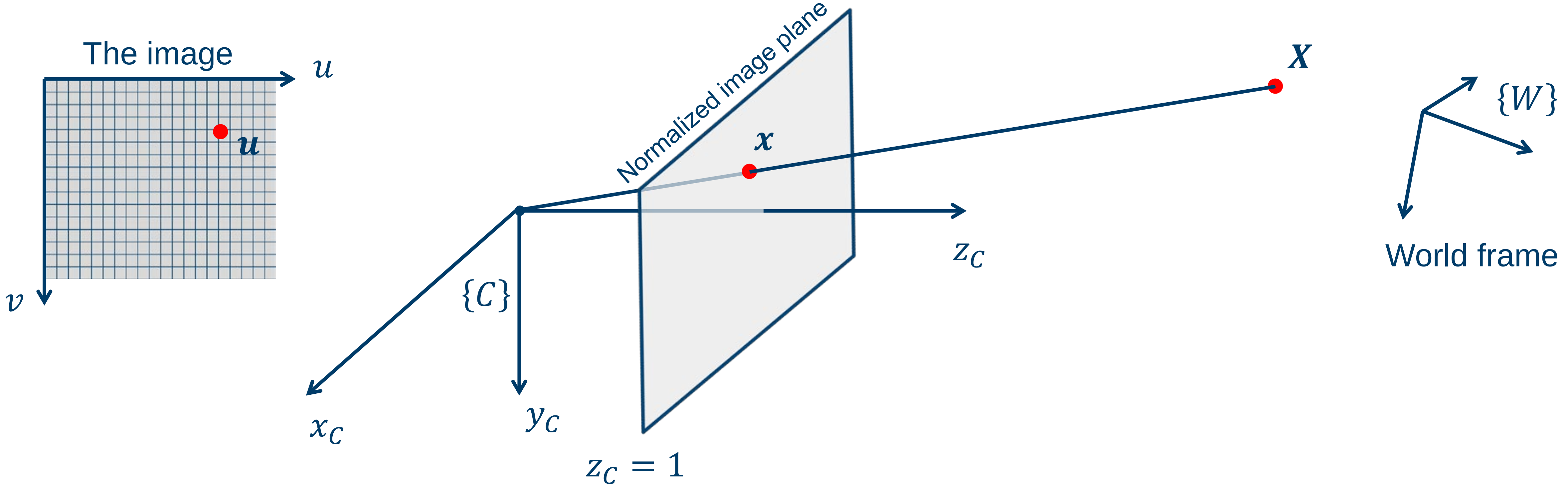
$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

- The camera calibration matrix K is homogeneous, so we are free to represent its parameters with the unit of our choice, but it is important that chosen unit is used consistently in K

$$u = f_u x + sy + c_u \quad \Rightarrow \quad [f_u] = \frac{[u]}{[x]}, [s] = \frac{[u]}{[y]}, [c_u] = [u]$$

$$v = f_v y + c_v \quad \Rightarrow \quad [f_v] = \frac{[v]}{[y]}, [c_v] = [v]$$

Recap



- The perspective camera model describes the correspondence between observed points in the world and points in the captured image

$$\tilde{\boldsymbol{u}} = K[R \quad \boldsymbol{t}]^W\tilde{\boldsymbol{X}} = \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} [R \quad \boldsymbol{t}]^W\tilde{\boldsymbol{X}} \qquad \text{where} \quad \begin{bmatrix} R & \boldsymbol{t} \\ \boldsymbol{0} & 1 \end{bmatrix} = {}^W\xi_C{}^{-1}$$

Comments

- The homogeneous 3×4 matrix that describes the correspondence between points in the world and points in the image is commonly referred to as *the camera matrix* or *the camera projection matrix* and denoted by P

$$\tilde{\boldsymbol{u}} = P\tilde{\boldsymbol{X}}$$

- Basic perspective camera

$$P = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} [R \quad \boldsymbol{t}]$$

- Finite projective camera

$$P = \begin{bmatrix} f_u & s & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix} [R \quad \boldsymbol{t}]$$

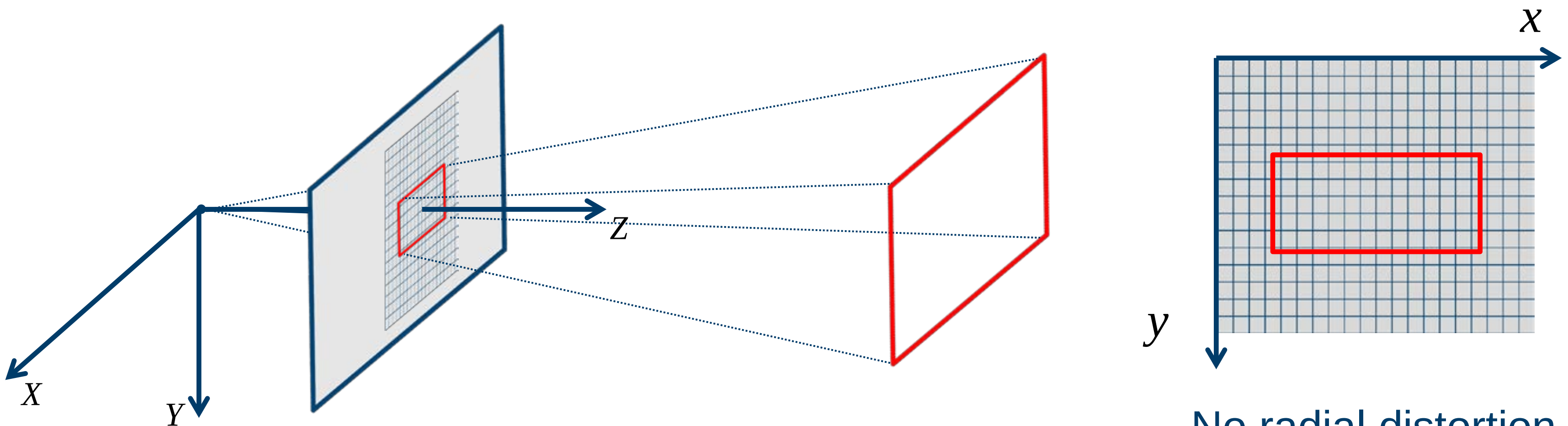
- General projective camera

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \quad \text{where } rank(P) = 2$$

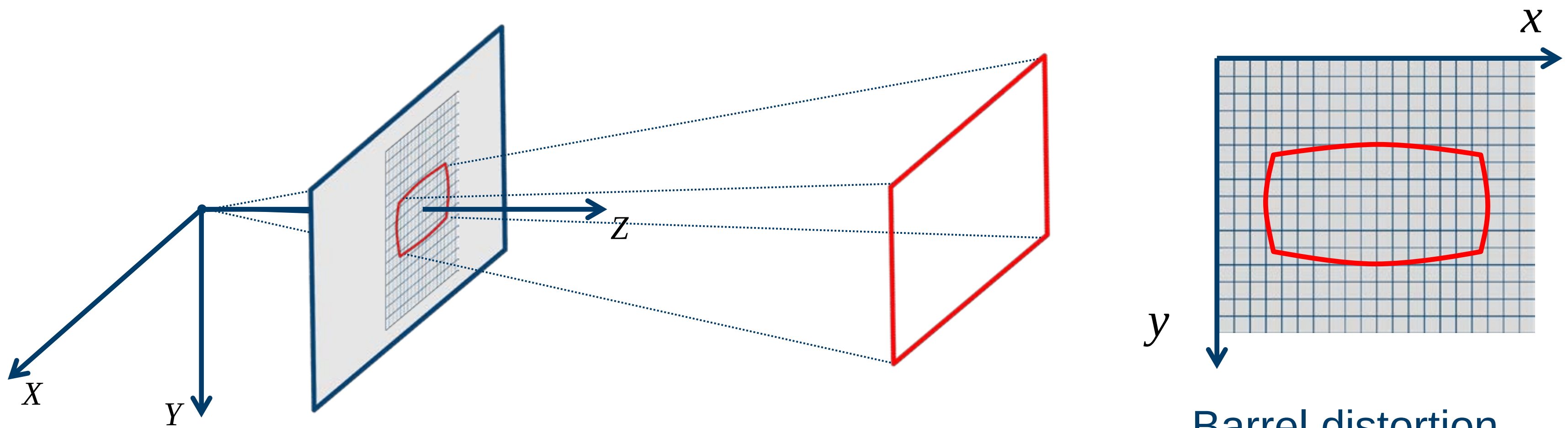
Lens distortions

- The geometry of the perspective camera is simple since we assume the pinhole to be infinitely small
- In reality the light passes through a lens that complicates the camera intrinsics
- Many wide-angle lenses have noticeable *radial distortion* which basically means that lines in the scene appear as curves in the image
- There are two types of radial distortion
  - barrel distortion
  - pincushion distortion

Lens distortions

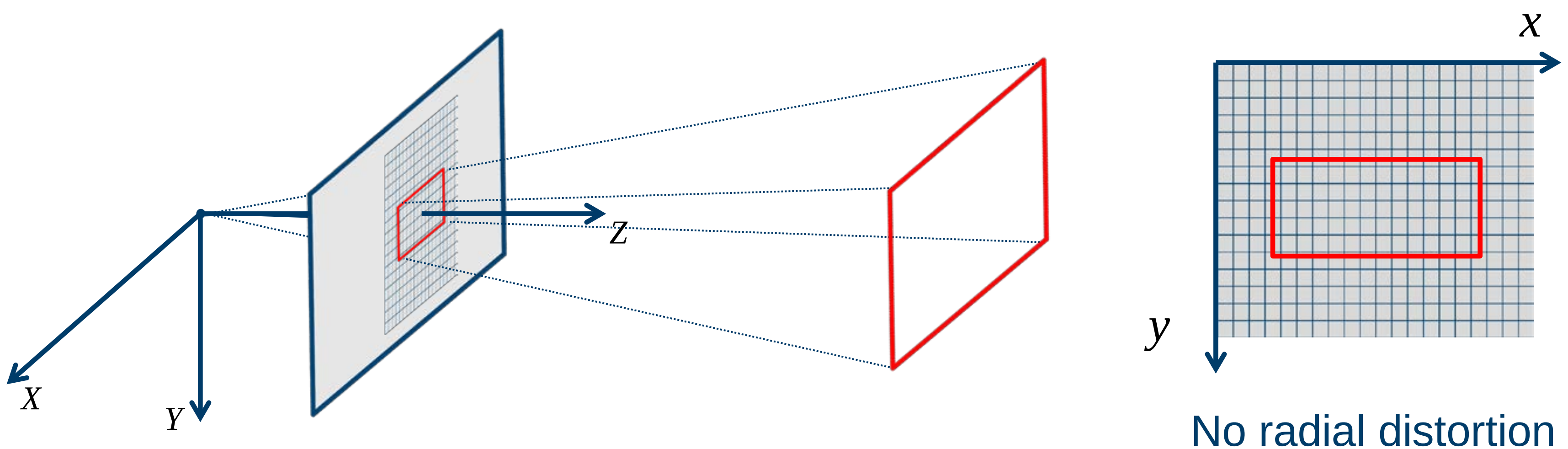


No radial distortion

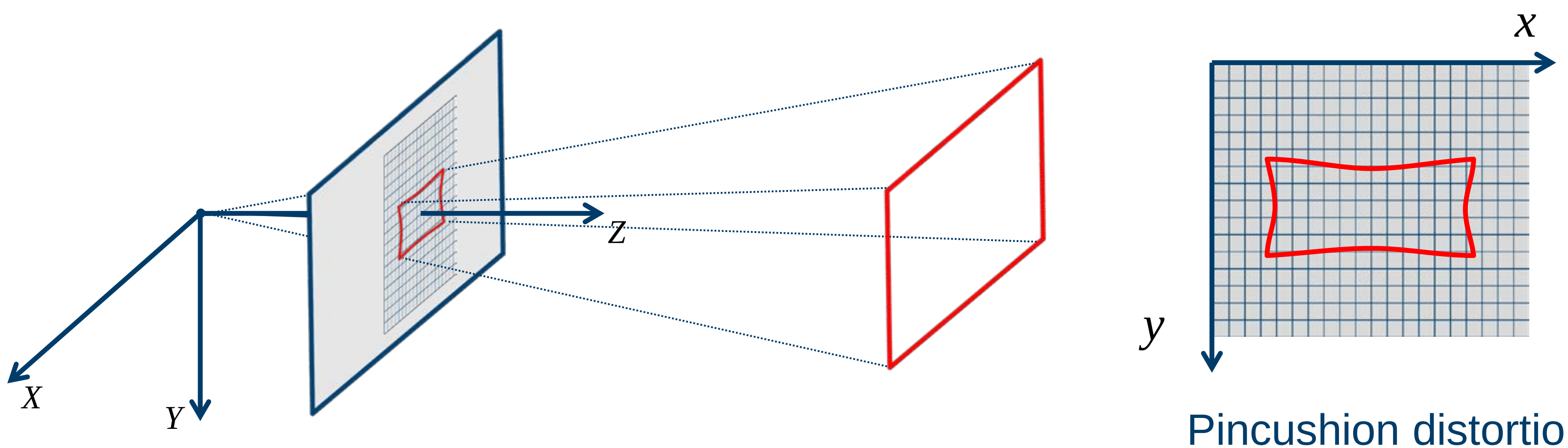


Barrel distortion

Lens distortions



No radial distortion



Pincushion distortion

Lens distortions

- A camera with radial distortion is not a perspective camera (lines are not preserved) and is not well described by the pinhole model
- Radial distortion can often be well described using a simple polynomial model, so the geometrical errors introduced by the lens is possible to correct
- The correction is performed on normalized image coordinates (x, y)
- Let $(\hat{x}, \hat{y})$ denote the corrected normalized image coordinates, then a simple radial distortion model can look like this:

$$\hat{x} = x(1 + \kappa_1(x^2 + y^2) + \kappa_2(x^2 + y^2)^2)$$
$$\hat{y} = y(1 + \kappa_1(x^2 + y^2) + \kappa_2(x^2 + y^2)^2)$$

where κ_1 and κ_2 are the radial distortion parameters

Lens distortions

- If we include radial distortion correction into our camera model, the full 3D to 2D transformation will look like this

$$\begin{bmatrix} {}^{W}X \\ {}^{W}Y \\ {}^{W}Z \\ 1 \end{bmatrix} \xmapsto{[R \quad \boldsymbol{t}]} \begin{bmatrix} {}^{C}x \\ {}^{C}y \\ 1 \end{bmatrix} \xmapsto{\substack{\text{radial} \\ \text{distortion} \\ \text{correction}}} \begin{bmatrix} {}^{C}\hat{x} \\ {}^{C}\hat{y} \\ 1 \end{bmatrix} \xmapsto{K} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$

- When we calibrate a camera, this usually includes the estimation of radial distortion

Summary

- The perspective camera model
  - $P = K[R, \boldsymbol{t}]$ – The camera matrix
  - Intrinsic: K – The camera calibration matrix
  - Extrinsic: $[R, \boldsymbol{t}]$

- Lens distortion
  - Radial distortion
  - Tangential distortion (often ignored)

- Additional reading:
  - Szeliski: 2.1.5, 2.1.6