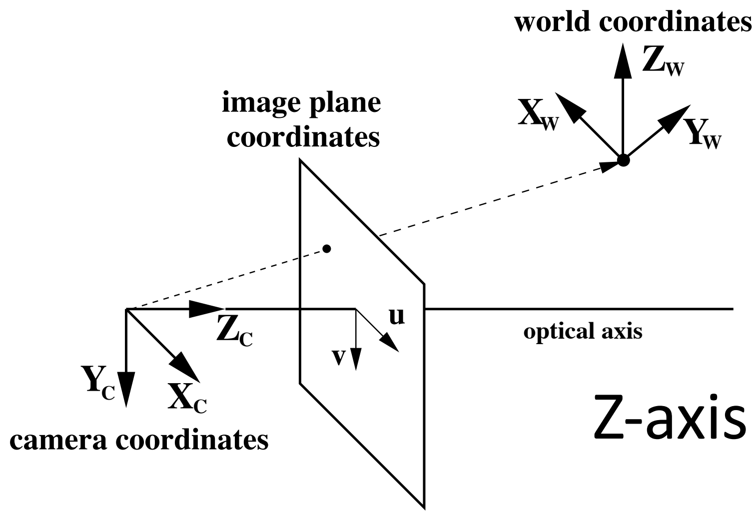
Camera Coordinate System



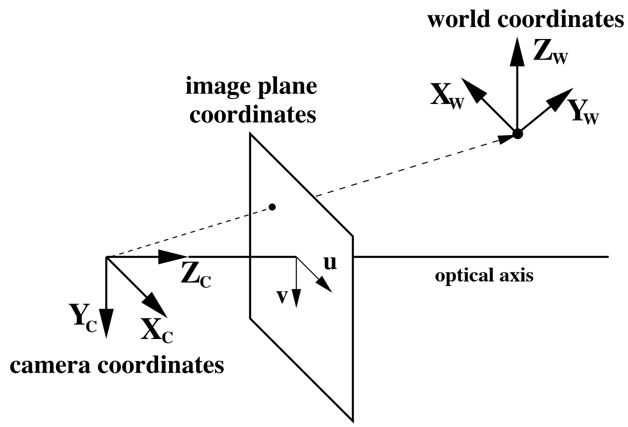
Z-axis is the optical axis

The image plane (u, v) is perpendicular to the optical axis.
Intersection of the image plane with the optical axis is the *image center* (u_o, v_o)

Projection in pixels $$u = f\frac{X_c}{Z_c} + u_o \quad v = f\frac{Y_c}{Z_c} + v_o.$$

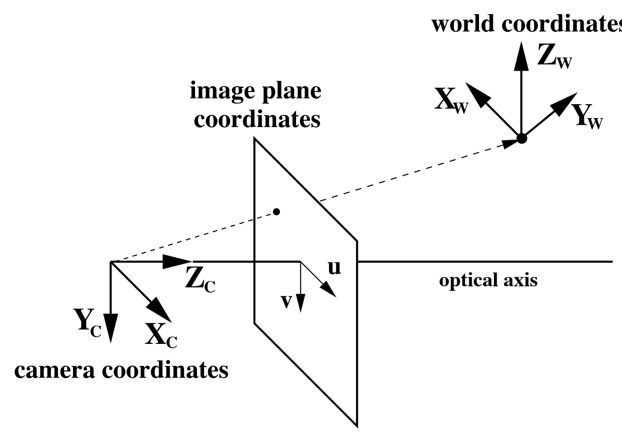
Perspective projection in matrix form

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & u_o \\ 0 & f & v_o \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix}$$



From camera to world

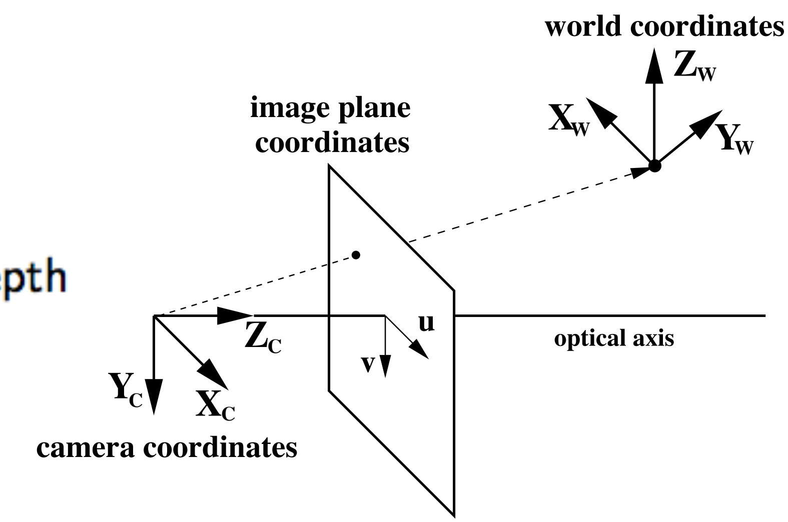
$$\begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix} = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$



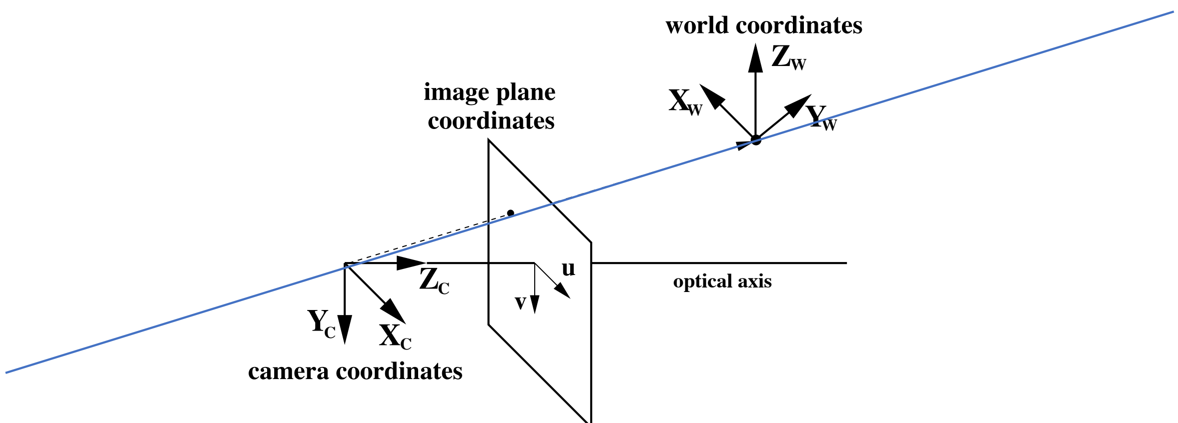
The 3x4 projection matrix P

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & u_o \\ 0 & f & v_o \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R & t \end{pmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} = P \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

λ is the unknown depth



The meaning of the projection equation:
It is the equation of a ray in world coordinates going through the camera center

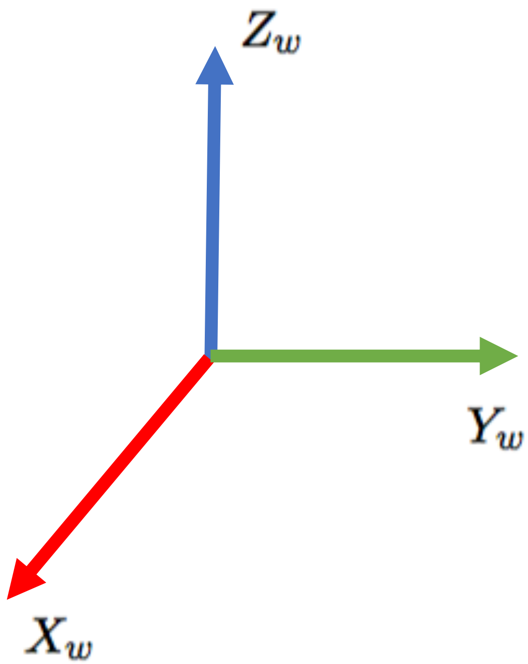


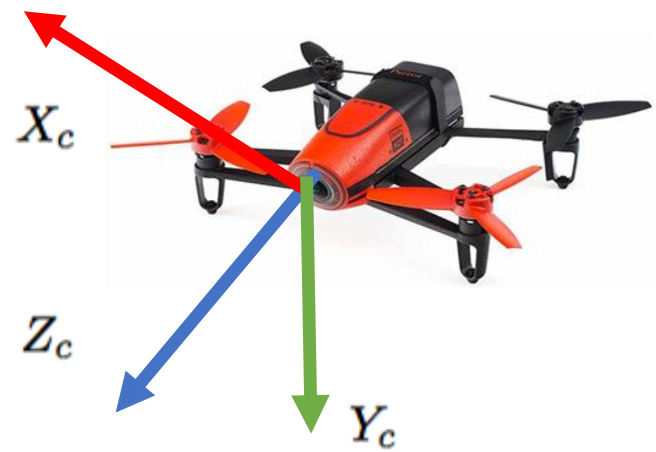
$$\lambda K^{-1}\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = R\begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} + t \Longrightarrow \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} = \underbrace{\boxed{-R^T T}}_{\text{Camera center}} + \lambda \underbrace{\boxed{R^T K^{-1}\begin{pmatrix} u \\ v \\ 1 \end{pmatrix}}}_{\text{Ray direction}}$$

Rotations and Translations

Kostas Daniilidis

Transformation between camera and world coordinate systems



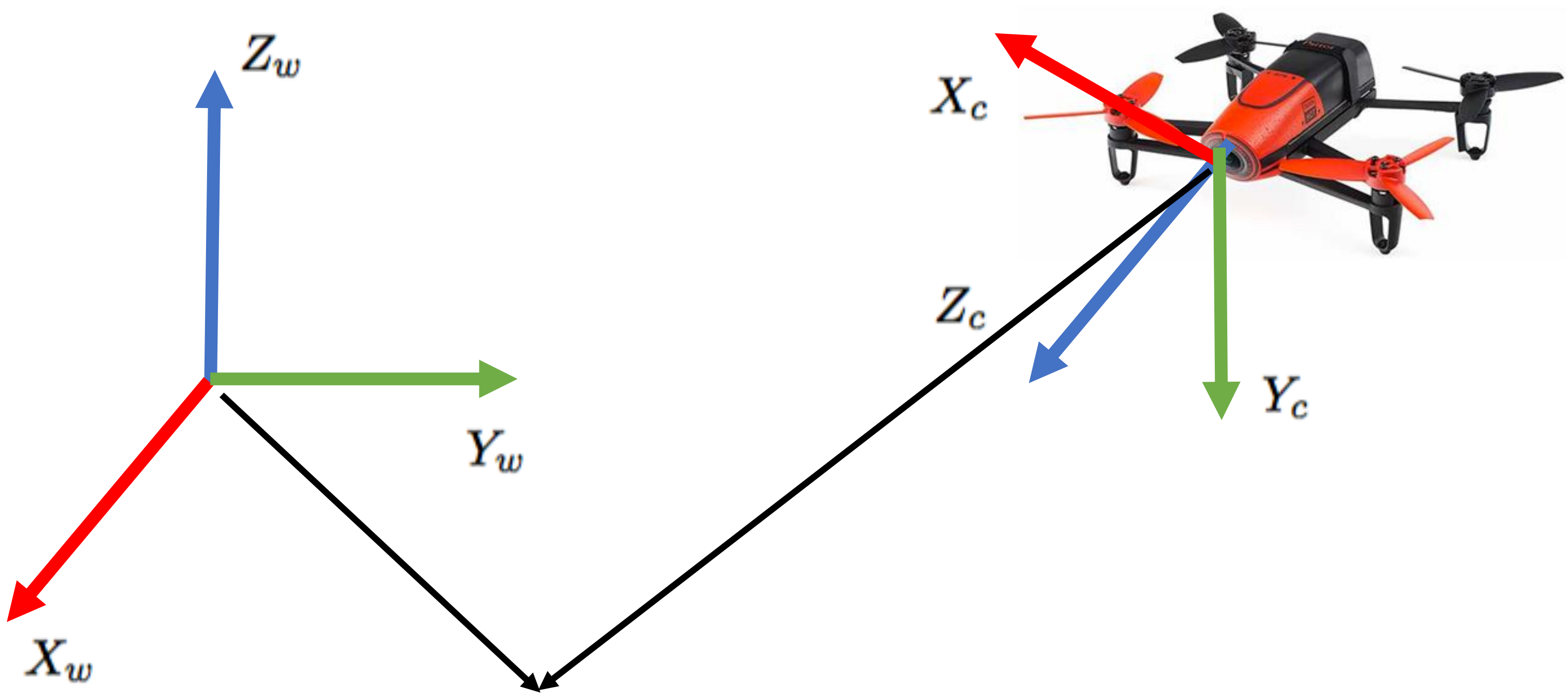


Red for X-Axis
Green for Y-Axis
Blue for Z-Axis

Remember
RGB is XYZ

$$ {}^{c}P = {}^{c}R_{w}\,{}^{w}P + {}^{c}T_{w} $$

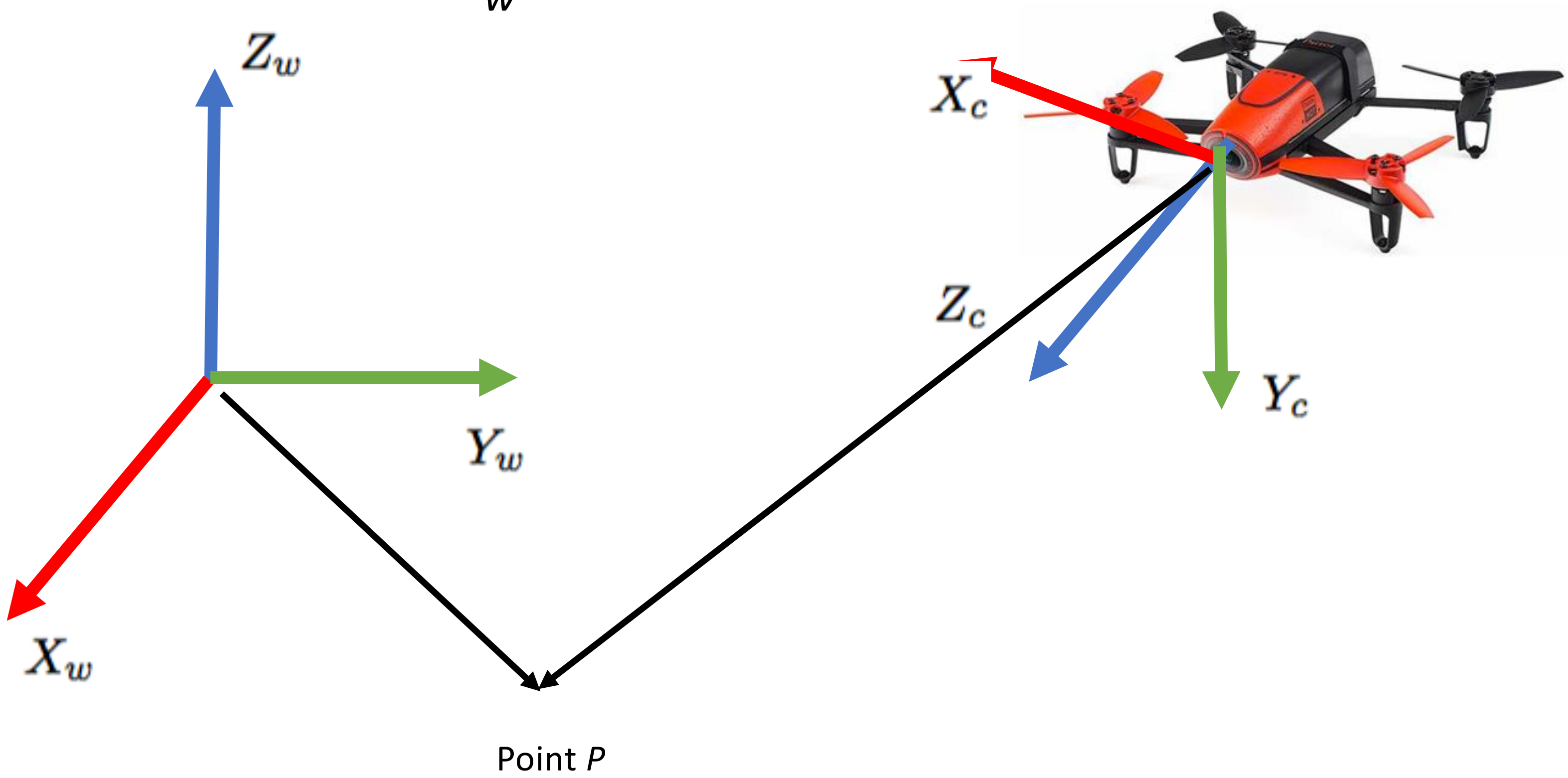
Point *P* can be expressed with respect to “w” or “c” coordinate frames



Point *P*

$$^{c}P = {^{c}R_w}\, {^{w}P} + {^{c}T_w}$$

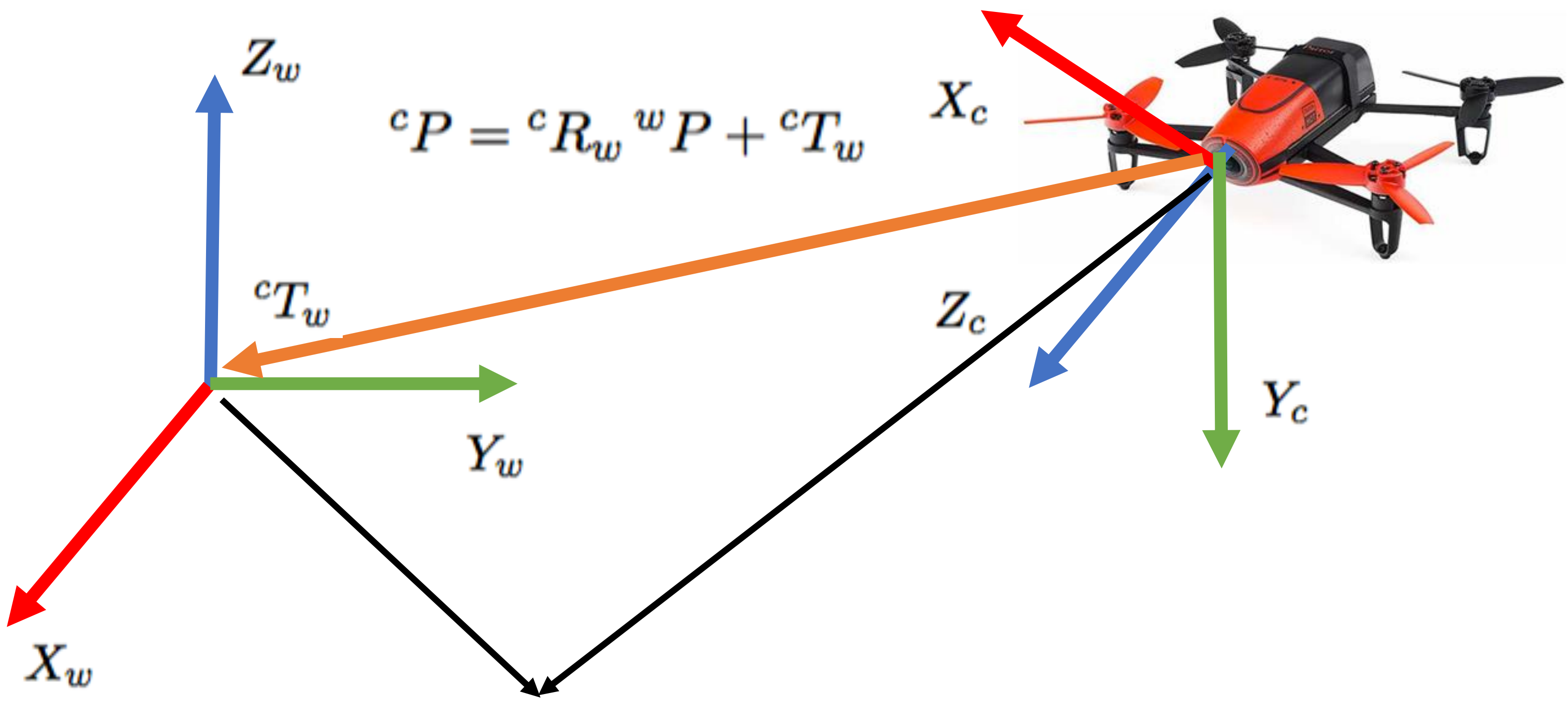
What is the geometric meaning of the rotation ${^{c}R_w}$ and the translation ${^{c}T_w}$?



What is the geometric meaning of the translation ${}^{c}T_{w}$?

This is easy to see if we set ${}^{w}P$ to zero.

Then, ${}^{c}P = {}^{c}R_{w}\,0 + {}^{c}T_{w}$ is the vector from camera origin to world origin:



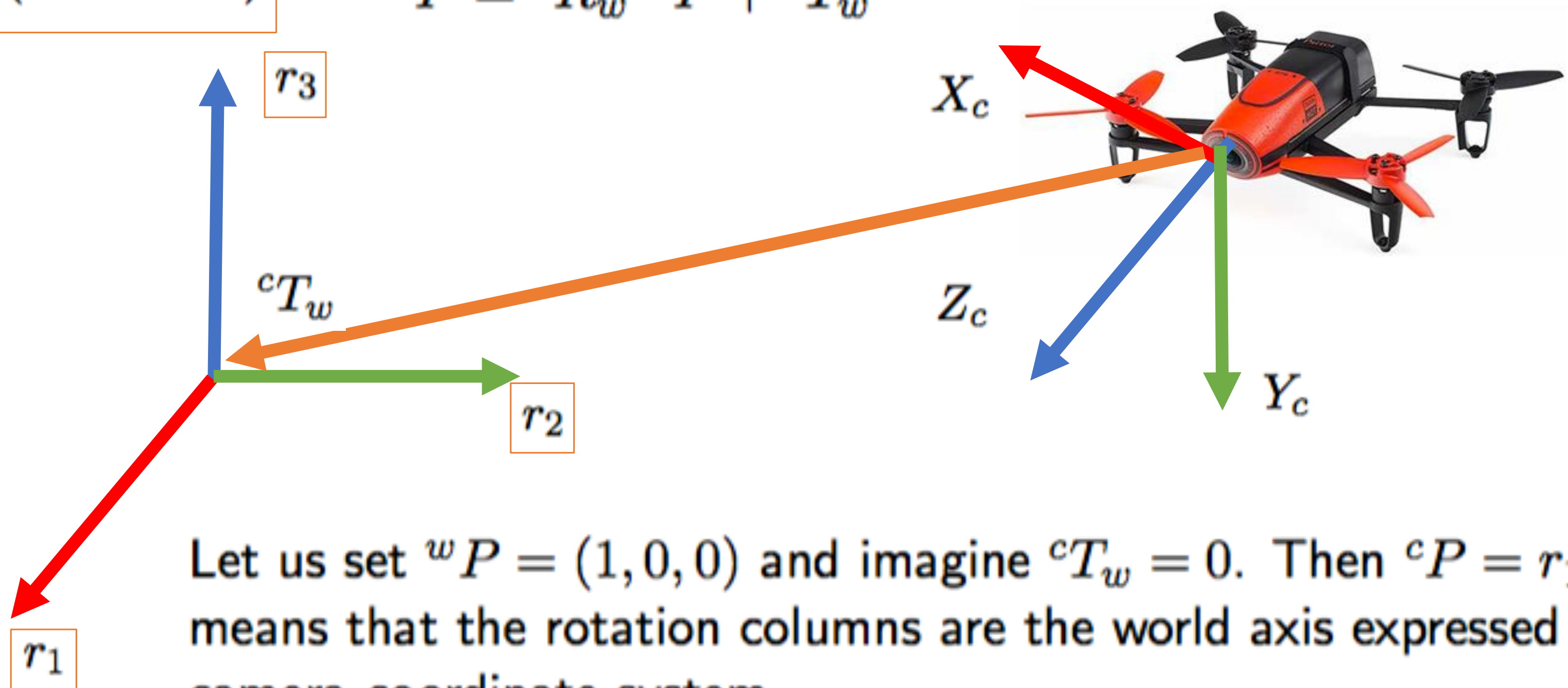
Point P

What is the geometric meaning of the rotation ${}^{c}R_{w}$?

Let the rotation matrix be written as 3 orthogonal column vectors:

$$\boxed{{}^{c}R_{w} = \begin{pmatrix} r_1 & r_2 & r_3 \end{pmatrix}}$$

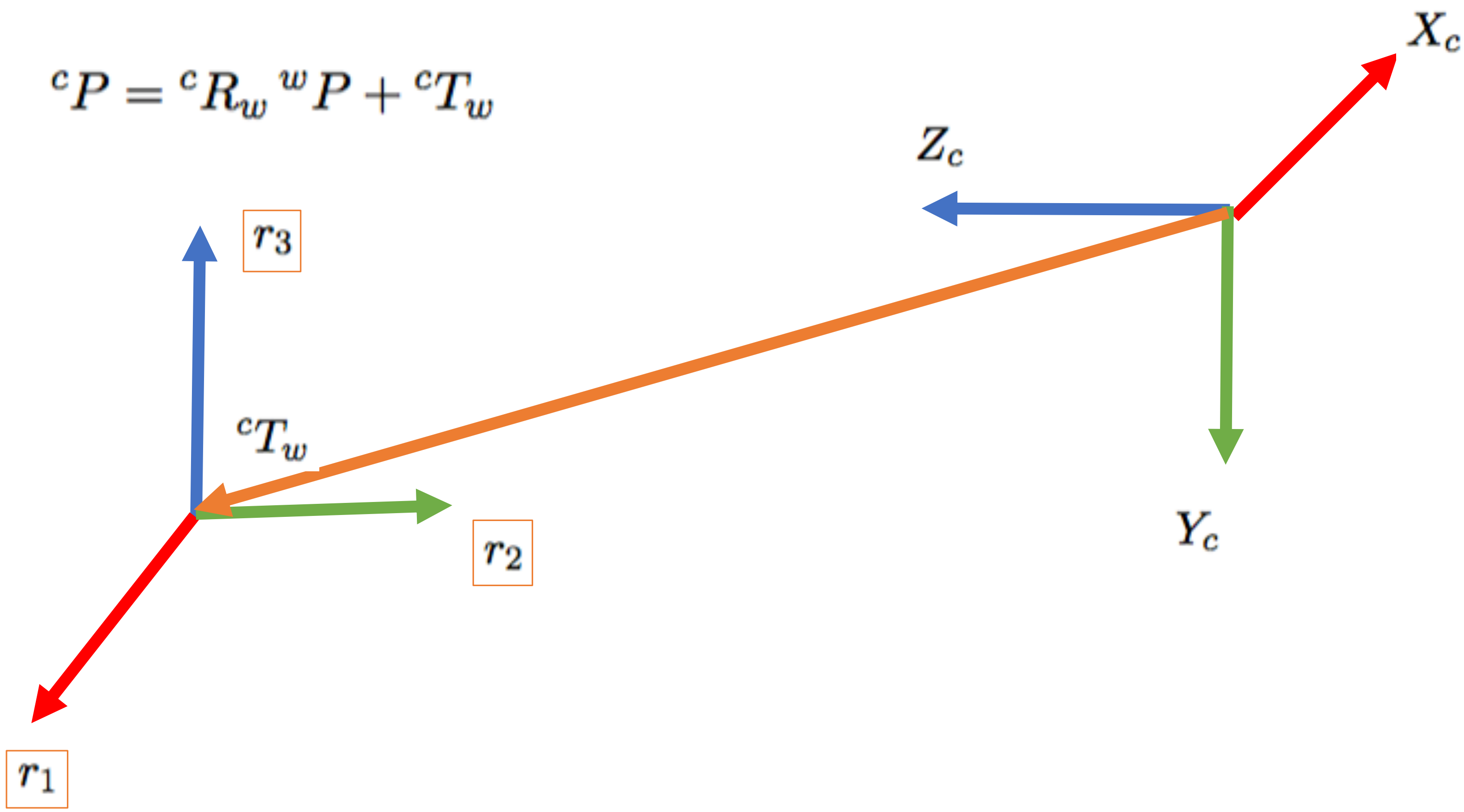
$${}^{c}P = {}^{c}R_{w}\, {}^{w}P + {}^{c}T_{w}$$



Let us set ${}^{w}P = (1, 0, 0)$ and imagine ${}^{c}T_{w} = 0$. Then ${}^{c}P = r_1$ which means that the rotation columns are the world axis expressed in the camera coordinate system.

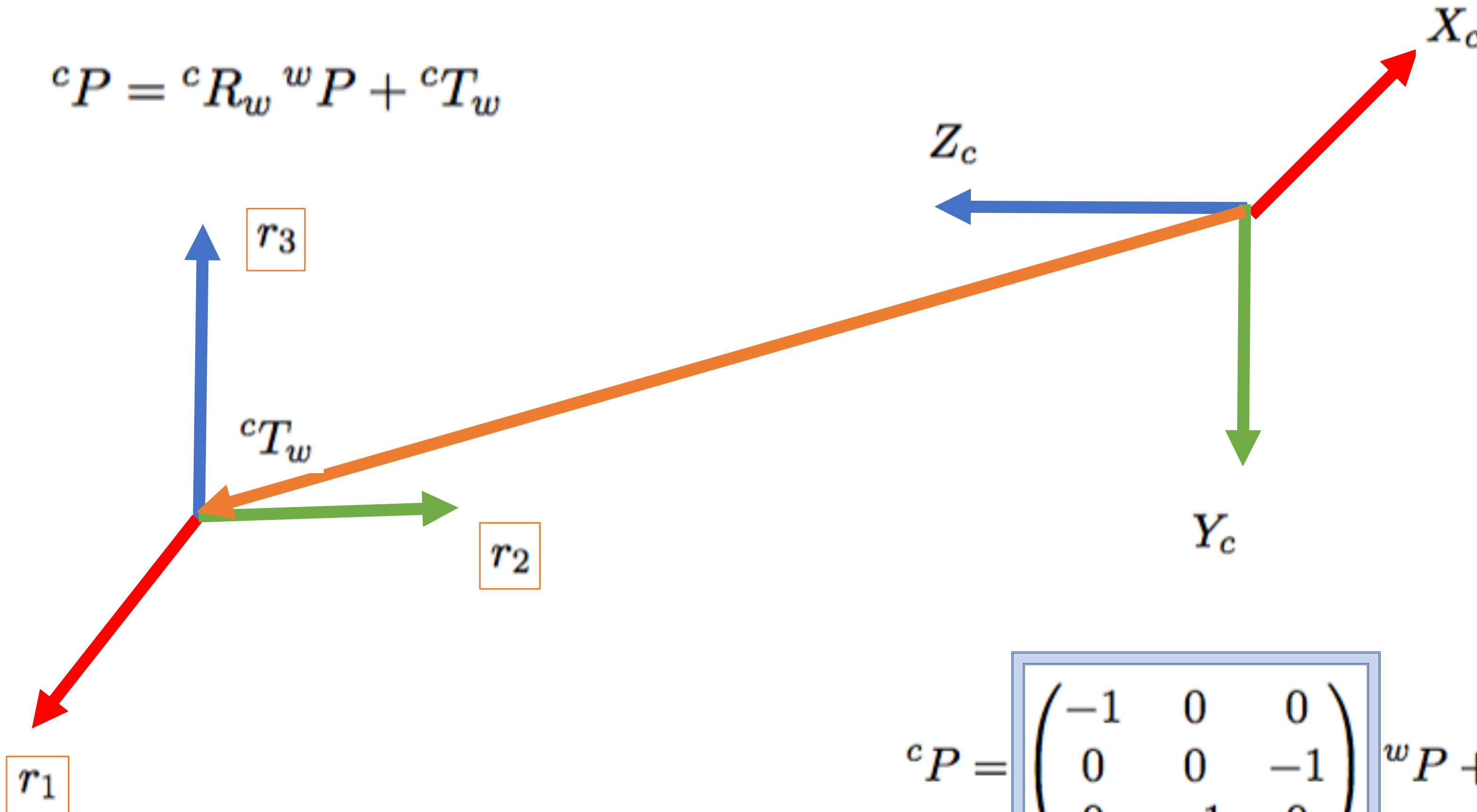
Let us look at the simple example:

$$^{c}P = {^{c}R_{w}}\,{^{w}P} + {^{c}T_{w}}$$



How does the rotation matrix read?

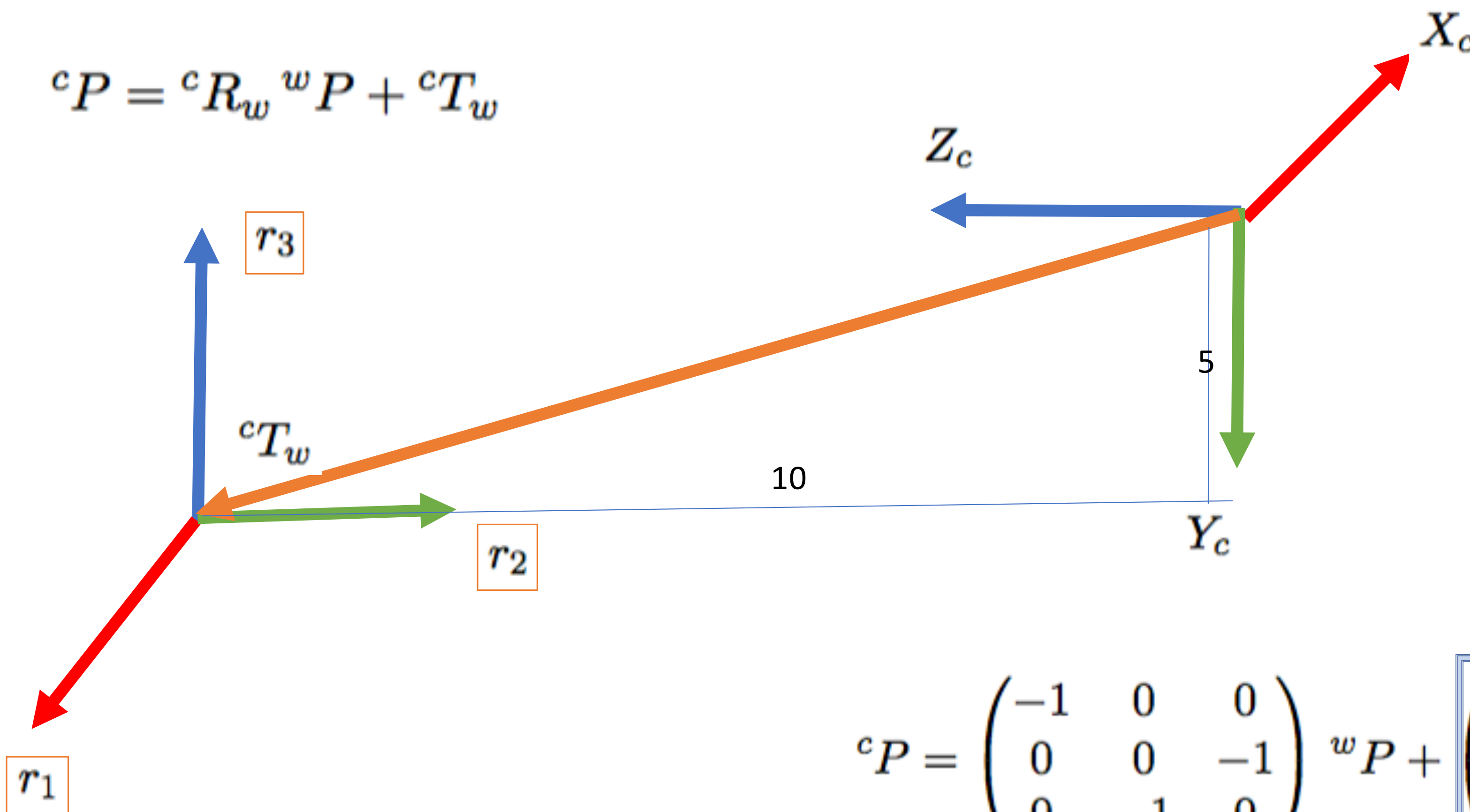
$$ {}^{c}P = {}^{c}R_{w}\, {}^{w}P + {}^{c}T_{w} $$



$$ {}^{c}P = \boxed{\begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}} {}^{w}P + \begin{pmatrix} ? \\ ? \\ ? \end{pmatrix} $$

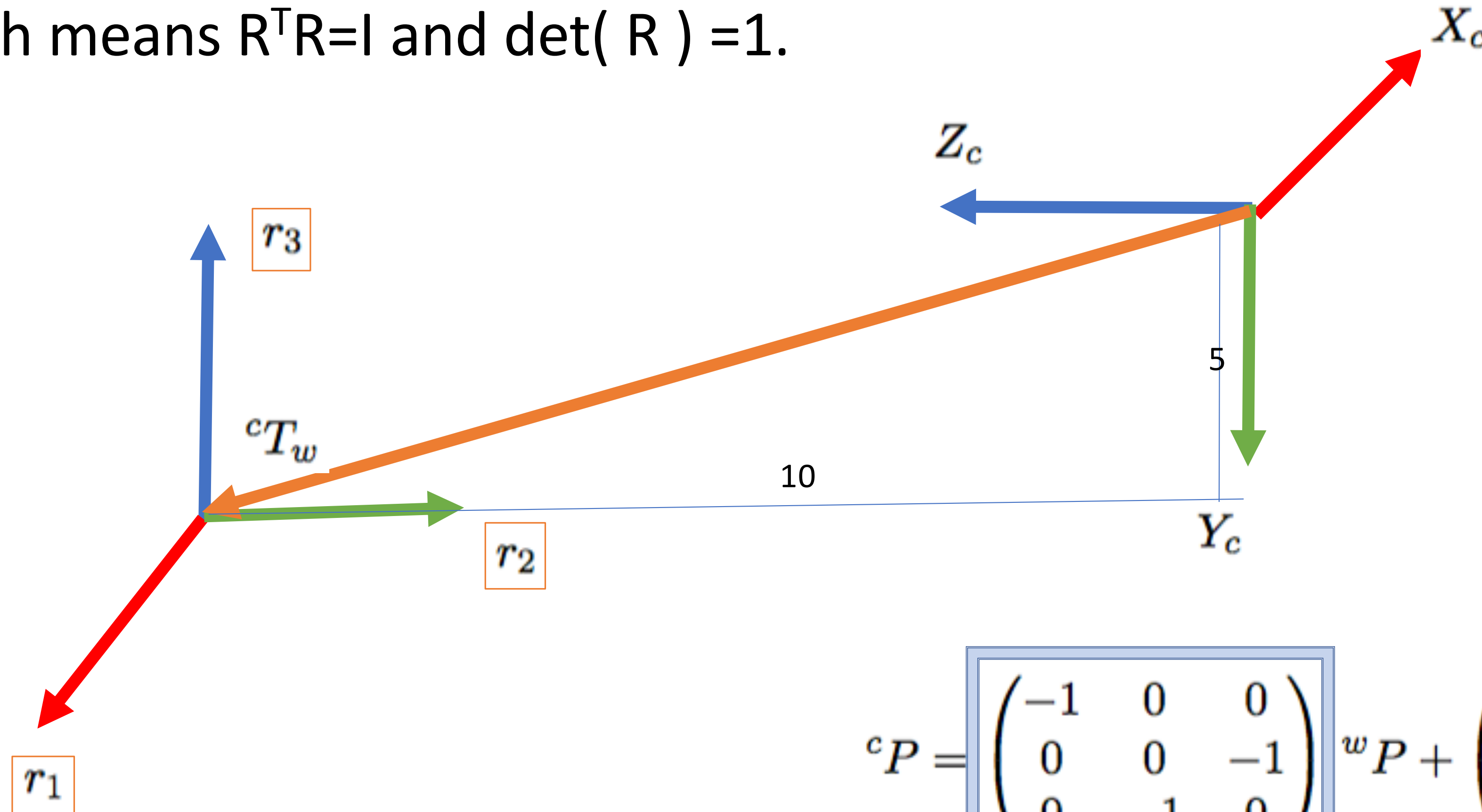
What about the translation:

$$ {}^cP = {}^cR_w\, {}^wP + {}^cT_w $$



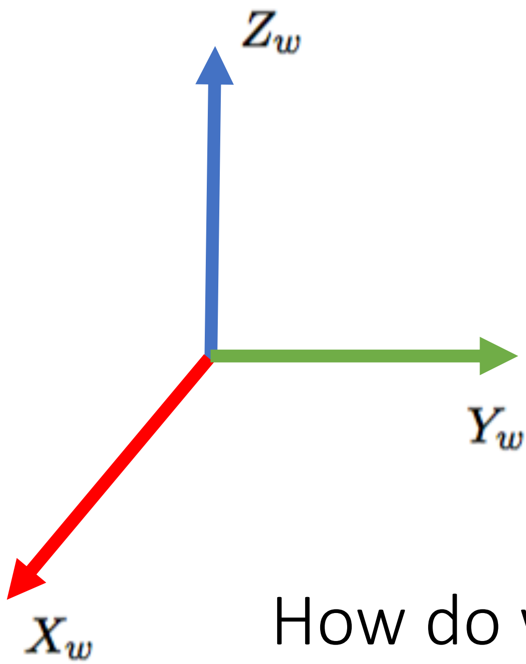
$$ {}^cP = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix} {}^wP + \boxed{\begin{pmatrix} 0 \\ 5 \\ 10 \end{pmatrix}} $$

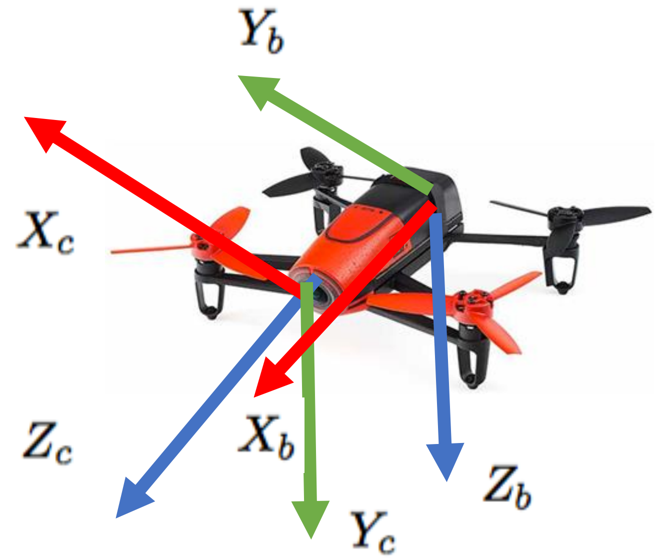
We have to make sure that the 3x3 matrix is a rotation matrix,
Which means $R^T R = I$ and $\det(R) = 1$.



$$
{}^cP = \boxed{\begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}} {}^wP + \begin{pmatrix} 0 \\ 5 \\ 10 \end{pmatrix}
$$

Now imagine one more coordinate frame: a body frame with axes corresponding to roll (X_b), pitch (Y_b), yaw (Z_b) angles.

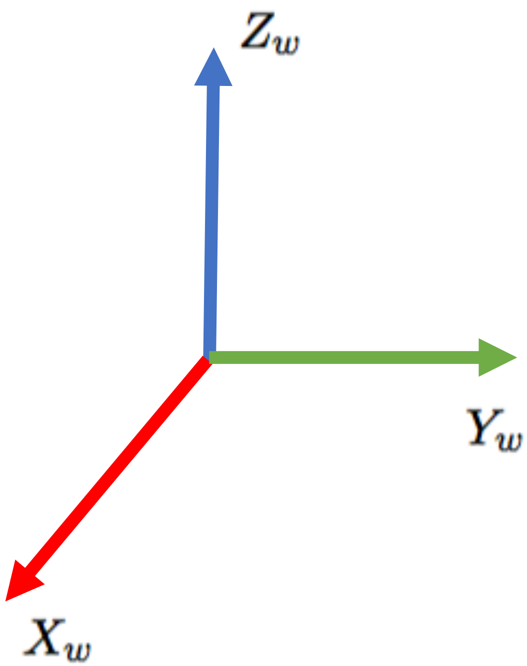


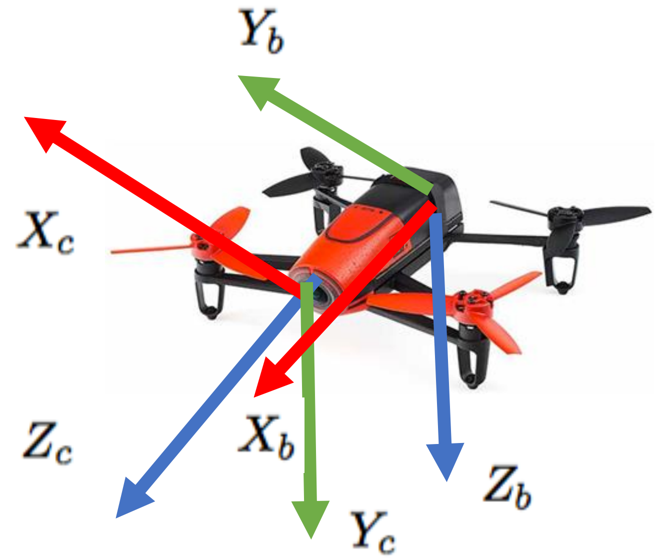


How do we compose transformations?

The easiest way to transform between coordinate systems is to use 4x4 matrices:

$$ {}^{c}M_{w} = \begin{pmatrix} & {}^{c}R_{w} & & {}^{c}T_{w} \\ 0 & 0 & 0 & 1 \end{pmatrix} $$



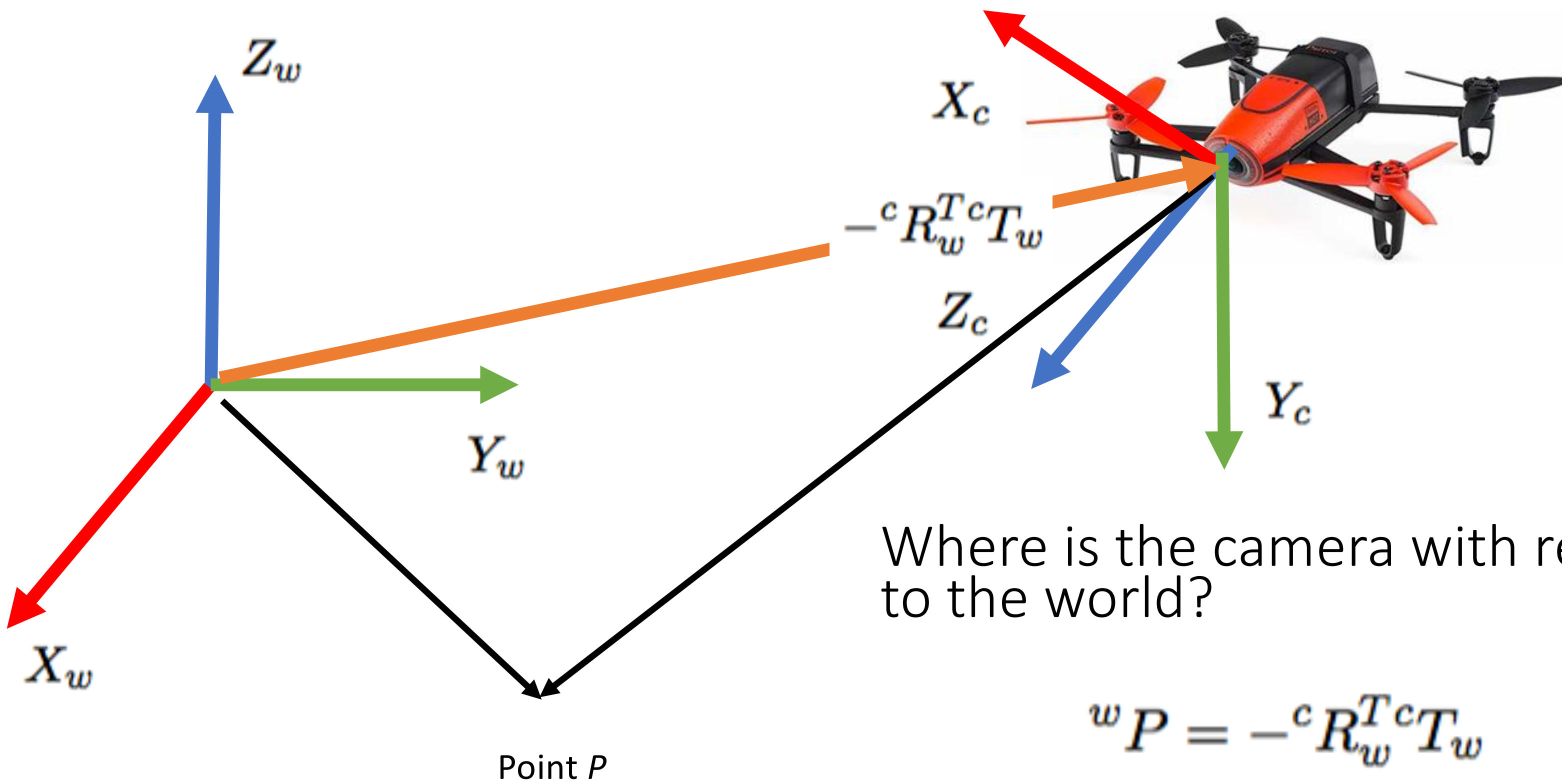


Then we just concatenate the 4x4 matrices

$$ {}^{w}M_{b} = {}^{w}M_{c}\,{}^{c}M_{b} $$

What about the inverse transformation?

$$^{w}M_c = \begin{pmatrix} & {}^{c}R_w^T & & -{}^{c}R_w^T\,{}^{c}T_w \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

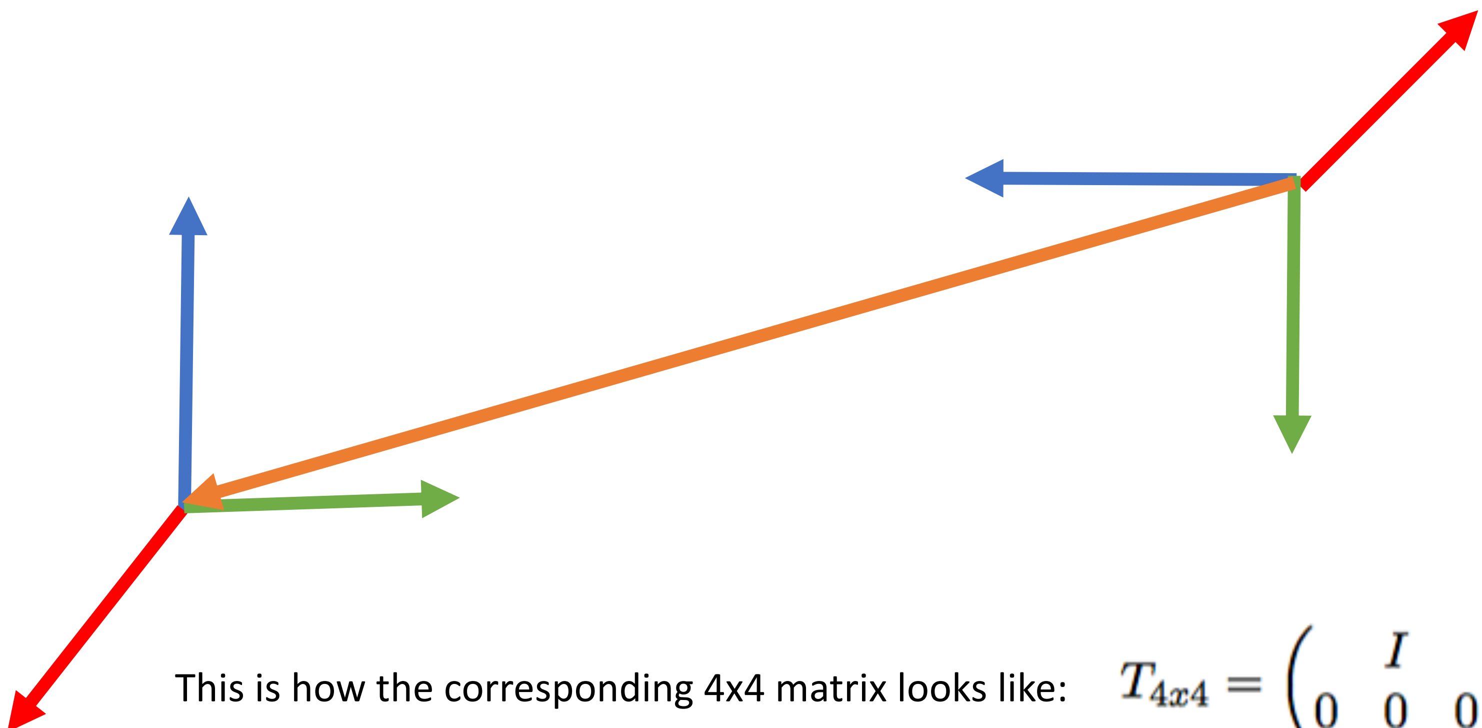


Where is the camera with respect to the world?

$$^{w}P = -{}^{c}R_w^T\,{}^{c}T_w$$

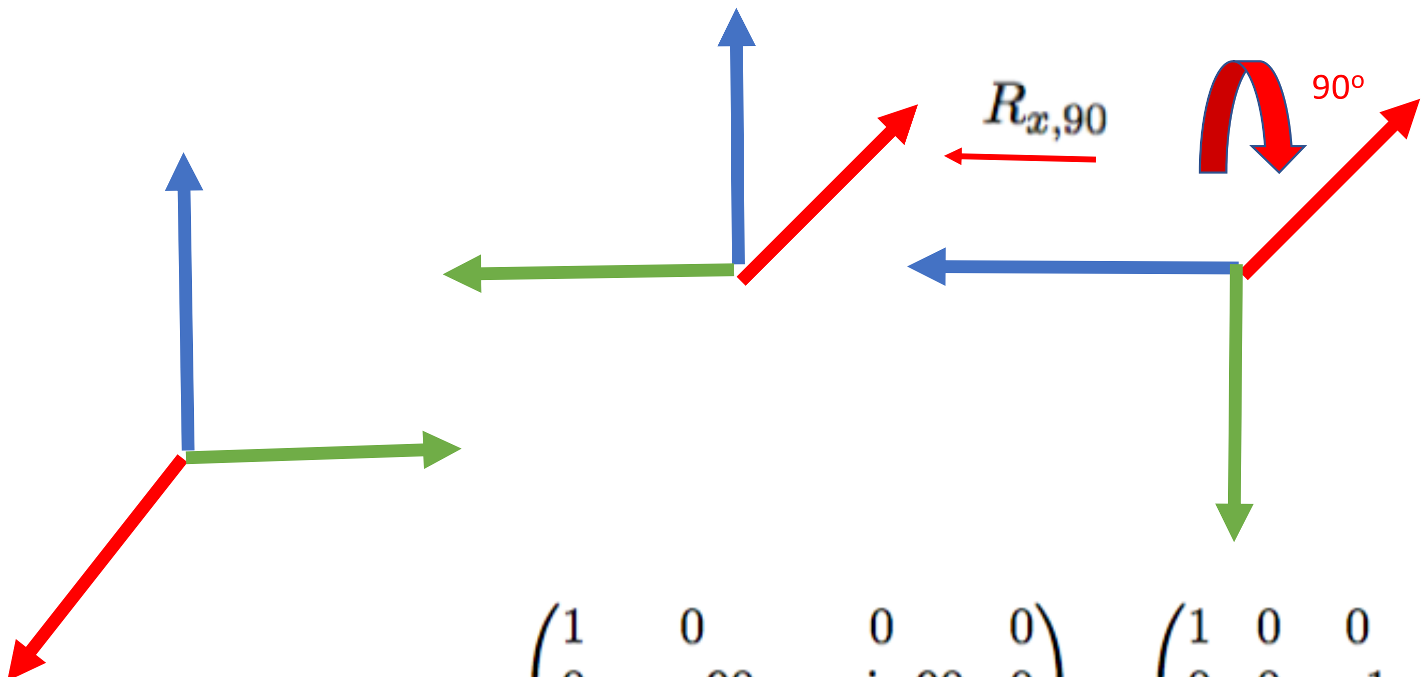
Alternative interpretation as a sequence of motions:

1. The camera frame first translates to the world



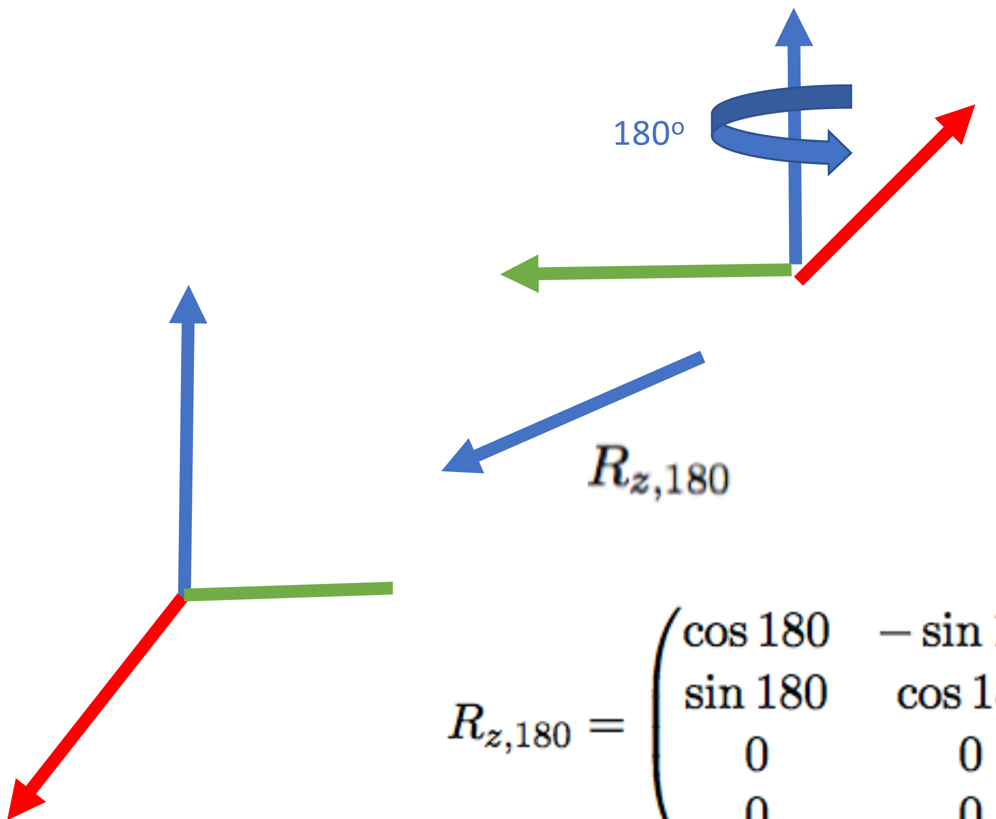
This is how the corresponding 4x4 matrix looks like: $T_{4x4} = \begin{pmatrix} & I & & {}^{c}T_{w} \\ 0 & 0 & 0 & 1 \end{pmatrix}$

2. The camera frame rotates 90 degrees around x



$$R_{x,90} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 90 & -\sin 90 & 0 \\ 0 & \sin 90 & \cos 90 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

A3. The camera frame rotates 180 degrees around z



$$R_{z,180} = \begin{pmatrix} \cos 180 & -\sin 180 & 0 & 0 \\ \sin 180 & \cos 180 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

How do we compose these motions? Golden rule: when we move coordinate frames and we refer to the most recent coordinate frame
we always postmultiply!

$$
\begin{aligned}
{}^{c}M_w &= T R_{x,90} R_{z,180} \\
&= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 10 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 5 \\ 0 & -1 & 0 & 10 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\end{aligned}
$$